Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 21 (2019), Issue 5   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Short Paper

Review

Data-Driven Blood Glucose Pattern Classification and Anomalies Detection: Machine-Learning Applications in Type 1 Diabetes

Ashenafi Zebene Woldaregay[1], MSc; Eirik Årsand[2], PhD; Taxiarchis Botsis[3], PhD; David Albers[4], PhD; Lena Mamykina[4], PhD; Gunnar Hartvigsen[1], PhD

[1]Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway
[2]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway
[3]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, United States
[4]Department of Biomedical Informatics, Columbia University, New York, NY, United States

Corresponding Author:
Ashenafi Zebene Woldaregay, MSc
Department of Computer Science
University of Tromsø – The Arctic University of Norway
Realfagbygget, Hansine Hansens vei 54
Tromsø,
Norway
Phone: 47 77646444
Email: ashenafi.z.woldaregay@uit.no

Abstract

Background: Diabetes mellitus is a chronic metabolic disorder that results in abnormal blood glucose (BG) regulations. The BG level is preferably maintained close to normality through self-management practices, which involves actively tracking BG levels and taking proper actions including adjusting diet and insulin medications. BG anomalies could be defined as any undesirable reading because of either a precisely known reason (normal cause variation) or an unknown reason (special cause variation) to the patient. Recently, machine-learning applications have been widely introduced within diabetes research in general and BG anomaly detection in particular. However, irrespective of their expanding and increasing popularity, there is a lack of up-to-date reviews that materialize the current trends in modeling options and strategies for BG anomaly classification and detection in people with diabetes.

Objective: This review aimed to identify, assess, and analyze the state-of-the-art machine-learning strategies and their hybrid systems focusing on BG anomaly classification and detection including glycemic variability (GV), hyperglycemia, and hypoglycemia in type 1 diabetes within the context of personalized decision support systems and BG alarm events applications, which are important constituents for optimal diabetes self-management.

Methods: A rigorous literature search was conducted between September 1 and October 1, 2017, and October 15 and November 5, 2018, through various Web-based databases. Peer-reviewed journals and articles were considered. Information from the selected literature was extracted based on predefined categories, which were based on previous research and further elaborated through brainstorming.

Results: The initial results were vetted using the title, abstract, and keywords and retrieved 496 papers. After a thorough assessment and screening, 47 articles remained, which were critically analyzed. The interrater agreement was measured using a Cohen kappa test, and disagreements were resolved through discussion. The state-of-the-art classes of machine learning have been developed and tested up to the task and achieved promising performance including artificial neural network, support vector machine, decision tree, genetic algorithm, Gaussian process regression, Bayesian neural network, deep belief network, and others.

Conclusions: Despite the complexity of BG dynamics, there are many attempts to capture hypoglycemia and hyperglycemia incidences and the extent of an individual's GV using different approaches. Recently, the advancement of diabetes technologies and continuous accumulation of self-collected health data have paved the way for popularity of machine learning in these tasks. According to the review, most of the identified studies used a theoretical threshold, which suffers from inter- and intrapatient variation. Therefore, future studies should consider the difference among patients and also track its temporal change over time. Moreover, studies should also give more emphasis on the types of inputs used and their associated time lag. Generally, we foresee that these developments might encourage researchers to further develop and test these systems on a large-scale basis.

(J Med Internet Res 2019;21(5):e11030) doi:10.2196/11030

KEYWORDS

type 1 diabetes; blood glucose dynamics; anomalies detection; machine learning

Introduction

Background

Diabetes mellitus is a chronic metabolic disorder that results in abnormal blood glucose (BG) regulation. The BG level is maintained close to normality through self-management practices, which involves actively tracking BG levels and taking proper actions including diet and insulin medications. The estimated number of people with diabetes aged between 20 and 79 years was 415 million (uncertainty interval: 340-536 million) in 2015 and is expected to reach 642 million (uncertainty interval: 521-829 million) by 2040 [1]. The global economic burden of diabetes in adults aged between 20 and 79 years was estimated to be US $1.31 trillion (95% CI 1.28-1.36) in 2015 [2]. The total number of deaths attributed to diabetes is estimated to be 5 million in people with diabetes aged between 20 and 79 years [1]. People with diabetes have a higher risk of getting infections as compared with the normal population, which potentially increases their morbidity and mortality [3]. The greater and frequent risk of infections is mainly correlated with a hyperglycemia environment [3,4]. Moreover, studies suggest a hypoglycemia episode could result in a higher hospitalization and mortality rate [5].

The individual's BG dynamic is affected by various factors, which are mainly categorized as common, individual, and unpredictable factors [6]. The common factors include amount of food intake, insulin intake, previous level of BG, pregnancy, drug and vitamin intake, smoking, and alcohol intake. The individual factors include dawn phenomena, physical exercise load, and menstruation. The unpredictable factors include stress, concomitant diseases, and infections [6]. Swings in BG dynamics, that is, hypoglycemia and hyperglycemia, could be generally categorized under a normal cause variation and special cause variation. The normal cause variation is regarded as caused by those common and individual factors, whereas the special cause variation is caused by those unpredictable factors. The underlying reason of the special cause variations is difficult to understand and remains a challenge for the patient during the incidences. For instance, during stress and infections, the patient usually struggles with hyperglycemia and injects frequent insulin to lower his or her BG levels.

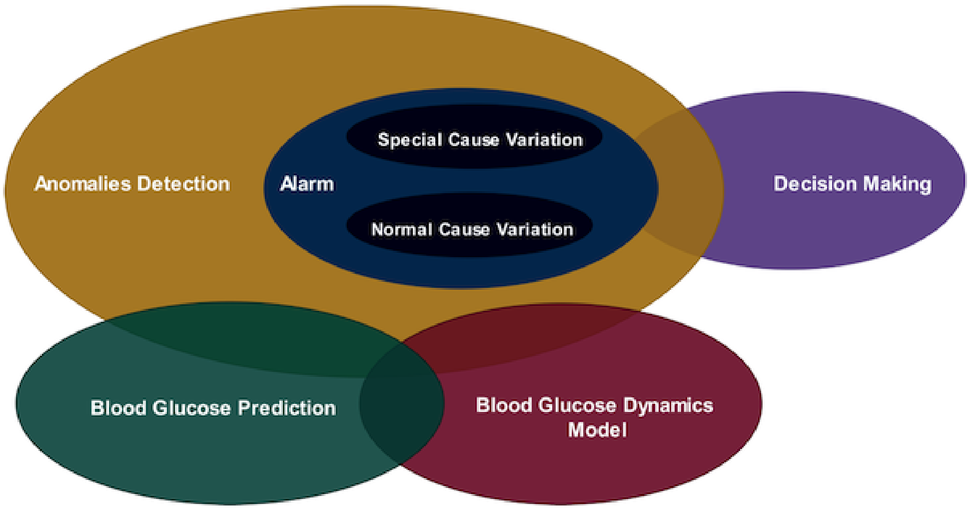
BG anomalies could be defined as any undesirable reading because of either a precisely known reason (normal cause variation) or an unknown reason (special cause variation) to the patient [7]. Even if the advancement in self-management applications and diabetes monitoring technologies has made things easier, the challenge of BG anomalies remains to be managed by the patient themselves. There are some technological developments in the direction of personalized decision systems and BG event alarms to provide an alert and decision support to the patient in the time of these challenges. Techniques such as classification and detection of glycemic variability (GV), hypoglycemia, and hyperglycemia, in particular, and BG anomalies, in general, are central to the development of these diabetes technologies. The ubiquitous nature and widespread use of mobile health (mHealth) apps, sensors and wearables, and other point-of-care (POC) devices for self-monitoring and management purposes have made possible the generation of automated and continuous diabetes-related data, which brought an opportunity for the introduction of machine learning and its application for intelligent and improved systems, which is capable of solving complex tasks within a dynamic knowledge and dynamic environment. In this regard, there are some reviews conducted toward the applications of artificial intelligence in diabetes-related tasks. For instance, Contreras et al [8] conducted literature reviews on the applications of artificial intelligence in the context of critical diabetes management issues such as BG prediction and strategies for BG control, adverse glycemic events detection, bolus calculators and advisory system, patient personalization (tailored features), and others [8]. Moreover, Rigla et al [9] also conducted a review to provide a general overview and popularity of artificial intelligence applications to diabetes problems. Generally, both Contreras et al [8] and Rigla et al [9] tried to demonstrate the potential of artificial intelligence with regard to all groups of people with diabetes focusing on general self-management issues. As far as our knowledge is concerned, there are almost no reviews conducted toward techniques of BG anomaly classification and detection focusing on various approaches, in general, and machine-learning applications, in particular. However, there were some reviews conducted to evaluate the significant effect of pattern management based on self-monitoring BG (SMBG) with regard to clinical practices [10]. Therefore, we suggest that there is a lack of reviews focusing on BG anomaly classification and detection. The objective of this review was to identify, assess, and analyze the state-of-the-art machine-learning strategies in BG anomaly classification and detection including GV, hyperglycemia, and hypoglycemia in people with type 1 diabetes. Moreover, it has presented the current modeling options of machine-learning applications and their hybrid systems. The review covers machine-learning approaches pertinent to personalized decision support systems and BG alarm events applications in type 1 diabetes.

Machine Learning Tasks in Type 1 Diabetes

Machine-learning approaches (tasks) are generally categorized as regression, prediction, classification, detection, and clustering, which are grouped either in supervised, semisupervised, unsupervised, or reinforcement learning based on the type of learning employed. Generally, reinforcement learning is out of the scope of this review, where we mainly focus on the other 3 categories. Machine learning–based data mining tasks could be categorized as descriptive or unsupervised (ie, clustering, association, and summarization), semisupervised (ie, classification and detection), and predictive or supervised learning (ie, classification and regression) [11]. In this regard, the most widely used machine learning–based data mining tasks in the literature are BG anomalies detection, BG prediction, modeling of BG dynamics, and decision making or education, as shown in Figure 1. In this review, we will focus on the typical applications of classification and detection tasks in diabetes research, specifically in BG anomaly detection within the context of a personalized decision support system and BG alarm events applications. The review considers various classes of machine learning algorithms: artificial neural network (ANN), decision trees (DTs), support vector machine (SVM), evolutionary algorithms (EAs), and others.

An ANN is a biologically inspired computational model consisting of a set of interconnected neurons and a scaled connection between them that is called weights [12]. On the basis of network topology, an ANN is mainly categorized as a feedforward ANN (single-layer perceptron (SLP), multi-layer perceptron (MLP), and radial basis function [RBF]) and feedback ANN (recurrent neural network [RNN], Elman net, Kohonen's self-organizing map (SOM), and Hopfield networks) [12]. The SVM works based on the theory of structural risk minimization principle [13]. Learning in the SVM occurs through finding an optimal hyperplane that can maximize the margin between the classes. The SVM has been widely exploited in numerous applications such as regression and prediction, pattern identification and recognition, categorization, and classification [13]. An EA is a biologically inspired approach to problem solving [14]. The 2 most used variants of EA are genetic programming (GP) and genetic algorithm (GA). Random forest (RF) or DTs are a kind of an ensemble approach of learning for different classification and regression applications, which mainly learns by constructing a multitude of DTs generating the mode of the class or mean of prediction. The hidden Markov model (HMM) is a variant of the statistical Markov model, where the system being modeled is assumed to follow a Markov property with unobserved states [15]. There are various versions of HMMs; however, in this review, we considered only those trained with a framework close to machine learning families. Hybridization is the process of combining 2 or more different approaches in parallel or serious connection, either at the preprocessing stage, feature extraction, or learning stage, when looking for an improved performance [16].

Figure 1. Most widely used machine learning–based data mining tasks based on self-recorded data in people with type 1 diabetes. The yellow shaded ellipse depicts the scope of this review.



Blood Glucose Anomaly Classification and Detection

Hawkins defined anomalies as "observations that deviate much from the other observations so as to arouse suspicions that it could be generated by a different process" [7,17,18]. There are terms that are often used interchangeably with anomalies, such as outliers, deviations, exceptions, rare instances, and irregularities. The problem of identifying and capturing anomalies in data can be supervised, semisupervised, and unsupervised tasks [19,20]. These strategies can roughly be categorized as classifier- or model-based (detection) approach. The semisupervised is better when anomalous instances are not easily available, whereas supervised techniques are more suitable when there are sufficient labeled instances of both normal and anomalous instances. The unsupervised approach does not require any reference data labels, where normal behaviors have to be determined dynamically, and the detections are mainly performed with regard to the entire datasets. The model-based strategies can be considered as a diagnosis of the system's behavior during abnormal situations through modeling and adequately characterizing the system's behavior during normal situations [19,21]. It uses a system's model to either estimate or predict the underlying system (process) dynamics to capture anomalies in the data. The most important design requirement in using a model includes discovering and characterizing what is to be considered a normal pattern of behaviors [22]. Unlike the classifier-based strategies, the model-based strategies do not require rigorous knowledge of the underlying expected anomalies, that is, to fully understand and characterize the shape and nature of the expected anomalies [22]. By simply defining what is the expected normal pattern the system should exhibit, the model-based anomaly detection is capable of detecting abnormal behavior, which is not considered as the normal behavior of the system. Defining and discovering what is *normal* is a challenging task especially for dynamic and complex systems, for example, BG dynamics. However, this is often tackled in a dynamic and complex system by relying on either a machine learning model trained on a large enough dataset or using an explicit mathematical model, for example, physiological model of BG dynamics, of the system if it exists already.

BG readings are time series data, and anomalies in BG levels could be regarded as any undesirable readings, as shown in Figure 1, because of either a predictable cause (normal cause variation) or an unpredictable cause (special cause variation). A normal cause variation could be defined as any hypoglycemia or hyperglycemia incidences with the underlying cause known to the patient herself or himself and also referred as predictable (patient controllable) factors such as insulin injection, diet intake, physical activity, and others. However, special cause variation refers to any hypoglycemia or hyperglycemia incidences with the underlying cause unknown to the patient and also called unpredictable (patient uncontrollable) factors such as stress, infections, insulin set failure, and others. The classifier, semisupervised (model)– and unsupervised-based approach could be used to solve the challenge of capturing BG anomalies caused by both the predictable factors (normal cause variation) and unpredictable cause (special cause variation). However, regarding the unpredictable factors (special cause variation), the classifier-based approach remains to be very challenging with limited feasibility as the classifier-based strategies require a thorough understanding and characterization of the nature, size, and shape of the anomalies, along with its inter- and intravariability among the patients. With the same token, the unsupervised approach could face the same challenge as it does not have any mechanisms for differentiating the one with special cause from the normal cause variations. However, the model-based (semisupervised) approach happens to be more appropriate given that it only requires to characterize what is considered to be normal so as to detect what is believed to be abnormal. For example, infection (stress)–related hyperglycemia and a diet-induced hyperglycemia are treated differently according to the model-based (semisupervised) anomaly detection strategies. In this regard, diet-induced hyperglycemia is treated as normal, as the model could describe the underlying cause (certain meal), but infection-related hyperglycemia is considered as an anomaly because the model cannot describe the underlying cause based on patient controllable variables (eg, meals and insulin).

GV measures the degree or the rate at which the patient's BG fluctuates between high and low levels [23]. GV is useful to provide all-inclusive information on one's self-management practices concerning postprandial spikes in BG, as well as episodes of hypoglycemic and hyperglycemic events [23,24], which are the main factors that contribute for a higher risk of cardiovascular events in people with diabetes. The evaluation of GV helps to comprehend and assess the effect of the patient's timely actions on the hypoglycemia and hyperglycemia incidence by associating out-of-target BG levels with patient-specific factors, such as insulin dosage, other medication, meals, activity, stress, and illness [23]. However, there is no gold standard approach for assessing GV, and despite its importance, it remains to be challenging.

Blood Glucose Prediction

BG prediction is about forecasting an individual's future BG levels using current and past information and is also an important constituent of BG anomaly classification and detection approaches. It mainly aims to provide crucial alarms for patients in advance with sufficient lead time so as to avoid further complications from hypoglycemia or hyperglycemia incidences. According to Oviedo et al [25], BG prediction models could be categorized into 3 main groups: physiological models, data-driven models, and hybrid models [25]. These categories are solely demarcated based on the necessity of extensive knowledge of the underlying BG dynamics: black box approach (data-driven model), intermediate knowledge (hybrid model), and extensive knowledge (physiological model). The data-driven model, which is mainly referred to as black box model, uses the patient's continuous glucose monitoring (CGM), insulin, dietary, and other relevant information to develop a prediction model, for example, machine learning and time series approaches. There are a variety of data-driven models developed and tested in the literature including machine learning (neural network, support vector regression, jump neural network, RNN, and others) and time series models (autoregressive [AR] with exogenous input, AR moving average with exogenous input, AR moving average, and others) [25]. Hybrid models make use of the advantages from the data-driven and physiological models [25]. Most of the hybrid models rely on the physiological model to compute meal and insulin information as input for the data-driven models [25]. Physiological models mainly rely on 3 sets of mathematical (differential) equations to describe the underlying dynamics: BG dynamics, insulin dynamics, and meal absorption dynamics. Physiological models are roughly grouped into lumped and comprehensive models based on the way the model treats each organ and tissue so as to develop the differential equations [26]. There are a variety of physiological models developed in the literature such as Berger, Hovorka, Cobelli, Lehmann and Deutsch model, and others [26]. Generally, there are plenty of models implemented in the literature on the prediction of BG levels [25,26]. However, BG prediction is not under the scope of this review, and we mainly focus on the data-driven BG pattern classification and anomaly detection approaches under the umbrella of machine learning.

Methods

Search Strategy

The objective of this review was to identify, assess, and analyze the state-of-the-art machine learning strategies and their hybrid system focusing on BG anomaly classification and detection including GV, hyperglycemia, and hypoglycemia in people with type 1 diabetes. The review covers machine learning approaches pertinent to personalized decision support systems and BG alarm events applications. Therefore, for the purpose of the study, a rigorous literature search was conducted between September 1 and October 1, 2017, through various Web-based databases including Google scholar, IEEE Xplore, DBLP Computer Science Bibliography, ScienceDirect, PubMed or Medline, Journal of Diabetes Science and Technology, and Diabetes Technology & Therapeutics. Additional search was also conducted between October 15 and November 5, 2018, on those databases to refine and update the records. Furthermore, the reference list of the selected articles was used to extract additional articles to get a complete overview of the field.

Peer-reviewed journals and articles published between 2000 and 2018 were considered. The inclusion and exclusion criteria were setup through rigorous discussion and brainstorming among the authors. Different combinations of terms such as *diabetes*, *intelligent system*, *hybrid system*, *machine learning*, *BG event indicators (hypo- and hyperglycemia prediction)*, *BG event alarm*, *BG personalized decision system*, *clinical*, *closed-loop system*, *hyperglycemia*, *hypoglycemia*, *GV*, and *personalized profile* were used during the search. The terms were combined using *AND/OR* for a better search strategy. Relevant articles were first identified by reviewing the title, keywords, and abstracts for a preliminary filter with our selection criteria, and then we reviewed full text articles that seemed relevant. Information from the selected literature was extracted based on some predefined categories, which were based on previous research, and further elaborated through brainstorming.

Inclusion and Exclusion Criteria

To be included in the review, the studies should have developed, implemented, tested, and discussed machine learning and any of its hybrid approaches in type 1 diabetes focusing on one or more of the following application areas:

- BG anomaly detection
- Hypoglycemia prediction, classification, or detection
- Hyperglycemia prediction, classification, or detection
- Glycemic or BG variability classification or detection

Therefore, the studies that reside outside of these stated scopes were excluded from the review including all articles written in other languages but English.

Data Categorization and Data Collection

Information was extracted from the selected studies based on predefined parameters (variables) and categories. The categories were defined based on rigorous brainstorming and discussion among the authors. These categories were demarcated solely to collect the relevant data and to assess, analyze, and evaluate the model’s characteristics and its experimental setup.

Application Scenario

This category defines the type of applications where the machine learning algorithm is being exploited. It can be hypoglycemia and hyperglycemia prediction, classification and detection, or GV classification and detection.

Type of Input

This category was defined to assess, analyze, and evaluate the type of inputs used to develop the algorithm. This includes the key diabetes parameters and other physiological parameters relevant for BG anomaly classification and detections: BG, heart rate variability, and others.

Data Format, Type, Size, and Data Source

This category was defined to assess, analyze, and evaluate the type of data format used as input to the algorithm. This depends on the basis of the type of diabetes technologies, mobile apps, and POC devices used for data collection and algorithm development. It includes different data formats such as from CGM devices, mHealth apps (ie, diabetes diary), heart rate monitoring devices, and others.

Input Preprocessing

This category defines the kind of preprocessing algorithm the system implements so as to avoid missing, sparse, and corrupted input data.

Class of Machine Learning

This category defines the class of machine learning algorithm used to train and test the BG anomaly classification and detection algorithm. It includes different classes of machine learning algorithm: ANN, SVM, Bayesian network, DT, and others.

Training or Learning Method and Algorithm

This category defines the class of learning algorithms used to train the model. It includes different training algorithms such as the backpropagation algorithm, kernel, optimization techniques, and others.

Performance Metrics or Evaluation Criteria

This category defines the type of evaluation metrics used to assess the accuracy of the classification and detection algorithm implemented. It includes different performance metrics such as specificity, sensitivity, receiver operating characteristic (ROC) curves, and others.

Literature Evaluation

The included literature was analyzed and evaluated based on the above defined categories and variables to uncover the state-of-the-art machine learning applications in hyperglycemia or hypoglycemia prediction, classification and detection, and GV classification and detection. It also tries to pinpoint their characteristics along with the experimental setup used to implement and test the algorithms. The first evaluation and analysis was carried out based on the type of input used to develop the algorithms to uncover the state-of-the-art inputs used in these circumstances. The second evaluation and analysis was carried out based on the various classes of machine learning used to develop these algorithms to uncover the rate of adoption and their suitability to the task. The third evaluation and analysis was carried out based on the performance metrics used to evaluate the performance of these algorithms.

Results

Relevant Literature

The initial hit was vetted using the title, abstract, and keywords and retrieved a total of 496 papers (DBLP Computer Science (20), Diabetes Technology & Therapeutics (23), Google Scholar (160), IEEE (215), Journal of Diabetes Science and Technology (22), PubMed Medlin (27), and ScienceDirect (29); see Figure 2). After removing duplicates from the list, 410 records remained. Then, we did an independent assessment of the articles and screening based on the inclusion and exclusion criteria, which eliminated another 215 papers, leaving 195 relevant papers. After a full-text assessment, 47 articles were left (hyperglycemia=5, glycemic variabilities=3, and hypoglycemia=39), which were critically analyzed as shown in

Figures 2 and 3. The interrater agreement was measured using a Cohen kappa test, and disagreements were resolved through discussion.

Figure 2. Flow diagram of the review process.

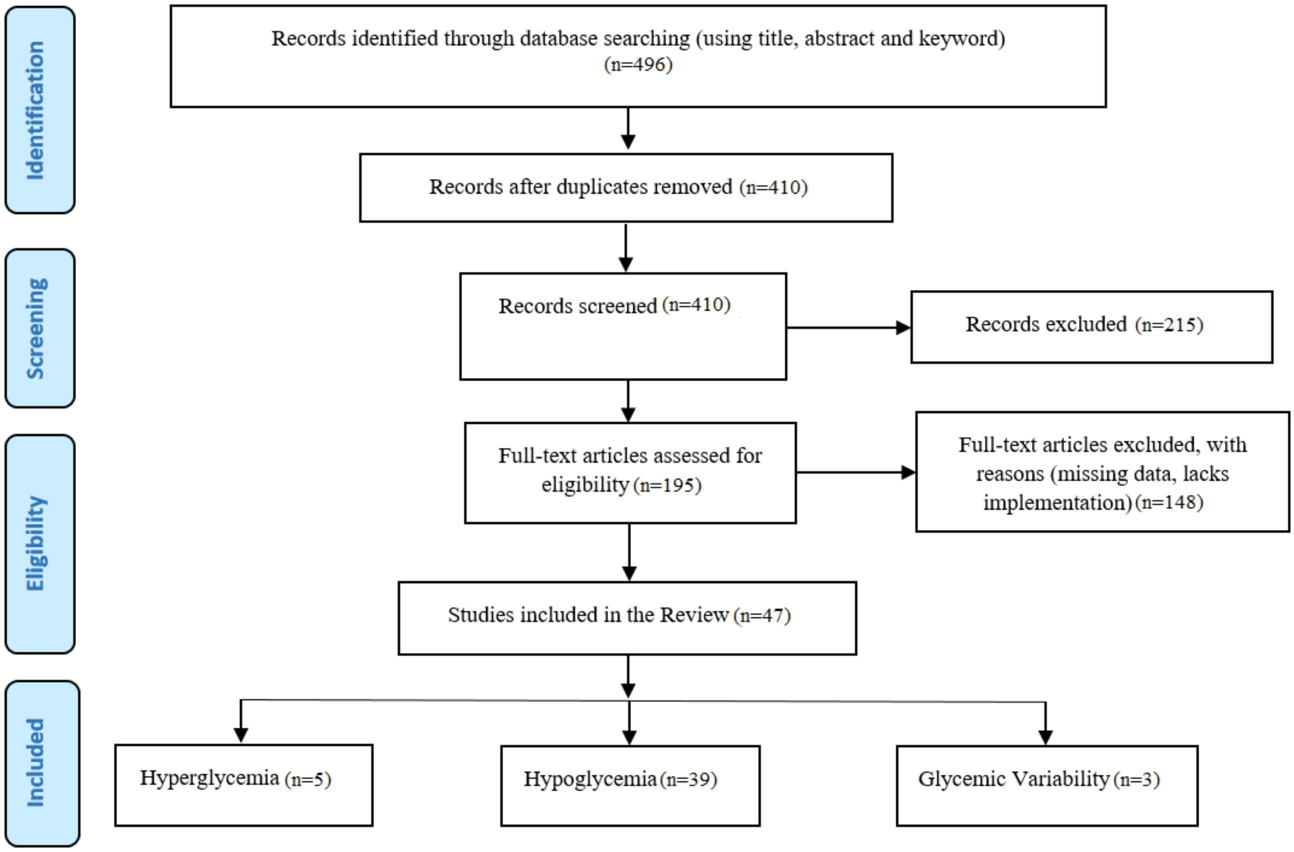
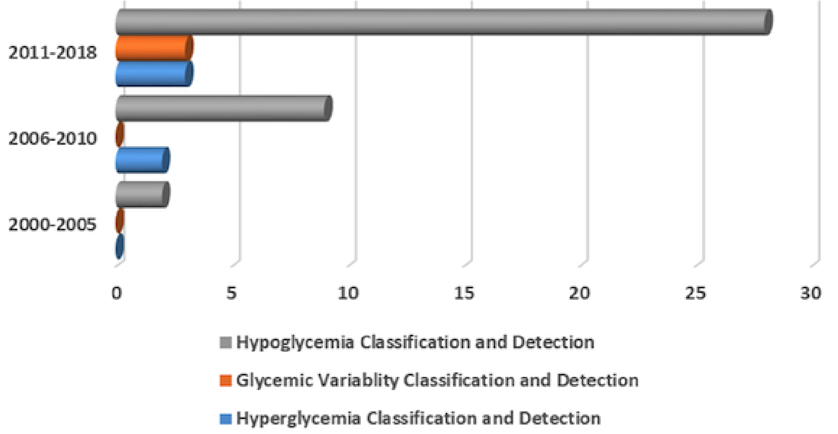


Figure 3. The number of articles published per year of publication.



Evaluation and Analysis of the Literature

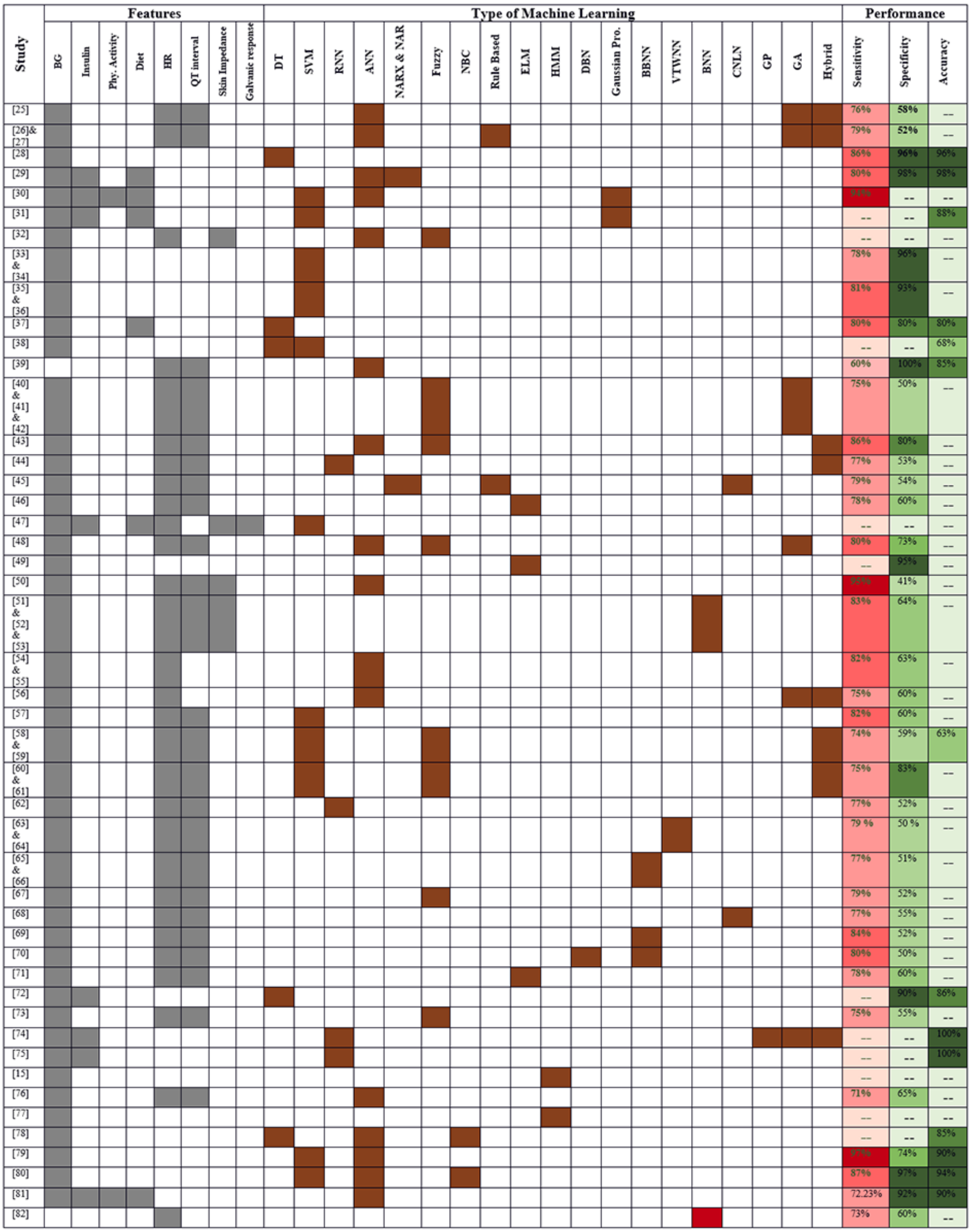
The literature, as described previously, was evaluated based on the type of machine learning used to develop the algorithm, the type of input used to train the system, and the performance metrics used to evaluate the algorithm performance based on the tables in Multimedia Appendices 1 and 2.

Data Characteristics and Input Parameters

Input Parameters

Selecting the proper types of input parameters is one of the crucial design strategies for successful classification and detection algorithm development. In this regard, the outer bigger ring, the middle ring, and the inner ring in Multimedia Appendix 1 depict the types of input used in hypoglycemia, hyperglycemia, and GV classification and detection algorithm, respectively. According to hypoglycemia classification and detection algorithm, BG, heart rate, and QT interval are the most used types of input parameters (25/39, 64%). BG alone is the second most used type of input parameter (4/39, 10%). BG and insulin are the third most used types of input parameters along with BG, insulin, diet, physical activity, and others (3/39, 8%). BG and diet alone, along with BG, insulin, and diet, and BG, heart rate, skin impedance, and BG, insulin, diet, heart rate, galvanic response, skin impedance are the fifth most used types of input parameters (1/39, 3%). According to hyperglycemia classification and detection algorithm, BG alone, and BG and insulin represent the most used types of input parameters (2/5, 40%). BG, heart rate, and QT interval represent the second most used types of input parameters (1/5, 20%). According to GV classification and detection algorithm, BG alone (3/6, 50%), and BG and insulin (3/6, 50%) are equally ranked as the most used types of input parameters, as shown in Figure 4.

Figure 4. Reported input features, machine learning class, and accuracy. ANN: artificial neural network; BBNN: block-based neural network; BG: blood glucose; BNN: Bayesian Neural Network; DBN: deep belief network; DT: decision tree; ELM: extreme learning machine; GA: genetic algorithm; GP: genetic programming; HMM: hidden Markov model; NAR: nonlinear autoregressive network; NARX: nonlinear autoregressive network with exogenous inputs; NBC: Naive Bayes classifier; RNN: recurrent neural network; SVM: support vector machine; VTWNN: variable translation wavelet neural network.



Study	Features								Type of Machine Learning																			Performance		
	BG	Insulin	Phy. Activity	Diet	HR	QT interval	Skin Impedance	Galvanic response	DT	SVM	RNN	ANN	NARX & NAR	Fuzzy	NBC	Rule Based	ELM	HMM	DBN	Gaussian Pro.	BBNN	VTWNN	BNN	CNLN	GP	GA	Hybrid	Sensitivity	Specificity	Accuracy
[25]	■				■	■						■														■	■	76%	58%	--
[26]& [27]	■				■	■						■				■										■	■	79%	52%	--
[28]	■								■																			86%	96%	96%
[29]	■	■		■								■	■															80%	98%	98%
[30]	■	■	■	■						■		■								■								94%	--	--
[31]	■	■		■						■										■								--	--	88%
[32]	■				■		■					■		■														--	--	--
[33] & [34]	■									■																		78%	96%	--
[35] & [36]	■									■																		81%	93%	--
[37]	■			■					■																			80%	80%	80%
[38]	■								■	■																		--	--	68%
[39]					■	■						■																60%	100%	85%
[40] & [41] & [42]	■				■	■								■												■		75%	50%	--
[43]	■				■	■						■		■													■	86%	80%	--
[44]	■				■	■					■																■	77%	53%	--
[45]	■				■	■							■			■								■				79%	54%	--
[46]	■				■	■											■											78%	60%	--
[47]	■	■		■	■		■	■		■																		--	--	--
[48]	■				■	■						■		■												■		80%	73%	--
[49]	■																■											--	95%	--
[50]	■				■	■	■					■																95%	41%	--
[51] & [52] & [53]	■				■	■	■																■					83%	64%	--
[54] & [55]	■				■							■																82%	63%	--
[56]	■				■							■														■	■	75%	60%	--
[57]	■				■	■				■																		82%	60%	--
[58] & [59]	■				■	■				■				■													■	74%	59%	63%
[60] & [61]	■				■	■				■				■													■	75%	83%	--
[62]	■				■	■					■																	77%	52%	--
[63] & [64]	■				■	■																■						79 %	50 %	--
[65] & [66]	■				■	■															■							77%	51%	--
[67]	■				■	■								■														79%	52%	--
[68]	■				■	■																		■				77%	55%	--
[69]	■				■	■															■							84%	52%	--
[70]	■				■	■													■		■							80%	50%	--
[71]	■				■	■											■											78%	60%	--
[72]	■	■							■																			--	90%	86%
[73]	■				■	■								■														75%	55%	--
[74]	■	■									■														■	■	■	--	--	100%
[75]	■	■									■																	--	--	100%
[15]	■																	■										--	--	--
[76]	■				■	■						■																71%	65%	--
[77]	■																	■										--	--	--
[78]	■								■			■			■													--	--	85%
[79]	■									■		■																97%	74%	90%
[80]	■									■		■			■													87%	97%	94%
[81]	■	■	■	■								■																72.23%	92%	90%
[82]					■																		■					73%	60%	--

Data Characteristics

Data Sources

Different kinds of data sources ranging from BG monitors, physical activity, electrocardiogram (ECG), and heart rate sensors have been used in the reviewed articles for hyperglycemia, hypoglycemia, and GV classification and detection algorithms. The reviewed articles relied on different kinds of data formats including SMBG (finger sticks), CGM, and ECG signals, as shown in Table 1. Generally, ECG signal is the most used type of data format (51%), followed by CGM (39%) and SMBG (10%). Specifically, hypoglycemia classification and detection involve (CGM (n=11), ECG (n=24), and SMBG (n=5)). Regarding, hyperglycemia classification and detection (CGM (n=5) and ECG (n=1)) and GV classification and detection (CGM (n=3)).

Table 1. Types of data formats used in the studies (N=49).

Data type/format	Count, n (%)
Continuous glucose monitoring	19 (39)
Self-monitoring blood glucose	5 (10)
Electrocardiogram signal	25 (51)

With regard to BG monitoring, different devices and brands have been exploited for developing hypo-/hyperglycemia and GV classification and detection algorithms, as shown in Table 2. Generally, Yellow Spring Instruments is the most used device (50%) followed by Guardian Real Time (MinMed CGM; 28%). GlycoMark (7%) is the third most used device followed by HemoCue Glucose 201 (5%) and Self-Monitored BG (5%). Specifically, the most used devices for hypoglycemia classification and detection are Guardian Real Time (MinMed, CGM; n=7), Yellow Spring Instruments (n=21), HemoCue Glucose 201 (n=2), Dexcom CGM system (n=1), Self-Monitored BG (SMBG; n=2), Medtronic Enlite CGM sensors (n=1), Medtronic insulin pump (n=4), SensorWear armband (physical activity; n=2), and Basis Peak fitness band (n=1). and Basis Peak fitness band (n=1). As for hyperglycemia classification and detection, Guardian Real Time (MinMed CGM; n=2) and Medtronic insulin pump (n=3) had been used. With regard to GV classification and detection, GlycoMark (n=3), Guardian Real Time (MinMed CGM; n=3), and Medtronic insulin pump (n=3) had been used.

Table 2. Types of devices used for the monitoring of blood glucose levels (N=42).

Devices	Count, n (%)
Guardian Real Time (MinMed, CGM[a])	12 (28)
HemoCue Glucose 201 (HemoCue)	2 (5)
Yellow Spring Instruments	21 (50)
Dexcom CGM system	1 (3)
Medtronic Enlite CGM sensors	1 (3)
GlycoMark	3 (7)
Self-Monitored Blood Glucose-unknown device	2 (5)

[a] CGM: continuous glucose monitoring.

Various brands of physiological monitoring (heart rate and ECG signals) devices have been exploited in the reviewed articles. Generally, as shown in Table 3, Compumedics system is the most used system (52%) followed by a customized device such as a battery-powered chest belt–worn device (22%). HypoMon is the third most used device (13%) followed by Basis Peak fitness band (9%) and a self-designed portable apparatus (4%). Specifically, for hypoglycemia classification and detection purposes, various devices have been used such as HypoMon (n=3), Basis Peak (n=2), Compumedics system (n=11), a battery-powered chest belt–worn (n=5), and self-designed portable apparatus (n=1). With regard to hyperglycemia classification and detection, only 1 article has used the Compumedics system (n=1), which indicates that heart rate and ECG signals have a limited use in this case.

Table 3. Types of devices used for the monitoring of physiological parameters (heart rate and electrocardiogram signals; N=23).

Devices	Count, n (%)
HypoMon	3 (13)
Basis peak fitness band	2 (9)
Compumedics system	12 (52)
A battery-powered chest belt–worn (customized)	5 (22)
Self-designed portable apparatus (customized)	1 (4)

Data Preprocessing

Data preprocessing is an important stage of any machine learning strategy. In this regard, there were various kinds of data preprocessing strategies used in the reviewed articles. The reviewed articles had relied on both BG and other physiological (heart rate, ECG, skin impedance, and others) data, which of course involves different preprocessing strategies depending on the data type under consideration. Regarding the BG data, various preprocessing approaches had been used including differencing (derivative) BG values [27,28], CGM data reconstruction, or smoothing using different methods such as spline interpolation [29-33], a rough feature elimination, such as fast *separability* and correlation analysis algorithm [28,29], representing BG temporal change information [34], feature selection and feature ranking [35], filtering using Pearson's correlation coefficient (PCC) and the *t* test, and the wrapper approach using greedy backward elimination [33]. The other physiological parameters (heart rate, ECG, skin impedance, and others) had been preprocessed using different methods such as normalization [36-38], feature extraction and selection [39,40], feature extraction using fast Fourier transform (FFT) [41], unsupervised restricted Boltzmann machine–based feature representation [42], filtering techniques such as Infinite impulse response high pass filter [41,43], correlation analysis [44-46], and transformation of frequency domain into time domain (FFT) [47].

Class of Machine Learning

Hypoglycemia Classification and Detection

Different classes of machine learning techniques have been adopted in hypoglycemia prediction, classification, and detection algorithms to predict, classify, and detect the incoming hypoglycemia incident in people with type 1 diabetes, as shown in Figure 4. Conventional feedforward ANN is the most adopted class of machine learning, which is used in 26% (17/65) of the studies , as shown in Multimedia Appendix 1. Hybridization of machine learning techniques with other approaches such as time series, fuzzy logic, and others are the second most adopted approach (12/65, 18%). The SVM ranked the third most adopted class of machine learning (9/65, 14%). DT ranked the fourth most adopted technique (4/65, 6%). GA, time delay ANN and time sensitive ANN, block-based neural network (BBNN), and adaptive neural fuzzy inference system (ANFIS) are the fifth most used classes of machine learning (3/65, 5%). Nonlinear autoregressive network with exogenous inputs (NARX) and nonlinear autoregressive network (NAR) along with Gaussian process regression, combinational neural logic network , and Bayesian neural network (BNN) ranked as the sixth most used classes of machine learning (2/65, 3%). Deep belief network (DBN), radial basis function neural network (RBFNN), and variable translation wavelet neural network (VTWNN) are the seventh most used classes of machine learning (1/65, 2%).

Hyperglycemia Classification and Detection

Hyperglycemia classification, prediction, and detection has been practiced less when compared with hypoglycemia, which might be linked because of its less severe short-term complications as opposed to hypoglycemia incidences. However, irrespective of this limitation, different types of machine learning techniques have been adopted, as shown in Figure 4. For example, ANN is the most used machine learning technique in 34% (3/9) of the studies (feedforward (1/9) and feedback RNN (2/9)), as shown in Multimedia Appendix 1 along with EA (3/9,34%) (GA (1/9) and GP (2/9)). The HMM (2/9, 22%) is the third most used followed by a hybrid approach (1/9, 11%).

Glycemic Variability Classification and Detection

GV detection is a recent development, which has great importance in quantifying factors associated with hypo-/hyperglycemia incidence. In this regard, there is some research and development involving machine learning techniques, as shown in Figure 4. For example, feedforward ANN is the most used class of machine learning (3/8, 37%), as shown in Multimedia Appendix 1. Naive Bayes classifier (NBC) and SVM are the second most adopted techniques of machine learning (2/8, 25%). DT is the third most used class of machine learning (1/8, 13%).

Performance Metrics

The performance metrics used in the evaluation of hypoglycemia, hyperglycemia, and GV classification and detection algorithms are depicted in the outer ring, the middle ring, and inner ring, respectively, as shown in Multimedia Appendix 1. According to hypoglycemia classification and detection, sensitivity, and specificity are the most used performance metrics (37/58, 64%). Accuracy and precision are the second most used performance metrics (9/58, 15%). Root mean square error and mean square error are the third most used performance metrics (4/58, 7%). Geometric mean is the fourth most used performance metric (3/58, 5%). Correlation coefficient is the fifth most used performance metric (2/58, 3%). Time lag (TL), recall, and ROC curve are the sixth most used performance metrics (1/58, 2%). According to hyperglycemia classification and detection, accuracy and precision, root mean square error and mean square error, time lag (TL), correlation coefficient, recall, and false positive rate are the most used performance metrics (2/15, 13%). ROC curve, geometric mean, sensitivity, and specificity are the third most used performance metrics (1/15, 7%). According to GV classification and detection, accuracy, and precision are the most used performance metrics (3/5, 60%). Sensitivity and specificity are the second most used performance metrics (2/5, 40%).

Discussion

Principal Findings

The objective of this review was to identify, assess, and analyze the state-of-the-art machine applications in BG pattern classifications and anomaly detection: hyperglycemia, hypoglycemia, and GV classification and detection. According to the reviewed literature, the anomaly classification and detection approach could be roughly categorized as either a classifier-based or a model-based approach [19,21]. The classifier-based approach mainly relies on using either a specified threshold or some kinds of rules to classify the BG levels as either normal or abnormal. The difference is that unlike the model-based approach, the classifier-based approach requires rigorous and deeper knowledge regarding the nature, size, and shape of the underlying anomalies under consideration so as to develop the necessary threshold or rule to capture them. However, the model-based approach only requires to demarcate the boundary of what is known to be normal so as to capture what is believed to be abnormal [21]. The model-based approach does not require rigorous knowledge of the underlying expected anomalies, that is, to fully understand and characterize the shape and nature of the expected anomalies [22]. By simply defining what is the expected normal pattern that the system should exhibit, a model-based approach is capable of detecting abnormal behavior, which is not considered as the normal behavior of the system. Defining and discovering what is *normal* is a challenging task especially for dynamic and complex systems, for example, BG dynamics. However, this is often tackled in a dynamic and complex system by either relying on a machine learning model trained on large enough datasets or using an explicit mathematical model of the system such as a physiological or compartmental BG dynamics model [21].

Various classes of machine learning algorithms have been adopted for the task. Regarding hypoglycemia classification and detection, feedforward ANN, hybrid systems, SVM, DT, GA, ANFIS, NARX, NAR, Gaussian process regression, DBN, and BNN have been developed and tested. These techniques have explored various kinds of input parameters notably BG, heart rate, QT interval, insulin, diet, physical activity, galvanic response, and skin impedance. Concerning hyperglycemia classification and detection, RNN, GP, HMM, feedforward ANN, GA, and hybrid systems have been developed and tested, exploring various types of input parameters including BG, insulin, heart rate, and QT interval. GV detection is a recent development, which has great importance in quantifying factors associated with hypoglycemia and hyperglycemia incidence. In this regard, there is some research and development involving machine learning techniques. For example, feedforward ANN, NBC, SVM, and DT have been tested up to the task using BG and insulin delivery profiles.

Generally, all of the studies have relied on either indirect indicator variables such as heart rate, QT interval, and others or a subset of input parameters that affects BG dynamics. The patient’s contextual information, for example, meals, physical activity, insulin, and sleep, have a significant effect on BG dynamics, and a proper anomaly classification and detection algorithm should consider the effects of these parameters. In this regard, however, the individual patient is expected to record meal, insulin, and physical activity data. One of the main limitations is meal modeling, where most of the algorithm depends on the individual estimation of carbohydrate, which is prone to errors and further aggravates the degradation to detection performance. With regard to physical activity, there are various wearables and sensors that can record the individual’s physical activity load and durations. However, there is the lack of a uniform approach among the studies with certain limitations on the way these signals are employed in the classification and detection algorithms. For example, there are some studies that consider levels of activity as low, moderate, and high and others consider descriptive features by summarizing the number, intensity, steps, exercise durations, and others to better quantify the effect of physical activities. Moreover, recording insulin dosage has its inherent limitations, which might affect the detection performance. For example, blockage of insulin flow from the insulin pump because of the infusion set failure and error incurred during manual registrations might pose a significant challenge in the performance of the detection system. Furthermore, CGM is becoming one of the most important components in these classification and detection algorithms. However, even if CGM advancement has enabled patients to have continuous estimation of their subcutaneous glucose levels, it has limitations when used in a personalized detection system (an alarm). In this regard, recent studies have showed that autocorrelation of the CGM reading vanishes after 30 min, making the detection performance to degrade afterward. These findings suggest that any classification and detection algorithms aiming for a better lead time should consider other patient’s contextual information and various features of the CGM itself. There are some studies that develop a model by assessing several features of the CGM signal so as to compensate for its inaccuracy. Moreover, CGM is found to be inaccurate during hypoglycemia episodes, that is, insulin-induced hypoglycemia versus spontaneous hypoglycemia. In this regard, insulin-induced hypoglycemia is found to be difficult to detect as compared with spontaneous hypoglycemia. Fast occurring hypoglycemia is difficult to detect because of the blood-interstitial delay, which makes them important features to be detected by a given model. Furthermore, CGM calibration frequency and timing also affects the performance of the classification and detection algorithm.

The reviewed studies are limited to and could be roughly categorized by age groups (children, young, adult, and old), time of the day (diurnal vs nocturnal) and configurations (online vs offline). For example, most of the studies consider nocturnal hypoglycemia detection, considering the fact that most of hypoglycemia crises occurred during nighttime and also the crises during this time have a bad consequence as compared with the diurnal period. Moreover, it is a fact that nocturnal detection is simpler as compared with the diurnal considering the dynamics of the patients. However, irrespective of these challenges, there are also studies that consider the diurnal period. However, there are limited studies that attempt to develop an algorithm that could detect anomalies in both of those contexts. With regard to the age group, most surveys reported that age group has a great effect on BG dynamics, which is typically related with the dynamics and active lifestyle adopted by each group. Therefore, it is deemed a necessary approach to consider a personalized algorithm for each age group. With regard to the configuration, there are fewer attempts of online (real-time) algorithms, where almost all of the algorithms were tested and implemented in the offline mode. In this regard, the most crucial issues concerning machine learning strategies could be the necessity of frequent retraining when subjected to a real-time and dynamic task.

In addition, the most important component in classification and detection algorithms is the threshold used to differentiate the normal from the abnormal. In this regard, almost most of the studies have used a static threshold based on suggestions either from the literature or physicians and other concerned bodies such as the American Diabetes Association. However, the critical issues in this approach are that the threshold might vary from patient to patient and also some patients might not feel any symptoms at the specified threshold (when using indirect indicators such as heart rate, QT interval, and others). However, there are some studies that employed a fuzzy logic–based approach by having a continuous decision space.

In principle, any future BG anomaly classification and detection algorithm should be expected to detect any upcoming anomalies as soon as possible (lead time—giving more response time), avoid any false alarm at any cost, perform in real-time (in an online fashion), adapt with the dynamics of BG evolution (learn continuously), automatically tune its parameters without user intervention, be able to perform throughout the day in a free living condition (diurnal and nocturnal periods), and incorporate as many input variables to better capture the dynamics. In this regard, for example, the most crucial issues concerning a real-time (online) machine learning algorithm could be the necessity of frequent retraining when subject to a real-time and dynamic task. Moreover, developing a model that considers a real-time and adaption-to-free-living condition needs to incorporate a wide range of parameters that affect BG dynamics. Furthermore, it should properly consider and address the inherent technological limitation that affects the performance of the detection algorithm. Almost all of the studies need a proper clinical validation to be integrated into a smartphone and CGM for a real-time application. This can be better described by looking at the number of samples used and their validation strategies (see Multimedia Appendix 2). Therefore, future studies should give more emphasis on clinical validation by taking a sufficient number of subjects in the development and testing phase so as to better quantify the inter- and intravariability among patients. In addition, the most crucial concept of justifying and reporting the underlying cause, as because of either patient controllable or patient uncontrollable parameters, for the detected anomalies is not addressed in any of the reviewed literature. For example, the underlying cause of hyperglycemia incidences could be patient controllable parameters such as diet or patient uncontrollable parameters such as stress and infections. Therefore, in this regard, a proper hyperglycemia classification and detection system might be expected to be able to identify and report the underlying cause, which has a greater significance to the patient especially during infection crises.

Summary of Existing Efforts: Machine Learning Techniques

Artificial Neural Network

There are various types of ANNs used in solving BG classification and anomaly detection tasks: hypoglycemia, hyperglycemia, and GV classification and detection. Regarding hypoglycemia classification and detection, for instance, Eljil et al [48], had proposed a special type of ANN known as the time-sensitive ANN and compared the result with a time delay neural network, NARX, distributed time delay neural network, and NAR. San et al [37,49] proposed an evolvable BBNN and compared the result with feedforward ANNs and multiple regression. Moreover, San et al [42] proposed a DBN and compared the result with a wavelet neural network, a feedforward ANN, and multiple regression models. Some of the studies have also investigated the advantage of having a separate feature extraction and classification unit. In this regard, for example, both Laione et al [47] and Nguyen et al [41,50] have proposed an ANN using FFT for data extraction. Nguyen et al [41] have further trained the network through a 2-step process that combines the advantage of GA and the Levenberg Marquardt algorithm. Chan et al and Yan et al [51,52] also proposed a neural network–based rule discovery system that consisted of a neural network–based classification unit and rule-based extraction unit. There are some studies that optimized the ANN parameters through a particle swarm optimization technique. For example, Ling et al [53], Phyo et al [36,54,55], and San et al [56] proposed a new hybrid rough neural network, a VTWNN, a normalized RBFNN, and a combinational neural logic network with the neural logic AND, OR, and NOT gates, respectively, where the design parameters of the network were optimized through a hybrid particle swarm optimization with wavelet mutation operation. Moreover, Nguyen et al [43,57] also proposed an ANN that is optimized through a standard particle swarm optimization strategy. Furthermore, some studies have investigated extreme learning machines (ELMs). For instance, Ling et al [58] and San et al [59] proposed a feedforward ANN trained through an ELM and compared the result with a feedforward ANN optimized through particle swarm optimization, multiple regression–based fuzzy inference system, fuzzy inference system, and linear multiple regression. Mo et al [60] have also used ELMs and regularized the ELMs on CGM data. In addition, Nguyen et al [61-63] and Ngo et al [64] had proposed an optimal BNN algorithm using feedforward ANN architecture. There are some studies that tried to integrate a physiological model with ANN. For instance, Bertachi et al [65] integrated the physiological model of an individual diabetes patient with an ANN to predict nocturnal hypoglycemia events. Regarding, hyperglycemia classification and detection, there is only 1 study by Nguyen et al [38] that uses a feedforward multilayer ANN trained using different training algorithms, that is, gradient descent, gradient descent with momentum, scaled conjugate gradient, and resilient back propagation. Regarding GV classification and detection, the reviewed studies had been performed either for detection purposes or for automated metrics purposes. For the detection purpose, for example, Wiley et al [33] proposed Naive Bayes (NB), multilayer perceptron (MLP) ANN, and SVM models to detect excessive GV on CGM data and compared the accuracy of the result with the other 2 diabetes experts. Concerning the automated metrics, Marling et al [66] had developed an NBC (probabilistic reasoning), an MLP ANN, and a logistic model tree (DT built using logistic regression), which could be used to monitor CGM data. Moreover, Marling et al [32] also proposed an MLP ANN and support vector regression to develop a consensus perceived GV metric.

Support Vector Machines, Kernel Function, and Gaussian Process Regression

SVM, kernel function (KF), and Gaussian process regression have been exploited for hypoglycemia classification and detection purposes in the reviewed literature. For example, Georga et al [67] developed a support vector regression for hypoglycemia prediction and compared the performance with a feedforward MLP ANN and Gaussian process regression. Georga et al [68] also proposed support vector regression and Gaussian process regression for BG prediction so as to indicate the daily incidences of hyperglycemia and hypoglycemia to the patients as well as provision of decision support to physicians in making the decision about treatment and risk of complications. Moreover, Jensen et al [29,30] developed an automatic pattern recognition system so as to detect hypoglycemia incidences retrospectively using CGM data and thereby to foster a thorough evaluation of past events and discussion with their caregivers. Jensen et al [28,69] also proposed a real-time pattern classification model by using several features from the CGM data so as to detect hypoglycemia incidences in real-time. Furthermore, Marling et al [70] proposed a hypoglycemia detection algorithm that incorporates noninvasive sensor data from fitness bands and also compared different kernels for the task: linear, Gaussian, and quadratic kernels. Nuryani et al [71] also proposed a swarm-based SVM algorithm using the repolarization variabilities as input so as to detect hypoglycemia incidences.

Genetic Programming and Genetic Algorithm

There is little visibility of GP and GA usage in their nonhybrid form for BG classification and anomaly detection tasks: hypoglycemia, hyperglycemia, and GV classification and detection. However, there are some studies that use these techniques in their hybrid form. For example, Ling et al [44,72,73] developed a hypoglycemia detection algorithm using a GA-based multiple regression coupled with a fuzzy inference system. The study exploited the GA so as to optimize the fuzzy rules, membership function of the fuzzy inference system, and also model parameters of the regression.

Random Forest

RF and DT have been mostly used in the context of hypoglycemia classification and detection tasks. For example, Eljil et al [27] proposed DTs using different techniques, namely, C4.5, J4.8, REPTree, bagging, and the cost-sensitive version of J4.8. Jung et al [74] also proposed DTs using new predictor variables using CGM data. Moreover, Jung et al [34] proposed a DT- and SVM-based prediction model using self-monitored BG. Zhang et al [35] also proposed a new approach using the classification tree to predict the occurrences of acute hypoglycemia during intravenous insulin infusion before the actual hypoglycemic events take place.

Hidden Markov Model

Generally, HMM is used to model an environment that could better describe the evolution of the individual BG dynamics. In this regard, there are some studies that use HMM to develop model-based BG anomaly classification and detection algorithms. For example, Zhu et al [15,75] studied an approach for automatic detection of anomalies in individual BG data, using a model trained with historical data containing daily normal measurements. The trained Markovian world tries to analyze the incoming BG data and flags if it deviates from what is known by the model.

Hybrid and Ensemble Models

Hybridization approaches have been extensively used when looking for performance improvement by exploiting the advantage from 2 or more different approaches [16]. In this regard, there are some attempts in the reviewed articles which tried merging different approaches for enhanced performance in hypoglycemia classification and detection. For example, hybridization of an ANN with other techniques is demonstrated in some of these studies. Chan et al [76] developed a hybrid system that consisted of an ANN and a GA and also compared the performance with MLP ANN and classical statistical algorithms. Ghevondian et al [77] proposed a novel hybrid system of a fuzzy neural network ANN estimator to predict the BG profile and hypoglycemia incidences. San et al [78] proposed a hybrid system using an ANFIS and compared the performance with the wavelet neural network, feedforward ANN, and multiple regression. There is also some literature that tries to hybridize the SVM with other techniques. For example, Nuryani et al [39,79] proposed a hybrid fuzzy SVM and investigated the applicability of 3 KFs: radial basis, exponential radial basis, and polynomial function for the task. Moreover, Nuryani et al [40,80] also further developed a novel strategy using a hybrid particle swarm-based fuzzy SVM technique. Fuzzy reasoning models are also tested in some of the studies. For example, Ling et al [81] developed a hybrid particle swarm-optimization–based fuzzy reasoning model, where the fuzzy rules and the fuzzy-membership functions are optimized through a hybrid particle swarm optimization with wavelet mutation. The model is also compared with feedforward ANN and multiple-regression models. Mathews et al [46] developed a hybrid model using a fuzzy inference system with multiple regression, where the fuzzy rules are optimized through a GA. The study also compares the performance of the developed system with an ANN whose parameters are optimized through particle swarm optimization. In addition, San et al [82] proposed a hybrid system based on rough sets concepts and neural computing. The study has compared various hybrid approaches trained through hybrid particle swarm optimization with wavelet mutation including the rough BBNN, BBNN, rough feedforward ANN, wavelet neural network, SVM with an RBF, and conventional feedforward ANN. Ling et al [45] also proposed an alarm system based on the hybrid neural logic network with multiple regression. Lai et al [83] developed a fuzzy inference system for hypoglycemia detection, where the system parameters are optimized through an intelligent optimizer.

Owing to the complexity of BG dynamics, it remains difficult to achieve an accurate result in every circumstance. One model can have better accuracy in some circumstances and the other model can achieve better accuracy where the first model fails to achieve a comparable result. Therefore, it is natural to look for possibilities to exploit the strengths from these different models to achieve better accuracy in most of the circumstances, which lead to ensemble approaches [16]. An ensemble approach is generally favored when one is interested to merge 2 or more different models for improved performance. In this regard, there are some studies that try to combine 2 different models looking for performance improvement in the overall system. In this regard, Daskalaki et al [84] proposed an early warning system, for both hyperglycemia and hypoglycemia, using RNN and AR with output correction module models. Moreover, the study investigated the performance improvement from the combined use of both RNN and AR with an output correction module. Moreover, Botwey et al [31] proposed combining an AR model with output correction and an RNN based on different data fusion schemes including the Dempster-Shafer evidential theory, GAs, and GP.

Conclusions

Despite the complexity of BG dynamics, there are many attempts to capture hypoglycemia and hyperglycemia incidences and the extent of an individual GV using different approaches. Recently, because of the ubiquitous nature of self-management mHealth apps, sensors and wearables have paved the way for the continuous accumulation of self-collected health data, which in turn contributed for the widespread research of machine learning applications in these tasks. In the reviewed articles, generally, the anomaly classification and detection approaches could be categorized as either model (process)–based or classifier (rule)–based approaches. Hypoglycemia classification and detection has been given more attention than hyperglycemia and GV detection, which might be because of its serious complication and the comparable complexity involved. The state-of-the-art indicates that various classes of machine learning have been developed and tested in these tasks. Regarding hypoglycemia classification and detection, feedforward ANNs, hybrid systems, SVM, DT, GA, adaptive neural fuzzy inference system, NARX, and NAR, Gaussian process regression, DBN, and BNN have been developed and tested. These techniques have explored various kinds of input parameters, notably BG, heart rate, QT interval, insulin, diet, physical activity, galvanic response, and skin impedance. Concerning hyperglycemia classification and detection, RNN, GP, HMM, feedforward ANN, GA, and hybrid systems have been developed and tested, exploring various types of input parameters including BG, insulin, heart rate, and QT interval. GV detection is a recent development, which has great importance in quantifying factors associated with hypoglycemia and hyperglycemia incidence. In this regard, there is some research and development involving machine learning techniques, for example, the feedforward ANN, NBC, and SVM.

Most of these studies have used a theoretical threshold suggested either by the literature or physicians and various concerned bodies such as the American Diabetes Association. However, the problem here is that some patients might feel no symptoms at the specified threshold, and it may vary from patient to patient. Therefore, a model should consider such differences among the patients (intra- and intervariability) and also track its temporal change over time. Moreover, the studies should give more emphasis on the TL and various types of inputs used. Furthermore, researchers should give proper emphasis to develop anomaly classification and detection models, which are capable of justifying and reporting the underlying cause, as either due to patient controllable or patient uncontrollable parameters. Generally, we foresee that these developments might encourage researchers to further develop and test these systems on a large-scale basis.

Acknowledgments

The project is funded by the University of Tromsø – The Arctic University of Norway and National Library of Medicine “Mechanistic machine learning” (grant number: LM012734).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Analysis of reported parameters, data characteristics, class of machine learning, and performance metrics.

[PDF File (Adobe PDF File), 201KB - jmir_v21i5e11030_app1.pdf]

Multimedia Appendix 2

Detail on reported accuracy, inputs and performance metrics used, and machine learning categorization.

[PDF File (Adobe PDF File), 281KB - jmir_v21i5e11030_app2.pdf]

References

1. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
2. Bommer C, Heesemann E, Sagalova V, Manne-Goehler J, Atun R, Bärnighausen T, et al. The global economic burden of diabetes in adults aged 20-79 years: a cost-of-illness study. Lancet Diabetes Endocrinol 2017 Dec;5(6):423-430. [doi: 10.1016/S2213-8587(17)30097-9] [Medline: 28456416]
3. Casqueiro J, Casqueiro J, Alves C. Infections in patients with diabetes mellitus: a review of pathogenesis. Indian J Endocrinol Metab 2012 Mar;16(Suppl 1):S27-S36 [FREE Full text] [doi: 10.4103/2230-8210.94253] [Medline: 22701840]

4. Knapp S. Diabetes and infection: is there a link?--A mini-review. Gerontology 2013;59(2):99-104 [FREE Full text] [doi: 10.1159/000345107] [Medline: 23182884]
5. McCoy R, van Houten HK, Ziegenfuss JY, Shah ND, Wermers RA, Smith SA. Increased mortality of patients with diabetes reporting severe hypoglycemia. Diabetes Care 2012 Sep;35(9):1897-1901 [FREE Full text] [doi: 10.2337/dc11-2054] [Medline: 22699297]
6. Bazaev N, Pletenev AN, Pozhar KV. Classification of factors affecting blood glucose concentration dynamics. Biomed Eng 2013 Jul 18;47(2):100-103. [doi: 10.1007/s10527-013-9344-7]
7. Ahmad S, Purdy S. arXiv. 2016. Real-Time Anomaly Detection for Streaming Analytics URL: https://arxiv.org/abs/1607.02480 [accessed 2019-04-02] [WebCite Cache ID 77Kc7RB7g]
8. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
9. Rigla M, García-Sáez G, Pons B, Hernando ME. Artificial intelligence methodologies and their application to diabetes. J Diabetes Sci Technol 2018 Dec;12(2):303-310 [FREE Full text] [doi: 10.1177/1932296817710475] [Medline: 28539087]
10. Choudhary J, Genovese S, Reach G. Blood glucose pattern management in diabetes: creating order from disorder. J Diabetes Sci Technol 2013 Nov 1;7(6):1575-1584 [FREE Full text] [doi: 10.1177/193229681300700618] [Medline: 24351184]
11. Banaee H, Ahmed MU, Loutfi A. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. Sensors (Basel) 2013 Dec 17;13(12):17472-17500 [FREE Full text] [doi: 10.3390/s131217472] [Medline: 24351646]
12. Shanmuganathan S, Samarasinghe S, editors. Artificial neural network modelling. In: Studies in Computational Intelligence Vol 628. Switzerland: Springer International Publishing; 2016.
13. Campbell C, Ying Y. Learning with Support Vector Machines: Synthesis Lectures on Artificial Intelligence and Machine Learning. San Rafael, California, USA: Morgan & Claypool Publishers; Feb 11, 2011:1-95.
14. Wong ML, Leung KS. Data mining using grammar based genetic programming and applications. In: Koza J, editor. Genetic Programming Series. New York: Kluwer Academic Publishers; 2002.
15. Zhu Y. Automatic detection of anomalies in blood glucose using a machine learning approach. J Commun Netw 2011 Apr;13(2):125-131. [doi: 10.1109/JCN.2011.6157411]
16. Georga EI, Fotiadis DI, Tigas SK. 6 - Nonlinear models of glucose concentration. In: Personalized Predictive Modeling in Type 1 Diabetes. London: Academic Press; 2018.
17. Hawkins DM. Identification of outliers. In: Monographs on Statistics and Applied Probability. London: Springer; 1987.
18. Lavin A, Ahmad S. Evaluating Real-Time Anomaly Detection Algorithms -- The Numenta Anomaly Benchmark. 2015 Presented at: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA); December 9-11, 2015; Miami, FL, USA p. 38-44. [doi: 10.1109/icmla.2015.141]
19. Mehrotra KG, Mohan CK, Huang H. Anomaly detection principles and algorithms. In: Subrahmanian VS, editor. Terrorism, Security, and Computation. Cham, Switzerland: Springer, Cham; 2017.
20. Chandola V, Banerjee A, Kumar V. Anomaly detection. ACM Comput Surv 2009 Jul 1;41(3):1-58. [doi: 10.1145/1541880.1541882]
21. Klerx T, Anderka M, Büning HK, Priesterjahn S. Model-Based Anomaly Detection for Discrete Event Systems. 2014 Presented at: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence; November 10-12, 2014; Limassol, Cyprus p. 665-672. [doi: 10.1109/ICTAI.2014.105]
22. Dunning T, Friedman E. In: Loukides M, editor. Practical Machine Learning: A New Look at Anomaly Detection. Sebastopol, CA: O'Reilly Media, Inc; 2014.
23. Frandes M, Timar B, Timar R, Lungeanu D. Chaotic time series prediction for glucose dynamics in type 1 diabetes mellitus using regime-switching models. Sci Rep 2017 Dec 24;7(1):6232 [FREE Full text] [doi: 10.1038/s41598-017-06478-4] [Medline: 28740090]
24. Suh S, Kim JH. Glycemic variability: how do we measure it and why is it important? Diabetes Metab J 2015 Aug;39(4):273-282 [FREE Full text] [doi: 10.4093/dmj.2015.39.4.273] [Medline: 26301188]
25. Oviedo S, Vehí J, Calm R, Armengol J. A review of personalized blood glucose prediction strategies for T1DM patients. Int J Numer Method Biomed Eng 2017 Dec;33(6):1-21. [doi: 10.1002/cnm.2833] [Medline: 27644067]
26. Balakrishnan N, Rangaiah GP, Samavedham L. Review and analysis of blood glucose (BG) models for type 1 diabetic patients. Ind Eng Chem Res 2011 Nov 2;50(21):12041-12066. [doi: 10.1021/ie2004779]
27. Eljil SK, Qadah G, Pasquier M. Predicting hypoglycemia in diabetic patients using data mining techniques. 2013 Presented at: 2013 9th International Conference on Innovations in Information Technology (IIT); March 17-19, 2013; Abu Dhabi, United Arab Emirates. [doi: 10.1109/Innovations.2013.6544406]
28. Jensen M, Mahmoudi Z, Christensen TF, Tarnow L, Seto E, Johansen MD, et al. Evaluation of an algorithm for retrospective hypoglycemia detection using professional continuous glucose monitoring data. J Diabetes Sci Technol 2014 Jan;8(1):117-122 [FREE Full text] [doi: 10.1177/1932296813511744] [Medline: 24876547]
29. Jensen M, Christensen TF, Tarnow L, Mahmoudi Z, Johansen MD, Hejlesen OK. Professional continuous glucose monitoring in subjects with type 1 diabetes: retrospective hypoglycemia detection. J Diabetes Sci Technol 2013 Jan 1;7(1):135-143 [FREE Full text] [doi: 10.1177/193229681300700116] [Medline: 23439169]

30. Jensen M, Christensen TF, Tarnow L, Johansen MD, Hejlesen OK. An information and communication technology system to detect hypoglycemia in people with type 1 diabetes. Stud Health Technol Inform 2013;192:38-41. [Medline: 23920511]
31. Botwey R, Daskalaki E, Diem P, Mougiakakou SG. Multi-model data fusion to improve an early warning system for hypo-/hyperglycemic events. Conf Proc IEEE Eng Med Biol Soc 2014;2014:4843-4846. [doi: 10.1109/EMBC.2014.6944708] [Medline: 25571076]
32. Marling C, Struble NW, Bunescu RC, Shubrook JH, Schwartz FL. A consensus perceived glycemic variability metric. J Diabetes Sci Technol 2013 Jul 1;7(4):871-879 [FREE Full text] [doi: 10.1177/193229681300700409] [Medline: 23911168]
33. Wiley M, Bunescu R, Marling C, Shubrook J, Schwartz F. Automatic Detection of Excessive Glycemic Variability for Diabetes Management. 2011 Presented at: 10th International Conference on Machine Learning Applications Workshops; December 18-21, 2011; Honolulu, HI, USA p. 148-154. [doi: 10.1109/ICMLA.2011.39]
34. Jung M. Toward Designing Mobile Software to Predict Hypoglycemia for Patients with Diabetes. 2016 Presented at: 2016 IEEE/ACM International Conference on Mobile Software EngineeringSystems (MOBILESoft); May 16-17, 2016; Austin, TX, USA. [doi: 10.1109/MobileSoft.2016.024]
35. Zhang Y. Predicting occurrences of acute hypoglycemia during insulin therapy in the intensive care unit. Conf Proc IEEE Eng Med Biol Soc 2008;2008:3297-3300. [doi: 10.1109/IEMBS.2008.4649909] [Medline: 19163412]
36. San PP, Ling SH, Nguyen HT. Optimized variable translation wavelet neural network and its application in hypoglycemia detection system. 2012 Presented at: 7th IEEE Conference on Industrial Electronics and Applications (ICIEA); July 18-20, 2012; Singapore, Singapore p. 547-551. [doi: 10.1109/ICIEA.2012.6360788]
37. San P, Ling SH, Nguyen HT. Block based neural network for hypoglycemia detection. Conf Proc IEEE Eng Med Biol Soc 2011;2011:5666-5669. [doi: 10.1109/IEMBS.2011.6091371] [Medline: 22255625]
38. Nguyen L, Su S, Nguyen HT. Neural network approach for non-invasive detection of hyperglycemia using electrocardiographic signals. Conf Proc IEEE Eng Med Biol Soc 2014;2014:4475-4478. [doi: 10.1109/EMBC.2014.6944617] [Medline: 25570985]
39. Nuryani N, Ling SH, Nguyen HT. Hypoglycaemia detection for type 1 diabetic patients based on ECG parameters using Fuzzy Support Vector Machine. 2010 Presented at: The International Joint Conference on Neural Networks (IJCNN); July 18-23, 2010; Barcelona, Spain p. 1-7. [doi: 10.1109/IJCNN.2010.5596916]
40. Nuryani N, Ling SH, Nguyen HT. Hybrid particle swarm-based fuzzy support vector machine for hypoglycemia detection. 2012 Presented at: IEEE International Conference on Fuzzy Systems; June 10-15, 2012; Brisbane, QLD, Australia p. 1-6. [doi: 10.1109/FUZZ-IEEE.2012.6250828]
41. Nguyen L, Nguyen AV, Ling SH, Nguyen HT. Combining genetic algorithm and Levenberg-Marquardt algorithm in training neural network for hypoglycemia detection using EEG signals. 2013 Presented at: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 3-7, 2013; Osaka, Japan p. 5386-5389. [doi: 10.1109/EMBC.2013.6610766]
42. San PP, Ling SH, Nguyen HT. Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes. Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:3503-3506. [doi: 10.1109/EMBC.2016.7591483] [Medline: 28269053]
43. Nguyen LB, Nguyen AV, Ling SH, Nguyen HT. A particle swarm optimization-based neural network for detecting nocturnal hypoglycemia using electroencephalography signals. 2012 Presented at: The International Joint Conference on Neural Networks (IJCNN); June 10-15, 2012; Brisbane, QLD, Australia. [doi: 10.1109/IJCNN.2012.6252745]
44. Ling SH, Nguyen H, Chan KY. Genetic algorithm based fuzzy multiple regression for the nocturnal Hypoglycaemia detection. 2010 Presented at: IEEE Congress on Evolutionary Computation; July 18-23, 2010; Barcelona, Spain. [doi: 10.1109/CEC.2010.5586315]
45. Ling S, San PP, Lam HK, Nguyen HT. Hypoglycemia detection: multiple regression-based combinational neural logic approach. Soft Comput 2015 Jul 25;21(2):543-553. [doi: 10.1007/s00500-015-1809-z]
46. Mathews S. Fuzzy inference system and multiple regression for the detection of hypoglycemia. Int J Comp Sci Appl 2012 Apr 30;2(2):37-50. [doi: 10.5121/ijcsa.2012.2204]
47. Iaione F, Marques JL. Methodology for hypoglycaemia detection based on the processing, analysis and classification of the electroencephalogram. Med Biol Eng Comput 2005 Jul;43(4):501-507. [doi: 10.1007/BF02344732] [Medline: 16255433]
48. Eljil SK, Qadah G, Pasquier M. Predicting hypoglycemia in diabetic patients using time-sensitive artificial neural networks. Int J Healthc Inf Syst Inform 2016;11(4):70-88. [doi: 10.4018/ijhisi.2016100104]
49. San P, Ling SH, Nguyen HT. Industrial application of evolvable block-based neural network to hypoglycemia monitoring system. IEEE Trans Ind Electron 2013 Dec;60(12):5892-5901. [doi: 10.1109/TIE.2012.2228143]
50. Nguyen H, Ghevondian N, Jones TW. Neural-network detection of hypoglycemic episodes in children with type 1 diabetes using physiological parameters. 2006 Presented at: International Conference of the IEEE Engineering in Medicine and Biology Society; January 17-18, 2006; New York, NY, USA p. 6053-6056. [doi: 10.1109/IEMBS.2006.259482]
51. Chan K, Ling S, Dillon T, Nguyen H. Diagnosis of hypoglycemic episodes using a neural network based rule discovery system. Expert Syst Appl 2011 Aug;38(8):9799-9808. [doi: 10.1016/j.eswa.2011.02.020]

52. Chan KY, Ling SH, Nguyen HT, Jiang F. A hypoglycemic episode diagnosis system based on neural networks for Type 1 diabetes mellitus. 2012 Presented at: 2012 IEEE Congress on Evolutionary Computation; June 10-15, 2012; Brisbane, QLD, Australia. [doi: 10.1109/CEC.2012.6256604]
53. Ling SH, San PP, Lam HK, Nguyen HT. Non-invasive detection of hypoglycemic episodes in Type 1 diabetes using intelligent hybrid rough neural system. 2014 Presented at: 2014 IEEE Congress on Evolutionary Computation (CEC); July 6-11, 2014; Beijing, China. [doi: 10.1109/CEC.2014.6900229]
54. San PP, Ling SH, Nguyen HT. Hybrid particle swarm optimization based normalized radial basis function neural network for hypoglycemia detection. 2012 Presented at: The International Joint Conference on Neural Networks (IJCNN); June 10-15, 2012; Brisbane, QLD, Australia p. 1-6. [doi: 10.1109/IJCNN.2012.6252743]
55. San P, Ling SH, Nguyen HT. Hybrid PSO-based variable translation wavelet neural network and its application to hypoglycemia detection system. Neural Comput & Applic 2012 Sep 27;23(7-8):2177-2184. [doi: 10.1007/s00521-012-1168-y]
56. San PP, Ling SH, Nguyen HT. Combinational neural logic systemits industrial application on hypoglycemia monitoring system. 2013 Presented at: IEEE 8th Conference on Industrial Electronics and Applications (ICIEA); June 19-21, 2013; Melbourne, VIC, Australia p. 947-952. [doi: 10.1109/ICIEA.2013.6566503]
57. Nguyen L, Nguyen AV, Ling SH, Nguyen HT. An adaptive strategy of classification for detecting hypoglycemia using only two EEG channels. 2012 Presented at: Annual International Conference of the IEEE Engineering in Medicine Biology Society; August 28-September 1, 2012; San Diego, CA, USA p. 3515-3518. [doi: 10.1109/EMBC.2012.6346724]
58. Ling S, San PP, Nguyen HT. Non-invasive hypoglycemia monitoring system using extreme learning machine for Type 1 diabetes. ISA Trans 2016 Sep;64:440-446. [doi: 10.1016/j.isatra.2016.05.008] [Medline: 27311357]
59. San PP, Ling SH, Soe NN, Nguyen HT. A novel extreme learning machine for hypoglycemia detection. Conf Proc IEEE Eng Med Biol Soc 2014;2014:302-305. [doi: 10.1109/EMBC.2014.6943589] [Medline: 25569957]
60. Mo X, Wang Y, Wu X. Hypoglycemia prediction using extreme learning machine (ELM) and regularized ELM. 2013 Presented at: 25th Chinese Control and Decision Conference (CCDC); May 25-27, 2013; Guiyang, China. [doi: 10.1109/CCDC.2013.6561727]
61. Nguyen H, Ghevondian N, Jones TW. Detection of nocturnal hypoglycemic episodes (natural occurrence) in children with Type 1 diabetes using an optimal Bayesian neural network algorithm. 2008 Presented at: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 20-25, 2008; Vancouver, BC. Canada p. 1311-1314. [doi: 10.1109/IEMBS.2008.4649405]
62. Nguyen H, Ghevondian N, Nguyen ST, Jones TW. Detection of hypoglycemic episodes in children with type 1 diabetes using an optimal Bayesian neural network algorithm. 2007 Presented at: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 22-26, 2007; Lyon, France p. 3140-3143. [doi: 10.1109/IEMBS.2007.4352995]
63. Nguyen H, Jones TW. Detection of nocturnal hypoglycemic episodes using EEG signals. 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology; August 31-September 4, 2010; Buenos Aires, Argentina p. 4930-4933. [doi: 10.1109/IEMBS.2010.5627233]
64. Ngo CQ, Truong BC, Jones TW, Nguyen HT. Occipital EEG Activity for the Detection of Nocturnal Hypoglycemia. 2018 Presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 18-21, 2018; Honolulu, HI, USA p. 3862-3865.
65. Bertachi A, Biagi L, Contreras I, Luo N, Vehí J. Prediction of Blood Glucose Levels And Nocturnal Hypoglycemia Using Physiological Models and Artificial Neural Networks. 2018 Presented at: 3rd International Workshop on Knowledge Discovery in Healthcare Data co-located with the 27th International Joint Conference on Artificial Intelligence and the 23rd European Conference on Artificial Intelligence (IJCAI-ECAI 2018); July 13, 2018; Stockholm, Schweden p. 85-90.
66. Marling C, Shubrook JH, Vernier SJ, Wiley MT, Schwartz FL. Characterizing blood glucose variability using new metrics with continuous glucose monitoring data. J Diabetes Sci Technol 2011 Jul 1;5(4):871-878 [FREE Full text] [doi: 10.1177/193229681100500408] [Medline: 21880228]
67. Georga E, Protopappas VS, Ardigò D, Polyzos D, Fotiadis DI. A glucose model based on support vector regression for the prediction of hypoglycemic events under free-living conditions. Diabetes Technol Ther 2013 Aug;15(8):634-643. [doi: 10.1089/dia.2012.0285] [Medline: 23848178]
68. Georga EI, Protopappas VC, Mougiakakou SG, Fotiadis DI. Short-term vs long-term analysis of diabetes data: application of machine learning and data mining techniques. 2013 Presented at: 13th IEEE International Conference on BioInformatics and BioEngineering; November 10-13, 2013; Chania, Greece. [doi: 10.1109/BIBE.2013.6701622]
69. Jensen M, Christensen TK, Tarnow L, Seto E, Johansen MD, Hejlesen OK. Real-time hypoglycemia detection from continuous glucose monitoring data of subjects with type 1 diabetes. Diabetes Technol Ther 2013 Jul;15(7):538-543. [doi: 10.1089/dia.2013.0069] [Medline: 23631608]
70. Marling C, Xia L, Bunescu R, Schwartz F. Machine learning experiments with noninvasive sensors for hypoglycemia detection. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 2016 Presented at: IJCAI Workshop on Knowledge Discovery in Healthcare Data; July 10, 2016; New York, NY.

71. Ling S, Nguyen HT. Ventricular repolarization variability for hypoglycemia detection. 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine Biology Society; August 30-September 3, 2011; Boston, MA, USA p. 7961-7964. [doi: 10.1109/IEMBS.2011.6091963]
72. Ling S, Nguyen HT. Genetic-algorithm-based multiple regression with fuzzy inference system for detection of nocturnal hypoglycemic episodes. IEEE Trans Inf Technol Biomed 2011 Mar;15(2):308-315. [doi: 10.1109/TITB.2010.2103953] [Medline: 21349796]
73. Ling SH, Nguyen HT, Leung FH. Hypoglycemia detection using fuzzy inference system with genetic algorithm. 2011 Presented at: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); June 27-30, 2011; Taipei, Taiwan. [doi: 10.1109/FUZZY.2011.6007319]
74. Jung M, Lee YB, Jin SM, Park SM. arXiv. 2017. Prediction of Daytime Hypoglycemic Events Using Continuous Glucose Monitoring Data and Classification Technique URL: https://arxiv.org/abs/1704.08769 [accessed 2019-04-02] [WebCite Cache ID 77KnFAH7R]
75. Zhu Y. Automatic detection of anomalies in blood glucose using a machine learning approach. 2010 Presented at: IEEE International Conference on Information Reuse & Integration; August 4-6, 2010; Las Vegas, NV. USA p. 92-97. [doi: 10.1109/IRI.2010.5558959]
76. Chan K, Ling SH, Dillon TS, Nguyen H. Classification of hypoglycemic episodes for Type 1 diabetes mellitus based on neural networks. 2010 Presented at: IEEE Congress on Evolutionary Computation; July 18-23, 2010; Barcelona, Spain. [doi: 10.1109/CEC.2010.5586320]
77. Ghevondian N, Nguyen HT, Colagiuri S. A novel fuzzy neural network estimator for predicting hypoglycaemia in insulin-induced subjects. 2001 Presented at: 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society; October 25-28, 2001; Istanbul, Turkey. [doi: 10.1109/IEMBS.2001.1020533]
78. San P, Ling SH, Nguyen HT. Intelligent detection of hypoglycemic episodes in children with type 1 diabetes using adaptive neural-fuzzy inference system. Conf Proc IEEE Eng Med Biol Soc 2012;2012:6325-6328. [doi: 10.1109/EMBC.2012.6347440] [Medline: 23367375]
79. Nuryani SL, Nguyen HT. Electrocardiographic T-wave peak-to-end interval for hypoglycaemia detection. Conf Proc IEEE Eng Med Biol Soc 2010;2010:618-621. [doi: 10.1109/IEMBS.2010.5627430] [Medline: 21096769]
80. Nuryani N, Ling SS, Nguyen HT. Electrocardiographic signals and swarm-based support vector machine for hypoglycemia detection. Ann Biomed Eng 2012 Apr;40(4):934-945. [doi: 10.1007/s10439-011-0446-7] [Medline: 22012087]
81. Ling S, Nguyen HT. Natural occurrence of nocturnal hypoglycemia detection using hybrid particle swarm optimized fuzzy reasoning model. Artif Intell Med 2012 Jul;55(3):177-184. [doi: 10.1016/j.artmed.2012.04.003] [Medline: 22698854]
82. San PP, Ling SH, Nuryani N, Nguyen H. Evolvable rough-block-based neural network and its biomedical application to hypoglycemia detection system. IEEE Trans Cybern 2014 Aug;44(8):1338-1349. [doi: 10.1109/TCYB.2013.2283296] [Medline: 24122616]
83. Lai J, Leung F, Ling S. Hypoglycaemia detection using fuzzy inference system with intelligent optimiser. Appl Soft Comput 2014 Jul;20:54-65. [doi: 10.1016/j.asoc.2013.12.015]
84. Daskalaki E, Nørgaard K, Züger T, Prountzou A, Diem P, Mougiakakou S. An early warning system for hypoglycemic/hyperglycemic events based on fusion of adaptive prediction models. J Diabetes Sci Technol 2013 May 1;7(3):689-698 [FREE Full text] [doi: 10.1177/193229681300700314] [Medline: 23759402]

Abbreviations

ANFIS: adaptive neural fuzzy inference system
ANN: artificial neural network
AR: autoregressive
BBNN: block-based neural network
BG: blood glucose
BNN: Bayesian Neural Network
CGM: continuous glucose monitoring.
DBN: deep belief network
DT: decision tree
EA: evolutionary algorithm
ECG: electrocardiogram
ELM: extreme learning machine
GA: genetic algorithm
GP: genetic programming
GV: glycemic variability
HMM: hidden Markov model
MLP: multilayer perceptron
NAR: nonlinear autoregressive network

NARX: nonlinear autoregressive network with exogenous inputs
NBC: Naive Bayes classifier
POC: point-of-care
RF: random forest
RNN: recurrent neural network
ROC: receiver operating characteristic
SLP: single-layer perceptron
SMBG: self-monitoring blood glucose
SVM: support vector machine
VTWNN: variable translation wavelet neural network

Edited by G Eysenbach; submitted 11.05.18; peer-reviewed by U Sartipy, J Vehi; comments to author 11.10.18; revised version received 27.11.18; accepted 30.01.19; published 01.05.19.

Please cite as:
Woldaregay AZ, Årsand E, Botsis T, Albers D, Mamykina L, Hartvigsen G
Data-Driven Blood Glucose Pattern Classification and Anomalies Detection: Machine-Learning Applications in Type 1 Diabetes
J Med Internet Res 2019;21(5):e11030
URL: https://www.jmir.org/2019/5/e11030/
doi:10.2196/11030
PMID:31042157

Viewpoint

Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of Responsible Health App Marketplaces

Til Wykes[1*], BSc (Hon), MPhil, DPhil; Stephen Schueller[2*], BA, MA, PhD

[1]King's College London, Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom
[2]Center for Behavioral Intervention Technologies, North Western University, Chicago, IL, United States
[*]all authors contributed equally

Corresponding Author:
Til Wykes, BSc (Hon), MPhil, DPhil
King's College London
Department of Psychology
Institute of Psychiatry, Psychology and Neuroscience
De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 20 7848 0596
Email: til.wykes@kcl.ac.uk

Abstract

The overselling of health apps that may provide little benefit and even harm needs the health community’s immediate attention. With little formal regulation, a light-touch approach to consumer protection is now warranted to give customers a modicum of information to help them choose from the vast array of *so-called* health apps. We suggest 4 guiding principles that should be adopted to provide the consumer with information that can guide their choice at the point of download. We call these the Transparency for Trust (T4T) principles, which are derived from experimental studies, systematic reviews, and reports of patient concerns. The T4T principles are (1) privacy and data security, (2) development characteristics, (3) feasibility data, and (4) benefits. All our questions are in a simple form so that all consumers can understand them. We suggest that app stores should take responsibility for providing this information and store it with any app marketed as a health app. Even the absence of information would provide consumers with some understanding and fuel their choice. This would also provide some commercial impetus for app developers to consider this requested information from the outset.

(J Med Internet Res 2019;21(5):e12390) doi:10.2196/12390

KEYWORDS

mobile health; digital health applications; consumer protection; ehealth; advertising standards; digital mental health interventions; digital health

Background

Digital therapeutics are being touted as having the potential to transform health care by improving people’s experience, increasing effectiveness, and reducing costs. A few digital aids are recommended and integral to health services, but much of the e-health field depends on overselling [1]. This business plan seems to be working as the digital health field was estimated to be worth $25 billion globally in 2017 [2,3]. One US survey found that 58% of smartphone users have downloaded at least one health app [4]. The overselling of health apps needs the health community’s immediate attention as many of these apps may provide little benefit and some apps may cause harm. With little formal regulation, alternative light-touch approaches to consumer protection are necessary to give customers a modicum of information as a basis for choosing from the vast array of *so-called* health apps.

We believe that simple, but informative, evidence should be available at the point of downloading and propose 4 succinct aspects that represent the critical information required for responsible health app marketplaces. We refer to these 4 principles—privacy and data security, development practices, feasibility, and health benefits—as the Transparency for Trust (T4T) principles. The goal of these principles is to operationalize efforts that can be made by app marketplaces to answer calls for better oversight and backing up products with data and research [3,5,6]. These T4T principles draw from several sources including patient and regulatory perspectives, recent systematic reviews, and experimental studies (eg [4,7-14]). We have used the definition of a health app from Innovate UK as those apps that contribute to the physical, mental, or social well-being of the user [15]. Our principles are designed to be applicable to the whole panoply of health apps from sleep apps to diabetes apps, symptom trackers to mindfulness interventions eg[16-18]. The fastest growing segment of digital therapeutics is health apps. Current estimates suggest that over 300,000 health apps exist with a recent yearly growth rate of 25% [17]. The rapid growth of products and lack of growth of information and regulation have resulted in very little information to separate quality health apps from those that are at best useless and at worst harmful. As a result, consumers are left to navigate the app stores alone. For one fast-growing sector, digital mental health, the likelihood of using an app is also affected by the relative lack of access or choice of mental health services as well as the stigma and discrimination experienced by sufferers and the interrelationship with physical health. We know, for instance, that people with depression face a higher risk of developing heart disease than individuals without depression and that following a heart attack, each additional depressive symptom that develops increases the risk of another heart attack by 15% [17]. Mental health problems also affect morbidity in other disorders such as rheumatoid arthritis and asthma [16]. We have used examples from mental health, but our principles are intended to inform and empower all health app users.

Why Do We Need Some Simple Principles?

Exponential Growth and Poor Regulation

The use of digital technologies to alleviate, prevent, or maintain health has been recognized for many years, but although it offers enormous potential to rethink how services are provided, there are large roadblocks in its way (eg [19]). A Lancet Psychiatry Commission suggested that digital therapeutics could provide benefits now to complement current mental health treatments and aid self-management [8]. However, evidence suggests that some apps are not only ineffective, unsafe, and hard to use, but do not meet users' privacy and security expectations [9,20,21].

Formal regulation is remarkably light and restricted to a narrow selection of health apps that provide formal diagnoses or treatment for specified medical conditions [3,22,23]. Even when an app is regulated, we cannot be sure that it will work. For example, the US Food and Drug Administration (FDA) recently approved the first behavioral health app, reSET, for the treatment of substance use by using evidence from a clinical trial of a Web-based version of the treatment, not the app itself [24]. In the United Kingdom, the Care Quality Commission issued guidance in 2017 for digital health care providers, but this concentrated on safety [25]. The Medicines and Healthcare Products Regulatory Agency (MHRA) provides Conformité Européenne marking for medical devices and Certificates of Free Sale [26] but leaves review to the National Information Board. As it stands, many health apps are marketed with few checks and even regulatory approval appears to offer little confidence on whether that specific product was ever directly evaluated.

More complex regulation has been proposed. The National Institute of Health and Clinical Excellence (NICE) is curating an NHS National Health Service app library. This is a burdensome process, and so few apps will be assessed in any year. Currently, the library has 78 apps, with only 18 for the fastest developing sector, mental health. This is a minute subset of the 325,000 available [27]. In the United States, the FDA launched a precertification pilot program involving 9 companies to speed the approval process, but this will evaluate the developers and their practices rather than focusing on the product [28]. Apple Inc has introduced additional requirements for medical apps for developers, but these focus mainly on measurement accuracy [29]. There is a middle way to fill this important, and now yawning, gap in consumer information. Health app marketplaces could take a lead by providing relatively simple guidance.

What is Wrong With Current Systems for Reviewing Health Apps?

Most proposed evaluations (eg, Mobile App Rating Scale, MARS [30], Enlight assessment tools [31]) assess usability, aesthetics, content, user engagement, and available research evidence, and others have been adding to this list [32]. These systems are useful because they facilitate multifaceted and thorough evaluations of apps, but they fall short of allowing clear recommendations. In fact, more recent evidence from Canada involving service users demonstrated that a high MARS rating would not on its own provide enough information to allow service users to form a decision on whether to download an app [33]. Advisory bodies such as NICE in the United Kingdom make a determination of what is likely to be cost-effective (effective, cost relative to benefit, and other comparison treatments), before they recommend its use in the UK NHS. Consumers, however, want to make choices based on simpler information. One for-profit company, ORCHA [34], provides reviews based on *current standards, regulation and good practice*, but their overall score does not allow a consumer to decide which components are important to them. PsyberGuide [35], a nonprofit organization, also provides reviews that include a service user focus but does not receive data directly from app developers. But apart from their lack of fulfilling all users' expectations, no method provides clear information at the point of sale, and a potential consumer would have to search in 2 places for the information they need, to make a choice. We propose only 4 aspects of apps that represent the critical information required for responsible health app marketplaces. These 4 principles, deemed the T4T principles, are privacy and data security, development characteristics, feasibility data, and benefits.

Transparency for Trust (T4T) Principles

Privacy and Data Security

Privacy and data security are a primary concern for patients and their clinicians [36-38], and its importance has only become more salient with recent events such as the Facebook and Cambridge Analytica scandal [39-41]. The European Union General Data Protection Regulation is strong and introduces new rights for people to access the information companies hold about them, obligations for better data management for businesses, and a new regime of fines across Europe. There are weaker regulations elsewhere, resulting in varying protections internationally. One review prompted the closure of the NHS app store when it was discovered that accredited apps were not encrypting data adequately and did not explicitly describe the personal data leaving the app [42]. Happtique, an early app certification company, met a similar fate when several of its certified apps were hacked, demonstrating the inadequacy of its processes to evaluate privacy and data security [43,44]. Many apps rely on selling the data they collect for their business plan, which jeopardizes personal privacy [42]. There is also evidence of poor practice resulting in fines for selling sensitive information to Lottery companies and fraudsters [45]. Privacy concerns change with the evolving technology, even though device operating systems are moving toward encryption on the device by default. Nevertheless, users need information about data leaving the app to make informed decisions about their willingness to provide sensitive health information [46,47].

Although full formal audits are needed to ensure apps follow their stated procedures [48], even requiring developers to list their privacy and data security procedures in simple terms would be a significant step forward on raising standards [49]. We propose 3 questions: (1) what data leave the device? (2) how are those data stored? (eg, de-identified, encrypted), and (3) who will have access to those data? It should be clear what, if any, data are being sold, to whom, and what steps are taken to ensure that users cannot be identified by those data.

Development Characteristics

Development characteristics describe how the app was developed, and our recommendations conceptually overlap with those of the FDA's precertification pilot and the MHRA in the United Kingdom. Good developmental practices would involve all stakeholders (clinicians and the target audience) as well as using evidence-based guidance from the beginning and at all stages of development and testing. The absence of the use of guidance or standards has recently been noticed for physical activity and fitness apps where very few of the thousands of Android apps provided any measurement or used any of the accepted guidance [50]. We especially emphasize including the target audience. This may seem obvious, but unfortunately, development practices often include clinicians and experts but more rarely involve the target audience until evaluation. Many studies rely on small numbers of participants or convenience samples, for example, soliciting feedback from stressed college students rather than individuals with depression [51]. Again, this may seem obvious, but independent usability evaluations have demonstrated that many popular commercial apps are frustrating and challenging for members of the intended audience, raising questions about their prior involvement and the potential for the app to benefit this community [52]. Recent evidence also suggests that good design contributes not only to usability but also engagement with health apps [53], and there are several authoritative descriptions of the processes for developing good design [54]. Developers should outline their design and development process and clearly describe how patients were involved. Our 3 questions are as follows: (1) how were target users involved in the initial design? (2) how were target users involved in usability evaluations? and (3) has usability been independently evaluated?

Feasibility

Feasibility evaluations should address how people use the app (usability and user experience), how long they use it (engagement), and whether any serious adverse concerns are discovered (safety). These aspects provide information on how people use the app, including expectations on the frequency and length of use. This information is also vital to assess benefits. It would not be possible to run a drug trial or market a drug without some concept of the dosing frequency and expected therapeutic dose, and the same should be true with health apps.

Again, we have 3 questions: (1) what proportion of users continue to use the app after 2 weeks? (2) what adverse events occurred and what was the rate of those events? and (3) has feasibility been independently evaluated? We propose a 2-week test not because it represents a likely therapeutic dose but because very few users persist in using a health app after the first week [55]. A standard metric, such as 2 weeks, could promote cross-app comparisons in engagement. Like usability testing, independent evaluation of apps is the key to promote transparency and confidence in findings. Despite the availability of engagement analytics, few are reported even in the clinical assessments of feasibility [56]. Independent evaluations could be carried out by service user groups, which could further strengthen service user involvement in the process of development and evaluation. Transparency could be further facilitated by making these datasets available to the research community.

Health Benefits

Health benefits are apparent from rigorous evaluations using standardized and accepted outcomes for the target condition that provides an indication of health benefits. Although many researchers have noted the mismatch between the development cycle for mobile apps and traditional randomized controlled trials [57,58], it is still the case that health apps presented as digital therapeutics require rigorous evaluation to back up their claims. The speed of development should not preclude such evaluations as suggested by some academics and designers [59]. We should be presented with direct evidence on an app's safety and effectiveness because they are not merely mobile versions of websites even if they have similar content. People use apps differently, including more frequently and in shorter bursts [60], and these differences could affect their impact. We have already mentioned that this is happening with the FDA-approved (and first) behavioral health app, reSET, using clinical trial evidence from a Web-based version of the treatment [32]. Although triangulation of different sorts of data has been suggested (eg, MindTech [61]), we believe that health apps should undergo a trial to determine their superiority to other treatment options, especially as many unsubstantiated claims have been made [36]. Advertising standards require evidence to support any claims made, so these data fulfill both commercial and patient needs. Evaluations should also consider opportunity cost as using a health app may delay treatments that could be more beneficial, or a delay could worsen the health condition, making it harder to treat. All these benefits and costs need to be weighed in the balance. Our 3 questions are: (1) what is the impact on the health condition? (2) what percentage of users received either no benefit or deteriorated? and (3) are there specific benefits that outweigh any costs?

What Would This Look Like in Practice?

We have inserted the information from 4 health apps, one of which was named the app of the year in the iTunes Store in 2017 (Calm) (see Table 1). We have extracted, where possible, the information on each of our principles from the information provided with the app. The differences are very clear, especially in privacy and health benefits. There could, of course, be more data available on benefits held elsewhere, but this was not available at the point of download. However, what these simple principles also provide is the ability for a consumer to trade off the attributes. Some may want to know that their data are totally secure, whereas others might want to allow some encrypted anonymized data to be transmitted if the effectiveness of the app is proven. Indeed, in a recent survey of participants recruited from a mood and anxiety disorder clinic, many respondents were willing to allow an app to collect data directly from one's phone, including global positioning system motion sensors, and screen state [37].

Table 1. Evaluating apps with the Transparency for Trust principles.

Transparency for Trust principle	Apps			
	BlueIce[a]	Calm[b]	My Fitness Pal[c]	Dario Diabetes Management[d]
Privacy and security				
1) What data leave the device?	No information leaves the device.	Device identifiers, user settings, device operating system, use of app, and location.	Data related to lifestyle (eg, sleeping habits), life events, dietary restrictions, fitness goals, height, weight, measurements, fitness level, heart rate, sleep data, body mass index, biometric data, similar types of data relating to physiological condition and activity, and personal data (name, email address, postal code, date of birth, and contact number).	Personal information (registration information, such as full name, gender, email address, phone number, and birth date; financial information, such as PayPal account or credit card number; voluntary information; health information, such as diabetes type; and device information) and nonpersonal information (nonidentifiable information such as software and hardware information).
2) How are those data stored?	All data are stored on the app and owned by the user.	No information.	All data are stored on the company's server	All data are stored on the company's server.
3) Who will have access to those data?	Only user of the device where BlueIce is installed has access.	Third party: service providers, marketers with Calm, other systems (Google Fit of HealthKit,), industry research, etc.	Partners and affiliates, service providers and vendors, and social network providers.	Unspecified third parties.
Development characteristics				
1) How were target users involved in the initial design of the app?	Coproduced by Oxford Health NHS Foundation Trust and young people with lived experience.	No information.	User testing on various app iterations, but no additional information provided.	User testing on various app iterations, but no additional information provided.
2) How were target users involved in usability evaluations?	Not provided.	Not provided.	User testing on various app iterations, but no additional information provided.	User testing on various app iterations, but no additional information provided.
3) Has usability been independently evaluated?	No independent usability evaluations were conducted.	Independently evaluated by PsyberGuide.org.	No independent usability evaluations were conducted.	Independently evaluated by Orcha.com; Food and Drug Administration approval.
Feasibility				
1) What proportion of users continue to use the app after 2 weeks?	93% of users kept using it.	No information provided.	No information provided.	No information provided.
2) What adverse events occurred in the test population and what was the rate of those events?	None found (clinicians did not withdraw user and users did not feel app use increased self-harm).	No information provided.	No information provided.	No information provided.
3) Has feasibility been independently evaluated?	No independent evaluations conducted.	No independent evaluations conducted.	No information provided.	No information provided.
Benefits				
1) What was the impact on clinical outcomes?	Significant reductions in depression and anxiety, and 73% reduced self-harm after 12 weeks.	No clinical outcomes research reported.	No clinical outcomes research reported.	No clinical outcomes research reported.
2) What percentage of users received no benefit or deteriorated?	27% reported no reductions in self-harm.	No description provided of nonresponders or users who deteriorated.	No description provided of nonresponders or users who deteriorated.	No description provided of nonresponders or users who deteriorated.

Transparency for Trust principle	Apps			
	BlueIce[a]	Calm[b]	My Fitness Pal[c]	Dario Diabetes Management[d]
3) Are the specific benefits worth the cost?	No information provided about the expected ratio of benefits to risks.	No information provided about the expected ratio of benefits to risks.	No information provided about the expected ratio of benefits to risks.	No information provided about the expected ratio of benefits to risks.

[a] BlueIce is a prescribed evidence-based app to help young people manage their emotions and to reduce urges to self-harm [62].

[b] Calm was the 2017 app of the year on the Apple store and is for meditation and sleep [63].

[c] My Fitness Pal is for logging and motivating physical activity and diet and was the top-rated app in the journal *Men's Health* in 2018 [64].

[d] Dario Diabetes Management monitors blood glucose history, allows carbohydrate counting, and will send messages to up to 4 people if your levels. Scored relatively highly by third-party app rater, Orcha [65].

Are These Principles Different From Those Suggested by Others?

As we have said, T4T principles were based on those suggested by others in recent years. However, we more clearly operationalize our principles into concrete questions that could be answered and made available to potential users. To do so, we considered information important to regulators, developers, and health services, as well as integrating patient viewpoints taken from a number of different studies [37,66]. Patients are, after all, the consumer group of interest. Their views do not necessarily coincide with the expert groups' views, as shown in the Delphi exercise by Zelmer et al [33]. Privacy and security feature in every assessment system and in regulations and are high on the list for patients, especially those with a mental health problem who may be more sensitive about information about them being shared [11]; So, it is included here, but in the simplest terms and not buried in an incomprehensible privacy statement. Our principle for a *fit-for-purpose* app includes development with patients. This principle is often suggested [19] but rarely incorporated into app assessment. As we know that some commercial apps are complex and hard to use by the patient group they were intended for, we have valued this section highly. Effectiveness is often mentioned in many assessment systems, but the promised effects are also dependent on the dose of the app and how intensively it is used. Patients need to consider this time constraint when deciding to make a purchase. Patients also want to know not just how effective it is, but also whether anyone does not receive any benefit. This is also important to clinicians, as patients who receive no benefit may view themselves as hopeless cases and not, as in the BlueIce exemplar, just part of the quarter of patients who report no advantage from following the app.

Our approach has, therefore, been to provide information to patients at the point of the download that allows them to make an informed decision, and which they can refer to later as part of their self-management plan.

Responsibility in the Health App Marketplaces

If these simple T4T principles are followed, then we will have gone some way toward protecting patients. Whose job is it to monitor the T4T principles? Our view is that formal regulation is not needed. We just need the information to allow patients (and patient groups) to make informed choices. Information that is not true can be picked up by advertising standards authorities. Recent examples of this process are the US Federal Trade Commission fining of Lumosity for deceiving consumers with unfounded claims about cognitive benefits [67] and Carrot Neurotechnology for claiming that their app, Ultimeyes, can improve users' vision [68]. Health apps are not a passing fad, and the low barrier of entry into current app marketplaces has resulted in an environment that at best confuses and at worst delays effective treatment. The problems have been highlighted but rarely have clear solutions like ours been proposed. Developers may be encouraged to produce these answers by commercial advantages, as apps with T4T principles might increase consumer comfort and produce unique revenue streams through increased adoption, not only from direct-to-patient markets, but also from health systems. They will also enjoy increased legitimacy among patient groups.

We also note that the contributions of these principles are that they are a small, yet informative, set of questions that could be adopted relatively simply. We suggest these principles as the first, but important step. Further steps could attempt to explore if these principles could be defined with more structure. However, this structure would likely require further empirical work and coordination between different stakeholders in the health app space, particularly, developers and purveyors.

Confidence in the efficacy and safety of these health apps is the least that patients should expect in making a choice to buy or use them. It is now time that existing commercial app stores, specifically the Google Play and Apple iTunes stores, step back from their libertarian ideology and adopt some rules for health app marketing. They should tighten up the definition of health apps and adopt a system, ours hopefully, to allow patients to understand what to expect from a health app. Although some might believe that this proposal is *other worldly*, starting somewhere is important. Health app marketplaces have a duty to, and health app developers a commercial advantage, from following our suggestions – we should not need to wait for another scandal or disaster before the Google Play or Apple iTunes Stores step up to the plate and help prevent worthless products being pressed on those with health needs.

Acknowledgments

TW acknowledges support from the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London and the NIHR Senior Investigator Award. SMS acknowledges the support of One Mind as well as the Implementation Research Institute at the George Washington University in St. Louis; support from the National Institute of Mental Health (5R25MH08091607) and the Department of Veterans Affairs, Health Services Research & Development Service, Quality Enhancement Research Initiative. The views expressed are authors' and not necessarily those of the NHS, the NIHR, the Department of Health, or National Institute for Mental Health .

Authors' Contributions

TW presented the outline and together with SS wrote the initial draft. Both authors contributed to the revisions. TW is the guarantor.

Conflicts of Interest

TW has developed a novel software intervention (CIRCuiTS) and sits on the PsyberGuide Scientific Advisory Board. She has not received any funding from companies involved in this field. SS has received funding from One Mind and serves as the Executive Director of PsyberGuide, a nonprofit funded by One Mind. SMS serves as a scientific advisor to Joyable, Inc, and Potentia Labs, Inc, and has received stock options for these contributions.

References

1. Wykes T, Brown M. Over promised, over-sold and underperforming? - e-health in mental health. J Ment Health 2016;25(1):1-4 [FREE Full text] [doi: 10.3109/09638237.2015.1124406] [Medline: 26732733]
2. The Lancet. Does mobile health matter? Lancet 2017 Nov 18;390(10109):2216. [doi: 10.1016/S0140-6736(17)32899-4] [Medline: 29165258]
3. Duggal R, Brindle I, Bagenal J. Digital healthcare: regulating the revolution. Br Med J 2018 Dec 15;360:k6. [doi: 10.1136/bmj.k6] [Medline: 29335296]
4. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
5. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 Dec 1;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
6. -. Is digital medicine different? Lancet 2018;392(10142):95. [doi: 10.1016/S0140-6736(18)31562-9]
7. Research2Guidance. 2017. 325,000 mobile health apps available in 2017 - Android now the leading mHealth platform URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2019-03-08] [WebCite Cache ID 76ie8Xisj]
8. Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. Lancet Psychiatry 2018 Mar;5(3):237-286. [doi: 10.1016/S2215-0366(17)30513-8] [Medline: 29482764]
9. S. Bhuyan S, Kim H, Isehunwa OO, Kumar N, Bhatt J, Wyant DK, et al. Privacy and security issues in mobile health: current research and future directions. Health Policy Technol 2017 Jun;6(2):188-191. [doi: 10.1016/j.hlpt.2017.01.004]
10. Academy of Medical Sciences. London: Academy of Medical Sciences; 2018 Nov 14. Our data-driven future in healthcare: People and partnerships at the heart of health related technologies URL: https://acmedsciacuk/file-download/74634438 [accessed 2019-03-08] [WebCite Cache ID 76ieMWDMu]
11. Castell S, Ashford H. Academy of Medical Sciences. London: Academy of Medical Sciences; 2018 Nov 14. Future data-driven technologies and the implications for use of patient data: Dialogue with public, patients and healthcare professionals URL: https://acmedsci.ac.uk/file-download/6616969 [accessed 2019-03-08] [WebCite Cache ID 76ieZeeju]
12. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
13. Federal Trade Commission. Washington: Federal Trade Commission; 2016. Interactive Mobile Health Apps Tool: Developing a mobile health app? URL: https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-apps-interactive-tool [accessed 2019-03-08] [WebCite Cache ID 76iekNRA0]
14. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
15. British Standards Institution. London: Innovate UK; 2015. Health and wellness apps: Quality criteria across the life cycle - Code of practice URL: https://shop.bsigroup.com/upload/271432/PAS%20277%20(2015)bookmarked.pdf [accessed 2019-03-08] [WebCite Cache ID 76ieq1A8B]

16. Naylor C, Knapp M, McDaid D. The Kings Fund. London; 2012. Long-term conditions and mental health: The cost of co-morbidities URL: https://www.kingsfund.org.uk/publications/long-term-conditions-and-mental-health [accessed 2019-03-08] [WebCite Cache ID 76if3LYCM]
17. Zuidersma M, Ormel J, Conradi HJ, de Jonge P. An increase in depressive symptoms after myocardial infarction predicts new cardiac events irrespective of depressive symptoms before myocardial infarction. Psychol Med 2012 Apr;42(4):683-693. [doi: 10.1017/S0033291711001784] [Medline: 22571951]
18. Ismail K, Winkley K, Stahl D, Chalder T, Edmonds M. A cohort study of people with diabetes and their first foot ulcer: the role of depression on mortality. Diabetes Care 2007 Jun;30(6):1473-1479. [doi: 10.2337/dc06-2313] [Medline: 17363754]
19. Bhugra D, Tasman A, Pathare S, Priebe S, Smith S, Torous J, et al. The WPA-Lancet Psychiatry Commission on the future of psychiatry. Lancet Psychiatry 2017 Oct;4(10):775-818. [doi: 10.1016/S2215-0366(17)30333-4] [Medline: 28946952]
20. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
21. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
22. Food and Drug Administration. Washington: Food and Drug Administration; 2015 Feb 9. Mobile Medical Applications: Guidance for industry and food and drug administration staff URL: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf [accessed 2019-03-08] [WebCite Cache ID 76ijZkMNC]
23. Food and Drug Administration. Washington: Food and Drug Administration; 2016 Jul 16. Mobile medical applications URL: https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-apps-interactive-tool [accessed 2019-03-08] [WebCite Cache ID 76iiqocVS]
24. Campbell AN, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. Am J Psychiatry 2014 Jun;171(6):683-690 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13081055] [Medline: 24700332]
25. Care Quality Commission. London: Care Quality Commission; 2017. Clarification of regulatory methodology: PMS digital healthcare providers URL: https://www.cqc.org.uk/file/1295582 [accessed 2019-03-08] [WebCite Cache ID 76iizMZ0G]
26. Department for Business EIS. London: Department for Business EIS; 2012 Oct 08. CE Marking: How a product complies with EU safety, health and environmental requirements, and how to place a CE marking on your product URL: https://www.gov.uk/guidance/ce-marking [accessed 2019-03-08] [WebCite Cache ID 76ijmm6yp]
27. National Health Service. London: National Health Service; 2018 Aug 25. NHS Apps Library: Find digital tools to help you manage and improve your health URL: https://www.nhs.uk/apps-library/ [accessed 2019-03-08] [WebCite Cache ID 76iqYsPsw]
28. Food and Drug Administration. Washington: Food and Drug Administration; 2019. Digital Health Software Precertification (Pre-Cert) Program URL: https://www.fda.gov/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/default.htm [accessed 2019-03-08] [WebCite Cache ID 76ik9N0Gb]
29. Apple Inc. California: Apple; 2019. App Store Review Guidelines URL: https://developer.apple.com/app-store/review/guidelines/ [accessed 2019-03-08] [WebCite Cache ID 76ikKLv3J]
30. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
31. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Dec 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
32. Drake G, Csipke E, Wykes T. Assessing your mood online: acceptability and use of Moodscope. Psychol Med 2013 Jul;43(7):1455-1464 [FREE Full text] [doi: 10.1017/S0033291712002280] [Medline: 23149120]
33. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. JMIR Mhealth Uhealth 2018 Jul 9;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
34. Orcha. Daresbury; 2018. Orcha: Unlocking the power of Digital Health for the population URL: https://www.orcha.co.uk/ [accessed 2019-03-08] [WebCite Cache ID 76ikWtXEV]
35. Psyberguide. Looking for a mental health app? URL: https://psyberguide.org/ [accessed 2019-03-08] [WebCite Cache ID 76ikZu58I]
36. Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, et al. Do “Brain-Training” Programs Work? Psychological Science in the Public Interest. Psychol Sci Public Interest 2016 Oct;17(3):103-186. [doi: 10.1177/1529100616661983] [Medline: 27697851]
37. Di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient willingness to consent to mobile phone data collection for mental health apps: structured questionnaire. JMIR Ment Health 2018 Aug 29;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]

38. Hendrikoff L, Kambeitz-Ilankovic L, Pryss R, Senner F, Falkai P, Pogarell O, et al. Prospective acceptance of distinct mobile mental health features in psychiatric patients and mental health professionals. J Psychiatr Res 2019 Feb;109:126-132. [doi: 10.1016/j.jpsychires.2018.11.025] [Medline: 30530207]
39. BBC. London: BBC; 2019 Feb 18. Facebook-Cambridge Analytica data breach URL: https://www.bbc.co.uk/news/topics/c81zyn0888lt/facebook-cambridge-analytica-data-breach [accessed 2019-03-08] [WebCite Cache ID 76ikha6n0]
40. Meyer R. The Atlantic. Washington: The Atlantic; 2018 Mar 20. The Cambridge Analytica Scandal, in Three Paragraphs. What it means for Facebook, for President Trump's world, and for every American URL: https://www.theatlantic.com/technology/archive/2018/03/the-cambridge-analytica-scandal-in-three-paragraphs/556046/ [accessed 2019-03-08] [WebCite Cache ID 76il0cwkf]
41. Wikipedia. 2019 Feb 21. Facebook - Cambridge Analytica data scandal URL: https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal [accessed 2019-03-08] [WebCite Cache ID 76imTDBQs]
42. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 7;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
43. Dan L. HIStalk. 2014 Dec 10. The Rise and Fall of Happtique, mHealth?s First App Prescribing Platform URL: http://histalkmobile.com/the-rise-and-fall-of-happtique-mhealths-first-app-prescribing-platform/ [accessed 2019-03-08] [WebCite Cache ID 76imeNGg9]
44. Dolan B. Mobi Health News. 2014 Dec 10. SocialWellth acquires health app certification company Happtique URL: https://www.mobihealthnews.com/38894/socialwellth-acquires-health-app-certification-company-happtique [accessed 2019-03-08] [WebCite Cache ID 76immiaT0]
45. Pharmacy 2U. 2015 Oct 20. Statement from Pharmacy2U URL: https://www.pharmacy2u.co.uk/news/statement-oct15/ [accessed 2019-03-08] [WebCite Cache ID 76inNzqKD]
46. Ennis L, Robotham D, Denis M, Pandit N, Newton D, Rose D, et al. Collaborative development of an electronic Personal Health Record for people with severe and enduring mental health problems. BMC Psychiatry 2014 Nov 18;14:305 [FREE Full text] [doi: 10.1186/s12888-014-0305-9] [Medline: 25403285]
47. Robotham D, Mayhew M, Rose D, Wykes T. Electronic personal health records for people with severe mental illness; a feasibility study. BMC Psychiatry 2015 Aug 6;15:192 [FREE Full text] [doi: 10.1186/s12888-015-0558-y] [Medline: 26245805]
48. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access 2018 Mar;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
49. Rosenfeld L, Torous J, Vahia IV. Data security and privacy in apps for dementia: an analysis of existing privacy policies. Am J Geriatr Psychiatry 2017 Aug;25(8):873-877. [doi: 10.1016/j.jagp.2017.04.009] [Medline: 28645535]
50. Kebede M, Steenbock B, Helmer SM, Sill J, Möllers T, Pischke CR. Identifying evidence-informed physical activity apps: content analysis. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e10314 [FREE Full text] [doi: 10.2196/10314] [Medline: 30563810]
51. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
52. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
53. Baumel K, Kane J. Examining predictors of real-world user engagement with self-guided eHealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
54. Edwards EA, Caton H, Lumsden J, Rivas C, Steed L, Pirunsarn Y, et al. Creating a theoretically-grounded, gamified health app: lessons from developing the Cigbreak smoking cessation mobile phone game. JMIR Serious Games 2018 Nov 29;6(4):e10252 [FREE Full text] [doi: 10.2196/10252] [Medline: 30497994]
55. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
56. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
57. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
58. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]

59. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
60. Oulasvirta A, Tamminen S, Roto VJK. Interaction in 4-second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In: Proceedings of the SIGCHI conference on Human factors in computing systems. SIGCHI Conference on Human Factors in Computing Systems; Portland, Oregon, USA: ACM; 2005 Presented at: CHI'05; April 2-7, 2005; Portland, Oregon p. 919-928. [doi: 10.1145/1054972.1055101]
61. Davies EB, Craven MP, Martin JL, Simons L. Proportionate methods for evaluating a simple digital mental health tool. Evid Based Ment Health 2017 Nov;20(4):112-117 [FREE Full text] [doi: 10.1136/eb-2017-102755] [Medline: 28993317]
62. Oxford Health NHS Foundation Trust. BlueIce app URL: https://www.oxfordhealth.nhs.uk/blueice/ [accessed 2019-03-12] [WebCite Cache ID 76oplpzty]
63. Calm. URL: https://www.calm.com/ [accessed 2019-03-12] [WebCite Cache ID 76opxSSDs]
64. MyFitnessPal. URL: https://www.myfitnesspal.com/ [accessed 2019-03-12] [WebCite Cache ID 76oq55UrV]
65. Dario. URL: https://mydario.com/ [accessed 2019-03-12] [WebCite Cache ID 76oqIVt44]
66. Schueller S, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
67. Federal Trade Commission. 2016 Jan 15. Lumosity to Pay $2 Million to Settle FTC Deceptive Advertising Charges for Its URL: https://www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges [accessed 2019-03-08] [WebCite Cache ID 76ipAICcf]
68. Federal Trade Commission. 2015 Sep 17. FTC Charges Marketers of 'Vision Improvement' App with Deceptive Claims URL: https://www.ftc.gov/news-events/press-releases/2015/09/ftc-charges-marketers-vision-improvement-app-deceptive-claims [accessed 2019-03-08] [WebCite Cache ID 76ipKE4WS]

Abbreviations

FDA: Food and Drug Administration
MARS: Mobile App Rating Scale
MHRA: Medicines and Healthcare Products Regulatory Agency
NICE: National Institute of Health and Clinical Excellence
NIHR: National Institute for Health Research
T4T: Transparency for Trust

Edited by G Eysenbach; submitted 04.10.18; peer-reviewed by S Davis, J Rose, K Stawarz; comments to author 05.01.19; revised version received 21.01.19; accepted 09.02.19; published 02.05.19.

Please cite as:
Wykes T, Schueller S
Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of Responsible Health App Marketplaces
J Med Internet Res 2019;21(5):e12390
URL: https://www.jmir.org/2019/5/e12390/
doi:10.2196/12390
PMID:31045497

Original Paper

Using Human-Centered Design to Build a Digital Health Advisor for Patients With Complex Needs: Persona and Prototype Development

Onil Bhattacharyya[1,2], MD, PhD; Kathryn Mossman[1], PhD; Lovisa Gustafsson[3], MBA; Eric C Schneider[3], MSc, MD

[1]Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, University of Toronto, Toronto, ON, Canada

[2]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[3]Commonwealth Fund, New York City, NY, United States

Corresponding Author:
Onil Bhattacharyya, MD, PhD
Women's College Hospital
Institute for Health Systems Solutions and Virtual Care
University of Toronto
WCH, 6th Fl
76 Grenville St
Toronto, ON,
Canada
Phone: 1 416 323 6400 ext 5217
Fax: 1 416 323 7323
Email: onil.bhattacharyya@wchospital.ca

Abstract

Background: Twenty years ago, a “Guardian Angel” or comprehensive digital health advisor was proposed to empower patients to better manage their own health. This is now technically feasible, but most digital applications have narrow functions and target the relatively healthy, with few designed for those with the greatest needs.

Objective: The goal of the research was to identify unmet needs and key features of a general digital health advisor for frail elderly and people with multiple chronic conditions and their caregivers.

Methods: In-depth interviews were used to develop personas and use cases, and iterative feedback from participants informed the creation of a low-fidelity prototype of a digital health advisor. Results were shared with developers, investors, regulators, and health system leaders for suggestions on how this could be developed and disseminated.

Results: Patients highlighted the following goals: “live my life,” “love my life,” “manage my health,” and “feel understood.” Patients and caregivers reported interest in four functions to address these goals: tracking and insights, advice and information, providing a holistic picture of the patient, and coordination and communication. Experts and system stakeholders felt the prototype was technically feasible, and that while health care delivery organizations could help disseminate such a tool, it should be done in partnership with consumer-focused organizations.

Conclusions: This study describes the key features of a comprehensive digital health advisor, but to spur its development, we need to clarify the business case and address the policy, organizational, and cultural barriers to creating tools that put patients and their goals at the center of the health system.

(J Med Internet Res 2019;21(5):e10318) doi:10.2196/10318

KEYWORDS

chronic disease; user-centered design; medical applications

Introduction

Background

Twenty years ago, Peter Szolovits and colleagues proposed a digital “Guardian Angel” built on the notion that information systems designers might shift focus from serving health institutions to empowering the individual patient [1]. The Guardian Angel (or digital health advisor) would collect individual patient data, monitor health conditions, interpret health information for patients, help customize treatment plans, share information with the care team, and provide reminders and alerts about medications and appointments. With recent technological advances, the capacity to create a digital health advisor is now within reach, but it has yet to emerge. One barrier to the development of a digital health advisor might be the lack of a clear set of requirements to inform its design. Although there are a range of tools that perform some related functions and a few studies that cover some needs, they have not been presented in a way that is actionable for developers [2]. The Commonwealth Fund, a foundation whose goal is to improve access to health care, has been exploring this opportunity through its program on the information technology (IT)-enabled consumer. The fund has a particular interest in models of care for high-need, high-cost (HNHC) patients and ways to enhance coordination as well as patient and caregiver engagement [3]. We sought to determine how the concerns of HNHC patients and their caregivers might be met through digital tools in order to communicate this to the stakeholders who can create and distribute these tools. This paper describes this strategy, highlighting the use of persona development and an illustrative prototype to engage key stakeholders in order to promote the creation of a digital health advisor for the people who could derive the most benefit.

Methods

Solutions to Improve the Care of High-Need, High-Cost Patients

Information and communication technologies (ICTs) have transformed many industries, often by reconfiguring services around the needs of consumers. But health care has lagged behind. Many of the current consumer-facing applications target people who are relatively healthy [4]. There has been less focus on the 5% of the population who account for nearly 50% of all system costs in the United States and in other regions [5]. These HNHC individuals often have multiple chronic conditions and are more likely to be lower income and face housing and food insecurity [6-9]. They are poorly served by a health system built around diseases and institutions that often results in fragmented and competing care [10,11]. Among population segments, frail elderly are the most likely (46.2%) to be high cost, followed by adults under 65 years with disability and adults with major complex chronic disease [12]. They all have to engage in a variety of activities to improve their health [13,14], and digital technologies may be able to assist. As the largest subgroup, seniors are increasingly adopting technology, with 76% using cell phones, 64% using computers, 43% using the internet, and 40% using email and texting [15]. They are already engaging with some of these new approaches in health care: 16% of seniors search the internet to obtain health information, 8% to fill prescriptions, 7% to contact physicians, and 5% to handle insurance matters [15]. However, while there are more than 165,000 health apps currently available, a 2016 review found that most available digital tools for chronic disease are piecemeal, have limited functionality, and do not address the needs of patients with complex chronic conditions [4,16]. Furthermore, a review of evaluations of apps for conditions associated with higher needs found that most studies were small and few assessed process or outcome measures [17]. Another review compiled proposed functions from different studies and highlighted that tools should be developed with user needs in mind [2].

Understanding and Communicating User Needs

Human-centered design (HCD) has been used to develop many transformative ICT solutions, is routine in a range of industries, and is increasingly applied in health care [10,18-27]. HCD is a problem-focused method that emerged from the fields of industrial design and (more recently) software development [28-30]. Strengths of design thinking include rapidly developing a deep understanding of user needs and then communicating them in ways that are emotionally engaging and actionable. This approach often involves interviews, observation, and immersion in a user’s context to develop user personas, or archetypes, and use cases [22,27]. A persona is a detailed description of a fictional person (often a composite of real individuals) used to communicate the key motivations, concerns, and interests of a user group [10,26,27]. Related to the persona is a use case or user story, which is a story with a plot describing the actions and decisions of a user in a particular context [26,27]. These representations can help designers foster empathy, better understand user needs, and develop new service options and tools [26]. Persona development has been used in a number of studies, but they have focused on specific chronic conditions (eg, older adults with diabetes [31], heart disease [32,33], or multiple sclerosis [34]) or groups like high users of the emergency department [35]. One study created personas of the “oldest old” but focused on implications for the design of computer interfaces [27]. None of these studies looked at a range of HNHC patients, and none were focused on designing a general tool to help achieve patient goals like the digital health advisor.

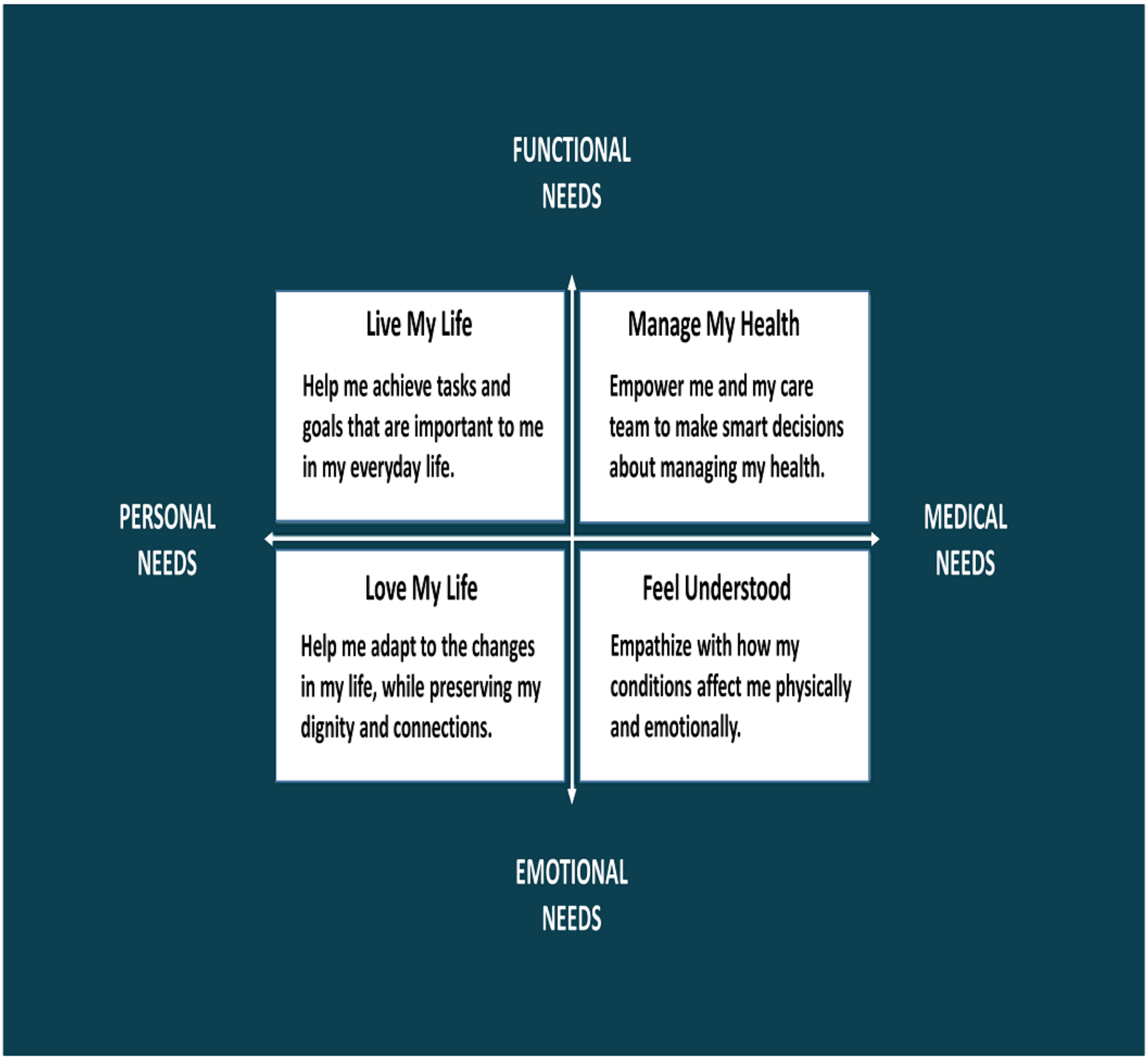
Designing a Prototype for a Digital Health Advisor

The Commonwealth Fund was interested in developing a vision of a digital health advisor that could meet the needs of HNHC patients and using this to encourage its development in the health care marketplace by developers, entrepreneurs, investors, health system providers, funders, and regulators. The Fund worked with the design firm gravitytank to identify the needs of HNHC patients and their caregivers and create a low-fidelity prototype, which simulates key features of interest to this group without building out any of the functions [36,37]. This involved interviews with experts, 8 patient-caregiver pairs of people with multiple chronic conditions or frail elderly, and mapping the workflow of 3 care managers with similar patients [36,37]. Based on analysis of these data, a rough prototype of a digital health advisor was created and shared with patients and caregivers so they could provide feedback on the usability and functionality of the tool. This informed different outputs to make the insights more actionable.

Results

First, the needs of patients with multiple chronic conditions and the frail elderly and their caregivers were characterized, focusing on the intersection of functional and emotional needs with medical and personal needs (see Figure 1). Patients felt that a tool that addressed functional needs could improve health outcomes, but if a tool addressed emotional needs they would be more likely to use it on an ongoing basis. One link between the functional and emotional was having a tool that could connect goals to functions and orient care so that it can better manage their conditions in order to increase their ability to do the things that make life worthwhile. They highlighted the challenges of communicating with family members about end of life, sharing goals, navigating apps, and approval processes. They also had difficulty understanding the wide range of uncoordinated advice they receive and following through on the subset of recommendations that address their needs. Based on these interviews, 4 patient personas and 2 caregiver personas were developed (see Textbox 1) along with 4 use cases (see Figure 2 for an example). Last, a low-fidelity prototype of a digital health advisor was developed with a number of key features (see Textbox 2, with screenshots in Multimedia Appendix 1) [38].

Figure 1. Key themes regarding patient needs.



Textbox 1. Overview of patient and caregiver personas.

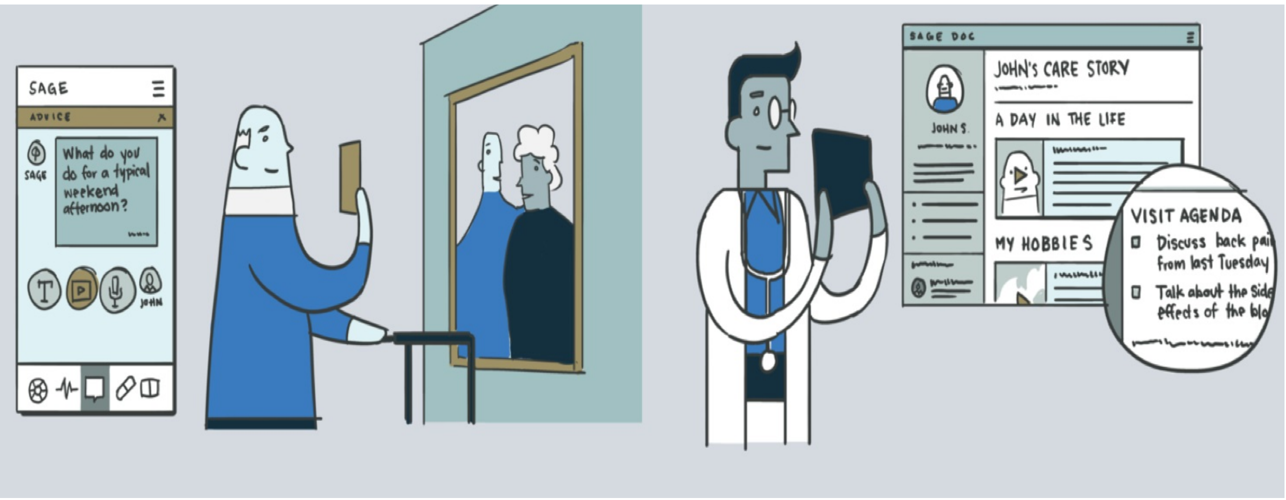
Patient personas

- “John”
    - An 87-year-old frail elderly man dealing with arthritis, hypertension, and mild cognitive impairment
    - He needs a tool to reach out to loved ones when he needs help, communicate his goals and current level of function, and give prompts for key health behaviors
- “Elizabeth”
    - A 70-year-old elderly patient with multiple chronic conditions
    - She needs help managing her medications, tracking her health status, and learning about activities to improve her health
- “Karen”
    - A 65-year-old patient with multiple chronic conditions
    - She needs help coordinating her care with her providers, managing her conditions, and accessing relevant health and social services
- “Jasmine”
    - A 44-year-old patient with a major chronic condition
    - She needs help communicating with her family about her health, managing ongoing symptoms, and learning about new treatments

Caregiver personas

- “Beth”
    - A 79-year-old caregiving spouse
    - She needs help managing her spouse’s medications and appointments and connecting with medical and community resources
- “Lisa”
    - A 45-year-old caregiver for her elderly mother
    - She needs help coordinating her mother’s care, remotely monitoring her health status, and communicating with family members regarding her mother’s health

Figure 2. Sample use case for digital health advisor: John is an 87-year-old man living in San Mateo, California, with hypertension, arthritis, and mild cognitive impairment. He has been feeling more tired, anxious, and clumsy recently. He lives with his wife, who is his primary caregiver, and gets help from his daughter. He has an upcoming appointment with a geriatrician and he’s concerned whether this new doctor will understand what his needs are. He would like to update his doctors and highlight his activities and goals to share a more holistic picture of his life with the care team. He also wants to manage his doctor appointments and keep track of what his doctor tells him.



Textbox 2. Key features of a digital health advisor for high-need, high-cost patients and their caregivers.

Tracking and insights:

- Metrics dashboard collects health data such as blood pressure, oxygen levels, and gait from sensors and connected devices or receives manual input on symptoms
- Tracks current state before a medical visit or patterns over time, provides recommendations to patients, and informs the care team of changes in health status
- Highlights patient's functional status and summarizes current symptom and risk factor control for different diseases so that care can focus on improving key functions

Advice and information:

- Personalized advice on health-related questions either through integration with a digital assistant or by connecting to a medical practitioner by text, voice, or video
- This advice would draw on information from medical records, personal metrics, insurance coverage, and available community resources
- The advice would either elicit or incorporate known preferences into recommendations

Holistic picture:

- A care journal and patient profile where patients can provide a brief image, a written summary of their story, and a set of goals and milestones they want to track
- This would include a checklist of care preferences to help create a comprehensive picture of their life to share with practitioners and foster deeper connections with their medical team and caregivers

Coordination and communication:

- A shared calendar with all appointments, a document center, and a task manager to improve coordination and communication between patients and their care team, including help with scheduling and transportation
- This would also include transparent communications between providers about a patient's care to highlight areas of ambiguity and ensure agreement between members of the team

The personas and prototypes of the digital health advisor were presented to a group of policymakers; regulators; clinicians; experts in informatics, advertising, and health services; patient advocates; technology industry executives; and health system leaders. They felt that most components are technically feasible using current technology, although further development is needed in decision support. Patients and participants from the advertising industry highlighted the need for relatability of the tool and a "stickiness" that made it the first thing patients turn to when they have a question about their health. They highlighted the need for a simple visual interface and opportunities for voice activation to benefit those with low literacy or cognitive challenges. Health system leaders highlighted how a dashboard with trends and the current state could make visits more efficient for people with complex conditions, by reducing the time needed to collect this information, leaving more time to discuss how to advance particular goals of care. They also mentioned how a tool to share detailed but actionable information on goals, functional status, physiologic parameters, and current burden of treatment could facilitate challenging therapeutic decisions involving multiple providers. Representatives from health care delivery organizations were felt to be critical to the development of a digital health advisor but stressed that it would also need to foster support in target communities, including patients, caregivers, families, and social networks. They stated that extensive collaboration is needed to develop a robust digital health advisor, which should be an integrated suite of tools. They also highlighted that novel analytics are not sufficient; data must be made available in a usable format, and the necessary data for real-time decision making (eg, cost, quality, availability of services) are currently held by many different groups in different formats. They suggested a useful digital health advisor is more likely to emerge in response to partnerships with organizations focused on consumer needs, such as consumer advocacy groups and large retail companies. Business cases and policy incentives need to be developed to encourage broad data sharing among community, government, and commercial initiatives.

Discussion

Designing Tools and Systems That Will Meet Patient Needs

The concept of a comprehensive digital health advisor for patients, which was proposed more than 20 years ago, is now technically feasible and attractive to HNHC patients in the United States. The patients highlight a range of functional, emotional, medical, and personal needs that might be addressed by this tool. They also list a series of key functions such as tracking, advice, providing a holistic picture of themselves, and coordination and communication. The use of human-centered design helps overcome the limitation that new designs in health care are often dreamed up by providers to be easy to implement but not necessarily use by people who rarely use health services and don't resemble the people they are trying to help. Personas can promote stakeholder engagement by providing detailed information on patient needs and how a tool might address those needs, making it easier to understand key functions, possible interactions with the health system, data requirements, and potential regulatory concerns.

While some studies have employed human-centered design to develop tools for patients with multiple chronic conditions [39,40], this approach tends to be applied to specific populations (eg, people with diabetes) [26,41,42] or specific uses (eg, communication, monitoring) [43,44]. Our work supplements the findings of a recent review of mobile technologies for older adults that summarized the following design features from different articles: graphs; notification systems; text and video messaging; scheduling; and vision, hearing, and memory aids [2]. Even though there are many digital health tools on the market, few comprehensively address the needs of HNHC patients and even those tend to focus on medically defined needs rather than act as general advisors. To go from an understanding of needs to a functioning tool that meets those needs in the context of someone's life is a stretch and it involves testing different use cases, functions, and target groups before finding a good fit. The personas and prototype helped elicit specific advice from providers, payers, regulators, developers, and investors by making the problems and options more tangible. This highlighted the range of challenges with trust, uptake, integration with workflows, value proposition for different stakeholders, and potential business models, going beyond a narrow focus on end user needs and features of the technology.

The learnings from this project about current HNHC patients could be applied to vulnerable populations at risk of becoming HNHC patients (another target group for the Commonwealth Fund) and even the general public. We found the highest users of the health system are interested in tools that can help them manage their health challenges but even more in helping them live their lives to the fullest. This is likely to be true of all patients, who not only value disease management but also see health as a means to identify and work toward their life goals. An individual's health is intertwined with their psychological, social, and economic context, and truly useful digital tools will help people manage their needs in a comprehensive and integrated way rather than focus on a disease or issue in isolation. This has not been a major goal for health systems or a major focus of quality improvement efforts; however, it could be supported by consumer-facing tools that engage and empower individuals.

Conclusion

Characterizing the needs of target users is essential to designing a comprehensive digital health advisor, but to realize this vision we need to address broader system and policy constraints that will impact its development. User-centered design can help draw attention to something that is technically feasible but has no business model and may not be compatible with many current regulations. It helps reframe problems and suggest solutions that may be independent of existing services or processes. In the area of models of care for HNHC patients, the Commonwealth Fund has commissioned rigorous evaluations of new models, synthesized evidence from the scientific literature, and created a playbook for health systems to help implement promising models and improve care for this group. We have also started using design thinking as a tool to generate a pipeline of new options that might constitute a breakthrough. We move this future-oriented work forward through our role as a convener and advocate for improving access to care. For the IT-enabled consumer, the next steps include exploring the type of collaboration needed between stakeholders such as health care delivery organizations, patients, caregivers, technology companies, government, and consumer advocacy groups. We also need to examine the incentives and business models needed to attract entrepreneurs and developers to work in this area and for health systems to engage with consumer-facing IT tools. System change is unlikely to come from those who already run it, so the Commonwealth Fund is working to address the policy and regulatory, business model, and cultural barriers to creating tools that put patients and their goals at the center of the health system.

Acknowledgments

The authors wish to acknowledge the following individuals at gravitytank (now Salesforce) for their contributions to this research: Susie Kang, Treanndis Hurst, Stephanie Szabo, and Dave Saunders. This research was conducted with financial support from the Commonwealth Fund and the Gordon and Betty Moore Foundation. The authors of this article are responsible for its contents. No statement in this article should be construed as an official position of The Commonwealth Fund, the Moore Foundation, or gravitytank (now Salesforce).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Patient and caregiver personas.

[PDF File (Adobe PDF File), 852KB - jmir_v21i5e10318_app1.pdf]

References

1. Szolovits P, Doyle J, Long W. Guardian Angel: patient-centered health information systems. Technical report MIT/LCS/TR-604.: Massachusetts Institute of Technology Laboratory for Computer Science; 1994. URL: http://groups.csail.mit.edu/medg/projects/ga/manifesto/GAtr.pdf [accessed 2019-03-01] [WebCite Cache ID 76YU4UkPH]
2. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
3. Designing a high-performing health care system for patients with complex needs: ten recommendations for policymakers—expanded and revised edition.: The Commonwealth Fund; 2018. URL: https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_fund_report_2017_aug_roland_10_recommendations_for_complex_patients_revisedexpanded.pdf [accessed 2018-01-08]
4. Aitken M, Lyle J. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. Parsippany: IMS Institute for Healthcare Informatics; 2015.
5. Cohen S, Yu W. The concentration and persistence in the level of health expenditures over time: estimates for the US population, 2008-2009.: Agency for Healthcare Research and Quality URL: https://meps.ahrq.gov/data_files/publications/st354/stat354.shtml [accessed 2019-02-19] [WebCite Cache ID 76JKBVtzc]
6. Reid R, Evans R, Barer M, Sheps S, Kerluke K, McGrail K, et al. Conspicuous consumption: characterizing high users of physician services in one Canadian province. J Health Serv Res Policy 2003 Oct;8(4):215-224. [doi: 10.1258/135581903322403281] [Medline: 14596756]
7. Guilcher SJT, Bronskill SE, Guan J, Wodchis WP. Who are the high-cost users? A method for person-centred attribution of health care spending. PLoS One 2016;11(3):e0149179 [FREE Full text] [doi: 10.1371/journal.pone.0149179] [Medline: 26937955]
8. Lemstra M, Mackenbach J, Neudorf C, Nannapaneni U. High health care utilization and costs associated with lower socio-economic status: results from a linked dataset. Can J Public Health 2009;100(3):180-183. [Medline: 19507718]
9. Fitzpatrick T, Rosella LC, Calzavara A, Petch J, Pinto AD, Manson H, et al. Looking beyond income and education: socioeconomic status gradients among future high-cost users of health care. Am J Prev Med 2015 Aug;49(2):161-171 [FREE Full text] [doi: 10.1016/j.amepre.2015.02.018] [Medline: 25960393]
10. Vaillancourt S, Shahin I, Aggarwal P, Pomedli S, Hayden L, Pus L, et al. Using archetypes to design services for high users of healthcare. Healthc Pap 2014;14(2):37-41. [Medline: 25880862]
11. Zulman DM, Asch SM, Martins SB, Kerr EA, Hoffman BB, Goldstein MK. Quality of care for patients with multiple chronic conditions: the role of comorbidity interrelatedness. J Gen Intern Med 2014 Mar;29(3):529-537 [FREE Full text] [doi: 10.1007/s11606-013-2616-9] [Medline: 24081443]
12. Joynt KE, Figueroa JF, Beaulieu N, Wild RC, Orav EJ, Jha AK. Segmenting high-cost Medicare patients into potentially actionable cohorts. Healthc (Amst) 2017 Mar;5(1-2):62-67. [doi: 10.1016/j.hjdsi.2016.11.002] [Medline: 27914968]
13. May C, Montori VM, Mair FS. We need minimally disruptive medicine. BMJ 2009;339:b2803. [Medline: 19671932]
14. Pefoyo AJK, Bronskill SE, Gruneir A, Calzavara A, Thavorn K, Petrosyan Y, et al. The increasing burden and complexity of multimorbidity. BMC Public Health 2015;15:415 [FREE Full text] [doi: 10.1186/s12889-015-1733-2] [Medline: 25903064]
15. Levine DM, Lipsitz SR, Linder JA. Trends in seniors' use of digital health technology in the United States, 2011-2014. JAMA 2016 Aug 2;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
16. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 01;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
17. Singh K, Drouin K, Newmark LP, Filkins M, Silvers E, Bain PA, et al. Patient-facing mobile apps to treat high-need, high-cost populations: a scoping review. JMIR Mhealth Uhealth 2016 Dec 19;4(4):e136 [FREE Full text] [doi: 10.2196/mhealth.6445] [Medline: 27993761]
18. Sy D. Adapting usability investigations for agile user-centered design. J Usability Stud 2007;2(3):112-132 [FREE Full text]
19. Olson GM, Olson JS. User‐centered design of collaboration technology. J Organ Comput 1991 Jan;1(1):61-83. [doi: 10.1080/10919399109540150]
20. Hassenzahl M, Tractinsky N. User experience: a research agenda. Behav Inf Technol 2006 Mar;25(2):91-97. [doi: 10.1080/01449290500330331]
21. Corry MD, Frick TW, Hansen L. User-centered design and usability testing of a web site: an illustrative case study. Educ Technol Res Dev Dec 1997 Dec;45(4):65-76. [doi: 10.1007/BF02299683]
22. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
23. Robert G, Cornwell J, Locock L, Purushotham A, Sturmey G, Gager M. Patients and staff as codesigners of healthcare services. BMJ 2015 Feb;350:g7714. [Medline: 25670179]
24. Zuckerman B, Margolis PA, Mate KS. Health services innovation: the time is now. JAMA 2013 Mar 20;309(11):1113-1114. [doi: 10.1001/jama.2013.2007] [Medline: 23512055]

25. Eckman M, Gorski I, Mehta K. Leveraging design thinking to build sustainable mobile health systems. J Med Eng Technol 2016;40(7-8):422-430. [doi: 10.1080/03091902.2016.1218560] [Medline: 27535325]
26. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
27. Turner AM, Reeder B, Ramey J. Scenarios, personas and user stories: user-centered evidence-based design representations of communicable disease investigations. J Biomed Inform 2013 Aug;46(4):575-584 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.006] [Medline: 23618996]
28. Brown T. Design thinking. Harv Bus Rev 2008 Jun;6:1-9 [FREE Full text]
29. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications; 2004.
30. Norman D, Draper S. User centered system design. In: New Perspectives on Human-Computer Interaction. Hillsdale: L. Erlbaum Associates Inc; 1986.
31. van der Molen P, Maas AH, Chen W, van Pul C, Cottaar EJ, van Riel NA, et al. Identifying user preferences for a digital educational solution for young seniors with diabetes. Diabetes Spectr 2017 Feb 09;30(3):182-187. [doi: 10.2337/ds16-0023]
32. Holden RJ, Kulanthaivel A, Purkayastha S, Goggins KM, Kripalani S. Know thy eHealth user: development of biopsychosocial personas from a study of older adults with heart failure. Int J Med Inform 2017 Dec;108:158-167. [doi: 10.1016/j.ijmedinf.2017.10.006] [Medline: 29132622]
33. Serio CD, Hessing J, Reed B, Hess C, Reis J. The effect of online chronic disease personas on activation: within-subjects and between-groups analyses. JMIR Res Protoc 2015 Feb 25;4(1):e20 [FREE Full text] [doi: 10.2196/resprot.3392] [Medline: 25720676]
34. Giunti G, Kool J, Rivera RO, Dorronzoro ZE. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
35. Jacob R, Wong ML, Hayhurst C, Watson P, Morrison C. Designing services for frequent attenders to the emergency department: a characterisation of this population to inform service design. Clin Med (Lond) 2016 Aug;16(4):325-329 [FREE Full text] [doi: 10.7861/clinmedicine.16-4-325] [Medline: 27481374]
36. Shah A, Gustafsson L, Bhattacharyya O. How a digital health advisor could help high-need, high-cost patients and their caregivers. New York: The Commonwealth Fund; 2016. URL: http://www.commonwealthfund.org/publications/blog/2016/dec/digital-health-advisor [accessed 2019-02-19] [WebCite Cache ID 76JKkeNUp]
37. Bhattacharyya O, Shah A, Gusafsson L. A digital health advisor for high-need, high-cost patients; exploring needs, functions and system constraints. 2017 Presented at: AMIA Annual Symposium; 2017; Washington.
38. Experience Sage. URL: http://sage-dha.squarespace.com/index/ [accessed 2019-02-19] [WebCite Cache ID 76JL85uzR]
39. Steele GC, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
40. Lyles CR, Altschuler A, Chawla N, Kowalski C, McQuillan D, Bayliss E, et al. User-centered design of a tablet waiting room tool for complex patients to prioritize discussion topics for primary care visits. JMIR Mhealth Uhealth 2016 Sep 14;4(3):e108 [FREE Full text] [doi: 10.2196/mhealth.6187] [Medline: 27627965]
41. Ahern DK, Parker D, Eaton C, Rafferty C, Wroblewski J, Goldman R. Patient-facing technology for identification of COPD in primary care. J Innov Health Inform 2016 Jul 15;23(2):824 [FREE Full text] [Medline: 27869583]
42. De Vito Dabbs A, Myers BA, McCurry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
43. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
44. Ferreira F, Almeida N, Rosa A, Oliveira A, Casimiro J, Silva S, et al. Elderly centered design for interaction: the case of the S4S medication assistant. Procedia Comput Sci 2014;27:398-408. [doi: 10.1016/j.procs.2014.02.044]

Abbreviations

HCD: human-centered design
HNHC: high-need, high-cost
ICT: information and communication technologies
IT: information technology

Edited by G Eysenbach; submitted 06.03.18; peer-reviewed by W Zhang, K Blondon; comments to author 05.08.18; revised version received 28.11.18; accepted 28.12.18; published 09.05.19.

Please cite as:
Bhattacharyya O, Mossman K, Gustafsson L, Schneider EC
Using Human-Centered Design to Build a Digital Health Advisor for Patients With Complex Needs: Persona and Prototype Development
J Med Internet Res 2019;21(5):e10318
URL: https://www.jmir.org/2019/5/e10318/
doi:10.2196/10318
PMID:31094334

Viewpoint

Health Data Processes: A Framework for Analyzing and Discussing Efficient Use and Reuse of Health Data With a Focus on Patient-Reported Outcome Measures

Niels Henrik Ingvar Hjollund[1,2], MD, PhD; José Maria Valderas[3], MD, PhD; Derek Kyte[4,5], PhD; Melanie Jane Calvert[4,5], PhD

[1]Occupational Medicine, University Research Clinic, AmbuFlex/WestChronic, Aarhus University, Herning, Denmark

[2]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark

[3]University of Exeter Collaboration for Academic Primary Care, Health Services & Policy Research Group, National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care (South West Peninsula), University of Exeter, Exeter, United Kingdom

[4]Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, United Kingdom

[5]National Institute for Health Research Birmingham Biomedical Research Centre and National Institute for Health Research Surgical Reconstruction and Microbiology Research Centre, University of Birmingham, Edgbaston, Birmingham, United Kingdom

Corresponding Author:
Niels Henrik Ingvar Hjollund, MD, PhD
Occupational Medicine, University Research Clinic
AmbuFlex/WestChronic
Aarhus University
Gl Landevej 61
Herning, DK-7400
Denmark
Phone: 45 25850902
Email: niehjo@rm.dk

Abstract

The collection and use of patient health data are central to any kind of activity in the health care system. These data may be produced during routine clinical processes or obtained directly from the patient using patient-reported outcome (PRO) measures. Although efficiency and other reasons justify data availability for a range of potentially relevant uses, these data are nearly always collected for a single specific purpose. The health care literature reflects this narrow scope, and there is limited literature on the joint use of health data for daily clinical use, clinical research, surveillance, and administrative purposes. The aim of this paper is to provide a framework for discussing the efficient use of health data with a specific focus on the role of PRO measures. PRO data may be used at an individual patient level to inform patient care or shared decision making and to tailor care to individual needs or group-level needs as a complement to health record data, such as that on mortality and readmission, in order to inform service delivery and measure the real-world effectiveness of treatment. PRO measures may be used either for their own sake, to provide valuable information from the patient perspective, or as a proxy for clinical data that would otherwise not be feasible to collect. We introduce a framework to analyze any health care activity that involves health data. The framework consists of four data processes (patient identification, data collection, data aggregation and data use), further structured into two dichotomous dimensions in each data process (level: group vs patient; timeframe: ad hoc vs systematic). This framework is used to analyze various health activities with respect to joint use of data, considering the technical, legal, organizational, and logistical challenges that characterize each data process. Finally, we propose a model for joint use of health data with data collected during follow-up as a base. Demands for health data will continue to increase, which will further add to the need for the concerted use and reuse of PRO data for parallel purposes. Repeated and uncoordinated PRO data collection for the same patient for different purposes results in misuse of resources for the patient and the health care system as well as reduced response rates owing to questionnaire fatigue. PRO data can be routinely collected both at the hospital (from inpatients as well as outpatients) and outside of hospital settings; in primary or social care settings; or in the patient’s home, provided the health informatics infrastructure is in place. In the future, clinical settings are likely to be a prominent source of PRO data; however, we are also likely to see increased remote collection of PRO data by patients in their own home (telePRO). Data collection for research and quality surveillance will have to adapt to this circumstance and adopt complementary data capture methods that take advantage of the utility of PRO data collected during daily clinical practice. The European Union’s regulation with respect to the protection of personal data—General Data Protection Regulation—imposes severe restrictions on the use of health data for parallel purposes, and steps should be taken to alleviate the consequences while still protecting personal data against misuse.

(J Med Internet Res 2019;21(5):e12412) doi:10.2196/12412

KEYWORDS

medical informatics; patient-reported outcome; patient-physician relationship; data collection

Introduction

Health information is central to all types of activities in the health care system, all of which involve collecting, analyzing, or using health information [1]. Securing personal data against misuse is the background for several legal initiatives, for instance, the implementation of the European Union’s General Data Protection Regulation (GDPR) [2]. One key element of this regulation is the principle that personal data collected for one purpose may not be immediately transferred and used for other purposes. However, while misuse of personal data poses a severe ethical problem, so does waste and duplicate collection of the same data from the same patients due to legal, organizational, and technical dysfunction. In addition, from the patient’s perspective, duplicate collection of data may be unnecessarily burdensome and time consuming, and the possibilities and advantages of alternative uses of health data should therefore be considered. We have discussed the patient’s perspective of joint use in more detail elsewhere [3].

Health information may be generated as an integrated part of health care activities, such as biochemical variables or entries in hospitals’ electronic health record (EHR) system, or it can obtained directly from the patient. The latter is the case for patient-reported outcome (PRO) measures, which have been defined by the US Food and Drug Administration as measurements “of any aspect of a patient’s health that comes directly from the patient, without interpretation of the patient’s responses by a physician or anyone else” [4]. This definition emphasizes the standardization of PRO data as opposed to unstructured clinician-reported summaries of patient history contained in the notes in patients’ health records.

The evaluation of treatment outcomes for each individual patient is typically captured by a combination of biological data, physical examination, and communication with the patient. However, evaluations of treatment outcomes at a group level (defined geographically, administratively, epidemiologically, or at the facility level) often focus solely on mortality; readmission; and, if available, data such as medicine use and other use of health services. Although these outcomes are undeniably important, they may fail to fully capture treatment outcomes. PRO measures can be used to complement such data as a primary or an additional distal outcome, or even serve as a proxy for an unmeasured clinical variable when collection of the latter is not feasible [5].

Health informatics aims to respond to the increasing demands of systematic collection and processing of data to inform individual patient care, service improvements, and precision medicine. A lot of effort and resources are expended on collecting, processing, storing, and retrieving health information (both PRO measures and other clinical measurements) such as in hospitals’ EHR systems. In parallel, an increasing number of research projects and initiatives independently collect health information for their own specific objectives. Although health informatics, as a discipline, engages with stakeholders from a wide range of professional backgrounds, roles, and interests, it mostly does so with a focus on one specific single application (clinical practice, clinical research, administrative purposes, surveillance, or computer science), as evidenced in textbooks and the relevant literature [1]. As a consequence, there is limited literature on the joint use of health information for several purposes.

Technical, legal, organizational, and other types of obstacles to the availability of data for multiple purposes result in inefficient use of resources among patients and clinicians as well as in the health care system and society. Where there is no additional benefit from repeating a measurement, the same health information should be collected only once. A typical example would be laboratory tests, which may be performed by the family doctor before referral to the hospital, but which may be repeated, in many circumstances, unnecessarily, once the patient arrives at the hospital. Similarly, clinicians frequently struggle to retrieve measurements needed for maintaining quality registers even though they may already be recorded in the EHR system. Similarly, PRO measurements may be repeatedly and independently collected in parallel for different purposes, such as clinical management, quality surveillance, and research projects. This may not just mean a waste of resources and an unnecessary burden to patients, but may also have implications for data quality, as response fatigue may lead to reduced response rates.

To qualify the discussion of efficient use of health data, we need a common language usable for all stakeholders, which does not exist. The aim of this paper is to propose a framework for analysis of use and reuse of health information, with a specific focus on the role of PRO measures in order to initiate and facilitate a more precise discussion.

Definitions

There is no consensus on the method to define health information and health data. All definitions rely on the concepts of *information* (facts about a situation, person, event, etc [6]) and the *organized* property of data. Data have been defined accordingly in various ways such as (1) “any organized information collected by a researcher” [7], (2) “information or knowledge represented or coded in some form suitable for better usage or processing” [8], or (3) “information, especially facts and numbers, collected to be examined and considered and used to help decision making” [6]. The first definition focuses on the collection process and excludes purposes other than research, while the second one relies on data structure only. The third one identifies three processes relevant to the health data: collection, examination, and use; it furthermore acknowledges that the nature of data is preserved even if they are only stored and not used, at least not immediately.

In this paper, we use the third definition and differentiate three data processes: *data collection*, *data aggregation*, and *data use*. As patients are the unit of observation for health data, we need to additionally consider a *patient-identification* process to define whose data will be collected. We will focus on persons who may have, or are under surveillance for, a health condition and use the term “patient” even though some may not have a medical diagnosis. A generic model for health data covering any patient-related health data *activity* is shown in Figure 1. Definitions are summarized in Textbox 1.

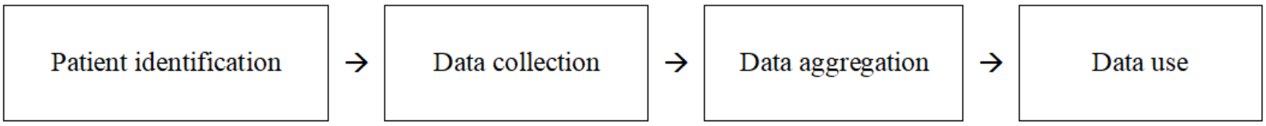
The *patient-identification* process corresponds to the definition of the patient or the group of patients that will be the ultimate source of data. The subsequent *data collection* process contains measurement methods for generating data for that patient or population of patients as well as logistic issues. In the *data aggregation* process, data are transferred, organized, and transformed to enable their subsequent use. Aggregation may include data logistic procedures like transmission, data reformatting, and data management procedures such as combining and merging with other data. The aggregation may be explicit during data management (eg, a specific data manager making the dataset ready for the researcher’s use) or implicit (eg, such as a clinical summary based on patient data in an emergency room). In the *data use* process, the aim for the actual health data activity is fulfilled (eg, publish the results or a clinical decision of a treatment plan). Any pair of consecutive data processes may be repeated and make take place simultaneously.

Two dimensions can be recognized across all four data processes: *level* and *timeframe*. Level may be either the *individual* patient or a defined patient *group* level (eg, patients admitted to a hospital department or patients with a specific health condition). Timeframe considers the scope of the health data activity and may be either *ad hoc* or part of a *systematic* planned process. Examples are provided below.

Based on these dimensions, 2 × 2 tables with four cells may be constructed for each of the four processes. Four basic health data activities may be defined, where the same cell is used in all the four data processes in Figure 1. Figure 2 shows such examples.

In most heath data activities, different cells are applied in the four health data processes, and these patterns will be analyzed to highlight their properties and differences.

Figure 1. The four data processes in the lifespan of patient-related health data. Patient identification process: Identification of patient(s) from whom data are to be collected. Data collection process: The actual collection of health data including logistic procedures. Data aggregation process: Management and organization of collected data for the data use process. Data use process: Use of the health data for the purpose of the specified activity. Each process may be repeated or may take place simultaneously with the previous process. Further information is provided in Textbox 1.



Textbox 1. Definitions of health data terms.

Health data: Health information about individual patients

Health data activity: An activity with a health-related aim that uses or produces health data

Health data processes: Any health data activity includes four processes:

- Patient identification process: Identification of patient(s) from whom data are to be collected
- Data collection process: The collection of health data
- Data aggregation process: Transfer or organization of health data in a way that enables data use
- Data use process: Use of health data for the purpose of a specified health data activity

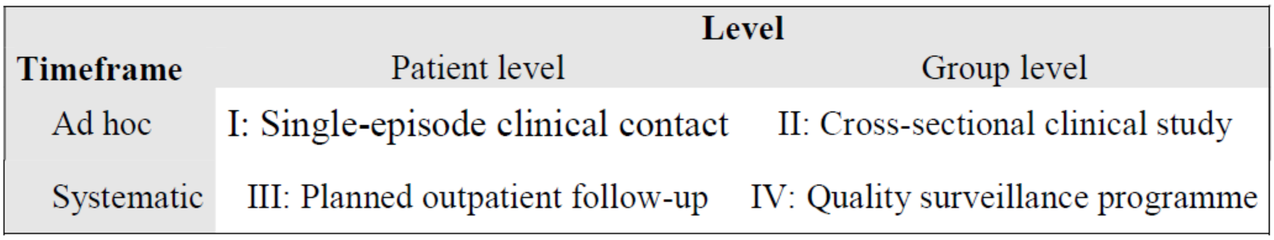
Timeframe: The timeframe of a health data process:

- Systematic: A planned or repeated health data process
- Ad hoc: A nonplanned process

Level: The level of a health data process:

- Patient level: The individual patient level
- Group level: A level with patients grouped according to some defined criteria

Figure 2. Examples of basic health data activities, where the same cell is used in all the four health data processes. I: The patient makes an appointment, and during the consultation, data are collected and aggregated to make a clinical decision and treatment plan (all four data processes ad hoc at the patient level). II: The target population is identified and data are collected, managed, analyzed, and published (all data processes ad hoc at a group level). III: An inpatient is discharged and referred for continuous planned outpatient follow-up and data are collected during follow-up, aggregated at each visit, and used at the visit (all data processes are systematic at the patient level) IV: Patient groups are identified repeatedly (eg, once a year) based on some criteria and data are collected, managed, and analyzed/reported (all data processes take place systematic at a group level).



	Level	
Timeframe	Patient level	Group level
Ad hoc	I: Single-episode clinical contact	II: Cross-sectional clinical study
Systematic	III: Planned outpatient follow-up	IV: Quality surveillance programme

Basic Health Data Activities

Each data process of basic health data activities (Figure 2) is described below and displayed in Table 1. A description of the contents of each process is shown in the Multimedia Appendix 1.

The Single-Episode Clinical Contact Activity

The patient makes an appointment with the general practitioner. During the consultation, data are collected by medical history and physical examination. These data and a general view of the patient and his/her resources are aggregated by the general practitioner into a conclusion and used for a clinical decision. Another single-episode example is an emergency room visit. The timeframe for all the four data processes listed in Figure 1 is *ad hoc*, and all take place at the *patient level*. The content of the data collection process may include systematic methods such as standardized blood tests or use of a specific validated questionnaire, in which case, the timeframe is *ad hoc*.

Table 1. Examples of basic and complex health data activities divided by level and timeframe. In basic health data activities, all four processes are in the same level/timeframe cell.

Health data process	Patient identification				Data collection				Data aggregation				Data use			
	Patient		Group		Patient		Group		Patient		Group		Patient		Group	
	Ad hoc	Sys[a]	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys
With basic data process patterns																
Single-episode clinical contact	✓[b]				✓[b]				✓[b]				✓[b]			
Planned patient follow-up		✓[b]				✓[b]				✓[b]				✓[b]		
Clinical research (cross-sectional)			✓[b]				✓[b]				✓[b]				✓[b]	
Quality surveillance program				✓[b]				✓[b]				✓[b]				✓[b]
With complex data process patterns																
Clinical research (cohort)			✓[b]					✓[b]			✓[b]				✓[b]	
Clinical guideline			✓[b]	✓				✓[b]			✓[b]	✓	✓[b]			
Individual prognosis forecast			✓[b]	✓				✓[b]	✓[b]		✓[b]		✓[b]			
Screening program				✓[b]				✓[b]		✓[b]			✓[b]			
Disease surveillance	✓	✓		✓[b]		✓[b]						✓[b]				✓[b]
Health care error surveillance	✓[b]				✓[b]				✓			✓[b]	✓			✓[b]
Primary health care, traditional	✓[b]	✓			✓[b]				✓[b]				✓[b]	✓		
Primary health care, new trend	✓[b]	✓		✓	✓[b]	✓		✓	✓[b]	✓		✓	✓[b]	✓		✓

[a] Sys: Systematic or repeated data process.

[b] The most frequently applied data processes.

The Planned Outpatient Follow-Up Activity

Many patients with a chronic disease have systematic follow-ups in an outpatient clinic. The patients are referred to outpatient follow-up in a systematic manner based on written formal or local informal guidelines. The data needed for the outpatient consultation (eg, medical history, laboratory tests, PRO measures, and physical examination) are collected, aggregated, and used at the patient level in relation to each visit. The timeframe for all the four data processes is *systematic* and takes place at the *patient level*.

The Cross-Sectional Clinical Research Activity

In a cross-sectional study, the group of patients is defined once, the data are collected and analyzed once, and the results based on condensed data are published once. Another example is a registry-based study. The timeframe for all four data processes is *ad hoc* and takes place at a *group level*.

The Quality Surveillance Program Activity

The quality surveillance program is an ongoing activity, where at a defined timepoint, (eg, once a year), data are sampled and subsequently analyzed with respect to differences between departments and hospitals. Reports are published and used for optimizing quality of care or to inform the patient's choice of health care provider. The timeframe for all four data processes is *systematic* with predefined intervals and take place at a *group level*.

Complex Health Data Activities

Although the abovementioned activities apply data processes in the same cell in all four data processes, most health data activities combine different cells. Table 1 (lower part) shows examples of such activities. The list is not comprehensive but represents examples of the possible combinations of data process patterns.

The Longitudinal Clinical Research Activity

The patient-identification process—the recruitment (eg, for a clinical trial)—takes place once, or in the case of an open cohort study, systematically over a long inclusion period. Data are collected *systematically* over time according to a defined study protocol. Aggregation (data management and analysis) and use (publication) take place only once. All data processes are *systematic* and take place at a *group level*. As discussed later, these data can be used for a range of other purposes.

The Clinical Guideline Activity

Clinical guidelines are based on meta-analyses of clinical trials and longitudinal studies collected at a *group level*. Data are aggregated to inform the guidelines and are published and implemented once or at regular intervals. The use of clinical guidelines is, however, most often *ad hoc* at the individual *patient level* when clinical decisions about diagnostic procedures and treatment are made at the "bedside" together with that specific patient. In many countries, the traditional *ad hoc* use of guidelines is being replaced by quality programs or pay-per performance systems with the purpose of implementing the guidelines for *all* relevant patients. This will move the data use process from *ad hoc* to *systematic* at the *patient level*.

The Individual Prognosis Activity

Like treatment guidelines, prognostic indicators rely on information collected at the *group level*. Prognostic forecasts are used at the individual *patient level* and use the experiences of cohorts of patients to provide information on individual prognosis. Prognostic information may also be used as decision support together with the patient, for example, to choose between two treatments. Two approaches that differ with respect to the aggregation process—model based or data based—may be distinguished. In the model-based approach, the data are aggregated once at the *group level* and published as, for example, an equation based on regression coefficients, while in the data-based approach, data are aggregated from the cohort data each time the prognosis is asked for, and the prognosis for a subgroup with characteristics similar to the patient is selected and displayed [9]. Traditionally, prognosis has been expressed in terms of clinical outcomes (survival, readmission etc), but PRO measures may be used to include outcomes such as symptom burden and functioning.

The Screening Program Activity

In a population screening program, citizens or patients to be invited are identified *systematically* based on risk factors such as age; gender; and at times, disease-specific risk factors. The data collection takes place at the *group level*, but the aggregation and use processes occur at the individual *patient level*, since each screening-positive citizen is referred and further diagnosed and treated individually.

The Disease Surveillance Activity

Registers for monitoring diseases have been known since the middle of the 19th century, when the first-known registry was established with the purpose of monitoring leprosy at the population level [10]. Relevant patients are preferably identified based on diagnosis codes in existing registers, but a number of disease registers still rely on reports from the individual clinician, as do etiological registers like worker's compensation registers. Secondary collection of data (eg, histologic type of cancer or treatment) is organized *systematically* at the *patient level*.

The Health Care Error Surveillance Activity

Health care error is, by nature, an *ad hoc* event at the *patient level*. In surveillance, patients are identified and data are collected *ad hoc* at the *patient level* and aggregated to statistics and reports at the *group level* (eg, hospital, department, or physician). In case of a serious error, the data may also be used at the individual level as a basis for audit, compensation, or even legal action.

Activities in Primary Health Care

Traditionally, all four data processes in primary health care have been *ad hoc* at the *patient level*, except for *systematic*, *group-level* programs like vaccination, pregnancy, and maternal care as well as some mandatory reporting of summary statistics to medical authorities. However, in some countries, primary care activities go from *ad hoc* to *systematically* framed processes at the *group level* (eg, chronic care programs), where the general practitioner is expected to identify patients with certain profiles, and primary health care quality surveillance programs based on *group-level* aggregation of clinical data are also being implemented.

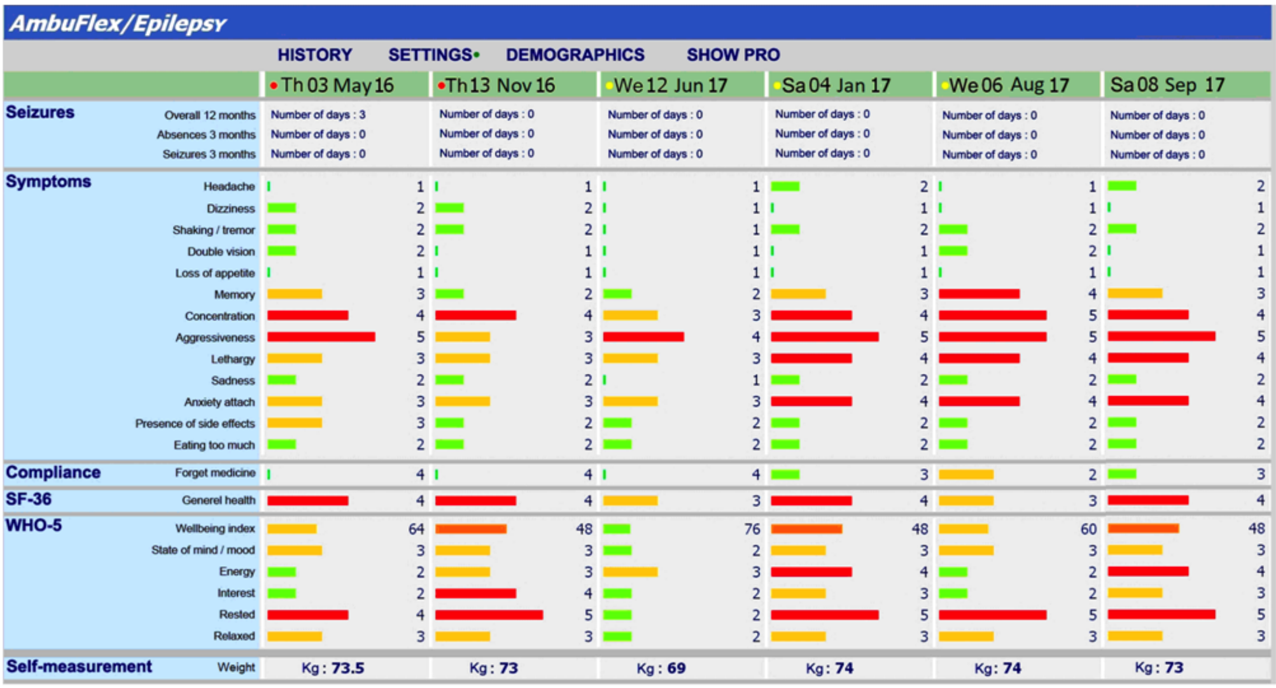
Patient-Reported Outcome Measures in the Data Collection Process

PRO-based health data are not essentially different from other sources of information with respect to the data processes of identification, aggregation, and use, but the data collection process has a number of features that are specific for PRO measures. First, without PRO measures, health data on symptoms and functioning are difficult to collect systematically and will be limited to observations and unsystematic clinician-reported subjective summaries of patient history, which frequently underestimate patient problems [11]. Second, PROs are often the only way to collect data from a patient at home (telePRO). A number of telehealth projects have tried to collect data from home with various hi-tech methods with impact limited to few specific diseases, whereas *telePRO* has shown robustness and been used in a range of chronic diseases [12]. In the following section, PRO-specific aspects of activities listed in Table 1 will be highlighted.

Patient-Reported Outcome in Patient-Level Activities

Paper-based patient questionnaires have been used in the clinical setting for decades to support the communication between the patient and physician. The PRO data are aggregated and used during the consultation as a tool to screen for a priori defined, critically important symptoms (red flags) and to prioritize issues based on the patient's preferences. This use of PRO measures has increased with the introduction of Web-based questionnaires, patient kiosks in the waiting area, etc. The effects on the consultation processes have been reviewed elsewhere [13-15]. During patient follow-up, PRO data are collected in connection with each scheduled visit and used to support a longitudinal overview of symptoms and functions over time and to provide real-time warnings of deterioration aimed at facilitating a prompt response from the care team (Figure 3). If patients complete the PRO remotely online, usually at home (*telePRO*), this information may be used as the base for demand-driven outpatient follow-up without prebooked visits, where disease-relevant PRO questionnaires filled in at home at fixed intervals are aggregated by a disease-specific algorithm that semiautomatically decides whether there is a need or wish for an outpatient visit [12]. This may solve the paradox that outpatient clinics may be drowning in patients even though a substantial part of the visits turn out to be unnecessary from both the patient's and clinician's point of view [16-18]. In Denmark, this principle has been implemented in chronic and malignant diseases including asthma, chronic obstructive pulmonary diseases, epilepsy, sleep apnea, prostatic cancer, and chemotherapy for a number of malignant diseases [19]. A national implementation of the principle is underway in Denmark for selected diagnostic groups.

Figure 3. Longitudinal overview of patient-reported outcome and self-reported measurements in outpatient follow-up (translated from Danish) [13].



		Th 03 May 16	Th 13 Nov 16	We 12 Jun 17	Sa 04 Jan 17	We 06 Aug 17	Sa 08 Sep 17
Seizures	Overall 12 months	Number of days : 3	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0
	Absences 3 months	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0
	Seizures 3 months	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0	Number of days : 0
Symptoms	Headache	1	1	1	2	1	2
	Dizziness	2	2	1	1	1	1
	Shaking / tremor	2	2	1	2	2	2
	Double vision	2	1	1	1	2	1
	Loss of appetite	1	1	1	1	1	1
	Memory	3	2	2	3	4	3
	Concentration	4	4	3	4	5	4
	Aggressiveness	5	3	4	5	5	5
	Lethargy	3	3	3	4	4	4
	Sadness	2	2	1	2	2	2
	Anxiety attach	3	3	3	4	4	4
	Presence of side effects	3	2	2	2	2	2
	Eating too much	2	2	2	2	2	2
Compliance	Forget medicine	4	4	4	3	2	3
SF-36	Generel health	4	4	3	4	3	4
WHO-5	Wellbeing index	64	48	76	48	60	48
	State of mind / mood	3	3	2	3	3	3
	Energy	2	3	3	4	2	4
	Interest	2	4	2	3	2	3
	Rested	4	5	2	5	5	5
	Relaxed	3	3	2	3	3	3
Self-measurement	Weight	Kg : 73.5	Kg : 73	Kg : 69	Kg : 74	Kg : 74	Kg : 73

Patient-Reported Outcome in Group-Level Activities

The clinical research, quality surveillance, clinical guideline, and individual prognosis activities (Table 1) rely on similar data and will be discussed together. PRO data collection has been applied for decades in clinical research based on the belief that outcomes cannot be evaluated on the basis of clinical measures only. Ideally, most group-based activities need longitudinal data with a long follow-up period, often beyond the time span of outpatient follow-up. Due to the increasing use of PRO measures for clinical purposes, isolated collection becomes problematic because the patient is often reluctant to answer more than one questionnaire, especially when the relevance is not clear and questions across measures overlap, leading to repeated questions with similar content [20].

The demand for data by the health care system will undoubtedly increase in the future for all described activities, with the cross-sectional study as a possible exception. Most of the listed data-demanding activities focus on longitudinal data, and the following discussion will focus on this and the role of PRO measures in a longitudinal follow-up.

Multiple Use of Data Collected as Part of Clinical Follow-Up

Of the four data processes, the data collection process is the main challenge with respect to costs as well as logistics. To reduce costs and workload among patients and clinicians, it is essential to focus on joint efforts of data collection with subsequent use in other health data activities. To some extent, this is already happening (eg, clinical research based on clinical quality databases).

The basic example of longitudinal activity is the patient follow-up, where information on the course of treatment, symptoms, and effect of the intervention is monitored, and, if necessary, treatment is adjusted. This activity is *systematic* at the *patient level*, and data are already stored for documentation purposes and may therefore potentially be reused in other activities. A schematic overview of principles in joint efforts where data collected from patient follow-up are used in other activities is shown in Table 2. For the activities listed in Table 1, data aggregation and data use are unchanged; only the processes in the alternative patient identification and data collection processes differ.

A model for joint use of health data based on data collected during patient follow-up with secondary identification of missing patients, observations, and variables for the alternative use is shown in Figure 4. The methods used for identification of missing patients, observations, and variables for the alternative use (Figure 4) depend on the timeframe of the alternative use. If the *ad hoc* method is used, data are exported to an external system where the completeness is analyzed with record linkage methods similar to those used in normal registry-based research, followed by additional *ad hoc* data collection. If the alternative use is to take place repeatedly in a systematic manner, this detection of missing data should preferably take place with online access to the environment in which the clinical data reside. In the Central Denmark Region, a central data warehouse has been established, which now contains clinical information on medication, diagnoses, and procedures, and more information is being collected [21]. These data are available for use in quality-improvement projects, but according to the GDPR, the use of data for research requires the patient to provide explicit permission, which reduces the possibility for joint use significantly.

Table 2. Examples of joint use of health data based on reuse of data routinely collected during patient follow-up with alternative patient identification, complementary data collection, alternative aggregations, and uses of data.

Examples	Patient identification				Data collection				Data aggregation				Data use			
	Patient		Group		Patient		Group		Patient		Group		Patient		Group	
	Ad hoc	Sys[a]	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys	Ad hoc	Sys
Clinical practice		✓[b]			Basis[c]	Basis				✓				✓		
Quality surveillance				✓		Reuse[d]	Comp[e]					✓				✓
Clinical research			✓			Reuse	Comp				✓				✓	
Individual prognosis			✓			Reuse	Reuse		✓		✓		✓			

[a] Sys: Systematic or repeated data process.

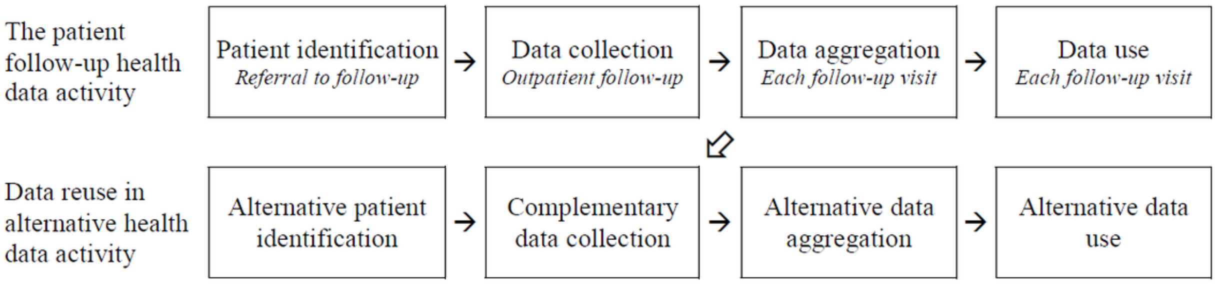
[b] All check marks indicate unchanged activity-specific processes (see Table 1).

[c] Basis: The routine collected follow-up data are the base for alternative uses.

[d] Reuse: Direct reuse of data collected in the cell above.

[e] Comp: Complementary data collection.

Figure 4. Joint use of health data based on data collected during patient follow-up. The oblique arrow indicates identification of missing patients, observations, and variables for alternative use.



Group-level ad hoc procedures are applied to identify missing patients, observations, and variables, with subsequent complementary data collection. In quality surveillance programs, identification takes place as a *systematic group-level* process, but data collection should rely mainly on data collected as part of normal clinical activity. In clinical research, the identification of patients to be included will be *ad hoc*, based on the specific research protocol, while data collection should rely on data collected as part of normal clinical activity as the primary source of data, supplemented with additional *ad hoc* data collection, when supplementary outcome assessment is needed. Individual prognosis will most often rely solely on data collection in research activities, but as the only other activity, data-based individual prognosis may entirely rely on data collected as part of clinical follow-up, given that the data are available for instantaneous on-the-fly aggregation and use [9].

Patient-Reported Outcome Data Collection Supplemental to Data Collected During Follow-up

Different health data activities may have different data demands with respect to timing and PRO content, which makes supplemental PRO data collection necessary. PRO collection in clinical practice does not cover the whole population of patients, since some patients may not attend follow-up and some patients may, for one reason or another, not complete a questionnaire. For clinical use, response rates over 90% are obtainable in *telePRO* [12,22]. For the traditional use of clinical PRO measures, typically collected in the waiting room area, PRO data are obtained only from patients who turn up at follow-up visits, and the response rate is dependent on local commitment, influenced by local population characteristics and configuration of the system (user interface, accessibility and compatibility) and integrated within the existing EHR, clinical pathways, and workflow.

Clinical data collection stops when the patient follow-up ends. For reuse in other activities, it may be necessary to apply supplemental data collection. It is possible to incorporate supplementary research items into PRO questionnaires used for clinical purposes as well as to extend the follow-up period longer than clinically relevant for data use in the activities of clinical research, quality surveillance, and individual prognosis. With respect to content, PRO questionnaires used for a clinical purpose may not be appropriate for other activities, but with respect to domains to be covered, the common set between the activities is often substantial. For example, the European Organisation for Research and Treatment of Cancer scales created for use at *group level* are often usable for *telePRO* [23] when supplemented with a few items, most importantly, the patient’s preference for contact with the health care provider.

High response rates are crucial for PRO data collection in any health data activity, and PRO data collection in clinical practices often has higher response rates than PRO data collected for use at *group level* [22], where the response rate is dependent on local coordination and commitment [24]. Response rates are dependent on how relevant the data appear to the patient, and clinical use seems nearest to the patient. PRO data collected in clinical practice may therefore yield higher quality of data than traditional surveillance studies. A known problem when PRO data are collected at a *group level* is what to do with alarming answers, the so-called PRO “alerts” such as high depression scores or signs of suicidal ideation [25]. This is feasible to deal with in a clinical setting but is very difficult when collecting data only for *group-level* use. While supplemental data collection of clinical data may be troublesome and expensive due to several reasons such as extra follow-up visits, PRO data collection processes may be centralized and automated if the relevant infrastructure is available [22].

Challenges in Reuse of Patient-Reported Outcome Data Collected During Follow-up

In order to achieve the anticipated potential of joint use of PRO data, some critical challenges should be addressed. The psychometric requirements may vary depending on the specific use (eg, level of reliability and sensitivity to change), but other requirements for the data collection process, such as high response rate, low attrition rate, and high data completeness, remain essentially the same regardless of the activity and type of data. All activities must meet challenges in terms of data collection logistics and management, and the demands for data security are typically also identical. Supplemental data collection requires close cooperation between PRO activities with real-time access to data, which raises some issues. The challenges are divided into three types.

Legal Challenges

Legislation issues have a bearing on all four data processes; therefore, the legal framework has to be precisely specified before any data collection can begin. Activities with *systematic* data collection may typically benefit from permanent permission from national data protection agencies, while *ad hoc* projects must apply for permission for a specified period. The fundamental problem is that all approvals are only valid for the specific activity (eg, quality surveillance or clinical research), which means that data cannot be used for other activities. The implementation of the European Union's regulation with respect to protection of personal data—GDPR [2]—will make it even more difficult to use data for other purposes without a specific consent from each patient. This will have a serious impact on joint use unless health data are given a differentiated treatment, such that the requirements for confidentiality can be maintained and individual approval can be collected in an efficient way (eg, through some form of umbrella approval process). For *group-level* use, analyses of personal data may be performed on a remote server where the researcher may upload a dataset and merge it with personal data using a unique personal identifier. The researcher has access to only aggregated data such as tables and outputs from statistical analyses [26]. Such a method of accessing personal data is available in Denmark and the Netherlands, but for now, few health data such as those on diagnoses and procedures are available for merging.

Technical Challenges

The principle of supplemental identification and data collection described above presupposes real-time access to relevant patient databases in the patient-identification process and in most cases, in the data collection process. Apart from that, there are substantial technical issues related to the aggregation process. Data may be collected and stored, but not available for the relevant alternative aggregation. A typical example is the quality surveillance activity, where the needed data may already exist in the patient's EHR, but an automated process of extracting and transporting data is not possible due to inadequate and incompatible information technology systems or a lack of relevant expertise. PRO data may already be collected but stored in a different system or format. A possible solution to the latter is proposed by the international Health Level Seven standards for transfer of clinical and administrative data between software apps used by various health care providers [27]. A special Health Level Seven section for PRO measures has recently been adopted.

Challenges Related to Content and Timing of Data Collection

The need for valid, reliable, and responsive measurement scales is common for data for any health data activity. For PRO data collected for making individual clinical decisions, measurement error is of particular importance. Although scales that have acceptable psychometric properties at the patient level will normally also perform well at a *group level*, the opposite is not true and the desirable content and length of a PRO questionnaire are likely to differ between *group-level* and *patient-level* activities. In routine patient follow-up visits, short instruments are often preferred and procedures that the clinician finds irrelevant for the actual patient may not be collected as prescribed (eg, a comprehensive time-consuming test of performance in a patient who has clinically completely recovered or a depression score in a patient who is clinically obviously not depressed). A possible solution to these contradictive interests may be application of item banks and computer-adaptive testing, which can achieve high reliability with the lowest-possible administration burden [28]. Timing of data collection poses another challenge for joint use, and the optimal timing of data collection may differ between activities. Quality surveillance and clinical research may prefer that data collection follow a fixed scheduled in compliance with a protocol, while outpatient clinical practice is focused on the practical arrangement of follow-up, and visits often have to be postponed for various reasons. Although from a clinician's perspective, it might be acceptable that patients who are doing well cancel their appointments, this may result in devastating selective attrition in *group-level* activities. From a resource point of view and the patient's perspective, a patient who does not need or want clinical attention should not go to follow-up visits just to deliver data for other purposes. A rational approach for addressing these problems with missing data for the alternative activity could be a supplemental real-time identification of patients with missing data combined with collection of PRO data on proxy variables.

Conclusions

We have introduced a model for health data with four data processes, each dividable with respect to timeframe and group level, which distinguishes properties relevant to the discussion of joint use across different purposes and supports consideration of the associated organizational and technological challenges. Based on this, we propose a model for joint use of health data, with data collected during follow-up as the backbone. In the future, clinical settings will be a prominent source of PRO data and data collection for research and quality surveillance will have to adapt to this circumstance and design ways of complementary data collection as and when necessary. Demands for health data will continue to increase, which will further add to the need for the concerted use and reuse of PRO data for parallel purposes due to financial, logistical, and ethical reasons. A number of legal, technical, and organization challenges must be addressed.

The risk of patients' information being accessed and used by people for whom it was not initially intended is real. For example, the use of health data by private insurance companies might restrict access to health coverage for vulnerable patients and those with a precondition. Additionally, access to private medical information by law enforcement agencies could be a risk for individuals and society. However, the current legal restriction on the joint use of health data imposed by the GDPR makes no distinction between these misuses and the uses described in this paper. Steps should be taken to alleviate the current legal restriction on the joint use of health data imposed by the GDPR while still protecting patient data against misuse.

Acknowledgments

We acknowledge Professor Henrik Toft Sørensen, Department of Clinical Epidemiology, Aarhus University Hospital, Denmark, for his valuable input and comments to an earlier version of the manuscript.

MC and DK are funded by the NIHR Birmingham Biomedical Research Centre and the NIHR Surgical Reconstruction and Microbiology Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. MC has received personal fees from Astellas, Takeda, and Glaukos outside the submitted work and grants from the NIHR and is a coinvestigator at Health Data Research UK Midlands. DK is supported by the NIHR postdoctoral fellowship scheme (PDF-2016-09-009). JV has received grants and fellowships from NIHR and Instituto de Salud Carlos III (Spain); grants from MRC, CRUK, Royal Devon, and Exeter Trust, Fondo de Investigaciones Sanitarias (Spain); and consultancy fees from the World Health Organization and Technische Krankenkasse (Germany) and currently is chairman (unpaid) of the International Consortium for Health Outcomes Measurement panel for the development of core sets of outcomes for Health Overall Adult Health. NH is the founder of AmbuFlex.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of basic health data activities.

[PDF File (Adobe PDF File), 70KB - jmir_v21i5e12412_app1.pdf]

References

1. Coiera E. Guide To Health Informatics. Boca Raton: CRC Press; 2015.
2. European Commission. Principles of the GDPR URL: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr_en [accessed 2019-01-25] [WebCite Cache ID 75gg8nd3Z]
3. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ 2019 Jan 24;364:k5267. [doi: 10.1136/bmj.k5267] [Medline: 30679170]
4. Food and Drug Administration. 2009 Dec. Guidance for Industry - Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf [accessed 2019-04-23] [WebCite Cache ID 6qqbfANOC]
5. Deshpande PR, Rajan S, Sudeepthi BL, Abdul NCP. Patient-reported outcomes: A new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144 [FREE Full text] [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
6. Cambridge Dictionary. Data URL: https://dictionary.cambridge.org/dictionary/english/data [accessed 2018-08-24] [WebCite Cache ID 71uQeqH8D]
7. Dictionary of Quality of Life and Health Outcomes Measurement. Milwaukee: International Society of Quality of Life (ISOQOL); 2015.
8. Wikipedia. Data URL: https://en.wikipedia.org/wiki/Data [accessed 2018-08-24] [WebCite Cache ID 71uQQeJdW]
9. Hjollund NHI. Individual Prognosis of Symptom Burden and Functioning in Chronic Diseases: A Generic Method Based on Patient-Reported Outcome (PRO) Measures. J Med Internet Res 2017 Dec 01;19(8):e278 [FREE Full text] [doi: 10.2196/jmir.8111] [Medline: 28765099]
10. Irgens LM, Bjerkedal T. Epidemiology of leprosy in Norway: the history of The National Leprosy Registry of Norway from 1856 until today. Int J Epidemiol 1973;2(1):81-89. [Medline: 4590337]
11. Laugsand EA, Sprangers MAG, Bjordal K, Skorpen F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. Health Qual Life Outcomes 2010 Sep 21;8:104 [FREE Full text] [doi: 10.1186/1477-7525-8-104] [Medline: 20858248]
12. Schougaard LMV, Larsen LP, Jessen A, Sidenius P, Dorflinger L, de Thurah A, et al. AmbuFlex: tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. Qual Life Res 2016 Mar;25(3):525-534 [FREE Full text] [doi: 10.1007/s11136-015-1207-0] [Medline: 26790427]
13. Porter I, Gonçalves-Bradley D, Ricci-Cabello I, Gibbons C, Gangannagaripalli J, Fitzpatrick R, et al. Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. J Comp Eff Res 2016 Dec;5(5):507-519. [doi: 10.2217/cer-2015-0014] [Medline: 27427277]
14. Johansen MA, Berntsen GKR, Schuster T, Henriksen E, Horsch A. Electronic symptom reporting between patient and provider for improved health care service quality: a systematic review of randomized controlled trials. part 2: methodological quality and effects. J Med Internet Res 2012 Oct 03;14(5):e126 [FREE Full text] [doi: 10.2196/jmir.2216] [Medline: 23032363]

15. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008 Mar;17(2):179-193. [doi: 10.1007/s11136-007-9295-0] [Medline: 18175207]
16. Bahiense-Oliveira M, Duarte D, Meira GGC, de Codes JJ, Ribeiro MZ. Referral to the nephrology outpatient clinic: inappropriate referral to the nephrologist. J Bras Nefrol 2010;32(2):145-148 [FREE Full text] [Medline: 21103673]
17. Donnellan F, Hussain T, Aftab AR, McGurk C. Reducing unnecessary outpatient attendances. Int J Health Care Qual Assur 2010;23(5):527-531. [doi: 10.1108/09526861011050556] [Medline: 20845682]
18. Lee JY, Jo M, Yoo W, Kim HJ, Eun SJ. Evidence of a broken healthcare delivery system in korea: unnecessary hospital outpatient utilization among patients with a single chronic disease without complications. J Korean Med Sci 2014 Dec;29(12):1590-1596 [FREE Full text] [doi: 10.3346/jkms.2014.29.12.1590] [Medline: 25469056]
19. Schougaard L. AmbuFlex – Experiences of PRO-based clinical decision-making. Qual Life Res 2017 Oct;26(Suppl 1):8. [doi: 10.1007/s11136-017-1658-6] [Medline: 28900889]
20. Philpot LM, Barnes SA, Brown RM, Austin JA, James CS, Stanford RH, et al. Barriers and Benefits to the Use of Patient-Reported Outcome Measures in Routine Clinical Care: A Qualitative Study. Am J Med Qual 2018 Dec;33(4):359-364. [doi: 10.1177/1062860617745986] [Medline: 29258323]
21. MIDT regionmidtjylland. Business Intelligence i Region Midtjylland (in Danish) URL: https://www.rm.dk/om-os/organisation/it/business-intelligence/ [accessed 2019-01-25] [WebCite Cache ID 75geZQOKI]
22. Hjollund NHI, Larsen LP, Biering K, Johnsen SP, Riiskjær E, Schougaard LM. Use of Patient-Reported Outcome (PRO) Measures at Group and Patient Levels: Experiences From the Generic Integrated PRO System, WestChronic. Interact J Med Res 2014 Feb 11;3(1):e5 [FREE Full text] [doi: 10.2196/ijmr.2885] [Medline: 24518281]
23. Snyder CF, Blackford AL, Okuyama T, Akechi T, Yamashita H, Toyama T, et al. Using the EORTC-QLQ-C30 in clinical practice for patient management: identifying scores requiring a clinician's attention. Qual Life Res 2013 Dec;22(10):2685-2691 [FREE Full text] [doi: 10.1007/s11136-013-0387-8] [Medline: 23532341]
24. Basch E, Dueck AC, Rogak LJ, Mitchell SA, Minasian LM, Denicoff AM, et al. Feasibility of Implementing the Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events in a Multicenter Trial: NCCTG N1048. J Clin Oncol 2018 Sep 11:JCO2018788620. [doi: 10.1200/JCO.2018.78.8620] [Medline: 30204536]
25. Kyte D, Draper H, Calvert M. Patient-reported outcome alerts: ethical and logistical considerations in clinical trials. JAMA 2013 Sep 25;310(12):1229-1230. [doi: 10.1001/jama.2013.277222] [Medline: 24065005]
26. Statistics Denmark. The Danish system for access to micro data URL: https://www.dst.dk/ext/645846915/0/forskning/Access-to-micro-data-at-Statistics-Denmark_2014--pdf [accessed 2019-01-25] [WebCite Cache ID 75gduIgZB]
27. HL7 International. Introduction to HL7 standards URL: http://www.hl7.org/implement/standards/index.cfm?ref=nav [accessed 2018-08-24] [WebCite Cache ID 71uQkEnLl]
28. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Reliability at group level and individual level. In: Measurement in medicine: A practical guide. Cambridge, UK: Cambridge University Press; 2011:142.

Abbreviations

EHR: electronic health record
GDPR: General Data Protection Regulation
PRO: patient-reported outcome

Edited by G Eysenbach; submitted 05.10.18; peer-reviewed by A Benetoli, N Rothrock, JS Howard; comments to author 20.12.18; revised version received 02.03.19; accepted 24.03.19; published 21.05.19.

Please cite as:
Hjollund NHI, Valderas JM, Kyte D, Calvert MJ
Health Data Processes: A Framework for Analyzing and Discussing Efficient Use and Reuse of Health Data With a Focus on Patient-Reported Outcome Measures
J Med Internet Res 2019;21(5):e12412
URL: https://www.jmir.org/2019/5/e12412/
doi:10.2196/12412
PMID:31115347

Viewpoint

Data Challenges With Real-Time Safety Event Detection And Clinical Decision Support

Eric Steven Kirkendall[1,2,3,4,5*], MD, MBI; Yizhao Ni[1,2*], PhD; Todd Lingren[1*], MS; Matthew Leonard[6*], BA; Eric S Hall[1,2,6*], PhD; Kristin Melton[2,6*], MD

[1]Department of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[2]Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, OH, United States
[3]Division of Hospital Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[4]James M Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[5]Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, NC, United States
[6]Division of Neonatology and Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[*]all authors contributed equally

Corresponding Author:
Eric Steven Kirkendall, MD, MBI
Department of Biomedical Informatics
Cincinnati Children's Hospital Medical Center
3333 Burnet Avenue
Cincinnati, OH, 45229
United States
Phone: 1 513 636 4200
Email: ekirkend@wakehealth.edu

Abstract

Background: The continued digitization and maturation of health care information technology has made access to real-time data easier and feasible for more health care organizations. With this increased availability, the promise of using data to algorithmically detect health care–related events in real-time has become more of a reality. However, as more researchers and clinicians utilize real-time data delivery capabilities, it has become apparent that simply gaining access to the data is not a panacea, and some unique data challenges have emerged to the forefront in the process.

Objective: The aim of this viewpoint was to highlight some of the challenges that are germane to real-time processing of health care system–generated data and the accurate interpretation of the results.

Methods: Distinct challenges related to the use and processing of real-time data for safety event detection were compiled and reported by several informatics and clinical experts at a quaternary pediatric academic institution. The challenges were collated from the experiences of the researchers implementing real-time event detection on more than half a dozen distinct projects. The challenges have been presented in a challenge category-specific challenge-example format.

Results: In total, 8 major types of challenge categories were reported, with 13 specific challenges and 9 specific examples detailed to provide a context for the challenges. The examples reported are anchored to a specific project using medication order, medication administration record, and smart infusion pump data to detect discrepancies and errors between the 3 datasets.

Conclusions: The use of real-time data to drive safety event detection and clinical decision support is extremely powerful, but it presents its own set of challenges that include data quality and technical complexity. These challenges must be recognized and accommodated for if the full promise of accurate, real-time safety event clinical decision support is to be realized.

(J Med Internet Res 2019;21(5):e13047) doi:10.2196/13047

KEYWORDS

decision support systems, clinical; clinical decision support; real-time systems; electronic medical records; electronic health records; medical records systems, computerized; informatics; data science; information science; patient safety

Introduction

Background

All of us as informaticians and, more broadly, users of data realize the common challenges and pitfalls of our substrate. The classic “garbage in, garbage out” and “90% of the time spent working with data is cleaning it” sayings represent some of the most popular conventional wisdom in informatics. Although many of the issues that have popularized these mantras are universal, the context within which the data are created and the context within which the data are manipulated and used often shuffle the focus on particular data challenges and introduce new ones. Challenges in the secondary use of electronic health record (EHR) data or claims data in research are well described [1-5]. This is also the case when considering the use of health data to detect events, particularly when trying to identify safety concerns and errors in a timely fashion.

The continued digitization and maturation of health care information technology has made access to real-time data easier and feasible for more health care organizations. Many EHR vendors have provided some amount of application programming interfaces and other means for real-time access to the data generated by their products, using standards and technologies such as Observational Medical Outcomes Partnership and Fast Health Interoperability Resources [6-9]. The push from many stakeholders for interoperability between EHRs has also helped in this regard. With this increased availability, the promise of using data to algorithmically detect health care–related events in real-time has become more of a reality [10-13]. However, as more researchers and clinicians have utilized real-time data delivery capabilities, it has become apparent that simply gaining access to the data is not a panacea, and some unique data challenges have emerged to the forefront in the process.

Objective

This brief report serves to highlight some of the challenges that are germane to real-time processing of health care system–generated data and the interpretation of the results. The list generated is by no means exhaustive but does represent some of the most common challenges (and solutions) we have encountered in our experience of designing, implementing, and evaluating numerous real-time event detection systems [14-25]. These challenges were cataloged during the project, highlighted in this report, and augmented with additional challenges from the experiences of the authors in their many other lines of research. Challenges were collected and sorted initially by the first author (EK) and then modified and edited by the other contributing authors. Similar challenges were sorted into high-level challenge categories, and examples from each specific challenge were described.

In this report, we will begin by covering some data issues that are clearly not limited to real-time use and progress to discussing challenges that are more unique to the real-time data use cases. We will offer specific examples of each type and subtype of challenges by referencing a current safety event and error detection project involving the use and joining of data from an EHR and smart infusion pumps and using that data to detect discrepancies between order, administration documentation, and smart pump infusion information. We will (1) briefly summarize the challenge category, (2) describe a specific challenge, and follow this by (3) sharing a specific example of the challenge and, where possible, some mitigation strategies. First, we will give the reader a very brief project overview to set the foundation for the specific examples that follow and to provide a context for the reader to ground in the challenges and examples.

Project Overview: Fusing Electronic Health Record–Based Medication Ordering, Medication Administration Record, and Smart Infusion Pump Data to Detect Potential Errors

Our research team has demonstrated that EHR data can be used to retrospectively detect discrepancies between how high-risk, continuously infused medications are ordered and how they were documented as being given in the medication administration record (MAR) in the EHR [20-22]. These discrepancies often represent errors, either clinically relevant errors (medication is being given at a rate not intended by the prescriber) or documentation errors (documentation in the MAR is not the dose actually being given). This line of research has evolved and matured to be capable of detecting these discrepancies prospectively, within minutes of the data being available in the EHR. In addition, the retrospective algorithms have been modified to incorporate data from smart infusion pumps, which gives us the capability of discriminating between clinical errors and documentation errors (the pump data are the *source of truth* for what the patient is receiving). The approach relies on chronologically aligning all of the data elements by their respective timestamps. Although simple in concept, in reality this exercise is more difficult than one would anticipate for many reasons. We now describe some of the more salient challenges encountered along this particular line of work; however, they have also been encountered in many other projects relying on real-time and device-related data.

Results

Data Challenge Categories, Specific Challenges, and Examples

This report is organized into data challenge categories, specific challenges, and examples. In total, 8 major types of challenge categories are reported, with 13 specific challenges and 9 specific examples detailed to provide a context for the challenges (Table 1).

Table 1. Major challenge types, specific challenges, and specific examples related to use of real-time data.

Challenge category	Specific challenge	Example
Selecting the correct data elements	Selecting the right action state and timestamp	Selecting the right action state in medication order data Selecting the right timestamp
Other timestamp-related considerations	Delayed documentation—verbal orders	—[a]
	Timestamp conversion and formatting	—
	Visualizing time series data to understand temporal patterns	Raw time data table versus data visualization
Metadata attributes	Misleading (meta)data labels	*Shown to user* data column label
Workflow imprints	Issues that affect performance and capabilities of algorithms	Patient deterioration in the Neonatal Intensive Care Unit necessitates verbal orders
	Delayed action on active orders	—
	Priming pumps: Speeding up infusion pump rates to prime may look like an error, but has no clinical consequence	—
Unstructured data entry	Complexity of human language	Free text dosing of Total Parenteral Nutrition
	Heterogeneity of human language	—
Fusing datasets and the role of device integration	Merging datasets from multiple sources requires valid linking identifiers	The nonintegration of smart pumps with electronic health records
	Clinical decision support blind spots	Use of smart infusion pump drug libraries
Retrospective versus prospective data or detection	Retrospective data and real-time data are processed and accessed differently	The order audit modification issue
Technical versus clinical validity	—	—

[a] Not applicable.

Challenge Category: Selecting the Correct Data Element—Multiple Action States and Timestamps Related to Similar Concepts

Modern EHRs contain very elaborate underlying data models, with simple clinical or technical concepts represented in very complex and granular ways. For example, the EHRs that support large health systems with tertiary or quaternary hospitals may have 50 to 100 different data elements that record systolic blood pressure measurements, each with a slightly different context related to workflows. Consequently, related data elements may be scattered across database tables, and despite data dictionaries suggesting relevance, specific data fields may not be populated as expected. Data dictionaries are often incomplete or suboptimal for other reasons, leaving a data consumer to rely on metadata labels and manual inspection of the data to guide accurate selection. Cross reference of the values and patterns between the operational system–derived data (ie, *front-end* views) and the data transformed into reporting databases (usually in a relational database, *back-end* views) is a common approach to guide the stakeholder, but large datasets or closely related data elements can make distinctions challenging and labor-intensive. Multiple locations for similar data may be convenient to both designers and users of the system, but this lack of parsimony can also lead to data completeness issues (reports only pulling data from one source when it should be from both sources) and conflicting values in the EHR (users enter data in multiple locations, and the data conflicts with each other).

Challenge: Selecting the Right Action State and Timestamp

Almost all EHR and medical device data are accompanied by one or more timestamps, pieces of data that track the date and time of a particular action or event occurred and were recorded. Although the timestamp denoting that documentation of an event is recorded (*documentation* or *file time*) may be automatically captured, clinicians may also record the time in which events actually occurred or were documented to have occurred (*action time*). Subsequent documentation may amend the originally documented action time, but each amending action would also generate a corresponding timestamp representing when the amendment was recorded (*amending time*). In addition, many data representations of clinical concepts or actions have multiple action states. Selecting the correct action state (including or excluding action states) and the right timestamp field can be difficult. For action states, it is not uncommon to have many closely related or overlapping enumerated list items that can make this selection complicated. For each of these action states that are recorded, there may be several timestamp fields to choose from. Selecting the wrong action state or timestamp can lead to misleading inferences from the data. Mitigation strategies for this challenge type include careful evaluation of all possibly relevant data fields related to the data request, reviews of existing data dictionaries, and careful analysis and validation of the output from data discovery activities. This analysis and validation should be from both the technical analyst and the clinical subject matter expert.

Example: Selecting the Right Action State in Medication Order Data

Poor communication between data requestors and data provisioners may lead to confusion between selecting and including order data (how things were ordered) versus MAR documentation data (how orders were acted or not acted upon) for use in reports or algorithms. In real clinical practice, not all medication orders are carried out for a variety of reasons and one should not assume that a medication order means that a patient actually received the medication. An action state of *Given* or *Not Given* in the MAR is more indicative of the clinical actions that actually took place. Smart infusion pumps generally log each action that a pump undergoes, whether it is an automated procedure or the result of user programming. Smart pumps may support dozens or hundreds of action states, with the most frequent states logged including *infusion started*, *infusion stopped*, and *infusion restarted*. If an analyst or researcher is attempting to count the number of times an infusion moved from an inactive (not infusing) to an actively infusing state and only counts the *infusion started* state, they will underestimate the count by missing the *infusion restarted* state (assuming the restarted state is not accompanied by logging an *infusion started* action simultaneously). This can obviously affect denominators in rate metrics. In one project centered around risk-stratifying and detecting acute kidney injury (AKI), early iterations of an algorithm based on medication profiles overestimated the population at risk for AKI because we incorporated medication orders and not medication administration data. Not all orders are acted upon and administered to patients. The accuracy of the algorithms improved with modification.

Example: Selecting the Right Timestamp

MAR documentation typically records a medication action state and several timestamps pertaining to the action state. The two most common are an action time and a documentation or file time. In many workflow instances, the time between the two may be trivial. In busy workflows or high clinical throughput scenarios, documentation of the action state may occur much later (eg, *back-documentation* by a nurse) and the interval between the 2 timestamps could be substantially large. The main challenges when utilizing timestamps in real-time applications are twofold: (1) interchanging the 2 values can affect time interval measurements by causing under- or overestimation inaccuracies and (2) back-documentation has the potential to undermine the power of real-time detection by delaying the availability of data. In the Medication Safety project described above, the delaying of documentation of MAR actions delays the ability of our algorithms to detect discrepancies between orders and actual administration, thereby limiting the ability to identify errors in real-time. In effect, a blind spot is created by the lack of timely documentation. Allowing for an appropriate time window to accommodate back-documentation is challenging because too wide a window enlarges the blind spot of vulnerability for prolonged time periods. However, establishing too narrow a window may increase false positive real-time error rates caused by delayed documentation. Understanding workflows is critical in fine-tuning the optimal tolerance for delayed documentation, which may vary from one unit or institution to another.

Challenge Category: Other Timestamp-Related Considerations

Timestamp data are central to the concepts discussed in this report, especially when merging and synchronizing disparate datasets. Even data representing a universal concept, such as time, can take on many forms that make processing it appropriately and making accurate inferences a challenge and is an exercise that is prone to error and limitations. The following are the specific challenges and examples related to timestamp data.

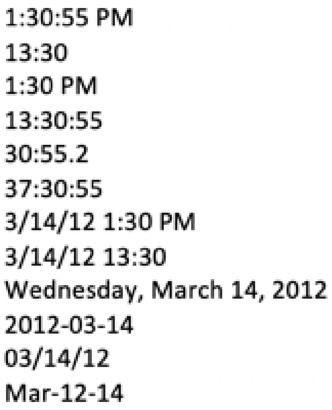
Challenge: Delayed Documentation—Verbal Orders

Verbal orders, by definition, are orders that are not documented in the EHR or other associated systems. Safety goals and regulatory mandates state that verbal orders be kept at a minimum, and in most cases, the expectation is that verbal orders be documented electronically at some point after they are clinically acted upon. This lag in time between the time they are carried out clinically and the time that they are documented electronically (if at all) undermines the capabilities of algorithms that depend on timely data streams to be functional. Technical issues that lead to delayed delivery of data can have the same effect. In many cases, this may lead to a false positive event detection, lowering the performance characteristics of an algorithm and leading to decreased confidence and buy-in from the recipients of the algorithm output. Mitigation strategies include allowing for a reasonable time window for verbal orders to appear. In our project, we allow 30 min for the new documentation of a verbal order to appear in the EHR and data feeds. Data from this allowance would then negate, or *call off*, the event detection and notification.

Challenge: Timestamp Conversion and Formatting

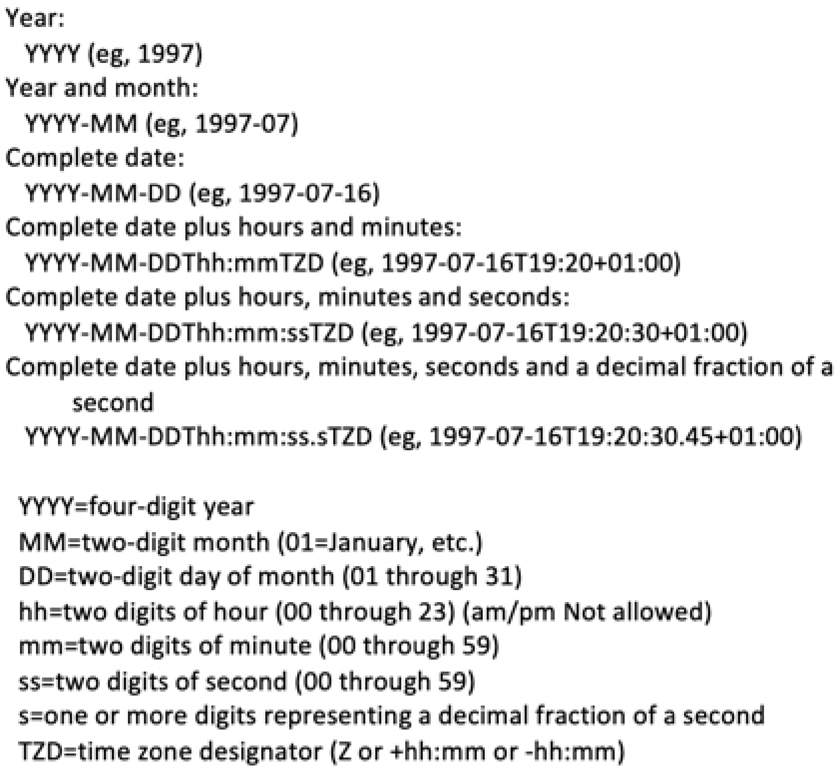
Timestamp data can be represented in many different formats. A thorough discussion of timestamps and their many challenges is out of scope for this report, but Figure 1 demonstrates a couple of different formats in which dates and times can be represented. Granularity and specificity of the timestamp is also an issue, as shown in Figure 2 [26]. Converting to an unexpected format during the extract-transfer-load (ETL) process (eg, converting YYYY-MM-dd HH:mm:ss to YYYY-MM-dd only) would cause unexpected errors in system outputs and subsequently pose a challenge to algorithm debugging. Other complicating factors include accounting for Daylight Saving Time adjustments, the selection of an epoch to count the duration of time (Unix-based systems use the time elapsed since 00:00:00 Coordinated Universal Time [UTC], Thursday, January 1, 1970), and the conversion of one time format to another (eg, from Eastern Standard Time to UTC). In particular, simple mistakes with timestamp entering and formatting, especially when fusing data from manually input sources, can cause pairing errors and have important negative effects in system processing. ISO 8601, drafted by the International Organization for Standardization (ISO), is intended to be the international standard for representation of dates and times but not all systems adhere to this representation [27].

Figure 1. Time and date represented in many different formats.



1:30:55 PM
13:30
1:30 PM
13:30:55
30:55.2
37:30:55
3/14/12 1:30 PM
3/14/12 13:30
Wednesday, March 14, 2012
2012-03-14
03/14/12
Mar-12-14

Figure 2. Representation of different amounts of granularity and specificity in timestamps, ranging from year only to fractions of a second, and including timezone information. Content from the World Wide Web Consortium.



Year:
YYYY (eg, 1997)
Year and month:
YYYY-MM (eg, 1997-07)
Complete date:
YYYY-MM-DD (eg, 1997-07-16)
Complete date plus hours and minutes:
YYYY-MM-DDThh:mmTZD (eg, 1997-07-16T19:20+01:00)
Complete date plus hours, minutes and seconds:
YYYY-MM-DDThh:mm:ssTZD (eg, 1997-07-16T19:20:30+01:00)
Complete date plus hours, minutes, seconds and a decimal fraction of a second
YYYY-MM-DDThh:mm:ss.sTZD (eg, 1997-07-16T19:20:30.45+01:00)

YYYY=four-digit year
MM=two-digit month (01=January, etc.)
DD=two-digit day of month (01 through 31)
hh=two digits of hour (00 through 23) (am/pm Not allowed)
mm=two digits of minute (00 through 59)
ss=two digits of second (00 through 59)
s=one or more digits representing a decimal fraction of a second
TZD=time zone designator (Z or +hh:mm or -hh:mm)

Challenge: Visualizing Time Series Data to Understand Temporal Patterns

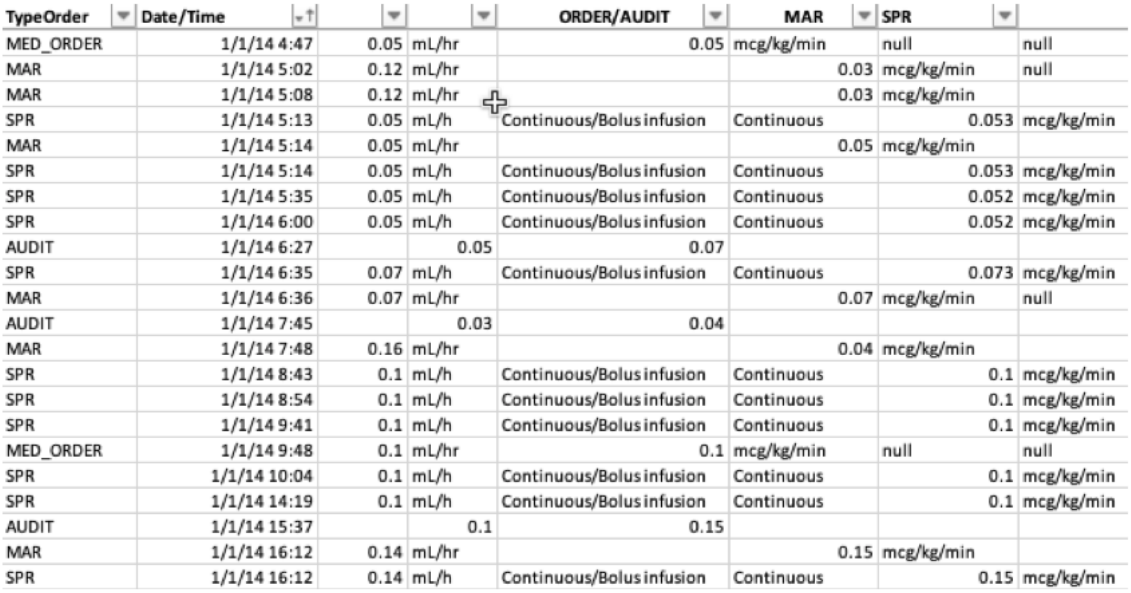
Working with raw data or simple nonvisual outputs of the data (such as in the tabular format) may be the simplest way to review time-based data, but it does not lend itself to appreciating chronologic events and judgements about the interval between 2 data points easily (eg, intervals such as elapsed time). This is usually most easily accomplished by a visualization, such as a timeline or other graphic representations, where one axis is linear time. Without some representation of the interval magnitude, reviewers will need to exert more cognitive effort to appreciate these measures. To mitigate this issue, we recommend the use of robust visualization analytics tools to appreciate these patterns. The following simple example demonstrates the utility of even rudimentary tools.

Example: Raw Time Data Table Versus Data Visualization

A comparison below demonstrates the 2 different data representation methods for the same dataset. In this example, data from several data element categories are represented in a tabular format and time series format.

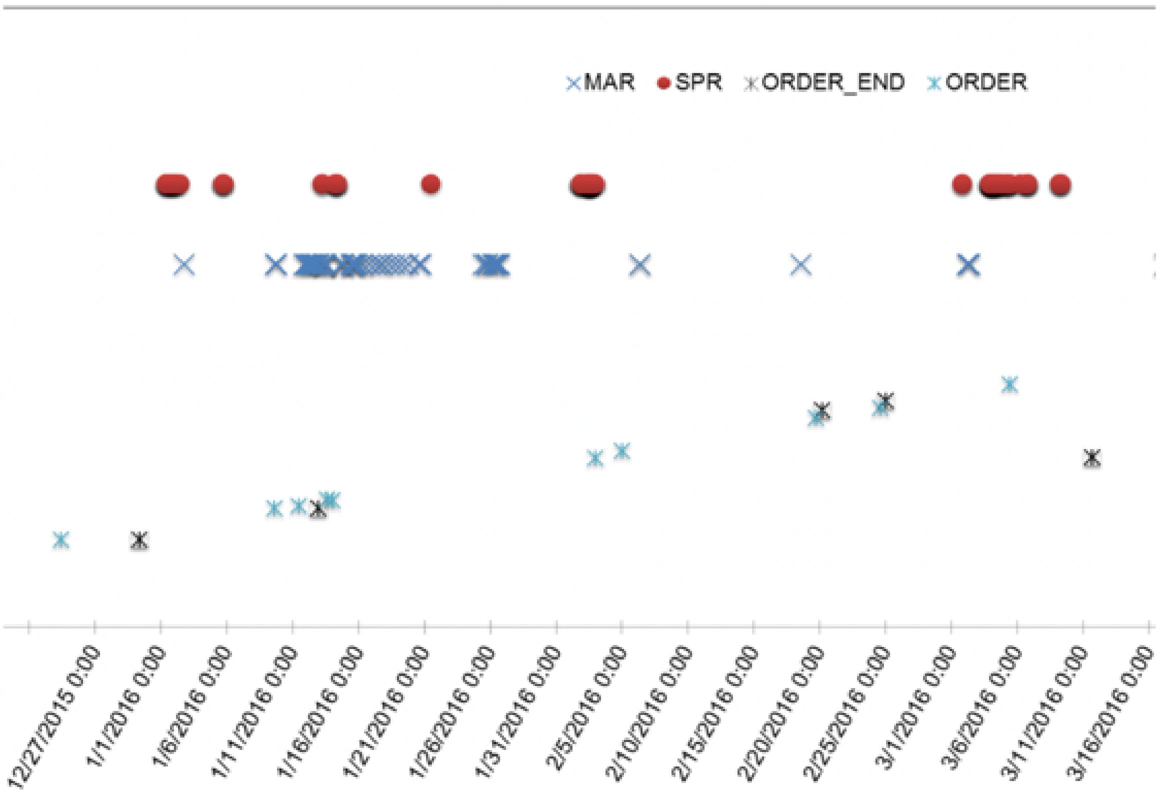
As shown in Figures 3 and 4, it is much more efficient to interpret time-series data in a visual format, which allows one to quickly understand the magnitude of events represented by the data, such as the time between events or overall relationship of several events. It also allows easier representation of a high number of events or occurrences by creating overlays for each event type of interest.

Figure 3. Tabular representation of the medication order, medication administration record (MAR), and smart pump record (SPR) data.



TypeOrder	Date/Time			ORDER/AUDIT	MAR	SPR	
MED_ORDER	1/1/14 4:47	0.05	mL/hr	0.05	mcg/kg/min	null	null
MAR	1/1/14 5:02	0.12	mL/hr		0.03	mcg/kg/min	null
MAR	1/1/14 5:08	0.12	mL/hr		0.03	mcg/kg/min	
SPR	1/1/14 5:13	0.05	mL/h	Continuous/Bolus infusion	Continuous	0.053	mcg/kg/min
MAR	1/1/14 5:14	0.05	mL/hr		0.05	mcg/kg/min	
SPR	1/1/14 5:14	0.05	mL/h	Continuous/Bolus infusion	Continuous	0.053	mcg/kg/min
SPR	1/1/14 5:35	0.05	mL/h	Continuous/Bolus infusion	Continuous	0.052	mcg/kg/min
SPR	1/1/14 6:00	0.05	mL/h	Continuous/Bolus infusion	Continuous	0.052	mcg/kg/min
AUDIT	1/1/14 6:27		0.05	0.07			
SPR	1/1/14 6:35	0.07	mL/h	Continuous/Bolus infusion	Continuous	0.073	mcg/kg/min
MAR	1/1/14 6:36	0.07	mL/hr		0.07	mcg/kg/min	null
AUDIT	1/1/14 7:45		0.03	0.04			
MAR	1/1/14 7:48	0.16	mL/hr		0.04	mcg/kg/min	
SPR	1/1/14 8:43	0.1	mL/h	Continuous/Bolus infusion	Continuous	0.1	mcg/kg/min
SPR	1/1/14 8:54	0.1	mL/h	Continuous/Bolus infusion	Continuous	0.1	mcg/kg/min
SPR	1/1/14 9:41	0.1	mL/h	Continuous/Bolus infusion	Continuous	0.1	mcg/kg/min
MED_ORDER	1/1/14 9:48	0.1	mL/hr	0.1	mcg/kg/min	null	null
SPR	1/1/14 10:04	0.1	mL/h	Continuous/Bolus infusion	Continuous	0.1	mcg/kg/min
SPR	1/1/14 14:19	0.1	mL/h	Continuous/Bolus infusion	Continuous	0.1	mcg/kg/min
AUDIT	1/1/14 15:37		0.1	0.15			
MAR	1/1/14 16:12	0.14	mL/hr		0.15	mcg/kg/min	
SPR	1/1/14 16:12	0.14	mL/h	Continuous/Bolus infusion	Continuous	0.15	mcg/kg/min

Figure 4. Graphic visualization representation of the medication order, medication administration record (MAR), and smart pump record (SPR) data in time-series format.



Challenge Category: Metadata Attributes

Challenge: Misleading (Meta)data Labels

EHR systems that generate data often do so for a primary purpose that is not research related and/or not with a priority on using precise, accurate, and nonoverloaded data labels. Written another way, clinical information systems are not usually designed and implemented with secondary uses of data as a priority. Data consumers are often left to guess about how the data are generated, based on the metadata information in data dictionaries (if available), database structures, attribute titles/description, and patterns within the data values. Imprecise or inaccurate metadata labels can lead to incorrect assumptions and interpretations. Resolution of this lack of (or missing) metadata requires both the technical analyst and clinical subject matter expert to work collaboratively to understand the data. In some cases, our teams have had to run simulations in sandbox environments of proprietary software and examine output to understand how the data are being generated by the system.

Example: Shown to User Data Column Label

Ambiguous metadata labels can be misleading. A binary data field (such as Boolean values) with a field name of *Shown_to_User* in a clinical decision support (CDS) database may lead a data consumer to guess that the field represents whether or not users of the system were exposed to CDS, for example, whether or not a user saw an alert. Data inspected after some reverse-engineering exercises involving simulation in a nonproduction environment, however, reveals that the data field represents a configuration setting (ie, whether the alert should be shown to user). This was discovered after running simulations in which the users conjure the alert in a scenario, but the Shown_to_User data for the corresponding records did not convert from a FALSE value to TRUE. Closer inspection reveals that the default setting for the alert is to not be shown to the user (Shown_to_User is FALSE) and that the data field is static (after viewing CDS, the data remain set to FALSE instead of converting to a TRUE value). Previous analyses using the Shown_to_User data have potentially been inaccurate, and any inferences based on this false assumption can be latently wrong and hard to detect.

Challenge Category: Workflow Imprints—Issues That Affect Performance and Capabilities of Algorithms

Appreciation of workflows is paramount to any clinical studies, and the same can be said for clinical informatics research with data generated by operations-supporting software and hardware. Unique clinical workflow idiosyncrasies can strongly affect the performance of real-time CDS algorithms and applications. These workflows can manifest in obvious and subtle ways in the data. The former is easy to detect but may or may not be easy to adjust for. The subtler cases, much like the metadata example above, can be harder to notice and accommodate for. As workflow is a universal and heterogenous concern, we will present several challenges and examples below. To minimize this challenge, data stakeholders must understand both the clinical workflows that generate the data and the technical systems that manipulate and store the data.

Example: Patient Deterioration in the Neonatal Intensive Care Unit Necessitates Verbal Orders

In the intensive care unit (ICU) environment, the clinical condition of patients can change quickly and abruptly. As such, medications are frequently ordered and administered in patterns that do not match the usual pattern of *order-then-administration*. For instance, in the case of a rapidly declining patient who needs more medication for blood pressure support (eg, an ionotrope), a physician or other prescriber may ask the bedside nurse to increase the dose of an already infusing medication before placing an electronic order for the rate in the EHR. This may happen because the prescriber is tending to other bedside needs in a critically ill patient and does not want to step away to place or document the order. The nurse may therefore document the MAR before the order which, in our algorithms, may be considered an erroneous dose as no electronic record of the order exists yet. To accommodate this challenge, we allow for a 30-min lag after the MAR is documented before calling the administration an error. This 30-min window is *waiting* for a verbal order to be documented (within a reasonable time frame) before triggering a notification that a potential error has occurred. This creates a trade-off between the *real-timeness* of the algorithms and the desire to decrease false positive error calls and notifications.

Challenge: Delayed Action on Active Orders

Total parenteral nutrition (TPN) is an intravenous nutrition alternative given to patients who cannot tolerate normal feeding, orally or through feeding tubes. It is a complex compounded liquid that takes time to prepare. TPN orders are often placed early in the day, with the knowledge that the new solution will not be administered until afternoon or evening. The rates on the orders, however, may be active at the time the order is placed and be changed over the day. Algorithms that use TPN data must take this factor into account and allow for a delayed action on the otherwise (technically) active order.

Challenge: Priming Pumps—Speeding Up Infusion Pump Rates to Prime May Look Like an Error But Has No Clinical Consequence

Before administering intravenous medications via an infusion pump, bedside staff must often *prime* the pump equipment by filling the lines/tubing with the medication substance to remove air and provide consistent delivery of the drug. The need to do so efficiently often leads to adjusting pump rates to high levels to accomplish priming quickly. This may manifest in the data as apparent erroneously high rates of infusion if there is no tagging of the priming action or other record in the data. Bolusing a medication (an intentional brief high rate of delivery of medication to the patient) over an already continuously infusing rate can also mislead data consumers to falsely believe an error has occurred. Inquiries to clinicians about their workflows in relation to the data patterns observed will often explain the digital manifestations and require refinements to real-time systems to account for these anomalies.

Challenge Category: Unstructured Data Entry

Approximately 30% to 50% of data entries useful for quality improvement are available only in an unstructured text format in modern EHRs [28]. The importance of this information has gained increasing recognition for quality improvement and patient safety. In the medication safety project described above, our clinicians often use both structured computerized provider order entries and free-text communication orders to prescribe allowable dose adjustments and ranges. As a best practice across the institution, the clinicians are encouraged to make dose adjustment via structured entries. However, the continuous infusion medications usually include instructions for frequent and complex dose changes related to the physiologic state of the patient. In this regard, the existence of free-text orders are artifacts that reflect the dynamic and changing clinical status of typical critical care patients. To identify information embedded in unstructured narratives, natural language processing (NLP) has become a critical component of computerized clinical support. Nevertheless, the complexity and heterogeneity of the human language makes its application and dissemination a challenging task. Conversion of unstructured data to structured data, when possible and favorable, minimizes this challenge.

Challenge: Complexity of Human Language

Different from structured entries that are enumerable, the complexity of human language creates an infinite space with countless linguistic variants. Accounting for the variants with NLP is therefore an onerous task.

Example: Free Text Dosing of Total Parenteral Nutrition

For instance, a physician could specify a dose adjustment explicitly as, “Please decrease TPN to 10 mL/hr”. They could also specify the dose adjustment implicitly as, “Please decrease TPN rate so that TPN rate + feeding rate=12 mL/hr”. In addition, modifiers are commonly used to adjust one’s meaning. For example, the physician could specify, “When the new bag arrives, please decrease the rate to 10 mL/hr”, suggesting that the rate is prescribed for when the next medication supply bag is delivered to the bedside. To accommodate language complexity, researchers tend to use flexible language constraints (eg, loose regular expressions) to parse narrative content. To address this challenge, we have evolutionarily modified our algorithms as we have identified false negatives and false positives and have found that necessary changes have decreased in frequency over time. Nevertheless, edge cases are identified in an ad hoc manner and require manual inspection constantly.

Challenge: Heterogeneity of Human Language

Although it is rarely noticed in single-institution applications, heterogeneity of the human language has been recognized as a major barrier to the dissemination of NLP-integrated CDS tools across health care institutions. For instance, the free-text narratives from a pediatric health care institution would primarily describe signs, symptoms, procedures, and medications for pediatric patients, whereas the narratives from a general hospital would primarily describe those for adult patients. The language heterogeneity therefore greatly affects the performance of CDS tools delivered from one institution to another. Customization is necessary for using NLP tools developed from external institutions and it requires both domain expertise and intensive resources.

Challenge Category: Fusing Datasets and the Role of Device Integration

Challenge: Merging Datasets From Multiple Sources Requires Valid Linking Identifiers

Many medical devices currently in use in health care facilities are not fully integrated in a *closed-loop* manner with EHRs and other systems, meaning that they are often not fully interoperable.

Example: The Nonintegration of Smart Pumps With Electronic Health Records

Infusion pump programming errors may result in a combination of missing patient and/or drug identifiers. For example, a patient’s medical record number (MRN) or encounter ID may be mis-entered or bypassed altogether. Similarly, a basic infusion rate may be selected without specifying the medication being administered. In the event of missing or incorrect patient or medication identifiers, linking smart pump generated data with order logs or MARs, particularly in real-time, becomes vastly more complicated. This is particularly true when the administered drug is commonly used and may have been ordered for several patients concurrently, which makes inference by the time of administration difficult. In our previous evaluation of NICU smart infusion pumps, we found in a convenience sample that although 89% of pump records included the medication ID, 76% contained a valid patient ID and only 68% contained both valid patient and medication IDs. Fortunately, only 3% were missing both identifiers (Table 2). Both data elements must be valid to accurately and efficiently link EHRs and infusion pump data.

Challenge: Clinical Decision Support Blind Spots

Almost all forms of CDS depends on some data elements to be present to drive the rules engines underlying the CDS platform. Lack of that data leads to unavailability of CDS and lost opportunities to inform and optimize treatment decision making.

Example: Use of Smart Infusion Pump Drug Libraries

Reviewing of the smart infusion pump-related literature reveals that one of the largest purported benefits of using smart pumps is to perform medication dose and administration checking at the time of pump programming by a bedside user. Smart pumps use drug libraries with various rules around allowable doses and rates of administration. However, use of the libraries requires, at a minimum, knowledge about what drug is about to be infused so it *knows* what drug library rules to enforce. A frequent workaround to programming this information into the pump is to use a generic drug profile, which (1) does not label the data with any kind of clear medication identifier and (2) effectively renders the potential CDS unusable, as the rules cannot be invoked. This leads to a CDS blind spot.

Challenge Category: Retrospective Versus Prospective Data/Detection

Challenge: Retrospective Data and Real-Time Data are Processed and Accessed Differently

There are fundamental differences in both how one accesses and processes data depending on, among other factors, whether it is coming from a data repository such as a relational database or from live data feeds from the operational source that is generating the data (such as via a real-time data interface). Retrospective data sources have frequently been subject to many ETL operations, any of which can and usually do alter the data in some fashion. These alterations are not always clear, and byproducts can arise in the data which are hard to detect and account for. ETL operations are also opportunities for errors to arise, many of which can be silent and lead to issues such as incomplete data extractions and incomplete datasets.

Table 2. Distribution of smart infusion pump data records with valid medication and patient IDs.

Medication ID	Patient ID, n (%)	
	Present	Missing
Present	5440 (68)	1680 (21)
Missing	640 (8)	240 (3)

Example: The Order Audit Modification Issue

A pattern of medication data was uncovered during retrospective data review that demonstrated the perils of taking data at face value. When an order was modified (recorded as an audit), the new value overwrote the original value. However, the data user viewing the data as an individual record would not notice this pattern and assume the value presented was the original value. The data quality issue would also not be noticed when analyzing the data in aggregate (such as performing descriptive statistics on doses ordered using the data). Real-time data interfaces would pull the data in real-time and send it for processing, which would mitigate the concern of using erroneous data. Review of these data obtained via retrospective reports (pulled from a data repository) would have different values from data obtained in a real-time, prospective manner but the difference would be extremely difficult to identify under normal inspection procedures. Only with rigorous examination would the issue be uncovered. We have applied postprocessing *patches* to our datasets to accommodate for ETL-related idiosyncrasies that we have noted. These patches revert the data to a state more reflective of reality.

Challenge Category: Technical Versus Clinical Validity

A primary purpose of real-time error detection should be to mitigate patient's exposure to harm. Detected errors which do not support that purpose should not necessarily be reported in real-time owing to their limited utility. For example, an error may be predictably detected based on improper documentation in which the correct dose was administered but the clinicians did not follow standards of practice in timely documentation of medication use (eg, placing an electronic order after a prolonged period of time following the corresponding verbal order). Although the detection algorithm performed as expected based on documentation, these errors have no potential to cause harm and are not clinically consequential. Similarly, overly sensitive algorithms may detect dosing administration errors that are technically incompatible with medication orders—though the dosing discrepancy is clinically meaningless and presents minimal risk of causing harm. An example would be a dosing discrepancy error where the dose given exceeds the prescribed dose by 1% or 2%. For almost all medications, this *overdose* is not clinically significant. To avoid inundating clinicians with unactionable alerts, some tolerance should be included in error detection algorithms to accommodate true errors that have little or no potential for clinical impact. A final example of an uninformative though technically accurate alert occurs when repeat errors are detected for the same underlying, persistent but inconsequential event.

Discussion

Practical Implications

In this paper, we have compiled many challenges unique to using real-time data. Most of the enumerated challenges and examples in this paper will be familiar to many data stakeholders, from requestors to the technical analysts who work diligently to provision complete, accurate datasets. The challenges were presented in both high-level overviews and grounded in specific examples, but the principles behind them are generalizable to almost any kind of data-based work. The different challenges should be seen as equally important as a standalone list, but some will be more important to monitor and address than others, depending on the use case and types of data involved. Various mitigation strategies were presented here, but the commonality among most of them is that to be effective, they must generally be built on a foundation of multi-disciplinary teamwork. These teams need to be collaborative in nature and be composed of members with strong data science knowledge, biomedical informatics skills, and familiarity with the clinical processes and workflows that are at the heart of the project in progress. As with any challenging process, these efforts take time and resources; this should be factored into any timelines and planning as appropriate. We typically recommend adding a time buffer to accommodate this in the initial stages of a project, as well as allocation of some resources for maintenance once the project is in a sustain phase. The systems that create and pass along the data are in constant flux—so too should be the vigilance of the downstream systems utilizing the data.

It should be noted that our list here should be seen as a starting point—certainly others in the field could easily add many other types of categories and examples from many different domains of the digital health domain and beyond. This report was meant to start the conversation and bring recognition to the fact that working with real-time data is not as easy to do practically as it is conceptually. This represents the first crucial step.

As with an early work, there are many ways we can envision advancing our knowledge in this area. Furthermore, research in these unique data challenges could focus on formalizing the list further and specifying the importance of each in different contexts. It would be particularly helpful if future investigators could continue to report on best practices and strategies of how to address specific challenges, as well as develop a systematic framework to assist in avoiding the issues in the first place or mitigating them once they have occurred.

Data are a powerful tool—but we must recognize and promote best practices if we are to get the most value out of data and not derive false insight from our work.

Conclusions

The use of real-time data to drive safety event detection and clinical decision support is extremely powerful but presents its own set of challenges including data quality and technical complexity. These challenges must be recognized and accommodated for if the full promise of accurate, real-time safety event CDS is to be realized.

Acknowledgments

Research reported in this publication was fully funded and supported by the National Library of Medicine of the National Institutes of Health under award number R01LM012230. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Previous research was supported by NIH/NICHHD award number R21HD072883.

Conflicts of Interest

Eric Kirkendall has a licensing agreement with Vigilanz Corp for real-time algorithms for detection of acute kidney injury developed at Cincinnati Children's.

References

1. Russo E, Sittig DF, Murphy DR, Singh H. Challenges in patient safety improvement research in the era of electronic health records. Healthc (Amst) 2016 Dec;4(4):285-290 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.06.005] [Medline: 27473472]
2. Rosenbloom ST, Carroll RJ, Warner JL, Matheny ME, Denny JC. Representing knowledge consistently across health systems. Yearb Med Inform 2017 Aug;26(1):139-147 [FREE Full text] [doi: 10.15265/IY-2017-018] [Medline: 29063555]
3. Preussler JM, Mau L, Majhail NS, Meyer CL, Denzen EM, Edsall KC, et al. Administrative claims data for economic analyses in hematopoietic cell transplantation: challenges and opportunities. Biol Blood Marrow Transplant 2016 Dec;22(10):1738-1746 [FREE Full text] [doi: 10.1016/j.bbmt.2016.05.005] [Medline: 27184624]
4. Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf 2012 Jan;21(Suppl 1):82-89. [doi: 10.1002/pds.2321] [Medline: 22262596]
5. Nelson JC, Cook AJ, Yu O, Dominguez C, Zhao S, Greene SK, et al. Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data. Pharmacoepidemiol Drug Saf 2012 Jan;21(Suppl 1):62-71. [doi: 10.1002/pds.2324] [Medline: 22262594]
6. Health Level Seven International. 2018 Dec 5. URL: http://www.hl7.org/ [accessed 2018-12-06] [WebCite Cache ID 74StWU7GV]
7. Health Level Seven International. 2018 Dec 5. HL7 Fast Healthcare Interoperability Resources Specification (aka FHIR®), Release 3 (STU) URL: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=449 [accessed 2018-12-06] [WebCite Cache ID 74SteFCjp]
8. Observational Health Data Sciences and Informatics. 2018 Dec 5. OMOP Common Data Model URL: https://www.ohdsi.org/data-standardization/the-common-data-model/ [accessed 2018-12-06] [WebCite Cache ID 74SttdjCH]
9. Klann JG, Phillips LC, Herrick C, Joss MA, Wagholikar KB, Murphy SN. Web services for data warehouses: OMOP and PCORnet on i2b2. J Am Med Inform Assoc 2018 Oct 01;25(10):1331-1338 [FREE Full text] [doi: 10.1093/jamia/ocy093] [Medline: 30085008]
10. Classen D, Li M, Miller S, Ladner D. An electronic health record-based real-time analytics program for patient safety surveillance and improvement. Health Aff (Millwood) 2018 Nov;37(11):1805-1812. [doi: 10.1377/hlthaff.2018.0728] [Medline: 30395491]
11. Barnett A, Winning M, Canaris S, Cleary M, Staib A, Sullivan C. Digital transformation of hospital quality and safety: real-time data for real-time action. Aust Health Rev 2018 Nov 02:- Epub ahead of print. [doi: 10.1071/AH18125] [Medline: 30384880]
12. Shaw SJ, Jacobs B, Stockwell DC, Futterman C, Spaeder MC. Effect of a real-time pediatric ICU safety bundle dashboard on quality improvement measures. Jt Comm J Qual Patient Saf 2015 Sep;41(9):414-420. [Medline: 26289236]
13. Shelov E, Muthu N, Wolfe H, Traynor D, Craig N, Bonafide C, et al. Design and implementation of a pediatric ICU acuity scoring tool as clinical decision support. Appl Clin Inform 2018 Jul;9(3):576-587. [doi: 10.1055/s-0038-1667122] [Medline: 30068013]
14. Deleger L, Brodzinski H, Zhai H, Li Q, Lingren T, Kirkendall ES, et al. Developing and evaluating an automated appendicitis risk stratification algorithm for pediatric patients in the emergency department. J Am Med Inform Assoc 2013 Dec;20(e2):e212-e220 [FREE Full text] [doi: 10.1136/amiajnl-2013-001962] [Medline: 24130231]
15. Goldstein SL, Kirkendall E, Nguyen H, Schaffzin JK, Bucuvalas J, Bracke T, et al. Electronic health record identification of nephrotoxin exposure and associated acute kidney injury. Pediatrics 2013 Sep;132(3):e756-e767. [doi: 10.1542/peds.2013-0794] [Medline: 23940245]

16. Goldstein SL, Mottes T, Simpson K, Barclay C, Muething S, Haslam DB, et al. A sustained quality improvement program reduces nephrotoxic medication-associated acute kidney injury. Kidney Int 2016 Dec;90(1):212-221. [doi: 10.1016/j.kint.2016.03.031] [Medline: 27217196]
17. Kirkendall ES, Spires WL, Mottes TA, Schaffzin JK, Barclay C, Goldstein SL. Development and performance of electronic acute kidney injury triggers to identify pediatric patients at risk for nephrotoxic medication-associated harm. Appl Clin Inform 2014;5(2):313-333 [FREE Full text] [doi: 10.4338/ACI-2013-12-RA-0102] [Medline: 25024752]
18. Stockwell DC, Kirkendall E, Muething SE, Kloppenborg E, Vinodrao H, Jacobs BR. Automated adverse event detection collaborative: electronic adverse event identification, classification, and corrective actions across academic pediatric institutions. J Patient Saf 2013 Dec;9(4):203-210. [doi: 10.1097/PTS.0000000000000055] [Medline: 24257063]
19. Kirkendall ES. Casting a wider safety net: the promise of electronic safety event detection systems. Jt Comm J Qual Patient Saf 2017 Apr;43(4):153-154. [doi: 10.1016/j.jcjq.2017.01.002] [Medline: 28325202]
20. Li Q, Melton K, Lingren T, Kirkendall ES, Hall E, Zhai H, et al. Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. J Am Med Inform Assoc 2014;21(5):776-784 [FREE Full text] [doi: 10.1136/amiajnl-2013-001914] [Medline: 24401171]
21. Li Q, Kirkendall ES, Hall ES, Ni Y, Lingren T, Kaiser M, et al. Automated detection of medication administration errors in neonatal intensive care. J Biomed Inform 2015 Oct;57:124-133 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.012] [Medline: 26190267]
22. Ni Y, Lingren T, Hall ES, Leonard M, Melton K, Kirkendall ES. Designing and evaluating an automated system for real-time medication administration error detection in a neonatal intensive care unit. J Am Med Inform Assoc 2018 May 01;25(5):555-563 [FREE Full text] [doi: 10.1093/jamia/ocx156] [Medline: 29329456]
23. Wellner B, Grand J, Canzone E, Coarr M, Brady PW, Simmons J, et al. Predicting unplanned transfers to the intensive care unit: a machine learning approach leveraging diverse clinical elements. JMIR Med Inform 2017 Nov 22;5(4):e45 [FREE Full text] [doi: 10.2196/medinform.8680] [Medline: 29167089]
24. Ni Y, Kennebeck S, Dexheimer JW, McAneney CM, Tang H, Lingren T, et al. Automated clinical trial eligibility prescreening: increasing the efficiency of patient identification for clinical trials in the emergency department. J Am Med Inform Assoc 2015 Jan;22(1):166-178 [FREE Full text] [doi: 10.1136/amiajnl-2014-002887] [Medline: 25030032]
25. Ni Y, Beck AF, Taylor R, Dyas J, Solti I, Grupp-Phelan J, et al. Will they participate? Predicting patients' response to clinical trial invitations in a pediatric emergency department. J Am Med Inform Assoc 2016 Dec;23(4):671-680 [FREE Full text] [doi: 10.1093/jamia/ocv216] [Medline: 27121609]
26. Consortium WWW. 2018 Dec 1. Date and Time Formats URL: https://www.w3.org/TR/NOTE-datetime [accessed 2018-12-06] [WebCite Cache ID 74SulwDEB]
27. International Organization for Standardization. 2018 Dec 1. ISO 8601: Data elements and interchange formats--Information interchange--Representation of dates and times URL: https://www.iso.org/standard/40874.html [accessed 2018-12-06] [WebCite Cache ID 74Su0wr4H]
28. Hicks J. RAND Corporation. 2018 Dec 6. The Potential of Claims Data to Support the Measurement of Health Care Quality URL: https://www.rand.org/pubs/rgs_dissertations/RGSD171.html [accessed 2018-12-06] [WebCite Cache ID 74StlfI34]

Abbreviations

AKI: acute kidney injury
CDS: clinical decision support
CPOE: Computerized Provider Order Entry
EHR: electronic health record
ETL: extract-transfer-load
ICU: intensive care unit
ISO: International Organization for Standardization
MRN: Medical Record Number
MAR: Medication Administration Record
NLP: natural language processing
NICU: neonatal intensive care unit
TPN: total parenteral nutrition
UTC: Coordinated Universal Time

Edited by G Eysenbach; submitted 06.12.18; peer-reviewed by A Staib, J Hefner, W Hackl; comments to author 07.02.19; revised version received 04.03.19; accepted 05.04.19; published 22.05.19.

Please cite as:
Kirkendall ES, Ni Y, Lingren T, Leonard M, Hall ES, Melton K
Data Challenges With Real-Time Safety Event Detection And Clinical Decision Support
J Med Internet Res 2019;21(5):e13047
URL: http://www.jmir.org/2019/5/e13047/
doi:10.2196/13047
PMID:31120022

Viewpoint

Participatory Disease Surveillance Systems: Ethical Framework

Lester Darryl Geneviève[1], MBChB, MSc; Andrea Martani[1], ML; Tenzin Wangmo[1], PhD; Daniela Paolotti[2], PhD; Carl Koppeschaar[3], MSc; Charlotte Kjelsø[4], MEd; Caroline Guerrisi[5], MPH; Marco Hirsch[6], MSc; Olivia Woolley-Meza[7,8], PhD; Paul Lukowicz[6], PhD; Antoine Flahault[9], MD, PhD; Bernice Simone Elger[1,10], MD

[1]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland
[2]Institue for Scientific Interchange Foundation, Torino, Italy
[3]De Grote Griepmeting, Science in Action BV, Amsterdam, Netherlands
[4]Statens Serum Institut, København, Denmark
[5]Sorbonne Université, Institut National de la Santé et de la Recherche Médicale, Institut Pierre Louis d'Épidémiologie et de Santé Publique, Paris, France
[6]German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
[7]ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland
[8]Novartis Pharma AG, Basel, Switzerland
[9]Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland
[10]University Center of Legal Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:
Lester Darryl Geneviève, MBChB, MSc
Institute for Biomedical Ethics
University of Basel
Bernoullistrasse 28
Basel, 4056
Switzerland
Phone: 41 61 207 17 82
Fax: 41 61 267 17 80
Email: lester.genevieve@unibas.ch

Abstract

Advances in information technology are changing public health at an unprecedented rate. Participatory surveillance systems are contributing to public health by actively engaging digital (eg, Web-based) communities of volunteer citizens to report symptoms and other pertinent information on public health threats and also by empowering individuals to promptly respond to them. However, this digital model raises ethical issues on top of those inherent in traditional forms of public health surveillance. Research ethics are undergoing significant changes in the digital era where not only participants' physical and psychological well-being but also the protection of their sensitive data have to be considered. In this paper, the digital platform of Influenzanet is used as a case study to illustrate those ethical challenges posed to participatory surveillance systems using digital platforms and mobile apps. These ethical challenges include the implementation of electronic consent, the protection of participants' privacy, the promotion of justice, and the need for interdisciplinary capacity building of research ethics committees. On the basis of our analysis, we propose a framework to regulate and strengthen ethical approaches in the field of digital public health surveillance.

(J Med Internet Res 2019;21(5):e12273) doi:10.2196/12273

KEYWORDS

ethics; research; influenza, human; smartphone; public health surveillance

Introduction

Advances in information technology are changing medical research [1] and public health at an unprecedented rate [2]. One of the most evident changes is that it has become easy for members of the general public to contribute to public health surveillance, practice, and policy [2] by sharing personal and health-related information through digital media. The pervasiveness of technology is underscored by the fact that as of 2018, almost 4 billion people are estimated to have access to the internet [3], and there are over 318,000 health-related mobile apps [4]. In public health surveillance (ie, public health data collection and analysis to inform public health practice [5]), vast real-time health data from informal sources (eg, health-related mobile apps and twitter) allow an early detection, prevention, and monitoring of public health threats and the potential for a prompt response from authorities to mitigate them. These informal sources have facilitated the reporting of diseases by complementing and reducing the time information is transmitted in multilevel public health infrastructures. Consequently, around the world, several early warning systems are now using this innovative approach [6].

Such activities have been termed *digital epidemiology* or *digital disease detection* [7,8]. Digital epidemiology can either be performed for nonresearch or research purposes. On the one hand, if the aims of the surveillance system are simply to monitor, control, and respond to health threats by producing data on the affected population, then this serves nonresearch purposes. On the other hand, if the purpose of the surveillance system is either to contribute to or to produce generalizable knowledge, potentially applicable to different populations and settings, then it serves research purposes [9].

Data in digital epidemiology can be obtained through 2 distinct approaches, with similar public health objectives but usually different challenges [10]. With a passive approach for data collection, data subjects are not directly informed that their everyday data (stored, eg, on social platforms, blogs, and Web search queries) are being mined and processed by advanced algorithms involved in *big data* analytics to monitor or predict disease outbreaks [10-12]. One of the first notable examples of such passive data collection approach to digital epidemiology was *Google Flu Trends* (GFT). There are numerous challenges to this approach, including *big data hubris* and unstable algorithm dynamics. *Big data hubris* states that big data are simply a replacement for data collected and analyzed by conventional means rather than an adjunct to traditional public health surveillance. Unstable algorithm dynamics refer to the continuous changes made by the company to the search algorithms, as a means of improving their searching capabilities by incorporating, for instance, new search terms. However, as the case of GFT demonstrated, these improvements led to biased estimates [13]. Nonetheless, *big data* surveillance offers an unprecedented opportunity to monitor in a timely manner the spatial and temporal evolutions of epidemics with increased granularity compared with more traditional surveillance systems, provided that the potential flaws of *big data* analytics are taken into consideration [12,13].

On the other hand, a *participatory disease surveillance* system has an active approach involving digital communities of volunteer citizens who consciously provide data. This can be done either interactively by reporting their symptoms and other relevant information through an appropriate interface or by *donating* sensor data (eg, location traces) from their digital devices. Such a participatory approach not only supports the detection of potential public health threats but also empowers individuals to reflect on them and adapt their behavior accordingly [10]. An example of such participatory disease surveillance system is the European Influenzanet Consortium, which monitors *influenza-like illness* (ILI) activity during flu seasons, with data from volunteer citizens using digital national platforms and in some cases, mobile apps [14,15]. Details on the Influenzanet Consortium and its inclusion as an adequate model to illustrate ethical issues pertinent to participatory disease surveillance systems have been covered in a previous publication [16].

In 2009, the Influenzanet Consortium was formed to standardize practices among the individual national *ILI* surveillance platforms to promote collaboration [17]. Recent research started exploring the use of crowdsourcing for detection of epidemic flu spreading, improving the self-reporting experience of symptoms with a more user-friendly mobile phone apps, and enriching the data with context data recorded by the phone's sensors [15]. The Consortium follows the top-down model of citizen participation [2], which guarantees the scientific requirements and integrity of the disease surveillance network while relying on volunteer citizens' data. This technology-driven public health surveillance has some benefits for its participants. Real-time information on *ILI* activity at local and national levels is provided to participants, who are also advised on strategies for disease prevention [17,18]. Importantly, the participants contribute directly to the ultimate goal of this public health initiative by providing real-time granular health-related data on *ILI* [19]. Such information complements the data of the *European Influenza Surveillance Network* (EISN) at finer levels, as EISN receives mainly epidemiological and virological data from its network of general practitioners [17,18]. The large cohorts of Influenzanet (eg, over 36,000 volunteer citizens for the flu season 2015/2016) also allow detection of even small epidemics of *ILI* [17,20]. This early detection [21] could potentially enable timely mitigation strategies to reduce the health burden of influenza and decrease health expenditures associated with increased hospitalization and treatment. In addition, Influenzanet enables research and the study of subgroups, for example, influenza vaccine effectiveness in vaccinated groups [22], attitudes toward vaccination [23], health status of population outside the health care system [24], and differences in medical care–seeking behavior across the European Union (EU) [20].

However, such top-down model of citizen participation in surveillance, specifically for research purposes, raises its own set of ethical issues on top of those inherent in traditional forms. Participants are involved actively in scientific research [2], but researchers have limited personal interaction with participants to ensure that they have indeed understood the research information provided on the national platforms or mobile apps and potential risks that their participation entails. In addition, participants could be influenced by the promise of expected benefits and the imperative of altruism. Therefore, it becomes a challenge to combine protection of research participants with the promotion of high-quality data collection for ethically acceptable research purposes [25]. In participatory disease surveillance systems, these are closely linked and mutually dependent on each other. On the one hand, ensuring participants' trust and engagement through adequate safeguards is crucial for the sustainability and quality of these surveillance systems. If participants perceive the risk of privacy violation, they might refrain from giving important information, thus, affecting the effectiveness of disease surveillance and subsequent future public health interventions [26]. On the other hand, if surveillance systems follow low-quality standards and operate outside an ethical framework, protection and collaborations of data subjects cannot be secured. Numerous ethical frameworks have been developed in the field of public health surveillance and the use of big data [27-29]. However, a 2017 systematic review on ethical issues of public health surveillance revealed that there is a need for more context-specific analyses to guide public health practice [5]. Consequently, providing an ethical framework for the regulation of such innovative participatory surveillance methods, using a real-world example, becomes of utmost importance.

In this paper, we use Influenzanet to illustrate challenges in protection of health and other sensitive information reported in participatory disease surveillance systems. We discuss and analyze challenges and needs of participant consent in surveillance and research using participant surveillance systems data. We argue that research ethics committee (RECs) should play an important role in this developing field. Finally, we propose a framework for the regulation of digital participatory disease surveillance systems, which strengthens protection of participants' data and privacy, while promoting the concept of justice.

Consent

Traditional Informed Consent in Internet-Based Surveillance

In public health surveillance, there are 2 antithetical forces. Although these systems pursue the improvement of population health through surveillance of diseases (such as in the case of Influenzanet, providing protection for vulnerable populations at risk of serious adverse outcomes of influenza infection), they must also safeguard individuals against any abuse of their data by researchers [30,31]. To strike a balance between the pursuit of societal welfare and protection of individual rights, consent from participants plays a fundamental role. Originating from the necessity to protect research subjects both physically and mentally, written informed consent is traditionally obtained for medical research, and its importance has emerged even more in the current era of data protection [32]. However, this type of consent seems to be poorly adapted to the collection and use of digital data in public health surveillance [2].

In light of the inadequacy of traditional informed consent for participatory public health surveillance, 1 potential response is to reject the need for further consent in these types of studies [33] because of the fact that participants enroll on their own and not in continued medical contact. The "no problem" solution rests also on the assumption that consent is not necessary, as in public health surveillance, individual interests may be put aside to protect the public good [33]. For example, in the United States, the issue of consent in public health surveillance is circumvented by considering the latter as public health practice instead of research, thus exempting it from institutional review boards' approval [34] and in most cases, of traditional informed consent requirements. Indeed, a participatory disease surveillance platform active in the United States, *Flu Near You*, received a waiver for informed consent [35]. However, this approach may not be the best solution for participatory disease surveillance where data are actively generated by participants, underscoring the urgent need to adapt the traditional model of informed consent more adequately to this type of surveillance system.

Informed consent in research was originally designed for studies involving a limited number of participants where it was practically and financially feasible for researchers to engage participants, provide details about the research, and obtain written informed consent before the beginning of the study [32]. A further problem with traditional informed consent is that it was designed to authorize the use of data only by those subjects and for those purposes according to which the data had been originally collected. It was, however, not intended to also cover retrospective research on samples or data. In the case of *big data* for surveillance of infectious diseases, which is often retrospective in nature (2 out of the 3 electronic data sources are medical encounter and nonhealth digital data) [12], obtaining traditional informed consent proves problematic, as it requires disclosing all potential risks of primary and retrospective research, but the latter are usually unknown at the time when data are collected [32].

There have been substantial efforts made by some Influenzanet national platforms at their outset to ensure some form of personal interaction with their participants to better explain the nature of the surveillance system. In 2003, the original Belgian/Dutch platform, called *de Grote Griepmeting* (ie, the Great Influenza Survey), received a lot of media attention, which led to the registration of tens of thousands of participants in 2003/2004. The participants' age distribution from youth to the elderly and their wide geographic spread and different levels of education made de Grote Griepmeting more accurate and quicker to signal the onset of a flu epidemic than the general practitioners' surveillance system organized by Netherlands Institute for Health Services Research (NIVEL). The Belgian/Dutch research team invited participants for an information, question and answer session, where they were provided with notes on the rationale of the survey questions to the flu survey study. Moreover, a forum was also created where participants could ask any remaining questions, and when specific virology questions were asked, consultation would follow with partners from the NIVEL and the National Institute for Health and Environment in the Netherlands. The team managed to answer all incoming questions from participants by email and during various local, national, and regional live radio interviews from people listening in. In 2009, the team started a public community on Facebook named *De Grote Griepmeting–Influenzanet*, where members would have their questions answered by the team. Such measures (ie, the information session and the team answering all additional questions received via email, through their forum, on the social platform Facebook, and during radio interviews) could be viewed as an equivalent solution to obtaining the informed consent of these participants (De Grote Griepmeting, email communication, April 3, 2018 and February 7, 2019).

The inadequacies of traditional informed consent have led to the development of many other ethically acceptable solutions. For instance, in retrospective research where risks are minimal, consent would not be necessary as long as the right to opt out and the right to be forgotten are preserved and enforced [36,37]. Alternatively, the requirement of informed consent upholds but is paired with waivers, which dispense researchers from requesting consent for secondary use of data, if the recontacting and reconsenting are unfeasible or would lead to nonrepresentative samples [38].

Another alternative to traditional informed consent is an *extended* version of consent, which is more suitable to public health/*big data* research, known as broad or general consent [2,32,39-41]. The key difference between traditional and broad or general consent is that data subjects provide their consent for entire classes of research [42]. This extended form of consent differs from blanket consent as data subjects do not give permission for any use of their data but rather define in broad terms the purposes of use [42]. Moreover, broad consent is only considered acceptable if 2 criteria are met. First, every new study needs to be approved by an REC or another competent entity [43]. Second, the right of participants to withdraw their consent at any time has to be maintained [32,44]. Despite the presence of these safeguards, consensus on whether broad consent can be considered truly informed is lacking [45-47]. The informative nature of broad consent rests on the assumption that autonomy is protected, as REC approval is necessary, and strategies to regularly update the data donor on ongoing opt-out opportunities are devised [48]. Furthermore, any modification to the research should automatically lead to reconsenting procedures [49]. However, broad consent cannot be entirely informed because of the unspecified nature of future research [48]. Although broad consent seems suitable for secondary uses in public health research involving digital communities of volunteer citizens or *big data*, it is uncertain whether broad consent represents the best solution in terms of respect for autonomy. Given the issues raised by broad consent and the fact that it requires initial face-to-face contact, seeking consent electronically could be an ethically satisfying alternative to traditional informed consent.

Electronic Consent, An Adaptation of Traditional Informed Consent

Electronic consent (e-consent) implies that participants give informed consent using an information technology (eg, digital technologies). In this sense, e-consent does not represent a new form of consent but simply an adaptation of informed consent to the electronic environment [50]. E-consent is currently being used in the Influenzanet Consortium and in similar participatory surveillance platforms such as *Flu Tracking* (Australia and New Zealand) as a valid form of consent for participants. Data subjects agree to the conditions, terms, and privacy policies when registering on their respective national platforms [16,51].

Although e-consent offers the substantial benefit of a tailored fit to the digital environment, it also has some inherent problems. A unique feature of internet-based research is the absence of personal interaction between researchers and participants, where researchers would traditionally be able to provide individually tailored information and answer any question participants might ask concerning the study and the collection of health data. Therefore, one of the major risks posed by e-consent is the provision of consent through automatic processes in the digital world, as parties are not directly involved. The provision of consent is rather based on a set of computer rules determining whether access to an individual’s data by researchers could be granted on reasonable grounds [50]. For this reason, it is possible that participants provide their consent without fully understanding—or even reading—the information, terms, and conditions that data collectors provide by simply clicking the relevant buttons in the digital forms [52]. We thus recommend that several precautions ought to be implemented when e-consent is obtained. For instance, e-consent should be designed in such a way that information is delivered through a simple PDF file where participants digitally put their signature (instead of clicking a button) to increase the likelihood of the document being read. Alternatively, other possibilities offered by information technology could be exploited to help verify participants’ understanding of the information provided during the e-consent process. These include tools such as the use of audio files, PowerPoint presentations, videos, pictures, or gamification (for instance, through quizzes and animation) [53].

Though the above recommendations could foster the informed nature of e-consent, the lack of personal interaction between study participants and researchers remains. Therefore, a properly implemented e-consent would be particularly beneficial in those studies where it is impossible to provide individual counseling and where the conditions, terms, and privacy policies would otherwise not be read [54,55]. In addition, one might even argue that participants are potentially less likely to consent under undue influence or constraints because of limited interaction with researchers. They can thus easily decline consent by signing out from the digital platforms whenever they feel the need to do so [52]. In this regard, e-consent increases the autonomy of participants.

Nonetheless, a further challenge raised by the absence of face-to-face contact is how to ensure that participants have the required legal capacity to legitimately give their e-consent after they electronically authenticate [52], because of the difficulty of verifying the participants’ identity. In this sense, even if measures were taken—such as quizzes or questionnaires—to ensure that participants have understood the research information, there would be little guarantee that those quizzes or study questionnaires are actually being completed by data subjects. A potential solution to this authentication issue could come from advances made in biometric identification technologies, commonly used for security purposes [52]. For instance, the use of face recognition technology [56] on computers and mobile phones or fingerprint recognition sensor in smartphones (eg, Touch ID by Apple Inc) could be used to verify the identity of participants throughout the process of e-consent. However, processing biometric data raises additional ethical and legal issues, in particular with respect to privacy. Biometric data, similar to genetic material, carries biological traits that are unique to data subjects and which could be easily used to reidentify them [57]. However, it must be noted that, although entailing sensitive personal information, the processing of biometric data can be lawful even without subjects’ consent if processing serves the public interest or scientific or statistical research purposes (eg, Article 9.2.(j) of the European General

Data Protection Regulation, GDPR) [58]. Ethical concerns with respect to biometric data might also be mitigated if, for example, the gathered biometric information was stored locally on the participants' computers or mobile devices and not transmitted to the research team or any third party.

In the case of multinational studies such as Influenzanet, an adequately implemented e-consent could consist of a standard informed consent information form [16], as a single PDF file, being delivered to participants at the time of their registration on the digital platform and followed by a new document each and every time new information is added on the single country websites. The information provided would have to be reader-friendly and succinctly summarized, thus nudging participants to read it thoroughly. Awareness of all potential ramifications because of their participation could be further improved through the provision of quiz questions. Grading of these quiz questions could then serve as a proxy to ensure adequate understanding of the informed consent information. This method has been employed at the *Harvard Personal Genome Project* (PGP) [59]. Participants were even provided with a study guide and were required to pass an enrollment test to be considered for the project. This additional burden to participation, which is justified for genetic research (*genetic exceptionalism*), should nonetheless remain minimal for Influenzanet to retain engagement of its participants. This is supported by the mildly sensitive nature of the gathered information and the low risks associated with this kind of surveillance. Indeed, the enrollment examination for the PGP was the main barrier to participation, with almost 60% of its users dropping out [59]. Digital signature of the consent form could also be a more personalized alternative, and it would also provide additional evidence on the identity of the participant, which altogether would enhance the informed nature of this e-consent procedure.

Protection of Subjects' Privacy

Epidemics forecasting studies and other public health research often gather useful and sensitive data on their participants, potentially interfering with their privacy. In the case of Influenzanet, protection of participants' privacy is secured by data anonymization and the use of a centralized database [18].

One might argue that full anonymization is not necessary for some public health surveillance, as part of the collected data is only mildly sensitive (eg, age group and gender) and thus poses only a minor threat to the fundamental rights and the privacy of participants even in case of misuse [60]. However, even nonpersonal information could be used to reveal much more sensitive information on data subjects if the former is coupled with additional geographical information, which is often collected by public health surveillance systems [61]. For instance, 1 of the core functions of Influenzanet is to map cases of *ILI* for the identification of hotspots of influenza outbreaks to model disease progression and implement effective prevention strategies. This spatiotemporal dimension of collected health data can enhance the privacy-invasive nature of epidemics forecasting research such as Influenzanet [62]. The collection of sensor and usage data from smartphones adds additional behavioral and context information, which, as shown in related work [15], has the potential to improve forecasting and risk analysis. Despite these potential benefits, even apparently nonpersonal data, such as a list of installed apps, can be an additional risk to the participant's privacy [63]. The Consortium took great care in protecting the privacy of its participants. In case of sensor data, information is processed directly on the user's device and only transmitted to Influenzanet in anonymized and highly aggregated form [15]. Location information of reported cases is, for example, never mapped to the individual level but rather to the postal code level [17], with only the aggregate number of cases shown. Some platforms went even further by randomizing virtual locations around the center of a large number of postal code areas taken together (eg, the De Grote Griepmeting platform). The grouping of postal codes areas was paramount for better protection of the privacy of participants, for example, in the case of a single participant in a postal code area (De Grote Griepmeting, email communication, April 3, 2018).

However, with increasing technological capabilities to integrate and analyze health data with local data, there are risks of leakage of sensitive information concerning participants' locations, which may lead to stigmatization of the particular locations as well as residents [62,64]. Even with full anonymization, cross-referencing of essential data gathered for epidemics research purposes (eg, sex, age, and medical conditions) with other databases could eventually lead to reidentification of data subjects [2,65]. For instance, 2 researchers showed it was feasible to reidentify individuals by matching a deidentified database on *Netflix* movie recommendations to available Web-based information (eg, Internet Movie Database) [66]. Hence, anonymization per se is not a sufficient measure to adequately protect privacy. The long life span of some anonymized datasets, which is often the case with epidemics forecasting studies, de facto increases the risks of reidentification and privacy breaches through repetitive data enrichment over time [54]. Consequently, reidentification should be considered a real risk for data subjects [2] even in case of anonymization. It is, thus, paramount to ensure ethical and accountable sharing of anonymized datasets between research institutions and to combine anonymization with other adequate data security measures to prevent misuse of data and unauthorized reidentification.

Justice

The concept of justice in research ethics is fully embodied by the policies of the Influenzanet Consortium as participation is free, open, voluntary, and nondiscriminative of any resident of the respective countries (except for Sweden, where some representativeness and comparison purposes are guaranteed by allowing participation through invitations only) [17]. This ensures a fair distribution of risks and benefits to all research participants and the public at large. If epidemics forecasting studies keep up with the high standards in terms of justice (ie, participation to the surveillance system is free, open, voluntary, and nondiscriminative) followed by the Influenzanet Consortium and similar platforms such as Flu Near You [35], the only remaining challenge would be dealing with those limitations on participation that are inherent to digital technologies. These limitations are commonly referred in the literature as the "*digital divide*" [2,67], and they concern both access to and proficiency with digital technologies. Access to digital technologies is also a product of many sociodemographic elements such as age, educational level of participants, ethnic groups, and their socioeconomic status [68,69]. This is reflected in the data collected by Influenzanet, which present an underrepresentation of younger age and elderly groups, an overrepresentation of the middle age group for both genders, and a higher educational level of participants in comparison with the general population [70]. In this respect, it could be claimed that epidemics forecasting studies such as Influenzanet are potentially empowering a more dynamic, informed societal group with a penchant for digital technology, whereas at the same time perpetuating the health inequalities between others [2]. However, it must also be stressed that public health surveillance benefits the public at large and not exclusively the participants. Public health surveillance, like biomedical research, is a public good, as the health benefits resulting from its interventions (based on knowledge generated from data subjects) are ultimately going to be shared with society [71]. In addition, the digital divide is decreasing annually, with technology becoming more and more pervasive [72]. Nevertheless, concerns about justice can be avoided only if results and disease prevention strategies are shared evenly and on a regular basis among all societal groups, something which the Influenzanet Consortium is promoting (eg, weekly national surveillance bulletins, regular press releases during the study, and radio broadcasting) [73,74]. Such regular results dissemination initiatives undeniably help in ensuring that expected benefits of research are shared more equally between societal groups. This could be further improved by granting access to more targeted and granular information on influenza activity to nonparticipants [74] under the concept of solidarity [75]. Such measures would allow a better protection of society as the spread of an influenza epidemic is an individual as well as a collective concern.

Capacity Building for Research Ethics Committees

Influenzanet and similar systems are faced with multifocal ethical and legal issues. For the safeguard of data subjects, appropriate oversight and specific regulation might be needed in the future. Currently, such oversight is beyond the governance capacity of RECs, as technological advances outpace national regulatory frameworks and undermine the definitions of those concepts—such as "anonymization," "encryption," and "personally identifiable information" [55]—upon which RECs rely. However, RECs should be actively involved in the design and implementation of public health research involving digital communities of volunteer citizens or big data. These RECs need to act as safety nets to fill the gaps of the current regulatory framework, which often dates back to an era where modern computational and technological capabilities were not foreseeable [55]. In this perspective, we recommend RECs to undergo interdisciplinary capacity building in those innovative research methods through mutual exchange of information and training with citizen science experts, big data researchers, data scientists, ethicists, legal experts, and sociologists. This would allow the identification of ethical and legal grey zones. Stakeholders could further anticipate potential conflicting situations resulting from the enactment of new legislation. This appears even more urgent as we have entered the GDPR era. This regulation came into force in May 2018, to replace the EU Data Protection Directive 95/46/EC [1,76]. The GDPR tries to harmonize EU data protection laws with the goal of guaranteeing the same level of freedom and protection to EU citizens, while protecting personal data during cross-border sharing with international organizations and third countries [76]. This legislative reform is likely to have a considerable impact on consent requirements and exemptions from obtaining consent [1]. This could affect the expected benefits that *big data* can bring to society by increasing the regulatory burden on public health surveillance studies [1]. Furthermore, as stressed in the study by Mittelstadt, the GDPR classifies "data concerning health" as a "special category of personal data" [77]. As this category includes any personal data that reveals information on the health status (physical or mental) of participants [77], health-related information gathered from Influenzanet participants or similar epidemics forecasting studies might—until properly anonymized—fall in this special category. It is, thus, possible that detailed limitations to health data usage are imposed in the future because of the protective stance endorsed by the GDPR [77]. Therefore, interdisciplinary capacity building and acquaintance of RECs with this innovative and developing research field will be paramount to proactively ensure an adequate protection of data subjects while preventing the development of additional research barriers. Such barriers could undermine the excellent contribution to the preservation of public health made by epidemics forecasting systems such as Influenzanet.

Ethical Framework for the Regulation of Participatory Disease Surveillance Systems

We propose the following 4 components ethical framework to provide guidance on how to ensure an adequate ethical oversight of participatory disease surveillance systems while safeguarding participants' privacy and eliminating barriers to the work of these surveillance platforms (Table 1).

Table 1. Ethical framework for the regulation of participatory disease surveillance systems.

Principle	Ethical component	Considerations
Autonomy of participants	Electronic consent	Standard, reader-friendly, and multilingual informed consent form with succinctly summarized information (eg, as a single PDF file) delivered at the time of registration and each time new information is added to the digital platforms Informed nature of consent can be fostered through the provision of a few quiz questions to reduce the risk of participants simply “clicking through” the consent process Require digital signature of the consent form to incentivize participants to read the information form and as evidence of their identity Making participants aware of the fact that despite best effort to protect their privacy, the residual risk of a privacy leak cannot be ruled out
Nonmaleficence	Protection of participants’ privacy	Anonymization of participants’ data should be combined with other data security measures such as a highly protected centralized database for storage of participants’ data Location data of participants should never be mapped to the individual level but rather to the postal code level to reduce the risk of reidentification in case of rare value entries Sensor data from mobile phones should only be transmitted in anonymized and highly aggregated form Ensure ethical and accountable sharing of anonymized datasets between research institutions to reduce reidentification risks for participants through database triangulation
Justice	Access to information on disease activity and prevention strategies	Free, open, and nondiscriminative participation should be offered to members of the general public Disease prevention strategies and results obtained through the participatory surveillance platforms should be disseminated on a regular basis to members of the public through various means
Beneficence and nonmaleficence	Research ethics committees (RECs)	Interdisciplinary capacity building of RECs is required to keep up with technological advances, thereby ensuring an adequate protection of data subjects RECs should play a proactive role in the design and implementation of public health research involving digital communities of volunteer citizens RECs should act as safety nets to prevent barriers to public health surveillance by identifying ethico-legal grey zones and anticipate potential conflicting situations resulting from the evolving legal landscape

Conclusions

In the developing field of participatory disease surveillance systems, the main ethical dilemma is how to ensure adequate protection of data subjects while at the same time obtaining the full benefits that public health surveillance directly involving digital communities of citizens could bring. In this complex situation, 1 of the key ethical safeguards proposed in our framework is a properly implemented e-consent. To pursue this objective, national platforms of the Influenzanet Consortium will put continuous effort in enhancing and adequately developing their e-consent procedures. Current e-consent procedures could be improved by providing standard, reader-friendly, multilingual information about the study, participants’ rights, the risks associated with their participation, and, in addition, a short series of quiz questions to verify proper understanding of the potential benefits and risks. Furthermore, requiring participants to digitally sign the Web-based consent form could both serve as a motivation for them to read properly the information provided and as a solution to allow personal identification. However, such additional burdens of participation need to remain minimal to ensure the sustainability of the platforms.

Acknowledgments

LDG, AM, TW, and BSE acknowledge the financial support provided by the Swiss National Science Foundation (SNF NRP-74 Smarter Health Care, grant number 407440_167356). The views expressed in this paper are those of the authors and not necessarily those of the funder. The authors would like to thank Ricardo Mexia for his contribution to the conceptualization of the manuscript.

LDG thanks Christopher Poppe and Maddalena Favaretto for their help. OWM was affiliated with ETH Zurich, the Swiss Federal Institute of Technology when the work for the publication was carried out and is currently affiliated with Novartis Pharma AG.

Authors' Contributions

All authors (except PL) contributed to the conceptualization of this paper. All authors contributed to the writing, editing, and critical evaluation of the manuscript. They approved the submission of the final version of the manuscript.

Conflicts of Interest

DP, CK, CK, CG, MH, OWM, PL, and AF are members of the Influenzanet Consortium. MH is cofounder of coneno, a software development company which is working together with the Influenzanet consortium since December 2018 as a technology partner to develop a new open source Influenzanet platform to be launched in the future. The development of the new platform is, at its current state mostly driven by volunteer work and partly funded by ISI foundation (part of Influenzanet Consortium).

References

1. Mostert M, Bredenoord AL, Biesaart MC, van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 2016 Jul;24(7):956-960 [FREE Full text] [doi: 10.1038/ejhg.2015.239] [Medline: 26554881]
2. Thiel MJ, Parkin J, Dratwa J, Halila R, Palazzani L, Górski A. Publications Office of the European Union. Luxembourg; 2016 Mar 18. The ethical implications of new health technologies and citizen participation URL: https://publications.europa.eu/en/publication-detail/-/publication/e86c21fa-ef2f-11e5-8529-01aa75ed71a1 [accessed 2017-10-01] [WebCite Cache ID 72Zh0cB7k]
3. International Telecommunication Union. 2018 Dec 07. ITU releases 2018 global and regional ICT estimates URL: https://www.itu.int/en/mediacentre/Pages/2018-PR40.aspx [accessed 2019-04-18] [WebCite Cache ID 77iofmwuq]
4. Aitken M, Clancy B, Nass D. IQVIA. 2017 Nov 07. The Growing Value of Digital Health URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-04-18] [WebCite Cache ID 77iphhud5]
5. Klingler C, Silva DS, Schuermann C, Reis AA, Saxena A, Strech D. Ethical issues in public health surveillance: a systematic qualitative review. BMC Public Health 2017 Apr 4;17:295 [FREE Full text] [doi: 10.1186/s12889-017-4200-4] [Medline: 28376752]
6. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, Keller M, et al. Global capacity for emerging infectious disease detection. Proc Natl Acad Sci U S A 2010 Dec 14;107(50):21701-21706 [FREE Full text] [doi: 10.1073/pnas.1006219107] [Medline: 21115835]
7. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 2013 Aug 1;369(5):401-404. [doi: 10.1056/NEJMp1307752] [Medline: 23822655]
8. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
9. Centers for Disease Control and Prevention. 2010 Jul 29. Distinguishing public health research and public health nonresearch URL: https://www.cdc.gov/od/science/integrity/docs/cdc-policy-distinguishing-public-health-research-nonresearch.pdf [accessed 2019-04-18] [WebCite Cache ID 77itmHtGr]
10. Smolinski MS, Crawley AW, Olsen JM, Jayaraman T, Libel M. Participatory disease surveillance: engaging communities directly in reporting, monitoring, and responding to health threats. JMIR Public Health Surveill 2017 Oct 11;3(4):e62 [FREE Full text] [doi: 10.2196/publichealth.7540] [Medline: 29021131]
11. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012 Jul 26;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]
12. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. J Infect Dis 2016 Dec 1;214(suppl_4):S375-S379 [FREE Full text] [doi: 10.1093/infdis/jiw400] [Medline: 28830113]
13. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
14. European Commission. 2014 Nov 27. EPIWORK: Ebola forecasting uses model developed by EU project URL: https://ec.europa.eu/programmes/horizon2020/en/news/epiwork-ebola-forecasting-uses-model-developed-eu-project [accessed 2018-06-01] [WebCite Cache ID 72Zjbbbft]
15. Hirsch M, Woolley-Meza O, Paolotti D, Flahault A, Lukowicz P. grippeNET App: Enhancing Participatory Influenza Monitoring Through Mobile Phone Sensors. : ACM; 2018 Oct 8 Presented at: The ACM International Joint Conference and International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers; October 08-12, 2018; Singapore p. 833-841. [doi: 10.1145/3267305.3274171]
16. Geneviève LD, Wangmo T, Dietrich D, Woolley-Meza O, Flahault A, Elger BS. Research ethics in the European Influenzanet Consortium: scoping review. JMIR Public Health Surveill 2018 Oct 10;4(4):e67. [Medline: 30305258]

17. Koppeschaar CE, Colizza V, Guerrisi C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: citizens among 10 countries collaborating to monitor influenza in Europe. JMIR Public Health Surveill 2017 Sep 19;3(3):e66 [FREE Full text] [doi: 10.2196/publichealth.7429] [Medline: 28928112]
18. Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clin Microbiol Infect 2014 Jan;20(1):17-21 [FREE Full text] [doi: 10.1111/1469-0691.12477] [Medline: 24350723]
19. Simonsen L, Gog JR, Olson D, Viboud C. Infectious disease surveillance in the big data era: towards faster and locally relevant systems. J Infect Dis 2016;214(suppl_4):S380-S385. [Medline: 28830112]
20. van Noort SP, Codeço CT, Koppeschaar CE, van Ranst M, Paolotti D, Gomes MG. Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. Epidemics 2015 Dec;13:28-36 [FREE Full text] [doi: 10.1016/j.epidem.2015.05.001] [Medline: 26616039]
21. Friesema IH, Koppeschaar CE, Donker GA, Dijkstra F, van Noort SP, Smallenburg R, et al. Internet-based monitoring of influenza-like illness in the general population: experience of five influenza seasons in The Netherlands. Vaccine 2009 Oct 23;27(45):6353-6357. [doi: 10.1016/j.vaccine.2009.05.042] [Medline: 19840672]
22. Debin M, Colizza V, Blanchon T, Hanslik T, Turbelin C, Falchi A. Effectiveness of 2012-2013 influenza vaccine against influenza-like illness in general population: estimation in a French web-based cohort. Hum Vaccin Immunother 2014 Mar 1;10(3):536-543 [FREE Full text] [Medline: 24343049]
23. Boiron K, Sarazin M, Debin M, Raude J, Rossignol L, Guerrisi C, et al. Opinion about seasonal influenza vaccination among the general population 3 years after the A(H1N1)pdm2009 influenza pandemic. Vaccine 2015 Nov 27;33(48):6849-6854. [Medline: 26322844]
24. Debin M, Turbelin C, Blanchon T, Bonmarin I, Falchi A, Hanslik T, et al. Evaluating the feasibility and participants' representativeness of an online nationwide surveillance system for influenza in France. PLoS One 2013 Sep 11;8(9):e73675 [FREE Full text] [doi: 10.1371/journal.pone.0073675] [Medline: 24040020]
25. Vayena E, Mastroianni A, Kahn J. Ethical issues in health research with novel online sources. Am J Public Health 2012 Dec;102(12):2225-2230. [doi: 10.2105/AJPH.2012.300813] [Medline: 23078484]
26. Myers J, Frieden T, Bherwani K, Henning K. Ethics in public health research: privacy and public health at risk: public health confidentiality in the digital age. Am J Public Health 2008 May;98(5):793-801. [doi: 10.2105/AJPH.2006.107706] [Medline: 18382010]
27. World Health Organization. WHO Guidelines on Ethical Issues in Public Health Surveillance. Geneva: World Health Organization; 2017.
28. Salerno J, Knoppers BM, Lee LM, Hlaing WM, Goodman KW. Ethics, big data and computing in epidemiology and public health. Ann Epidemiol 2017 May;27(5):297-301. [doi: 10.1016/j.annepidem.2017.05.002] [Medline: 28595734]
29. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb 9;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
30. Woolley JP. How data are transforming the landscape of biomedical ethics: the need for ELSI metadata on consent. In: Mittelstadt B, Floridi L, editors. The Ethics of Biomedical Big Data. Cham: Springer; 2016:171-197.
31. Quinn SC. Ethics in public health research: protecting human subjects: the role of community advisory boards. Am J Public Health 2004 Jun;94(6):918-922. [Medline: 15249289]
32. Elger BS, Caplan AL. Consent and anonymization in research involving biobanks: differing terms and norms present serious barriers to an international framework. EMBO Rep 2006 Jul;7(7):661-666 [FREE Full text] [doi: 10.1038/sj.embor.7400740] [Medline: 16819458]
33. Ploug T, Holm S. Meta consent–a flexible solution to the problem of secondary use of health data. Bioethics 2016;30(9):721-732 [FREE Full text] [doi: 10.1111/bioe.12286] [Medline: 27628305]
34. Otto JL, Holodniy M, DeFraites RF. Public health practice is not research. Am J Public Health 2014 Apr;104(4):596-602. [doi: 10.2105/AJPH.2013.301663] [Medline: 24524499]
35. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/AJPH.2015.302696] [Medline: 26270299]
36. Olsen J. Meta consent - a workable procedure in the area of Big Data? Br Med J 2015 Jun 1;350:h2146 [FREE Full text] [doi: 10.1136/bmj.h2146]
37. Elger B, Mauron A. A presumed-consent model for regulating informed consent of genetic research involving DNA banking. In: Knoppers BM, editor. Populations and Genetics: Legal Socio-Ethical Perspectives. Leiden: Martinus Nijhoff; 2003:269-295.
38. Gefenas E, Dranseika V, Cekanauskaite A, Serepkaite J. Research on human biological materials: what consent is needed, and when. In: Lenk C, Sándor J, Gordijn B, editors. Biobanks and Tissue Research: The Public, the Patient and the Regulation. Dordrecht: Springer; 2011:95-110.
39. Ioannidis JP. Informed consent, big data, and the oxymoron of research that is not research. Am J Bioeth 2013;13(4):40-42. [doi: 10.1080/15265161.2013.768864] [Medline: 23514395]

40. Colledge F, Elger BS. Impossible, impractical, and non-identifiable? New criteria regarding consent for human tissue research in the Declaration of Helsinki. Biopreserv Biobank 2013 Jun;11(3):149-152. [doi: 10.1089/bio.2013.0006] [Medline: 24850091]
41. Colledge F, Persson K, Elger B, Shaw D. Sample and data sharing barriers in biobanking: consent, committees, and compromises. Ann Diagn Pathol 2014 Apr;18(2):78-81. [doi: 10.1016/j.anndiagpath.2013.12.002] [Medline: 24485935]
42. Helgesson G. In defense of broad consent. Camb Q Healthc Ethics 2012 Jan;21(1):40-50. [doi: 10.1017/S096318011100048X] [Medline: 22152457]
43. Hansson MG, Dillner J, Bartram CR, Carlson JA, Helgesson G. Should donors be allowed to give broad consent to future biobank research? Lancet Oncol 2006 Mar;7(3):266-269. [doi: 10.1016/S1470-2045(06)70618-0] [Medline: 16510336]
44. Elger B. Withdrawal of consent and destruction of samples. In: Elger B, Biller-Andorno A, Mauron A, Capron AM, editors. Ethical Isues in Governing Biobanks: Global Perspectives. London: Ashgate; 2008:131-166.
45. Sheehan M. Can broad consent be informed consent? Public Health Ethics 2011 Nov;4(3):226-235. [doi: 10.1093/phe/phr020] [Medline: 22102849]
46. Elger B. Consent and use of samples. In: Elger B, Biller-Andorno N, Mauron A, Capron AM, editors. Ethical Issues in Governing Biobanks: Global Perspectives. London: Ashgate; 2008:57-88.
47. Elger B. Consent to research involving human biological samples obtained during medical care. In: Elger B, Biller-Andorno N, Mauron A, Capron AM, editors. Ethical Issues in Governing Biobanks: Global Perspectives. London: Ashgate; 2008:89-120.
48. Steinsbekk KS, Myskja BK, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur J Hum Genet 2013 Sep;21(9):897-902 [FREE Full text] [doi: 10.1038/ejhg.2012.282] [Medline: 23299918]
49. Steinsbekk KS, Solberg B. Biobanks - when is re-consent necessary? Public Health Ethics 2011 Nov 2;4(3):236-250. [doi: 10.1093/phe/phr031]
50. Coiera E, Clarke R. e-Consent: the design and implementation of consumer consent mechanisms in an electronic environment. J Am Med Inform Assoc 2004;11(2):129-140. [Medline: 14662803]
51. FluTracking. About Flutracking URL: https://www.flutracking.net/Join [accessed 2019-01-24] [WebCite Cache ID 75f0e69fe]
52. Grady C, Cummings SR, Rowbotham MC, McConnell MV, Ashley EA, Kang G. Informed consent. N Engl J Med 2017;376(9):856-867. [doi: 10.1056/NEJMra1603773] [Medline: 28249147]
53. Simon CM, Klein DW, Schartz HA. Traditional and electronic informed consent for biobanking: a survey of U.S. biobanks. Biopreserv Biobank 2014 Dec;12(6):423-429. [doi: 10.1089/bio.2014.0045] [Medline: 25496155]
54. Metcalf J, Crawford K. Where are human subjects in big data research? The emerging ethics divide. Big Data Soc 2016 Jun 17;3(1):1-14. [doi: 10.1177/2053951716650211]
55. Tene O, Polonetsky J. Big data for all: privacy and user control in the age of analytics. Nw J Tech Intell Prop 2013;11(5):240-273 [FREE Full text]
56. Voth D. Face recognition technology. IEEE Intell Syst 2003 May;18(3):4-7. [doi: 10.1109/MIS.2003.1200719] [Medline: 27295638]
57. Alterman A. "A piece of yourself": ethical issues in biometric identification. Ethics Inf Technol 2003;5(3):139-150 [FREE Full text] [doi: 10.1023/B:ETIN.0000006918.22060.1f]
58. EUR-Lex. 2016 Apr 27. Regulation (EU) 2016/679 of the European Parliament and of the Council URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC [accessed 2018-06-03] [WebCite Cache ID 72ZlDVdRz]
59. Ball MP, Bobe JR, Chou MF, Clegg T, Estep PW, Lunshof JE, et al. Harvard Personal Genome Project: lessons from participatory public research. Genome Med 2014 Feb 28;6(2):10 [FREE Full text] [doi: 10.1186/gm527] [Medline: 24713084]
60. Article 29 Data Protection Working Party. 2011 Apr 4. Advice paper on special categories of data ("sensitive data") URL: https://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf [accessed 2019-04-19] [WebCite Cache ID 77kFCvEZg]
61. Zwitter A. Big data ethics. Big Data Soc 2014 Nov 20;1(2):1-6. [doi: 10.1177/2053951714559253]
62. Ölvingson C, Hallberg J, Timpka T, Lindqvist K. Ethical issues in public health informatics: implications for system design when sharing geographic information. J Biomed Inform 2002 Jun;35(3):178-185 [FREE Full text] [Medline: 12669981]
63. Welke P, Andone I, Blaszkiewicz K, Markowetz A. Differentiating Smartphone Users by App Usage. 2016 Presented at: UBICOMP'16; September 12-16, 2016; Heidelberg, Germany p. 519-523. [doi: 10.1145/2971648.2971707]
64. Vicente CR, Freni D, Bettini C, Jensen CS. Location-related privacy in geo-social networks. IEEE Internet Comput 2011 May;15(3):20-27. [doi: 10.1109/MIC.2011.29]
65. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for responsible sharing of clinical trial data. N Engl J Med 2013 Oct 24;369(17):1651-1658. [Medline: 24144394]
66. Narayanan A, Shmatikov V. Robust De-anonymization of Large Sparse Datasets. 2008 Presented at: IEEE Symposium on Security and Privacy; May 18-22, 2008; Oakland, CA, USA p. 111-125. [doi: 10.1109/SP.2008.33]

67. van Dijk J, Hacker K. The digital divide as a complex and dynamic phenomenon. Informa Soc 2003;19(4):315-326. [doi: 10.1080/01972240309487]
68. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. New Media Soc 2014 Jun 12;18(2):313-331. [doi: 10.1177/1461444814538648]
69. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
70. Cantarelli P, Debin M, Turbelin C, Poletto C, Blanchon T, Falchi A, et al. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. BMC Public Health 2014 Sep 20;14:984 [FREE Full text] [doi: 10.1186/1471-2458-14-984] [Medline: 25240865]
71. Schaefer GO, Emanuel EJ, Wertheimer A. The obligation to participate in biomedical research. J Am Med Assoc 2009 Jul 1;302(1):67-72. [Medline: 19567441]
72. International Telecommunication Union. Geneva, Switzerland: ITU; 2017 Jul. ICT FACTS AND FIGURES 2017 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [accessed 2019-04-19] [WebCite Cache ID 77kIi654d]
73. Perrotta D, Bella A, Rizzo C, Paolotti D. Participatory online surveillance as a supplementary tool to sentinel doctors for influenza-like illness surveillance in Italy. PLoS One 2017 Jan;12(1):e0169801 [FREE Full text] [doi: 10.1371/journal.pone.0169801] [Medline: 28076411]
74. Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory syndromic surveillance of influenza in Europe. J Infect Dis 2016 Dec 1;214(suppl_4):S386-S392. [doi: 10.1093/infdis/jiw280] [Medline: 28830105]
75. Dawson A, Jennings B. The place of solidarity in public health ethics. Public Health Rev 2012 Jun 12;34(1):65-79. [doi: 10.1007/BF03391656]
76. Preite F, Salardi S, Gesuita R, Villani S, Trerotoli P, Guardabasso V, et al. The new European regulation on personal data protection: significant aspects for data processing for scientific research purposes. Epidemiol Biostat Public Health 2017;14(2):e12286-1-e1228613. [doi: 10.2427/12286]
77. Mittelstadt B. Designing the health-related internet of things: ethical principles and guidelines. Information 2017;8(3):77. [doi: 10.3390/info8030077]

Abbreviations

e-consent: electronic consent
EISN: European Influenza Surveillance Network
EU: European Union
GDPR: General Data Protection Regulation
GFT: Google Flu Trends
ILI: influenza-like illness
NIVEL: Netherlands Institute for Health Services Research
PGP: Personal Genome Project
REC: research ethics committee

Edited by G Eysenbach; submitted 20.09.18; peer-reviewed by A Crawley, S Barteit, J Duggan; comments to author 08.01.19; revised version received 08.03.19; accepted 29.03.19; published 23.05.19.

Please cite as:
Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelsø C, Guerrisi C, Hirsch M, Woolley-Meza O, Lukowicz P, Flahault A, Elger BS
Participatory Disease Surveillance Systems: Ethical Framework
J Med Internet Res 2019;21(5):e12273
URL: https://www.jmir.org/2019/5/e12273/
doi:10.2196/12273
PMID:31124466

Original Paper

Development and Evaluation of a Web-Based Resource for Suicidal Thoughts: NowMattersNow.org

Ursula Whiteside[1,2], MS, PhD; Julie Richards[3,4], MPH; David Huh[2,5], PhD; Rianna Hidalgo[1,6], BA; Rebecca Nordhauser[1], MS; Albert J Wong[1], BS; Xiaoshan Zhang[1], MS; David D Luxton[1,2], PhD; Michael Ellsworth[7], BA; DeQuincy Lezine[1,8], PhD

[1]NowMattersNow.org, Seattle, WA, United States
[2]School of Social Work, University of Washington, Seattle, WA, United States
[3]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States
[4]Department of Health Services, University of Washington, Seattle, WA, United States
[5]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States
[6]Berkeley School of Law, University of California, Berkeley, CA, United States
[7]Civilization, Seattle, WA, United States
[8]Prevention Communities, Fresno, CA, United States

Corresponding Author:
Ursula Whiteside, MS, PhD
NowMattersNow.org
2331 E Madison St
Seattle, WA, 98112
United States
Phone: 1 206 679 6349
Email: ursulawhiteside@gmail.com

Abstract

Background: Nearly half of people who die by suicide see a health care provider in the month before their death. With the release of new care guidelines, detection of suicidal patients will likely increase. Providers need access to suicide-specific resources that can be used as part of immediate, brief interventions with a suicidal patient. Web-based suicide prevention resources have the potential to address this need.

Objective: This study aimed to describe the development of the NowMattersNow.org website as a resource for individuals with suicidal thoughts and to evaluate the utility of the site via user experience surveys.

Methods: NowMattersNow.org is an online video-based free public resource that provides evidence-based teachings, examples, and resources for managing suicidal thoughts and intense emotions focused largely around skills from dialectical behavior therapy. Developed with assistance from mental health consumers, it is intended to address gaps in access to services for suicidal patients in health care systems. Visitors stay an average of a minute and a half on the website. From March 2015 to December 2017, a user experience survey measured self-reported changes on a 1 (not at all) to 5 (completely overwhelming) scale regarding intensity of suicidal thoughts and negative emotions while on the website. Longitudinal regression analyses using generalized estimating equations evaluated the magnitude and statistical significance of user-reported changes in suicidal ideation and negative emotion. In secondary analyses, user-reported changes specific to subgroups, including men aged 36 to 64 years, mental health care providers, and other health care providers were evaluated.

Results: During the period of analysis, there were 138,386 unique website visitors. We analyzed surveys (N=3670) collected during that time. Subsamples included men aged 36 to 64 years (n=512), mental health providers (n=460), and other health care providers (n=308). A total of 28% (1028/3670) of survey completers rated their suicidal thoughts as a 5 or “completely overwhelming” when they entered the website. We observed significant reductions in self-reported intensity of suicidal thoughts (–0.21, $P<.001$) and negative emotions (–0.32, $P<.001$), including decreases for users with the most severe suicidal thoughts (–6.4%, $P<.001$), most severe negative emotions (–10.9%, $P<.001$), and for middle-aged men (–0.13, $P<001$). Results remained significant after controlling for length of visit to website (before the survey) and technology type (mobile, desktop, and tablet).

Conclusions: Survey respondents reported measurable reductions in intensity of suicidal thoughts and emotions, including those rating their suicidal thoughts as completely or almost completely overwhelming and among middle-aged men. Although results from this user-experience survey administered at one point in time to a convenience sample of users must be interpreted with caution, results provide preliminary support for the potential effectiveness of the NowMattersNow.org website as a tool for short-term management of suicidal thoughts and negative emotions.

(J Med Internet Res 2019;21(5):e13183) doi:10.2196/13183

KEYWORDS

dialectical behavior therapy; suicide; internet; help-seeking behavior; behavior therapy; crisis intervention; primary care; integrated health care systems

Introduction

Background

Nearly half of people who die by suicide in the United States see some type of health care provider in the month before their death [1,2]. Newly released screening and care guidelines for suicidal patients [3] have the potential to increase the number of suicidal patients detected across health care settings. Unfortunately, most providers, particularly those in primary care settings where the majority of patients are seen before death by suicide, generally are not trained to work with suicidal patients [4,5]. There is also a lack of widely disseminated resources available for immediate support of suicidal patients, and the vast majority of people who die by suicide never receive specialized mental health care [6], with men being particularly unlikely to seek mental health care but far more likely to die by suicide than women [7,8].

Providers need access to suicide-specific resources that can be used as part of immediate, brief interventions with a suicidal patient. Web-based suicide-prevention resources are available around the clock and have the potential to address several major challenges to caring for suicidal patients [9,10]. Web-based services can assist providers as they are providing care by giving them a free resource to share, support suicidal patients during the time between identification and receipt of specialty mental health care, and serve as an adjunct intervention for suicidal patients who are not able to access specialty metal health care or prefer not to receive this type of care (eg, males). However, in reviewing the existing literature on Web-based prevention strategies for suicidal individuals, Jacob et al explicitly noted the lack of evidence-based resources and called for further development and evaluation of these tools [11]. A recent review of the small number of studies on the effectiveness of Web-based intervention websites (and mobile apps) reported some promise in terms of reductions in suicidal thoughts [12], although many of the interventions reviewed included barriers to access such as the need to set up an account or were difficult or impossible to access outside of the specific research trial.

Web-based resources that do not require the user to sign-in (ie, provide contact information or other details about their identity) can reduce important obstacles to use, including concerns about privacy. In the United States, examples of highly used, free Web-based resources include sites from the American Foundation for Suicide Prevention [13] and the National Suicide Prevention Lifeline [14]. Both of these sites provide important resources and self-help tips for suicide prevention but are not specifically tailored for a patient and clinician audience. The effectiveness of educational Web resources for suicide and self-directed violence prevention has historically been an area of research need [15]. Recent research indicates positive findings, for example, Till et al [16] reported results from a randomized trial that indicated that users with higher baseline scores of suicidality experienced an immediate reduction of suicidal thoughts compared with persons who visited a control website, thus indicating promise for the use of educative websites for brief intervention. Moreover, free Web-based resources may be particularly useful for middle-aged men in the United States who are unlikely to seek mental health care [17] and disproportionately impacted by suicide death [18].

Objectives

NowMattersNow.org is a website that was developed based on the evidence provided above, for the purpose of supporting suicidal patients and providers who care for them ([19], see Multimedia Appendix 1). Central to the educative website are first-person stories of successful coping with and recovery from suicidal thoughts and painful emotions using strategies from a highly effective treatment for suicidal individuals (dialectical behavior therapy, DBT) [20-22]. Indeed, research indicates that video portrayals of recovery from suicide-related crisis may be beneficial to those experiencing higher rates of suicidal thoughts [23]. The purpose of this report was to describe the development and evaluation of this Web-based resource. For our evaluation, we present data from a brief user-experience survey collected from NowMattersNow.org users. The survey measures changes in suicidal thoughts and negative emotions while on the website and also collects demographic data on a high-risk group (men aged 36-64 years) and health care provider status. We hypothesized that survey completers, including those who rated their suicidal thoughts and negative emotions at the most intense end of the scale, would report reduced intensity after viewing the website. Secondary stratified analyses evaluated the impact on a smaller and higher risk group (middle age men) and providers (mental health and other health care providers) who may be using the website to help support suicidal patients rather than self-help. In addition, we investigated length of time to take the survey as an indicator of website “dose” as the length of exposure to the website may be associated with ratings of intensity of suicidal thoughts and negative emotions, and the impact of viewing the website on a mobile phone versus other device to inform our recommendations for website use.

Methods

NowMattersNow.org Origins

DBT is a robust evidence-based treatment that includes skills and self-management strategies (DBT skills) to change and tolerate emotions that drive suicidal thoughts [20-22,24]. However, frontline clinical providers such as those working in primary care have few resources to explain or connect patients to specialized mental health treatment, including DBT. Mental health consumer advocates have created websites to make DBT resources more widely available, for example, dbtselfhelp.com [25]. When brief DBT-based interventions began to demonstrate promise for treatment of suicidal patients [26-28], the first and second authors (motivated as researchers but also by personal and professional experiences–related suicidal patients, their providers, and perceived lack of resources) initiated research to begin the development of a DBT-inspired Web-based resource [19,29]. The results were used to design the NowMattersNow.org website. This was done in collaboration with Team Now Matters Now, a group of mental health consumers with suicidal experiences recruited for the project and the website design firm Civilization [30]. Most of the website content consists of videos of Team Now Matters Now members discussing their lived experience including using DBT skills to manage suicidal thoughts, thus providing social models of success [31]. The NowMattersNow.org website was launched on World Suicide Prevention Day in 2014 (September 10) and had more than 250,000 unique visitors as of December 2018.

NowMattersNow.org Design

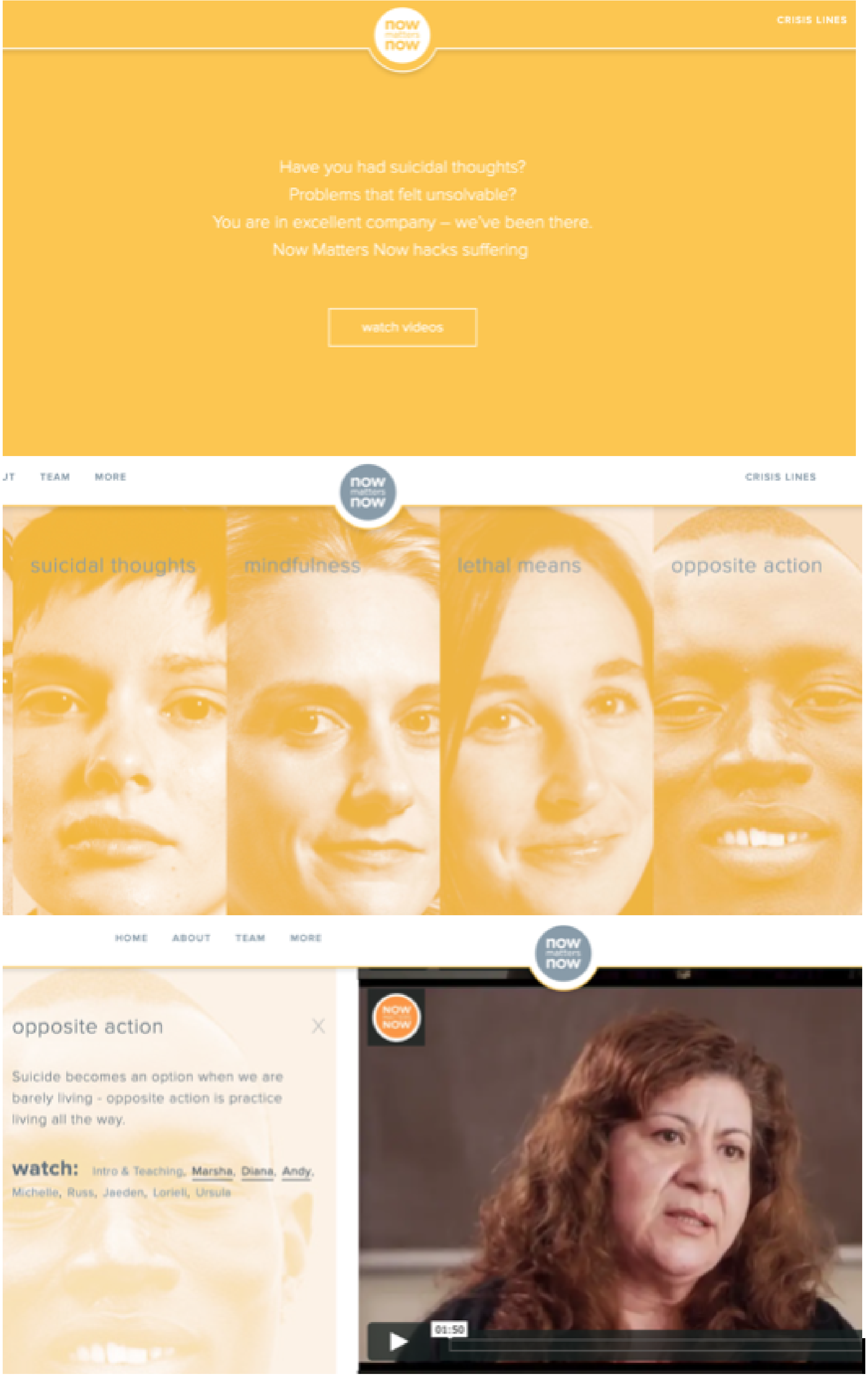
The website is designed to provide people experiencing suicidal thoughts and health care providers with DBT skills and to give health care providers a resource for supporting suicidal patients. The long-term goal is having providers use the skills in their own life and teach them to patients to help prevent and manage suicidal crises. As such, the website was developed with the provider audience in mind as a tool to help provide better care to their suicidal patients. The NowMattersNow.org landing page states “Have you had suicidal thoughts? Problems that felt unsolvable? You are in excellent company—we’ve been there.” This statement was intended to reduce stigma and signify that others having similar experiences were part of the website. Users proceed via a button to the home page. This page presents a panel of images of individuals with lived experience who have videos on the website. Images lead to topics such as DBT skills (opposite action, mindfulness, mindfulness of current emotion, paced breathing), suicidal thoughts (videos linking suicidal thoughts to difficulties managing emotions and DBT skills), and lethal means and caring messages (topics central to current care guidelines; Figure 1) [32]. Clicking on text in individual panels leads to videos on the topic with personal stories and didactic content. Home page menu options are About (text-based description of the website and a brief video provide an overview of the website), Team (bios and photos of Team Now Matters Now members allowing visitors more context about those featured in the videos), and More (eg, additional engagement options such as social media links, email sign-up, and downloadable “caring message” cards that providers can share with patients to direct them to the site). The home page also has a Crisis Lines option leading to national and international phone, short message service text message, and instant messaging options (also part of care guidelines). NowMattersNow.org is built on WordPress and instrumented with Google Analytics, which collects client-side statistics for how users find and interact with the site.

User Experience Survey

NowMattersNow.org regularly collects survey data during a user’s session for quality improvement purposes and to determine to what degree certain populations use the website and if it helps them. The 7-item survey was built in WordPress using HTML elements triggered by JavaScript, is not advertised, and open (no log-in required). The header is, “We’d appreciate your confidential feedback to improve our site.” The survey asks users to retrospectively rate their “Intensity of negative emotions” and “Intensity of suicidal thoughts” as 1 to 5 (“Not at all” to “Completely Overwhelming;” Figure 2) for “When you entered this site” (baseline) and “What level are they now?” (post website use). Only 1 number can be selected for each question. These survey items were adapted from DBT skills “diary cards” used by patients to track their suicidal thoughts, mood, and related behaviors [20-22]. The survey also asks about demographics as nonexclusive checkboxes (ie, male aged 36-64 years, mental health provider, and health care provider). Survey responses were not assessed for completeness nor required for survey submission (although only complete surveys were used in analyses), and once submitted (using a “submit” button), all responses were final. Data that were entered, but not submitted, were not recorded. The methods of survey development were similar to best practices recommended by the US Department of Health and Human Services for developing Web-based surveys such as simplicity, clarity, and user centeredness [33]. The survey asks about suicidal thoughts and negative emotions, which likely correspond [34].

The survey displays after 1 min on any page viewed on mobile phones (timing determined by the survey software). On desktops and tablets, it originally displayed after 3 min (the approximate length of 1 short video) on a page and changed to 8 min (essentially the length of 2 short videos or 1 longer video) on August 7, 2015 to allow for longer time on the website before surveying users. We know that participants visited a page for the minimum time to receive the survey, but they could have visited several pages before triggering the survey. Without using cookies, we could track only how long users were on a page, not the website, and we are reluctant to use tracking features because of the sensitive website content. Once the survey surfaces, users cannot proceed until they fill out the survey, close its window using a button in the right-hand corner, or choose another option on the navigation bar. Once displayed (whether completed or not), the survey does not surface for that user for at least 24 hours. We do not know the number of times the survey surfaced and was not completed. The survey was tested by Civilization website developers before release and before data collection for this study started. We had no contact with the responders other than their survey submission.

Figure 1. Website.



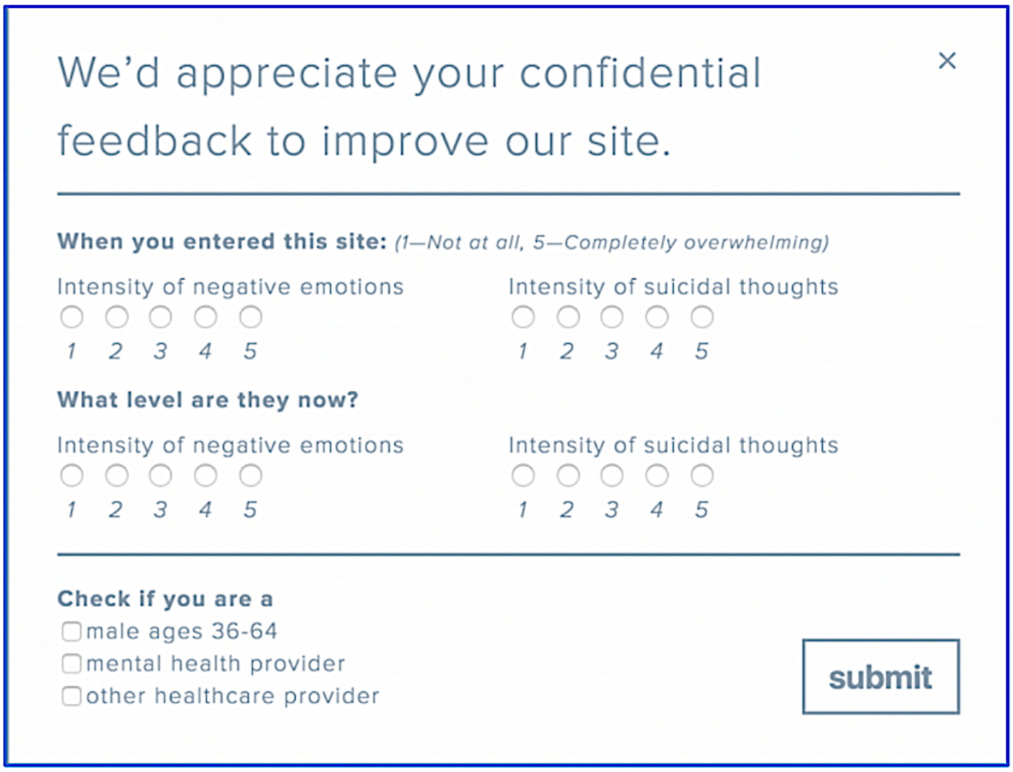
Have you had suicidal thoughts?
Problems that felt unsolvable?
You are in excellent company – we've been there.
Now Matters Now hacks suffering

opposite action

Suicide becomes an option when we are barely living - opposite action is practice living all the way.

watch: Intro & Teaching, Marsha, Diana, Andy, Michelle, Russ, Jaeden, Lorieli, Ursula

Figure 2. Survey.



We'd appreciate your confidential feedback to improve our site.

When you entered this site: *(1—Not at all, 5—Completely overwhelming)*

Intensity of negative emotions
○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Intensity of suicidal thoughts
○ 1 ○ 2 ○ 3 ○ 4 ○ 5

What level are they now?

Intensity of negative emotions
○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Intensity of suicidal thoughts
○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Check if you are a

☐ male ages 36-64
☐ mental health provider
☐ other healthcare provider

Study Sample and Study Data

This study analyzed a convenience sample of data from surveys completed from March 5, 2015, the first full day the survey was posted, to December 3, 2017, when data were retrieved. The personal information of Internet Protocol (IP) address snippets was removed from the dataset after assigning each participant a unique identification number. The final dataset was stored in Google Drive, to which authors UW, RN, AJW, XZ, DH had access. We did not track how long survey completers stayed on the website beyond the survey. The Allendale Institutional Review Board (Old Lyme, Connecticut) approved this study.

Statistical Analyses

Primary Outcome

Longitudinal regression analyses used generalized estimating equations (GEEs; [35]) to evaluate the magnitude and statistical significance of user-reported changes in (1) suicidal ideation and (2) negative emotion. An exchangeable working correlation matrix in conjunction with cluster robust standard errors was estimated to account for correlation of repeated measures (baseline and post) within participants. A total of 2 sets of analyses were performed evaluating outcomes for both suicidal thoughts and negative emotions as (1) a continuous measure and (2) a dichotomous indicator of *high* (4 or 5) versus *low* ratings (3 or less). Logistic and Gaussian GEEs were used for binary and continuous variables, respectively. In the primary analyses evaluating changes in outcomes across all participants, each outcome variable was regressed on time point (post vs baseline).

Men Aged 36 to 64 Years and Health Care Providers

To evaluate if changes in outcomes differed by these demographic groups, primary analyses were extended to include indicator variables for (1) middle-aged male (=1), (2) mental health provider (=1), (3) other health provider (=1), and (4) the interaction of each demographic variable with time (post vs baseline). In stratified analyses, each outcome variable was regressed on middle-aged male, mental health provider, other health provider, middle-aged male by time (post vs baseline), mental health provider by time, and other health provider by time. The statistical test of the subgroup effects was the magnitude and statistical significance of each demographic variable by time interaction.

Mobile Versus Other Device and Delay to Survey

To evaluate if outcomes differed by electronic device type or delay before being presented with the survey, primary analyses were extended to include indicator variables for (1) device type (nonmobile vs mobile) and (2) survey delay interval (1, 3, 8 min). For device-type analysis, outcome variables were regressed on device type, time (post vs baseline), and the device type by time interaction. In delay-interval analysis, outcome variables were regressed on delay interval, time, and the delay interval by time interaction. Delay interval was divided into 2 planned contrasts of (1) 8 versus 1 min and (2) 3 versus 1 min. Statistical tests of device-type effects and delay-related effects were magnitude and statistical significance of (1) device type by time and (2) delay interval by time interactions.

Results

Website Visitors

During the time in which surveys were analyzed, 167,878 sessions (website visits) and 138,386 unique visitors were recorded. Of those sessions, 52% came to the website via "organic search terms," 21% came directly to the website (bookmark or typing in the URL), 9% came via a link from social media, 9% came via paid advertisements for the site (Google AdWords), and 9% came via referrals (link from another website). On average, users viewed 2.1 pages per session and stayed on the website for 1 min 31 seconds. The average bounce rate (leaving after viewing 1 page) was 61%, and 18% of sessions were from returning visitors. People arrived via organic search terms when they searched for something that led to NowMattersNow.org being listed, and then clicked on our link, and the link was not a paid advertisement. To better understand how people found the website, Google search term data collection was enabled from December 11, 2017 to February 3, 2018 (54 days), which provided results that indicated that a number of organic search terms related to suicide and suicide help (eg, "crisis chat," "crisis text line uk," "suicide chat," and "now matters now," each individually making up no more that 4% of searches) represented the search terms leading to NowMattersNow.org. The Google Adwords were phrases we thought people who were suicidal might Google ("I want to die" and "how do I kill myself?"). If someone Googled such a term, they might see the website as an advertisement option. Geographically, visitors were largely from the United States (61%), Canada (15%), United Kingdom (9%), and Russia (3%), with Australia, Brazil, New Zealand, India, France, and Germany each making up 2% or less.

Survey Responders

We received 3931 complete responses. Of these, 261 were removed because they were likely from people who previously completed the survey, identified from duplicate IP address snippets (keeping only earliest dated record). Surveys with incomplete data were not included in the analyses as the amount of missing data was minimal (6.7%) [36]. Of the 3670 unique survey responses, 514 (14% of survey sample) identified as men aged 36 to 64 years. Of survey completers, 460 (13%) identified as mental health professionals and 308 (8%) as other health care professionals, with 40 (1%) identifying as both. Users completed surveys after 1-min (2096; 57%), 3-min (388; 11%), or 8-min (1186; 32%) delays. Mobile users (57%) all received the survey at 1 min, therefore nonmobile device users comprised the remaining 43% (1574) of the sample (Table 1).

Survey ratings for suicidal thoughts pooled at the low and high ends of the scale, clustering toward low (1, not at all) and high (4 and 5, completely overwhelming). For baseline, the greatest number of responses were a score of 5 (28%), followed by 1 (28%), 4 (21%), 3 (17%), and 2 (7%). Post scores were 1 (32%), 2 (29%), 5 (23%), 4 (19%), and 3 (17%). In our sample, of those who reported suicidal thoughts (2644) at baseline, 29% (n=763) reported a 1 or more point reduction in suicidal thoughts pre versus post visit, the vast majority stayed the same (63%, n=1666), and 8% (n=215) rated suicidal thoughts 1 or more points worse (8%). Of those who reported worse scores, 2% (n=51) were more than a point worse, 1% (n=21) were 3 or 4 points worse.

Primary Outcome Analyses

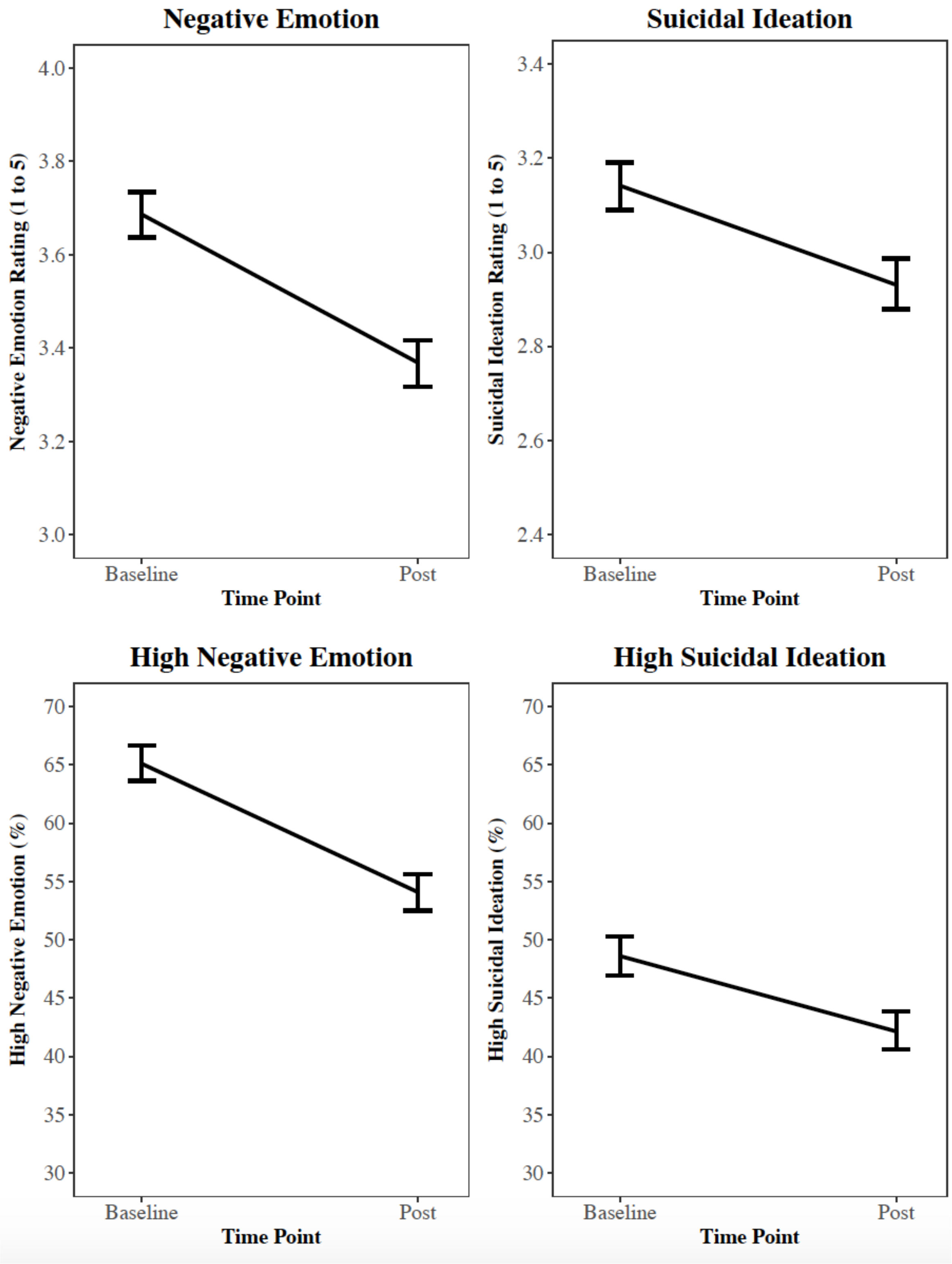
The overall changes reported for all participants for the 4 main study outcomes (continuous and dichotomous for baseline vs post) is shown in Figure 3. Survey participants reported an average 0.21-point reduction (95% CI –0.24 to –0.19; *P*<.001; *d*=0.13) in suicidal ideation and a 0.32-point reduction (95% CI –0.35 to –0.29; *P*<.001; *d*=0.21) in negative emotion ratings (both 1 to 5 scale) after viewing the website. The average visitor's probability of endorsing high suicidal ideation (4 or 5) decreased 6.4 percentage points (95% CI –7.55 to –5.33; *P*<.001; *d*=0.13) after viewing the website. The average visitor's probability of endorsing high negative emotion (4 or 5 at baseline) decreased by 10.9 percentage points (95% CI –12.20 to –9.64; *P*<.001; *d*=0.23) after viewing the website.

Table 1. Demographic groups, electronic device type, and delay to survey (N=3670).

Outcome variable	Total	Men aged 36 to 64 years (n=512)	Mental health provider (n=460)	Other health care provider (n=308)	Mobile (all 1 min to survey) (n=2096)	8 min to survey (n=1186)	3 min to survey (n=388)
Baseline suicidal thoughts							
Intensity, mean (SD)	3.1 (1.6)[a]	3.3 (1.6)[a]	1.8 (1.5)	2.8 (1.7)	3.6 (1.4)[a]	2.7 (1.6)[a]	2.2 (1.5)[a]
Percentage low (score=1), n (%)	1012 (28)	131 (26)	331 (72)	130 (42)	315 (15)	486 (41)	215 (55)
Percentage high (score=4-5), n (%)	1784 (49)[a]	277 (54)[a]	86 (19)	122 (40)	1235 (59)[a]	448 (38)[a]	101 (26)[a]
Baseline negative thoughts							
Intensity, mean (SD)	3.7 (1.5)[a]	3.7 (1.5)[a]	2.2 (1.6)[a]	3.1 (1.7)[a]	4.1 (1.3)[a]	3.3 (1.6)[a]	2.8 (1.6)[a]

[a] Indicates significant reduction from baseline to post *P*<.001, not applicable from baseline suicide ideation.

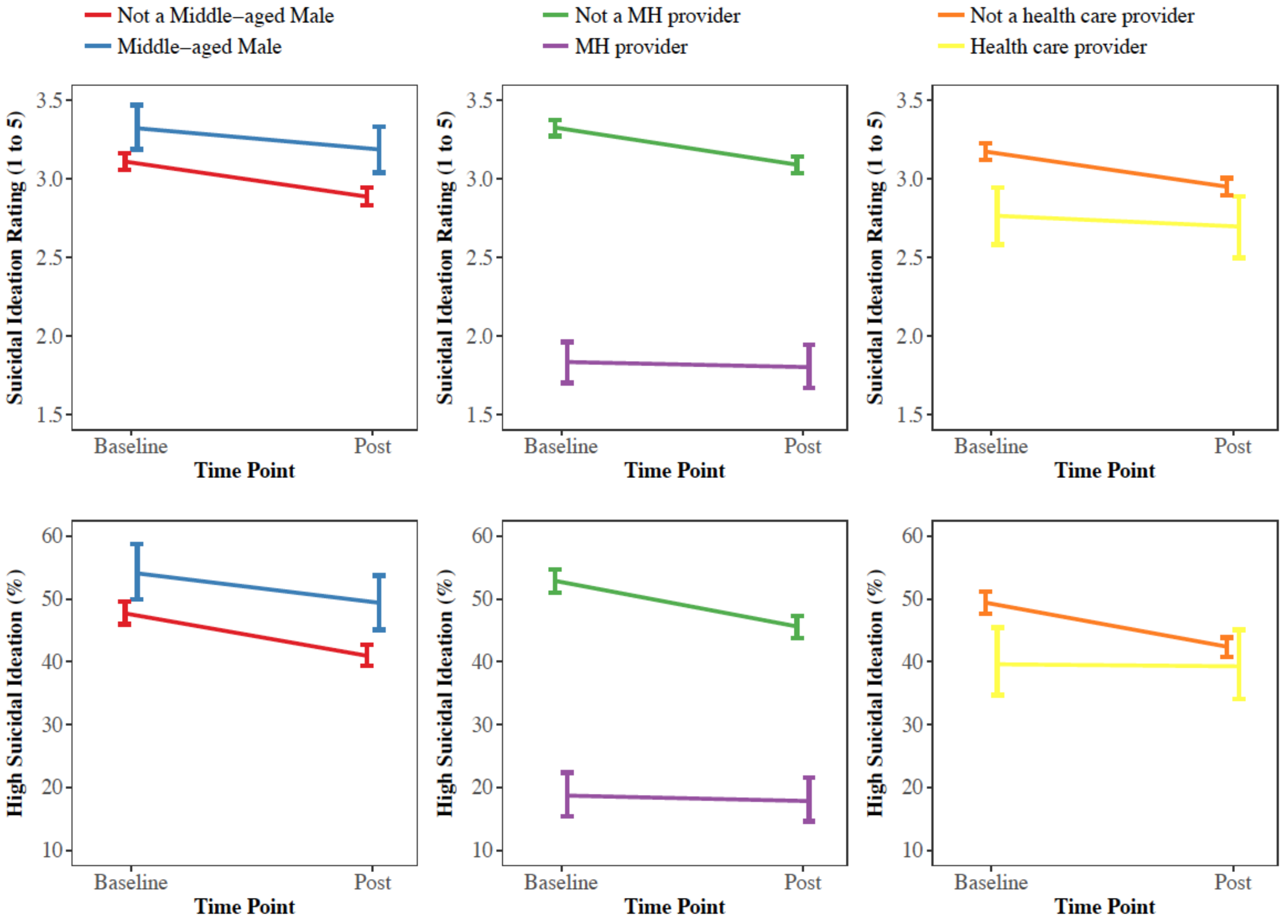
Figure 3. Change in negative emotion and suicidal ideation (continuous and dichotomous) from baseline to post website for all participants. Error bars indicate SDs.



Men Aged 36 to 64 Years

Men aged 36 to 64 years reported their overall severity of suicidal ideation decreased on the 1 to 5 scale (Estimate, est=–0.13; 95% CI –0.19 to –0.07; *P*<.001; *d*=–0.08), although the decrease was 0.09 points greater (*d*=0.06; *P*<.001) among nonmiddle-aged males (est=-0.22; 95% CI –0.25 to -0.20; *P*<.001; *d*=0.14). High suicidal ideation decreased for middle-aged males (%=–4.74; 95% CI –7.38 to –2.10; *P*<.001; *d*=0.09). Results were similar and significant for overall negative emotions (est=–0.21; 95% CI –0.29 to –0.14; *P*<.001; *d*=0.14) and for the probability of high negative emotion (%=–6.90; 95% CI –10.70 to –3.11; *P*<.001; *d*=0.14; Figure 4).

Figure 4. Average suicidal ideation and high suicidal ideation ratings from baseline to post website by demographic group. MH: mental health.



Health Care Providers

For mental health providers, there were no changes observed for overall severity of suicidal ideation from baseline to post (est=–0.03; 95% CI –0.07 to 0.01; *P*=.10; *d*=0.02) or for the probability of endorsing high suicidality (%=–1.36; 95% CI –2.86 to 0.14; *P*=.08; *d*=0.03). This finding was consistent with other health care providers for suicide ideation overall (est=–0.07; 95% CI –0.15 to 0.02; *P*=.12; *d*=0.04) and high suicidal ideation (%=–0.54; 95% CI –3.95 to 2.87; *P*=.76; *d*=0.01). However, significant improvements were seen in negative emotions overall for mental health providers (est=–0.18; 95% CI –0.24 to –0.12; *P*<.001; *d*=0.12) and high negative emotions (%=–4.17; 95% CI –6.39 to –1.95; *P*<.001; *d*=0.09) as well as other health care providers (Overall: est=–0.18; 95% CI –0.29 to –0.07; *P*=.001; *d*=0.12 and high rating: %=–5.59; 95% CI –10.04 to –1.14; *P*=.01; *d*=0.12; Figure 4).

Mobile Versus Other Device and Delay to Survey

For website platform (mobile vs other), suicidal ideation severity decreased for both desktop or tablet users (est=–0.21; 95% CI –0.25 to –0.17; *P*<.001; *d*=0.13) and mobile users (est=–0.21; 95% CI –0.25 to –0.18; *P*<.001; *d*=0.13). High suicidal ideation decreased for both desktop or tablet users (%=–6.99; 95% CI –8.59 to –5.38; *P*<.001; *d*=0.14) and mobile users (%=–6.06; 95% CI –7.59 to –4.52; *P*<.001; *d*=0.12). Similarly, negative emotion severity decreased for both desktop or tablet users (est=–0.34; 95% CI –0.38 to –0.29; *P*<.001; *d*=0.23) and mobile users (est=–0.30; 95% CI –0.34 to –0.26; *P*<.001; *d*=0.20). In addition, high negative emotions decreased for both desktop or tablet users (%=–10.67; 95% CI –12.55 to –8.79; *P*<.001; *d*=0.22) and mobile users (percentage=–11.26; 95% CI –13.07 to –9.45; *P*<.001; *d*=0.24).

For survey delay interval (1, 3, 8 min), suicidal ideation severity decreased for those surveyed after delays of 1 min (est=–0.21; 95% CI –0.25 to –0.18; *P*<.001; *d*=0.13), 3 min (est=–0.18; 95% CI –0.24 to –0.12; *P*<.001; *d*=0.11), and 8 min (est=–0.22; 95% CI –0.26 to –0.18; *P*<.001; *d*=0.14). High suicidal ideation also decreased for those surveyed after delays of 1 min (%=–6.06; 95% CI –7.59 to –4.52; *P*<.001; *d*=0.12), 3 min (%=–5.67; 95% CI –8.38 to –2.96; *P*<.001; *d*=0.11), and 8 min (%=–7.42; 95% CI –9.36 to –5.48; *P*<.001; *d*=0.15). Similarly, negative emotion severity decreased for surveyed after delays of 1 min (est=–0.30; 95% CI –0.34 to –0.26; *P*<.001; *d*=0.20), 3 min (est=–0.29; 95% CI –0.37 to –0.21; *P*<.001; *d*=0.19), and 8 min (est=–0.35; 95% CI –0.40 to –0.30; *P*<.001; *d*=0.24). High negative emotions also decreased for those surveyed after delays of 1 min (%=–11.26; 95% CI –13.07 to –9.45; *P*<.001; *d*=0.24), 3 min (%=–10.05; 95% CI –13.37 to –6.74; *P*<.001; *d*=0.21), and 8 min (%=–10.88; 95% CI –13.12 to –8.63; *P*<.001; *d*=0.23).

Sensitivity Analyses Correcting for Multiple Statistical Tests

All findings ($P<.05$) remained statistically significant after accounting for multiple testing using the Benjamini-Hochberg procedure [37].

Discussion

Principal Findings

The results of our evaluation of NowMattersNow.org suggest that this Web-based resource has the potential to address several major challenges to caring for suicidal patients. First, NowMattersNow.org was developed to address a historic need for suicidal patients [15] by providing tailored content to support suicidal individuals and the providers who care for them [19]. During our 32-month evaluation period, there were 138,386 unique visitors to NowMattersNow.org, indicating this site was indeed receiving a fair amount of visitation. Next, over 70% of survey completers recalled having some degree of suicidal thoughts when they arrived at the website. Survey participants, overall, also reported significant reductions in suicidal thoughts. Highly suicidal respondents reported significant decreases in intensity of suicidal thoughts from website entrance to survey. Reductions in suicidal thoughts were matched by similar declines in reported intensity of negative emotions while visiting the website. This result indicated that the site attracted people who are experiencing suicidal thoughts, often at a high level of intensity. The results also suggest that website visits were associated with reductions in suicidal thoughts and negative emotions.

Finally, we examined different types of visitors and found that significant reductions in suicidal thoughts and negative emotions were largely consistent across subgroups. Middle-aged men reported significant reductions in suicidal thoughts and negative emotions. Given that this group makes up 38% of all suicides [35], these results are promising. However, reductions for middle-aged men were not as large as for the rest of the sample; therefore, the website could use greater tailoring toward men. For example, male identification on the survey could trigger an additional box requesting suggestions for improvements or their contact information for follow-up interview or survey about the same topic. About one-fifth of responders identified as mental health or other health care providers, which was overrepresented compared with the general population. These responders reported lower levels of suicidal thoughts and negative emotions. Health care providers likely have visited the website for reasons other than self-help, such as for providing a referral for a patient, as the website is part of clinician training efforts. These findings suggest that the website content is most useful for those for whom it is directly intended: those experiencing suicide ideation. Given the intentional simplicity of the survey, we could not glean information about the motivations for initial visits to the website (eg, experiencing severe negative emotions vs visiting for work). In the future, the survey could be developed to have an opt-out option for health care providers or those visiting the website for reasons not related to their own mental health, but this may also create unintentional stigma toward having suicidal thoughts. When we looked separately at user platform (mobile, desktop, and tablet), positive reductions were found in each group. We found this result to be promising because it may mean the website was showing benefit for youth, given that mobile users tend to be younger [38]. Indeed, the demographic category “age 12-18” was added to the user-experience survey in 2018, and results of these are forthcoming. Interestingly, mobile device users had higher severity ratings across both time points. Therefore, we may seek ways to target mobile audiences who, based on self-report, may be at greater risk for suicide. When we looked at time on website pages before the survey surfaced (1, 3, or 8 min), positive reductions were also consistent across times. This result meant that some users may find benefit in as little as 1 min on the website. We cannot separate the 1-min survey and mobile findings, but both were positive for reducing suicidal ideation.

Limitations

There are important limitations to this study. The primary limitation is that user-experience data were collected at one point in time from individual users, without a control-group comparison. Distraction alone may have had an effect, or the passage of time may be associated with similar reductions. In addition, we do not know if experienced improvements lasted beyond the brief period that the survey covered. In addition, nonresponse bias may account for the findings—these findings may not generalize to people who saw the survey but did not complete it or people who never saw the survey who may not have had positive responses. It is possible that the nature of the survey topic (suicide) may have made some individuals less likely (eg, among those with greater self-stigma or experienced discrimination) or more likely (eg, experiencing the survey as highly relevant) to participate, including those potentially at higher risk of suicide. Another limitation is that participants were asked to rate their pre and post website suicidal thoughts and negative emotions on a single survey. Although this format was designed to maximize survey response, reported reductions in distress may be at least partially attributable to recall bias. Furthermore, the survey has face validity, but other types of validity have not been examined, and we do not know if survey completers represent typical suicidal patients seen in health care settings, our target audience. Finally, although changes in suicidal thoughts and emotions might occur through a number of mechanisms (eg, didactic information, exposure to evidence-based interventions, distraction from emotional suffering, referral to crisis services, return to baseline, placebo effect, and viewing videos of those with lived experiencing discussing stories of recovery), we could not evaluate these mediators given survey limits. Future randomized trials need to examine if this website (or which elements of this website), compared with other sites, is associated with similar or greater reductions in suicidal thought reductions. An in-progress randomized controlled trial comparing risk of suicide attempt (over 18 months) among patients reporting suicidal ideation will shed additional light on the utility of its content. Patients invited to a skills training intervention, utilizing a guided interactive version the NowMattersNow.org website, will be compared with patients randomized to usual care [39].

Conclusions

NowMattersNow.org is a Web-based resource with the potential to help address gaps in suicide prevention and treatment resources for mental health professionals. Our findings supported use of the website as originally intended—as a support tool for providers who identify suicidal patients in their practice. In light of our results indicating website use is associated with reductions in suicidal thoughts, we suggest that health care providers can, with some confidence, refer suicidal people for short-term management of suicidal thoughts and negative emotions. For example, when suicidal patients are discovered as part of routine screening for depression, providers could share this website with the patient as part of safety planning procedures, which includes a list of options for managing suicidal thoughts [40]. Indeed, a number of health care providers are directed to the website through the Zero Suicide initiative, which has trained over 500 health systems to provide better suicide care. The resource is shared in nearly all of these trainings [41].

Acknowledgments

The authors would like to gratefully acknowledge the great number of volunteer hours that resulted in this work. Funding for the development of the website came from an R-34 grant from the National Institutes of Mental Health (MH097836) and a Postdoctoral Fellowship (011811) from the American Foundation for Suicide Prevention awarded to Dr UW. Funding for Ms JR’s time was provided by a development fund award from Kaiser Permanente Washington Health Research Institute. Funding for Ms RH and Dr DH was provided by small private donations to NowMattersNow.org.

Conflicts of Interest

NowMattersNow.org is a free public resource. It also is a 501c3 nonprofit in the United States. Authors UW, RN, DL, AJW, XZ, and DQL volunteer for this organization and none have ever been paid. RH and DH are contract employees who worked on this paper. In 2014, ME’s company (Civilization) helped develop the website and branding. There is no conflict of interest, but there are co-interests of better access to support for people at risk for suicide.

Multimedia Appendix 1

NowMattersNow.org introduction video.

[MP4 File (MP4 Video), 31MB - jmir_v21i5e13183_app1.mp4]

References

1. Ahmedani BK, Simon GE, Stewart C, Beck A, Waitzfelder BE, Rossom R, et al. Health care contacts in the year before suicide death. J Gen Intern Med 2014 Jun;29(6):870-877 [FREE Full text] [doi: 10.1007/s11606-014-2767-3] [Medline: 24567199]
2. Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. Am J Psychiatry 2002 Jun;159(6):909-916 [FREE Full text] [doi: 10.1176/appi.ajp.159.6.909] [Medline: 12042175]
3. National Action Aliance for Suicide Prevention. 2018. Transforming Health Systems Initiative Work Group URL: https://theactionalliance.org/sites/default/files/action_alliance_recommended_standard_care_final.pdf [accessed 2019-04-21] [WebCite Cache ID 77oPA1AbW]
4. Jahn D, Quinnett P, Ries R. The influence of training and experience on mental health practitioners’ comfort working with suicidal individuals. Prof Psychol Res Pr 2016;47(2):130-138. [doi: 10.1037/pro0000070]
5. Schmitz WM, Allen MH, Feldman BN, Gutin NJ, Jahn DR, Kleespies PM, et al. Preventing suicide through improved training in suicide risk assessment and care: an American Association of Suicidology Task Force report addressing serious gaps in U.S. mental health training. Suicide Life Threat Behav 2012 Jun;42(3):292-304. [doi: 10.1111/j.1943-278X.2012.00090.x] [Medline: 22494118]
6. Simon GE, Rutter CM, Peterson D, Oliver M, Whiteside U, Operskalski B, et al. Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? Psychiatr Serv 2013 Dec 1;64(12):1195-1202 [FREE Full text] [doi: 10.1176/appi.ps.201200587] [Medline: 24036589]
7. Parent M, Hammer J, Bradstreet T, Schwartz E, Jobe T. Men's mental health help-seeking behaviors: an intersectional analysis. Am J Mens Health 2018 Jan;12(1):64-73 [FREE Full text] [doi: 10.1177/1557988315625776] [Medline: 29226771]
8. Centers for Disease Control and Prevention, National Center for Health Statistics. Suicide and Self-Inflicted Injury. 2019. URL: https://www.cdc.gov/nchs/fastats/suicide.htm? [accessed 2019-03-13] [WebCite Cache ID 76pj3vQBG]
9. Till B, Tran US, Voracek M, Niederkrotenthaler T. Beneficial and harmful effects of educative suicide prevention websites: randomised controlled trial exploring Papageno v Werther effects. Br J Psychiatry 2017 Aug;211(2):109-115. [doi: 10.1192/bjp.bp.115.177394] [Medline: 28522433]
10. van Spijker BA, Werner-Seidler A, Batterham PJ, Mackinnon A, Calear AL, Gosling JA, et al. Effectiveness of a web-based self-help program for suicidal thinking in an Australian community sample: randomized controlled trial. J Med Internet Res 2018 Feb 14;20(2):e15 [FREE Full text] [doi: 10.2196/jmir.8595] [Medline: 29444769]

11. Jacob N, Scourfield J, Evans R. Suicide prevention via the internet: a descriptive review. Crisis 2014 Jan 1;35(4):261-267. [doi: 10.1027/0227-5910/a000254] [Medline: 24984890]
12. Witt K, Spittal MJ, Carter G, Pirkis J, Hetrick S, Currier D, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. BMC Psychiatry 2017 Dec 15;17(1):297 [FREE Full text] [doi: 10.1186/s12888-017-1458-0] [Medline: 28810841]
13. American Foundation for Suicide Prevention. URL: http://afsp.org [accessed 2011-05-10] [WebCite Cache ID 5yaNQ37Zw]
14. National Suicide Prevention Lifeline. URL: https://suicidepreventionlifeline.org/ [accessed 2018-12-19] [WebCite Cache ID 74lmbMQKK]
15. Luxton DD, June JD, Fairall JM. Social media and suicide: a public health perspective. Am J Public Health 2012 May;102(Suppl 2):S195-S200. [doi: 10.2105/AJPH.2011.300608] [Medline: 22401525]
16. Till B, Tran US, Voracek M, Niederkrotenthaler T. Beneficial and harmful effects of educative suicide prevention websites: randomised controlled trial exploring Papageno v. Werther effects. Br J Psychiatry 2017 Aug;211(2):109-115. [doi: 10.1192/bjp.bp.115.177394] [Medline: 28522433]
17. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Leading Causes of Death Reports, 1981-2017. 2016. URL: https://webappa.cdc.gov/sasweb/ncipc/leadcause.html [accessed 2019-04-21] [WebCite Cache ID 77oPKVVp5]
18. House J, Marasli P, Lister M, Brown JS. Male views on help-seeking for depression: a Q methodology study. Psychol Psychother 2018 Dec;91(1):117-140. [doi: 10.1111/papt.12144] [Medline: 29087607]
19. Whiteside U, Richards J, Ellsworth M. NowMattersNow. 2014. URL: https://www.nowmattersnow.org/ [accessed 2018-12-18] [WebCite Cache ID 74lJybEMh]
20. Linehan MM. DBT Skills Training Handouts And Worksheets. New York: The Guilford Press; 2014.
21. Linehan MM. Cognitive-behavioral Treatment Of Borderline Personality Disorder. New York: The Guilford Press; 1993.
22. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry 2006 Jul;63(7):757-766. [doi: 10.1001/archpsyc.63.7.757] [Medline: 16818865]
23. Till B, Strauss M, Sonneck G, Niederkrotenthaler T. Determining the effects of films with suicidal content: a laboratory experiment. Br J Psychiatry 2015 Jul;207(1):72-78. [doi: 10.1192/bjp.bp.114.152827] [Medline: 25906792]
24. Linehan MM, Korslund KE, Harned MS, Gallop RJ, Lungu A, Neacsiu AD, et al. Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: a randomized clinical trial and component analysis. JAMA Psychiatry 2015 May;72(5):475-482. [doi: 10.1001/jamapsychiatry.2014.3039] [Medline: 25806661]
25. Dietz L. DBT Self Help. 2003. URL: http://www.dbtselfhelp.com/ [accessed 2018-12-18] [WebCite Cache ID 74lKEunmX]
26. Ward Ciesielski EF. An open pilot feasibility study of a brief dialectical behavior therapy skills-based intervention for suicidal individuals. Suicide Life Threat Behav 2013 Jun;43(3):324-335. [doi: 10.1111/sltb.12019] [Medline: 23409778]
27. Whiteside U. A Brief Personalized Feedback Intervention Integrating a Motivational Interviewing Therapeutic Style and Dialectical Behavioral Therapy Skills for Depressed Or Anxious Heavy Drinking Young Adults. Seattle: University of Washington; 2009.
28. McMain SF, Guimond T, Barnhart R, Habinski L, Streiner DL. A randomized trial of brief dialectical behaviour therapy skills training in suicidal patients suffering from borderline disorder. Acta Psychiatr Scand 2017 Feb;135(2):138-148. [doi: 10.1111/acps.12664] [Medline: 27858962]
29. Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, et al. Designing messaging to engage patients in an online suicide prevention intervention: survey results from patients with current suicidal ideation. J Med Internet Res 2014 Feb 7;16(2):e42 [FREE Full text] [doi: 10.2196/jmir.3173] [Medline: 24509475]
30. Ellsworth M. Civilization Design Firm. URL: https://builtbycivilization.com/ [accessed 2018-12-18] [WebCite Cache ID 74lKxMmTw]
31. Bandura A. Principles of Behavior Modification: Social Learning Theory. New York: Holt, Rinehart and Winston, Inc; 1969.
32. National Action Alliance for Suicide Prevention. Washington, DC: Education Development Center Recommended standard care for people with suicide risk: making health care suicide safe URL: https://theactionalliance.org/sites/default/files/action_alliance_recommended_standard_care_final.pdf [accessed 2019-04-21] [WebCite Cache ID 77oPpiQW0]
33. US Department of Health & Human Services. Online Surveys URL: https://www.usability.gov/how-to-and-tools/methods/online-surveys.html [accessed 2019-02-06] [WebCite Cache ID 761Yt7rua]
34. Hinkin T. A brief tutorial on the development of measures for use in survey questionnaires. Organ Res Methods 2016 Jun 29;1(1):104-121. [doi: 10.1177/109442819800100106]
35. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
36. Schafer J. Multiple imputation: a primer. Stat Methods Med Res 1999 Mar;8(1):3-15. [doi: 10.1177/096228029900800102] [Medline: 10347857]
37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B 2018 Dec 5;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]

38. Anderson M, Jiang J. Public Services Alliance. 2018. Teens, social media & technology URL: http://publicservicesalliance.org/wp-content/uploads/2018/06/Teens-Social-Media-Technology-2018-PEW.pdf [accessed 2019-04-19] [WebCite Cache ID 77leTDki4]
39. Simon GE, Beck A, Rossom R, Richards J, Kirlin B, King D, et al. Population-based outreach versus care as usual to prevent suicide attempt: study protocol for a randomized controlled trial. Trials 2016 Dec 15;17(1):452 [FREE Full text] [doi: 10.1186/s13063-016-1566-z] [Medline: 27634417]
40. Stanley B, Brown G. Safety planning intervention: a brief intervention to mitigate suicide risk. Cogn Behav Pract 2012;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
41. Goldstein-Grumet J. Suicide Prevention Resource Center. Zero Suicide URL: http://zerosuicide.sprc.org/ [accessed 2019-04-08] [WebCite Cache ID 77WMT6Qqp]

Abbreviations

DBT: dialectical behavior therapy
GEE: generalized estimating equation
IP: Internet Protocol

Edited by G Eysenbach; submitted 19.12.18; peer-reviewed by SH Kwon, E Bradshaw, C Wilks, E Ward-Ciesielski; comments to author 17.01.19; revised version received 14.03.19; accepted 12.04.19; published 02.05.19.

Please cite as:
Whiteside U, Richards J, Huh D, Hidalgo R, Nordhauser R, Wong AJ, Zhang X, Luxton DD, Ellsworth M, Lezine D
Development and Evaluation of a Web-Based Resource for Suicidal Thoughts: NowMattersNow.org
J Med Internet Res 2019;21(5):e13183
URL: http://www.jmir.org/2019/5/e13183/
doi:10.2196/13183
PMID:31045498

Original Paper

Effects, Adherence, and Therapists' Perceptions of Web- and Mobile-Supported Group Therapy for Depression: Mixed-Methods Study

Raphael Schuster[1], PhD; Inanna Kalthoff[1], MSc; Alexandra Walther[1], MSc; Lena Köhldorfer[2], MSc; Edith Partinger[2], MSc; Thomas Berger[3], Prof Dr; Anton-Rupert Laireiter[1,2], Prof Dr

[1]Center for Clinical Psychology, Psychotherapy and Health Psychology, Department of Psychology, University of Salzburg, Salzburg, Austria
[2]Department of Applied Psychology: Health, Development and Promotion, Faculty of Psychology, University of Vienna, Vienna, Austria
[3]Department of Clinical Psychology and Psychotherapy, University of Berne, Berne, Switzerland

Corresponding Author:
Raphael Schuster, PhD
Center for Clinical Psychology, Psychotherapy and Health Psychology
Department of Psychology
University of Salzburg
Hellbrunnerstraße 34
Salzburg, 5020
Austria
Phone: 43 6644156146
Email: raphael.schuster@stud.sbg.ac.at

Abstract

Background: Blended group therapy (bGT) has been investigated a several times for anxiety and depression, but information on patients' adherence to and therapists' perception of the novel format is nonexistent. Furthermore, many studies investigated mainly female and highly educated populations, limiting the validity of previous findings.

Objective: This study aimed to reduce the gaps and limitations of the previous findings by evaluating an integrated internet- and mobile-supported bGT format.

Methods: A total of 27 patients diagnosed with major depression (14/27, 52% female and 7/27, 25.9% compulsory education) participated in a 7-week treatment at a university outpatient clinic. Furthermore, 8 novice therapists participated in semistructured interviews and a subsequent cross-validation survey.

Results: Primary symptom reduction was high (d=1.31 to 1.51) and remained stable for the follow-up period. Therapists identified advantages (eg, patient engagement, treatment intensification, and improved therapeutic relation) and disadvantages (eg, increased workload, data issues, and undesired effects) of bGT. The required online guidance time was 10.3 min per patient and week, including guidance on exercises (67% or 6.9 min) and intimate communication (33% or 3.4 min). Concerning patients' adherence to bGT, tracked completion of all Web-based and mobile tasks was high and comparable with group attendance.

Conclusions: Results suggest high feasibility of bGT in a gender-balanced, moderately educated sample. bGT provides group therapists with tools for individual care, resulting in an optimization of the therapy process, and high completion rates of the implemented bGT elements. The limited work experience of the involved therapists restricts the study findings, and potential drawbacks need to be regarded in the development of future bGT interventions.

(J Med Internet Res 2019;21(5):e11860) doi:10.2196/11860

KEYWORDS

eHealth; mobile phone; computer-assisted therapy; monitoring; group therapy; depression; professional-patient relations

Introduction

Background

Depression is one of the most prevalent mental disorders and a leading cause of disability. It imposes suffering and high costs on individuals, societies, and health systems [1]. In line with international research priorities [2,3], different forms of mobile- and internet-based interventions constitute innovative and efficient strategies to deliver evidence-based psychological treatments for common mental health disorders [4-8].

Among their most frequent formats, mobile and Web-based interventions offer flexible and anonymous access to mental health services, resulting in low social barriers and low risk of stigmatization [9,10]. Owing to the high degree of automatization, those interventions guarantee standardized treatments to a highly scalable extent. These properties make them attractive for mental health care organizations and have led to the launch of the first routine online clinics [11-13].

However, Web-based and mobile interventions also exhibit limitations as they do not meet all patients' needs and preferences, and therapist contact usually is restricted to a wide degree. Furthermore, therapeutic guidance frequently is associated with better treatment outcomes and reduced dropout rates [14,15]. Finally, many therapists lack experience with this novel approach and hold more cautious attitudes toward Web-based interventions [16]. Comparable levels of caution and awareness have been found among different interest parties (eg, mental health care providers and policy makers), contributing to a frequently discussed retardation of dissemination efforts [17]. Therefore, it is crucial to gain further insight into therapists' perception and acceptance of technology-aided treatments.

The techniques developed in the field of Web-based therapy [7,18-21] can also be harnessed to improve existing forms of face-to-face therapy, resulting in a continuum of blended treatments (Figure 1). Within blended interventions (synonymous computer- or mobile-supported interventions), the spectrum of possible applications ranges from adjuncts to psychotherapy [22], which can be applied before, after, or during treatment [23,24], to more integrated forms of therapy in which Web-based or mobile elements and personal sessions are more deeply intertwined into one treatment rationale [25-27]. A growing number of studies show that blended interventions can lead to shortened treatments in which less therapist time is needed to achieve substantial effects [23,26,28]. Simultaneously, Web-based or mobile elements can be deployed to optimize the therapeutic process, to foster transfer, and to boost effects of classical treatments. In this regard, first studies in routine care found additive effects of traditional face-to-face therapy augmented with Web-based therapy elements [29,30]. According to therapists, patients can profit from blended interventions in the form of increased treatment accessibility and flexibility, as well as from the improvement of patients' self-management and the optimal use of face-to-face sessions [31,32]. Additionally, mental health care providers, policy makers, and other such organizations seem to have a more positive conception of blended therapy compared with pure Web-based therapy [17], and therapists seem to prefer the blended format because it is associated with less risks (eg, diagnostic process) [16]. Among the potential disadvantages of the integrated format, therapists frequently remark that blended therapy is not feasible for all patients and that the format at times could hamper the therapeutic process—in particular, the establishment of the therapeutic alliance [31,32]. Consequently, those issues should be investigated in more detail in patient- and therapist-related studies.

While most blended research focuses on individual therapy [23], less is known about its potential for group therapy. Psychological groups have a broad range of applications in inpatient and outpatient settings [33]; and the spectrum ranges from informational groups, over psychoeducational groups, to group counseling and group psychotherapy [34]. So far, the feasibility and effects of blended group therapy (bGT; synonymous computer- or app-supported group therapy) have been investigated in terms of brief interventions for depression and anxiety. For example, computer-based relaxation, cognitive restructuring, and self-control desensitization have been found to be supportive in the treatment of generalized anxiety disorder (GAD) [35]. Furthermore, first evidence for the efficiency of brief bGT for social anxiety disorder and GAD was found in small comparative trials, leading to significant symptom reductions in a comparably short time period [36,37]. As for depression, several feasibility studies investigated the merits of supportive computer- and mobile-based elements. For example, a tablet-guided behavioral activation (BA) intervention was found to be feasible for the treatment of major depressive disorder [38]. In another study, Aguilera et al found beneficial effects of group therapy augmented with monitoring and text messaging [39]. Furthermore, a brief resource-oriented bGT intervention was developed by our workgroup to address depression by means of a low-threshold, stigma-free treatment strategy. The results revealed high feasibility in terms of client satisfaction and observed between-group treatment effects. Moreover, assessed parameters of treatment adherence (eg, self-reported exercising) indicated high acceptability of bGT elements [40-41]. In a subsequent qualitative investigation (including 13 patients of this study), the use of technology was described as a therapeutic factor, facilitating insight, exercising, and treatment transfer [42].

Figure 1. Blends of Web-based and face-to-face therapy.

Objectives

This study wants to carry this work forward by investigating an integrated bGT intervention based on Acceptance and Commitment Therapy (ACT) principles [43], complemented with elements of BA [44]. As in previous studies, patients' self-reported depressiveness and general health, as well as ACT-specific variables and standardized measures for service satisfaction and usability, were assessed. For the first time, log data were tracked to provide reliable information on completion rates of computer- and mobile-based elements. Focusing on the therapist-related feasibility of bGT, this study includes therapist interviews and a subsequent follow-up survey. As a related aspect, the amount of weekly online guidance was recorded to ascertain therapists' between-session workload.

Methods

Participants

The trial was preregistered at the German trial register (DRKS Number: DRKS00010888), and the regional ethics committee of the University of Salzburg approved the study procedure. Participants were recruited via a multimodal recruitment strategy by handing out flyers in public health centers and densely populated public areas and by advertisements on depression-related Web pages. After registering on the study platform participants obtained detailed information about the procedure and goals of the study and were asked to give informed consent.

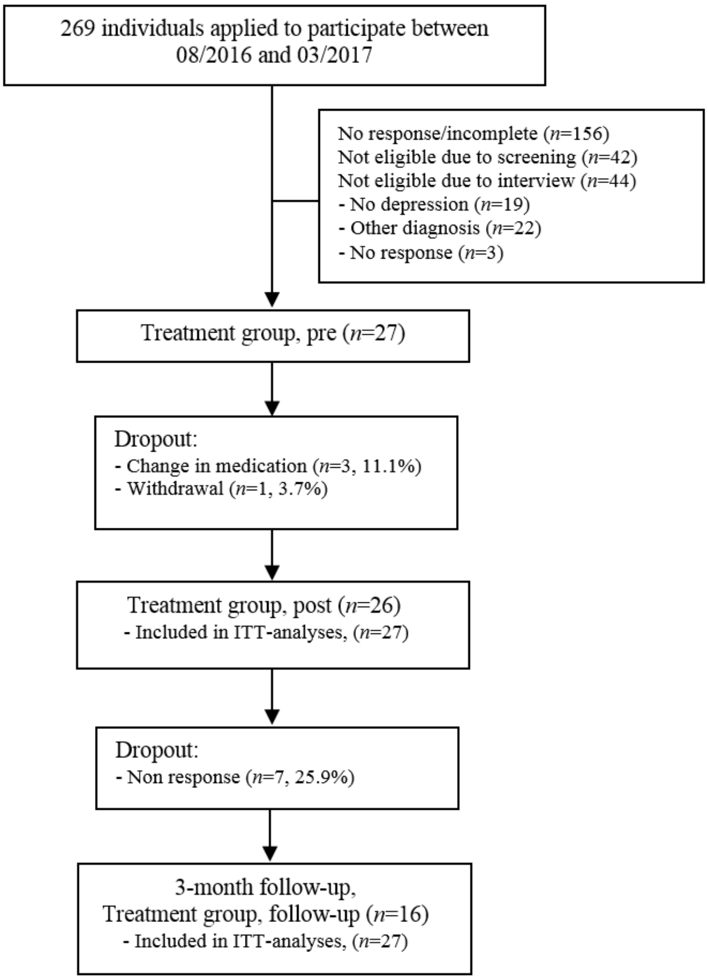
The selection of participants followed 2 steps. Participants were asked to fill out a short screening questionnaire. This included the short version of the Center of Epidemiologic Studies Depression (CES-D) scale [45,46] and additional questions regarding current and past psychological or medical treatment. Participants reporting at least mild levels of depression (CES-D>17) and no suicidal ideation, critical drinking, or past or recent history of severe psychiatric conditions were invited to take part in a diagnostic interview.

Personal clinical interviews were conducted by 3 independent and experienced psychologists, applying the German Mini-Diagnostic Interview for Psychological Disorders (DIPS) [47]. The Mini-DIPS is a 30- to 45-min version of the German DIPS [48], based on the International Classification of Diseases-10 depression criteria. Participants were deemed eligible if the following criteria applied: aged between 18 and 65 years, suffering from mild-to-moderate levels of major depression and/or dysthymia; and/or mild-to-moderate comorbid anxiety, as well as familiarity with the use of personal computers and possession of a smartphone. According to clinical judgement, participants were excluded if they suffered from severe depression (>7 criteria, including main symptoms), severe anxiety disorder, bipolar disorder, any schizoaffective disorder, severe psychiatric and psychotic conditions, substance abuse, suicidal ideation, or if they exhibited low German language and/or computer skills. Participants were also excluded if they currently underwent psychotherapy. Psychiatric medication was tolerated but has been kept constant for at least 3 months before study onset. Figure 2 presents the flowchart demonstrating the recruitment and research procedure in detail.

Procedure

After preassessment, participants were provided with access to the internet platform (Minddistrict) and scheduled to one of 2 weekly groups, depending on personal preferences. To provide personal support in case of technical problems, the app-based diary was installed at the end of the first group session. Group meetings lasted 7 weeks, and each session was preceded by a preparatory Web-based module. The therapist gave supportive feedback after completion of a given Web-based session and occasionally gave reminders to participants by sending out prompts via the platform. The app-based diary complemented the blended treatment with a focus on the transfer of previously learned techniques into daily life. Participants were free to logon to the platform after treatment had ended but did no longer receive therapist guidance. As recommended by several guidelines [33], group sessions were held in a double trainer format which lasted 90 min each. One week after the last group session, the Web-based post assessment had to be filled out and follow-up assessment took place 3 months later.

Figure 2. Study flow chart. ITT: intention to treat.



Intervention

The 7 weeks intense group treatment was based on the ACT and BA principles. ACT [43] is one of the several new treatments originating from cognitive behavioral therapy (CBT). The core principles of this contemporary approach can be divided into mindfulness and acceptance techniques (acceptance, cognitive defusion, and self as context) and behavior change techniques (contact with the present moment, values, and committed action). Even though ACT and BA diverge regarding certain theoretical assumptions (ie, proposed mechanisms of action) [49], they also share many communalities (eg, clarification of goals or strong emphasis on behavioral techniques). Therefore, ACT-based behavior change techniques can be complemented by BA principles [50]. The current treatment rationale was recreated based on a previous intervention, merging ACT and BA into one integrated rationale [51]. Detailed information on intervention content and design can be obtained from Table 1 and Figure 3.

With regard to the use of computer- and mobile-based elements, the patients' weekly routine consisted of 3 steps. First, a preparatory Web-based module, featuring video clips, text-based tasks, and an asynchronous therapist chat, had to be completed. Afterward, patients received individualized feedback from the assigned therapist (if applicable within 2 days). Second, patients participated in the weekly reunions, which again were partially complemented by modern media (ie, short clips or PowerPoint presentations). As a last step, patients were guided by weekly mobile phone diary tasks, which were scheduled for 7 days following the weekly group session. All reminders and prompts were modifiable according to personal preferences, and wherever possible, therapists were instructed to balance media and personal treatment elements according to patient needs and their professional judgement. If patients did not adhere to the Web-based tasks, therapists were instructed to send out a prompt in the middle of the treatment week and, again, once on the day before the forthcoming group session. If patients complained about the number of reminders, the prompts were reduced or stopped.

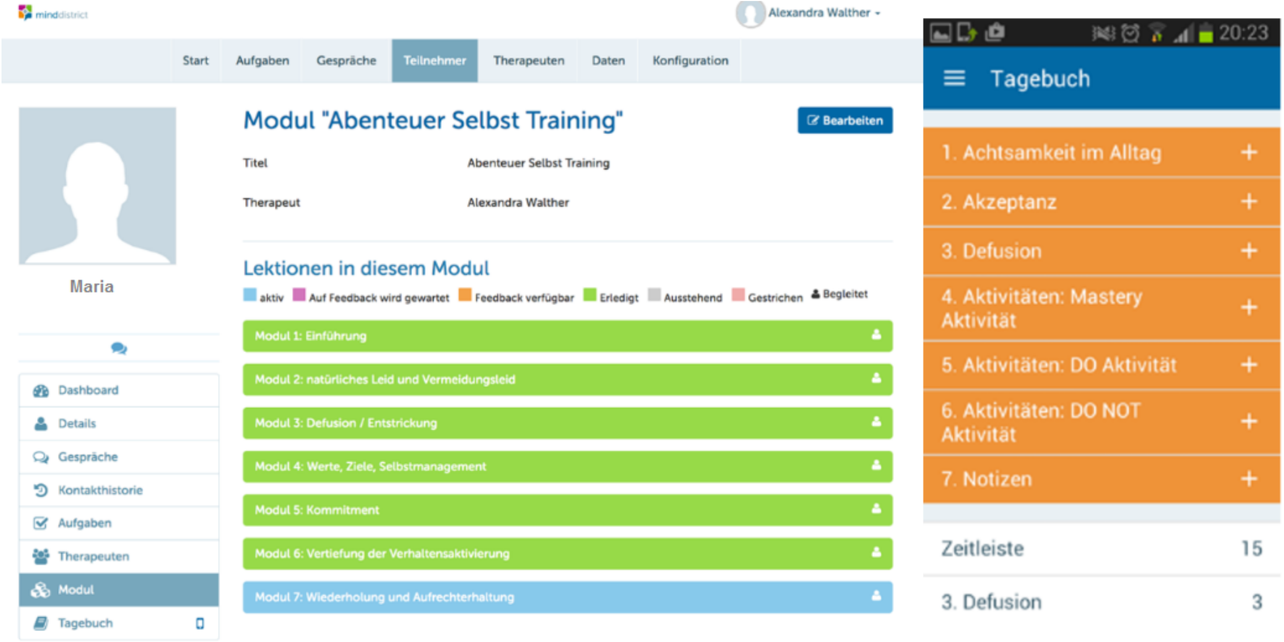
Table 1. Group sessions and computer and multimedia elements of the intervention.

Week	Web-based module	Group session	App	Workbook
1	Introduction into mindfulness	Introduction into ACT[a], mindfulness	Feature 1: Mindfulness in daily life	List of mindful activities
2	Natural suffering and suffering through avoidance	Avoidance and acceptance	Feature 2: Acceptance	Acceptance of a difficult situation, topic, character trait, or conflict; Reflection on mindfulness
3	Defusion	Fusion and defusion	Feature 3: Defusion	Typical examples of defusion
4	Values, goals, and self-management	Values, mastery, and self-management	Feature 4: Mastery activities	Bull's-eye exercise; Example and sheet for SMART[b] principle; Activity planning
5	Commitment	Commitment and positive reinforcement	Feature 5a: "Do activities" Feature 5b: "Do not activities"	Determination, ranking, and planning of do- and do not activities; Self-management; Activity planning
6	Expansion of behavioral activation	Expansion of behavioral activation	Continuation of previous features of the app	Contracts
7	Review and transfer	Transfer and conclusion	Continuation of previous features of the app	Plan for relapse

[a] ACT: Acceptance and Commitment Therapy.

[b] SMART: frequent self-management principle.

Figure 3. User interfaces of the web-based platform and the smartphone app.



Modul "Abenteuer Selbst Training"

Titel	Abenteuer Selbst Training
Therapeut	Alexandra Walther

Lektionen in diesem Modul

Modul 1: Einführung
Modul 2: natürliches Leid und Vermeidungsleid
Modul 3: Defusion / Entstrickung
Modul 4: Werte, Ziele, Selbstmanagement
Modul 5: Kommitment
Modul 6: Vertiefung der Verhaltensaktivierung
Modul 7: Wiederholung und Aufrechterhaltung

1. Achtsamkeit im Alltag +
2. Akzeptanz +
3. Defusion +
4. Aktivitäten: Mastery Aktivität +
5. Aktivitäten: DO Aktivität +
6. Aktivitäten: DO NOT Aktivität +
7. Notizen +

Zeitleiste	15
3. Defusion	3

Therapists

A total of 8 novice therapists (2 male and 6 female) conducted the groups in a double trainer setting. Of the 8 therapists, 2 finished their master’s degree (MSc) or Doctor of Science in psychology and underwent tertiary training in psychotherapy (CBT), clinical psychology (CBT), or medicine at the time of the intervention. The remaining 5 therapists were in their final year of clinical psychology (MSc) and had clinical experience with conducting classical forms of individual or group therapy, as well as with drafting psychological expert reports. None of the therapists had previous experience with conducting bGT or any other form of Web-based therapy. All participating therapists underwent previous training (minimum 40 hours), including a 6-digital versatile disc ACT-series (ACT in Action) and 2 textbooks [52,53] encompassing sections on difficult situations in the ACT. Therapists also protocolled their weekly group sessions. Adherence to the foreseen treatment course was supported by in- and between-session media and technology elements. Of the 8 therapists, 2 therapists participated in a previous bGT study [41] and 6 participated in this study. At the time of the therapist interviews (3 to 12 months after study end), all except 1 therapist were in tertiary clinical trainings for psychotherapy (3 therapists; CBT and client-centered therapy), clinical psychology (3 therapists; CBT), or medicine (1 therapist).

Outcome Measures

Primary Outcomes

The principal outcome of the study was reduction of depressed mood. It was measured by the short version of the German translation of the CES-D scale [45,46]. This questionnaire measures interactive, cognitive, and somatic symptoms, as well as emotions and motor functions related to depression. The 16 items are rated on a 4-point Likert scale. Any value above 17 is interpreted as critical. The German version’s critical threshold (>17) has high discriminative validity, pointed out by a sensitivity of 90%, a specificity of 87% [45], and an area-under-the-curve value of 0.94 [54]. The reliability of the CES-D has been shown to be high [46]. The Cronbach alpha in this study was .90.

As a more general self-report questionnaire that measures psychological distress, nonspecific current mental health, and the risk of developing psychological disorders, the General Health Questionnaire-12 (GHQ-12 [55]) was used. The questionnaire has shown solid reliability [56] and good intercultural validity [57]. The Cronbach alpha in this study was .84.

Secondary Outcomes

Psychological flexibility (ie, acceptance of unpleasant feelings, worry, and control agendas) is the central psychological construct of the ACT and was measured by the Fragebogen zu Akzeptanz und Handeln II [58]. This is the German version of the Acceptance and Action Questionnaire-II (AAQ-II) by Bond et al [59]. The 7 items are rated on a 7-point Likert scale. The Cronbach alpha of the this study was .89.

Anxiety was measured with the Anxious Thoughts Inventory (AnTi) [60] (German translation [61]). It analyzes 3 dimensions of worry: social worry, physical health worry, and meta-worry (worry about worries). The 22 items are rated on a 4-point Likert scale. The Cronbach alpha in this study was .87.

Finally, worry was measured with the PSWQ-3 [62], a short form of the Penn State Worry Questionnaire [63]. It is a questionnaire that assesses self-reported key aspects of worry in GAD. The 3 items are rated on a 5-point Likert scale. The Cronbach alpha in this study was .74.

Client Satisfaction and System Usability

System usability of applied app and Web elements was measured by the System Usability Scale (SUS) [64]. The SUS is a robust questionnaire with 10 items rated on a 5-point Likert scale. The sum score ranges from 1 to 100. SUS scores >85.5 classify excellent usability; scores ≤85.5 and >71.4 classify as good, scores ≤71.4 and >50.9 as OK, scores ≤50.9 and >35.7 as poor, and scores ≤35.7 and >20.3 as awful [65]. The Cronbach alpha in this study was .78.

The ZUF-8 (Fragebogen zur Patientenzufriedenheit) [66], the German version of the Client Satisfaction Questionnaire-8 [67], was used to assess several aspects of participants' overall treatment satisfaction. The 8 items are rated on a 4-point Likert scale. The total score can range from 8 to 32, with a cut-off value of 24 [68] to grade a person as dissatisfied. The Cronbach alpha in this study was .94. For the mobile-based app, adherence was defined as more than 3 weekly entries.

Statistical Analyses

SPSS 24 (SPSS Inc) was used to carry out the analyses. Significant differences between pre, post, and follow-up were analyzed by linear mixed models, with compound symmetry as covariance type and restricted maximum likelihood estimation. Missing outcome values were analyzed according to the intention-to-treat (ITT) principle. Individual pre to post changes served as a base for the reliable change indexes (RCIs) [69]. We used internal consistency as a parameter for RCI reliability [70]. The reliable change criteria were 5.87 scale points for the CES-D and 4.87 for the GHQ-12. For the assessment of change, within-group effect sizes were calculated with pooled SD and reported in Cohen *d* [71]. Power analysis was executed with G*Power [72]. We assumed that the effect size for the secondary outcomes may only lie in the medium range. Thus, an estimated sample size of *N*=22 was calculated for a medium within-subjects effect size of *d*=0.65 (alpha error=.05; power *beta*=.90).

Qualitative Analyses

On the basis of a structured interview guide (Multimedia Appendix 1), audiotaped therapist interviews were conducted by the first author (RS). Interviews lasted between 28 and 56 min (*mean* 44) and were transcribed by 2 independent psychologists who also analyzed the material obtained. MAXQDA was used to conduct the analysis. Analysts were blind to the outcomes and identity of participants. The qualitative content analysis [73] served as the method of information extraction by applying a deductive extraction based on the addressed research questions. After analyzing one third of the transcript, both psychologists and the first author (RS) jointly revised the code system to reach agreement on the applied coding system. Principal codes closely related to the structured interview guide were then specified into further emerging subthemes. After content analysis, a set of follow-up questions was surveyed anonymously to depict the degree of consensus on particular findings among the interviewed therapists. Of the 30 items, 10 items related to design aspects will be reported in a further publication on bGT design. The complete list of follow-up questions was translated by a bilingual psychologist and is presented in (Multimedia Appendix 2).

Results

Participants

A comprehensive overview of participant characteristics at baseline is provided in Table 2. Men and women were equally represented (52%, 14/27 female,), with a mean age of 37.7 years (*SD* 13.7), and relatively low levels of education and employment status. Furthermore, 1 patient withdrew from treatment, resulting in a completion rate of 96% (26/27). During the study period, 3 patients reported changes in medication. According to ITT principles, those patients remained in the analyses. Detailed information on participants' enrolment and participation throughout the study can be seen in Figure 2.

Table 2. Demographic, behavioral, and clinical characteristics of the sample at pretreatment (N=27).

Characteristic	Statistics
Age (years), mean (SD)	37.70 (13.66)
Gender, female n (%)	14 (51.9)
Education, n (%)	
≥9 years (compulsory school)	7 (25.9)
≥12 years (A level)	12 (44.4)
≥any tertiary education (eg, university)	8 (29.6)
Employment, n (%)	
Full time	11 (40.7)
Part time	6 (22.2)
None/marginally	5 (18.5)
Currently in education	5 (18.5)
Current psychopharmacological treatment, n (%)	3 (12)
Previous psychotherapeutic treatment, n (%)	14 (54)
Computer experience, n (%)	
Daily use	25 (92.6)
Weekly use	2 (7.4)
Diagnosis, n (%)	
F32.0 (mild depressive episode), n (%)	3 (11.1)
F32.1 (moderate depressive episode), n (%)	8 (29.6)
F33.0 (recurrent depressive episode, current episode mild), n (%)	10 (37.0)
F33.1 (recurrent depressive episode, current episode moderate), n (%)	4 (14.8)
F33.4 (recurrent depressive disorder, in remission—elevated levels of depression), n (%)	2 (7.4)
Comorbidities, n (%)	
F10.1/2 (harmful use of alcohol/addiction)	1 (3.7)
F40.0 (agoraphobia without panic disorder)	1 (3.7)
F40.1 (social phobia)	2 (7.4)
F40.2 (specific phobia)	1 (3.7)
F41.1 (generalized anxiety disorder)	3 (11.1)
F43.2 (adjustment disorder)	1 (3.7)
F50.2 (bulimia nervosa)	1 (3.7)

Primary and Secondary Outcomes

Linear mixed models unveiled significant changes in all outcome measures, and pre- to posteffect sizes for primary outcomes were large to very large (*d*=1.31 to 1.51). The primary outcome CES-D showed a statistically significant decrease in self-reported depressiveness, with an *F* value of $F_{2,43.323}$=18.94, *P*<.001. For the CES-D, 74% (20/27) of participants exhibited RCIs from pre to post assessment (deteriorations=3.7% [1/27]). Self-reported psychological distress, measured by the GHQ-12, decreased significantly, $F_{2,41.616}$=12.04, *P*<.001, and RCI was found in 63% (17/27) of participants (deterioration=0% [0/27]). Estimated means, SDs, effect sizes, and RCIs of both scales are depicted in Table 3.

For applied secondary outcomes, the treatment resulted in less pronounced effects (*d*=0.38 to *d*=0.71). The AAQ-II revealed a significant change over time, $F_{2,39.710}$=10.41, *P*<.001, and an effect size of *d*=0.59. A comparable pattern was found with regard to the AnTi, $F_{2,39.450}$=12.68, *P*<.001, and *d*=0.72, and with regard to the PSWQ-3, $F_{2,39.447}$=4.11, *P*<.001, and *d*=0.37. For further information on estimated means, SDs, and effect sizes, see Table 3.

Table 3. Means, SDs, effect sizes (Cohen *d*), and reliable change for primary and secondary outcomes (N=27).

Questionnaire	Estimated mean (SD)			Effect sizes (estimated mean [95% CI]), pre to post effect size	Reliable change	
	Pre	Post	Follow-up		Pre to post RCI[a]	Pre to follow-up RCI
CES-D[b]	22.44 (5.18)	13.56 (6.48)	12.19 (7.94)	1.51 (0.89 to 2.09)	74	78
GHQ-12[c]	16.07 (5.41)	9.63 (4.39)	11.94 (7.12)	1.31 (0.70 to 1.87)	63	52
AAQ-II[d]	26.15 (8.87)	20.71 (8.85)	18.63 (9.71)	0.59 (0.02 to 1.14)	—[e]	—
AnTi[f]	44.33 (10.22)	36.46 (10.45)	36.25 (11.59)	0.72 (0.14 to 1.27)	—	—
PSWQ-3[g]	7.63 (2.50)	6.67 (2.76)	6.50 (2.97)	0.37 (−0.19 to 0.91)	—	—

[a] RCI: reliable change index.

[b] CES-D: Center for Epidemiological Studies-Depression scale.

[c] GHQ-12: general health questionnaire (12-item version).

[d] AAQ-II: Acceptance and Actions Questionnaire.

[e] Not applicable.

[f] AnTi: Anxious Thoughts Inventory.

[g] PSWQ-3: Penn State Worry Questionnaire (ultra-short version).

Maintenance of Treatment Effects

After a follow-up period of 3 months, the reduction of self-reported depression (CES-D) remained stable ($F_{1,23.556}$=29.98; *P*<.001) and 78% (21/27) of participants exhibited RCI (deteriorations=7.4% [2/27]). With regard to self-reported psychological distress (GHQ-12), participants indicated significant effects from pre to follow-up, $F_{1,22.758}$=4.82, *P*=.04, and RCI was found in 52% of participants (14/27) (deteriorations=11.1% [3/27]). Contrary to self-reported depressiveness, treatment effects on psychological distress regressed slightly during the follow-up period. However, these reductions failed to rise above the level of statistical significance (contrast: t_{26}=1.39; *P*=.17). Stable treatment effects were also found for the 3 secondary outcomes: psychological flexibility (AAQ-II) $F_{1,18.867}$=12.59, *P*=.002; anxious thoughts (AnTi) $F_{1,17.771}$=12.04, *P*=.003; and worry (PSWQ-3) $F_{1,18.825}$=4.60, *P*=.04. Further information can be obtained from Table 3.

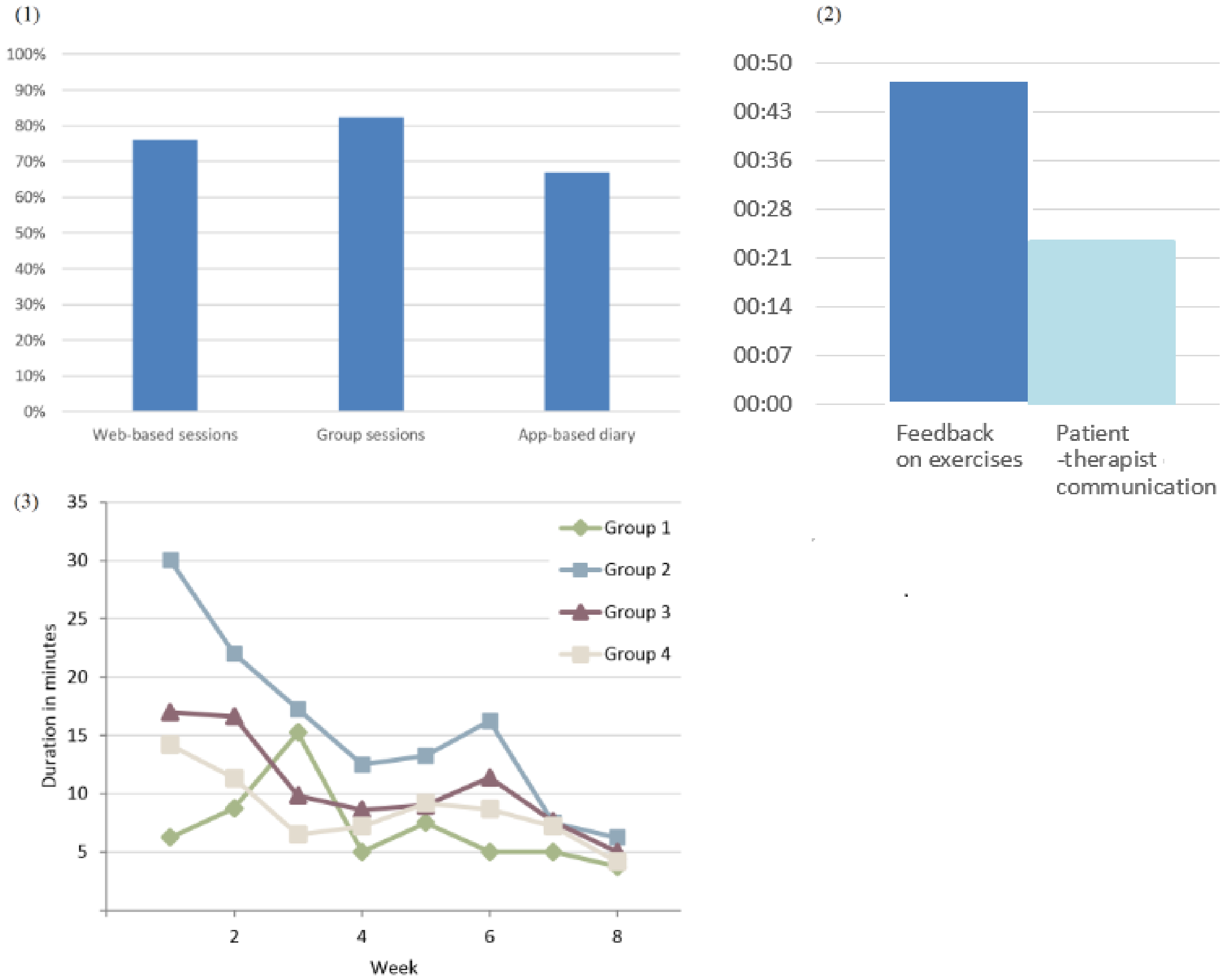
Client Satisfaction and System Usability

System usability of applied app and Web elements, measured by the SUS [64], unveiled an average system usability of 65.33 (SD 18.95) of 100 possible scale points. Accordingly, system usability can be classified as OK to good [65]. Participant's service satisfaction, measured by the ZUF-8 [66], assessed an average satisfaction of mean 26.43 (*SD* 4.80) on a 32-point scale, indicating "good" client satisfaction. However, according to the weekly documentation of group sessions, group coherence in 1 group was low, and the group climate would have profited from including personality disorders (Cluster A and B) in the diagnostic procedure.

Intervention Usage and Therapeutic Guidance

Besides group attendance=82.4% (5.9/7 sessions), usage of digital elements was high (Figure 4): completion rate of Web-based modules 76% (5.3/7 modules), and 67% (14/21 entries) for the mobile-based diary app. However, the average number of app entries during treatment (mean 33) exhibited great variety (range 0 to 246). The average time therapists spent in the guiding of weekly Web-based modules was *mean* 10.3 min per patient, including guidance on accomplished exercises (67% or 6.9 min) and lateral patient-therapist communication (33% or 3.4 min). Thus, two-third of the total guidance was dedicated to the supervision of Web-based tasks, whereas intimate patient-to-therapist communication constituted the remaining third. There was a trend toward a reduction of guidance as the study progressed, and the single groups differed in the required guidance time.

Figure 4. (1) Patients' completion rates of all intervention elements. (2) Therapists' average guidance time per patient during entire treatment; a total of 24 min was spent on personal topics, whereas 48 min was spent on feedback on specific exercises. (3) Therapists' weekly Web-based guidance by single group.



Therapist Interviews and Subsequent Follow-Up Survey

The therapists' experiences with and attitudes toward bGT can be described as cautiously positive to positive. Important themes concerned the functionality and applicability of bGT and patients' interaction with the format, as well as the general appraisal of bGT. Table 4 depicts main themes, subthemes, and frequently assigned codes of the interviews. With Cohen kappa=0.49, interrater agreement was sufficiently high. Interview results were subsequently validated by an anonymous follow-up survey, which was based on the content of the interviews (Table 5).

In the wider perspective, therapists agreed that bGT can have a positive impact on current forms of group therapy and that they had more positive attitudes toward bGT after applying this format. Perceived merits of bGT were augmented monitoring, in addition to patients' responsiveness to given online reminders in terms of increased treatment adherence. Most therapists agreed that patients would profit from the technology-aided treatment transfer and from the repeated presentation of therapy materials (platform, app, and group sessions). Furthermore, they agreed that Web-based modules would prepare patients for subsequent group reunions. A total of 6 out of 8 therapists reported that particular patients disclosed more openly via private Web-based communication (online disinhibition effect) compared with the group meetings. Individual differences emerged in the preference of particular treatment elements. Although some therapists emphasized the added value of between-session elements, others underpinned the merits of applied in-session tools.

With regard to the potential risks of bGT, a consensus emerged that in-session media should be applied cautiously (eg, overloaded sessions) and that the intervention at times may have hampered some of the desired group dynamics (eg, too little time for discussions). In this context, the preservation of technology-free group sessions was suggested. Furthermore, 2 therapists also advocated a cautious use of Web-based reminders and prompts to prevent less interested patients from feeling overwhelmed or discouraged. During the interview, 1 therapist expressed serious concerns about data safety.

Table 4. Main themes, subthemes, and frequent codes of therapist interviews.

Main theme and subtheme	Frequent codes
Advantages	
Patients	Content repeatable; greater learning effect; increased engagement with therapy tasks
Therapists	Additional information through monitoring; helpful for younger therapists; guiding thread
Interaction	Patients more open (online disinhibition); building relationship through intimate Web-based communication
Disadvantages	
General	Additional effort; data security; limited management of acute crisis; predefined treatment course
Specific	Effects on group climate and cohesion; sessions overloaded
General evaluation	
Positive	Contemporary; suitable for in-patient settings; improved handling with increased routine
Negative	Preference toward classic therapy; more training than therapy; technical issues; initial skepticism
Web-based communication	
Online reminders	Require organized working style; increase compliance; unwanted effects
Online feedback	Important feature; needs to be short in duration
Patients’ differences	
Optional classic treatment path	Adaptation to patient preferences; possible side effects
Differences in patients	Not for severe depression; amount of required guidance time; differences in media affinity; requires openness and compliance

Table 5. Benefits and drawbacks of blended group therapy (bGT) according to interview follow-up survey (n=8).

Statement[a]	Percentages		Mean (SD)
	Agree (rather agree)	Disagree (rather disagree)	
I am more open after experience with bGT[a]	25 (75)	0 (0)	3.25 (0.46)
I am more critical after experience with bGT[a]	0 (0)	13 (88)	1.86 (0.36)
I have serious concerns about data safety[a]	0 (25)	25 (50)	2.00 (0.76)
bGT may also be feasible for in-patient treatment[a]	13 (75)	0 (13)	3 (0.53)
Advantages of more flexible working hours because of Web-based guidance[a]	50 (38)	13 (0)	3.25 (1.03)
Computer elements[b] should be used for in-session support[a]	50 (25)	0 (25)	3.25 (0.89)
Overuse of in-session media can hamper group dynamics[a]	50 (38)	0 (13)	3.38 (0.75)
Overuse of in-session media did hamper dynamics in my groups[a]	0 (25)	50 (25)	1.75 (0.87)
Computer elements[c] should be used for between-session support[a]	63 (38)	0 (0)	3.63 (0.52)
Platform prepares patients optimally for group reunions[a]	38 (63)	0 (0)	3.38 (0.52)
Repeated application of therapy content fosters abilities (CE, app, and session)[a]	38 (63)	0 (0)	3.38 (0.52)
Reminders increased compliance with Web-based tasks[a]	13 (75)	0 (13)	3.00 (0.53)
bGT cannot increase treatment transfer[a]	0 (13)	38 (50)	1.75 (0.71)
Reminders did exert a lot of pressure on some patients[a]	13 (50)	0 (38)	2.75 (0.71)
Additional between-session therapist time needs to be reimbursed[a]	88 (13)	0 (0)	3.88 (0.35)
Patients shared additional private concerns over platform (online disinhibition)[a,d]	50 (34)	0 (17)	3.33 (0.82)
Between-session contact made me feel more connected with clients[a,d]	17 (83)	0 (0)	3.17 (0.41)
Between-session contact does not promote relationship with client[a,d]	0 (0)	33 (67)	1.67 (0.52)

[a] Exact wording is provided in Multimedia Appendix 2.

[b] Slides and videos.

[c] Platform, app, and monitoring.

[d] Optional questions only applied to 6 therapists.

Discussion

Principal Findings

This study investigated the feasibility of a mobile- and Web-supported bGT for depression, with a focus on therapists' perception of and patients' adherence to the novel format. High effects on self-reported depressiveness and general health, as well as beneficial effects on ACT-related secondary outcomes, were observed. Effects remained stable over a short follow-up period. Therapist interviews revealed high treatment applicability and perceived benefits concerned treatment availability and monitoring and transfer, as well as the establishment of the therapeutic relation. On an average, therapists spent 10 min per patient per week with online guidance, with decreasing guidance over the course of time and variation between individual groups. Regarding patients' system usage, participants almost equally engaged in weekly group reunions and Web-based tasks. Usage patterns of the mobile-based diary varied to some extent.

Applied primary outcome measures indicated substantial effects on self-reported depressiveness and general health after the outpatient treatment had ended. Observed effects correspond to earlier bGT depression studies [40,41,74], to benchmarking meta-analyses on group therapy [75-77], and to recent group therapy trials in routine care [78,79]. As guideline-based group CBT usually entails 15- to 20-hour sessions [33], high treatment effects where achieved in a comparably short period of time. Although the most observed effects remained stable, self-reported general health decreased slightly but non significantly at follow-up. To further increase treatment success, different forms of online aftercare [23,24] could easily be integrated into bGT and flexible care solutions, such as discontinuous groups, booster sessions, or online groups [80,81], can be facilitated by bGT. As a related aspect, long-term effects of bGT need to be studied in future trials.

This study adds a first therapist-related perspective to the growing evidence on bGT. Retrospectively, novice therapists described the format as contemporary, featuring patient- and therapist-related, as well as interactional, advantages. They reported patients to engage intensely with the bGT tasks, leading patients to be well prepared for the next group session. Furthermore, they appreciated the format for providing flexible working hours, as well as information about the individual treatment progress. Even though not all therapists were initially fond of the novel format, personal experience increased the self-reported willingness to work with the novel approach [82,83]. As for the perceived disadvantages, therapists mentioned the additional between-session workload and some preferred a more classical format. In this context, some therapists stated that it may be difficult to attract experienced or less-interested therapists and that the treatment had a more training-like character. Furthermore, patients should not feel overwhelmed by the use of technology or the intensity of treatment (eg, reminders).

With regard to the reported improvement of therapeutic alliance, the therapist back-end system allowed personalized feedback on completed tasks (two-third of the time), as well as intimate lateral communication between therapists and clients (one-third of the time). Interviewed therapists appreciated both functionalities, and according to the therapists, patients responded to online prompts, resulting in an increased completion of outstanding therapy tasks. As a last consideration, all therapists that used software with implemented confidential communication (6 out of 8 therapists) reported that some of their patients disclosed more openly via intimate lateral communication. This phenomenon can be classified as a form of the online disinhibition effect [84]. In a previous study, the online disinhibition in bGT seemed to be fostered by the perceived intimacy between the patient and therapist in the absence of an additional audience [42].

As another important feasibility criterion, the amount of additional workload because of Web-based guidance is from particular relevance [85]. Beyond doubt, the time required by therapists depends on the implemented tasks of a given intervention. We found a moderate amount of additional workload in an intervention designed to provide close between-session guidance. Most therapists expected further reductions of required guidance time with a growing routine in conducting bGT. Therapist support is frequently associated with improved treatment adherence and lower dropout rates [15]. Here, bGT can be a reasonable alternative to existing formats, such as Web-based therapy or blended individual therapy.

Treatment flexibility is of particular interest in outpatient groups, as the scheduling of group sessions is usually restricted to evening hours on a specific weekday. On one hand, technology-induced treatment flexibility is appreciated by patients [32,42]. On the other hand, more flexible working hours may also prove to be attractive for certain therapists. Here, increases in flexibility are achieved by moving working hours toward Web-based guidance between sessions. In a double trainer setting with a group size of 8 patients, the expected Web-based guidance for 4 patients takes around 45 min per therapist and week. In this regard, surveyed therapists uniformly emphasized the relevance of reimbursement for Web-based guidance time. Even though this additional workload can easily be compensated by shortening the overall treatment duration [28], such shortenings should be carried out carefully and in accordance with patient needs [86] (eg, time to establish trust in the group).

bGT takes a special position in the field of internet interventions. First, bGT can be a cost-efficient treatment option situated between guided Web-based interventions and blended individual therapy (Figure 1). Compared with Web-based interventions, bGT preserves real-world contact at slightly higher costs. When compared with individual therapy, however, bGT can lead to similar cost savings as known from classical group therapy. Second, group phenomena could be harnessed to support therapist efforts to promote compliance with Web-based tasks [42,87]. Compared with Web-based interventions that sometimes suffer from low adherence rates [15,88], patients engaged to a wide extent in the featured Web-based tasks, as results indicate comparable adherence to group sessions and technology-based elements. Finally, bGT blurs distinctions between individual and group therapy, as it brings a high degree of individualized care to the group format. For example, it opens new ways for intimate patient-to-therapist communication, and it routinely provides therapists with individual information on treatment progress or potential problems [74].

With regard to potential disadvantages of bGT, therapists mentioned that certain participants may feel overwhelmed by the close monitoring of between-session activities or by the number of set reminders. For this reason, the intensity of monitoring and Web-based activities should be adaptable to patient needs. As a second aspect, 2 therapists expressed concerns about data safety. These concerns should be treated with high priority to prevent therapists from being deterred. Third, extensive in-session media use was described as a risk factor, potentially dampening desired group dynamics. Although observable incidences were reported less frequently (Table 5), bGT interventions can profit from a cautious implementation of in-session technology. Fourth, some therapists stated that it may be more difficult to attract experienced or less interested therapists. Thus, incentives seem important to make bGT a workable approach (eg, reimbursement of Web-based guidance time, flexibility of working hours, and balance of work tasks). As a last aspect, 1 therapist mentioned the limited management of acute crisis, which theoretically may be induced by Web-based elements or between-session tasks. Here, technology can provide new ways of emergency management too, for example, by the installation of an emergency button, as seen in a blended app-supported problem-solving treatment for patients with intentional self-harm [89].

This study has several noteworthy strengths and limitations. First, this study adds a first therapist-related perspective to previous findings on bGT [37-42]. Second, it applies a multimodal research strategy (eg, triangulation of quantitative and qualitative methods and implementation of log data) to investigate feasibility in a more holistic way. Third, in accordance with recommendations on the documented use of technology [90], this study provides detailed and objectively measured information on Web-based and mobile app completion rates. Fourth, compared with previous bGT depression studies [40,41] and studies on Web-based interventions [91,92], the current sample composition is more balanced with regard to gender and the level of education. Finally, the study reports deterioration rates and possible risks associated with the novel format.

Among its most important limitations, this study was designed and powered to investigate the feasibility of bGT for depression. The study design, therefore, does not allow any conclusions about technology-induced increases in efficiency or effectiveness. Together with blended individual therapy trials [29,30], future research will have to determine the merits of bGT in terms of augmented treatment effects. Second, many different constellations of blended therapy exist and heterogeneity within the field is high [23]. At hand, findings primarily represent the more integrated forms of blended therapy, whereas less-integrated forms (eg, adjunct Web-based programs) may differ in the therapist's guidance, the flexibility of treatment, or the intensity of treatment. In this context, bGT concepts for group psychotherapy (>15 to 20 sessions), as well as blends of internet interventions with telegroup therapy [93,94], and discontinuous groups should be developed. Third, even though conducted in an outpatient clinic, the study setting restricts generalizability, as groups were held at an affiliated university center for psychotherapy and counseling, and the sample was self-selected. Therefore, it is probable that clients were more interested in this kind of treatment. Furthermore, the treatment was carried out by novice therapists. Although some study aspects appear less prone to introducing bias (ie, Web-based guidance time or log files), it is likely that novice therapists are more adaptable to innovations. More ample evaluations of therapist views exist in neighboring fields, such as individual blended therapy, tele therapy, and Web-based therapy [16,31,93,94].

Conclusions

This study adds a first therapist perspective to previous research on bGT. Feasibility was supported within a university outpatient setting, treating a demographically balanced sample with a short but intense ACT-based group intervention. Even though the intervention entailed a variety of Web- and app-based elements, the amount of online guidance was manageable, and guidance resulted in more flexible working hours. The Web-based platform was appreciated for the implementation of between-session monitoring and the establishment of therapeutic alliance. According to therapists, compliance with CBT tasks can be fostered by prompts via the Web-based platform, resulting in high adherence rates. Potential negative effects of blending should be regarded in the design and implementation of bGT interventions.

Acknowledgments

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

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guide.

[PDF File (Adobe PDF File), 55KB - jmir_v21i5e11860_app1.pdf]

Multimedia Appendix 2

Subsequent questionnaire based on previous therapist-interviews.

[PDF File (Adobe PDF File), 135KB - jmir_v21i5e11860_app2.pdf]

References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
2. Tomlinson M, Rudan I, Saxena S, Swartz L, Tsai AC, Patel V. Setting priorities for global mental health research. Bull World Health Organ 2009 Jun;87(6):438-446 [FREE Full text] [doi: 10.2471/BLT.08.054353] [Medline: 19565122]
3. Wykes T, Haro JM, Belli SR, Obradors-Tarragó C, Arango C, Ayuso-Mateos JL, ROAMER consortium. Mental health research priorities for Europe. Lancet Psychiatry 2015 Nov;2(11):1036-1042. [doi: 10.1016/S2215-0366(15)00332-6] [Medline: 26404415]
4. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
5. Cuijpers P, Riper H, Andersson G. Internet-based treatment of depression. Curr Opin Psychol 2015 Aug;4(4):131-135. [doi: 10.1016/j.copsyc.2014.12.026]
6. 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 Dec;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]

7. Josephine K, Josefine L, Philipp D, David E, Harald B. 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]
8. Rathbone A, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017 Dec 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
9. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
10. Emmelkamp P, David D, Beckers T, Muris P, Cuijpers P, Lutz W, et al. Advancing psychotherapy and evidence-based psychological interventions. Int J Methods Psychiatr Res 2014 Jan;23(Suppl 1):58-91. [doi: 10.1002/mpr.1411] [Medline: 24375536]
11. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot Clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
12. Hedman E, Ljótsson B, Kaldo V, Hesser H, El Alaoui S, Kraepelien M, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
13. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
14. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-869. [doi: 10.1586/ern.12.63] [Medline: 22853793]
15. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
16. Schuster R, Pokorny R, Berger T, Topooco N, Laireiter AR. The advantages and disadvantages of online and blended therapy: survey study amongst licensed psychotherapists in Austria. J Med Internet Res 2018 Dec 18;20(12):e11007 [FREE Full text] [doi: 10.2196/11007] [Medline: 30563817]
17. 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]
18. Ly K, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 09;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
19. Ly K, Janni E, Wrede R, Sedem M, Donker T, Carlbring P, et al. Experiences of a guided smartphone-based behavioral activation therapy for depression: a qualitative study. Internet Interv 2015 Mar;2(1):60-68. [doi: 10.1016/j.invent.2014.12.002]
20. Nahum M, Van Vleet TM, Sohal VS, Mirzabekov JJ, Rao VR, Wallace DL, et al. Immediate Mood Scaler: tracking symptoms of depression and anxiety using a novel mobile mood scale. JMIR Mhealth Uhealth 2017 Apr 12;5(4):e44 [FREE Full text] [doi: 10.2196/mhealth.6544] [Medline: 28404542]
21. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
22. Krieger T, Meyer B, Sude K, Urech A, Maercker A, Berger T. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: design of a pragmatic randomized controlled trial. BMC Psychiatry 2014 Oct 08;14:285 [FREE Full text] [doi: 10.1186/s12888-014-0285-9] [Medline: 25298158]
23. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Dec 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
24. Ebert D, Tarnowski T, Gollwitzer M, Sieland B, Berking M. A transdiagnostic internet-based maintenance treatment enhances the stability of outcome after inpatient cognitive behavioral therapy: a randomized controlled trial. Psychother Psychosom 2013;82(4):246-256. [doi: 10.1159/000345967] [Medline: 23736751]
25. 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 Dec 03;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
26. Ly K, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]

27. Urech A, Krieger T, Möseneder L, Biaggi A, Vincent A, Poppe C, et al. A patient post hoc perspective on advantages and disadvantages of blended cognitive behaviour therapy for depression: A qualitative content analysis. Psychother Res 2018 Jan 31:1-13 (forthcoming). [doi: 10.1080/10503307.2018.1430910] [Medline: 29385964]
28. Thase M, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani GK, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Dec 01;175(3):242-250. [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
29. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial. J Affect Disord 2018 Dec;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
30. Zwerenz R, Becker J, Knickenberg RJ, Siepmann M, Hagen K, Beutel ME. Online self-help as an add-on to inpatient psychotherapy: efficacy of a new blended treatment approach. Psychother Psychosom 2017;86(6):341-350. [doi: 10.1159/000481177] [Medline: 29131090]
31. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: a qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164 [FREE Full text] [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
32. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
33. DeLucia-Waack J, Kalodner CR, Riva M. Handbook of Group Counseling and Psychotherapy. London, UK: Sage Publications; 2013.
34. Conyne R. The Oxford Handbook of Group Counseling (Oxford Library of Psychology). Oxford, UK: Oxford University Press; 2011.
35. Przeworski A, Newman MG. Palmtop computer-assisted group therapy for social phobia. J Clin Psychol 2004 Feb;60(2):179-188. [doi: 10.1002/jclp.10246] [Medline: 14724925]
36. Gruber K, Moran PJ, Roth WT, Taylor CB. Computer-assisted cognitive behavioral group therapy for social phobia. Behav Ther 2001;32(1):155-165. [doi: 10.1016/S0005-7894(01)80050-2]
37. Newman MG, Przeworski A, Consoli AJ, Taylor CB. A randomized controlled trial of ecological momentary intervention plus brief group therapy for generalized anxiety disorder. Psychotherapy (Chic) 2014 Jun;51(2):198-206 [FREE Full text] [doi: 10.1037/a0032519] [Medline: 24059730]
38. Samaan Z, Dennis BB, Kalbfleisch L, Bami H, Zielinski L, Bawor M, et al. Behavioral activation group therapy for reducing depressive symptoms and improving quality of life: a feasibility study. Pilot Feasibility Stud 2016;2:22 [FREE Full text] [doi: 10.1186/s40814-016-0064-0] [Medline: 27965841]
39. Aguilera A, Bruehlman-Senecal E, Demasi O, Avila P. Automated text messaging as an adjunct to cognitive behavioral therapy for depression: a clinical trial. J Med Internet Res 2017 Dec 08;19(5):e148 [FREE Full text] [doi: 10.2196/jmir.6914] [Medline: 28483742]
40. Schuster R, Leitner I, Carlbring P, Laireiter A. Exploring blended group interventions for depression: randomised controlled feasibility study of a blended computer- and multimedia-supported psychoeducational group intervention for adults with depressive symptoms. Internet Interv 2017 Jun;8:63-71 [FREE Full text] [doi: 10.1016/j.invent.2017.04.001] [Medline: 30135830]
41. Schuster R, Fichtenbauer I, Sparr VM, Berger T, Laireiter AR. Feasibility of a blended group treatment (bGT) for major depression: uncontrolled interventional study in a university setting. BMJ Open 2018 Mar 12;8(3):e018412 [FREE Full text] [doi: 10.1136/bmjopen-2017-018412] [Medline: 29530905]
42. Schuster R, Sigl S, Berger T, Laireiter AR. Patients' experiences of web- and mobile-assisted group therapy for depression and implications of the group setting: qualitative follow-up study. JMIR Ment Health 2018 Jul 11;5(3):e49 [FREE Full text] [doi: 10.2196/mental.9613] [Medline: 29997106]
43. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies - republished article. Behav Ther 2016 Nov;47(6):869-885. [doi: 10.1016/j.beth.2016.11.006] [Medline: 27993338]
44. Lejuez CW, Hopko DR, Acierno R, Daughters SB, Pagoto SL. Ten year revision of the brief behavioral activation treatment for depression: revised treatment manual. Behav Modif 2011 Mar;35(2):111-161. [doi: 10.1177/0145445510390929] [Medline: 21324944]
45. Hautzinger M, Bailer M. Allgemeine Depressions Skala (ADS). Manual. Weinheim, Germany: Beltz; 1993.
46. Hautzinger M, Bailer M, Hofmeister D, Keller F. Allgemeine Depressionsskala. Manual (2nd ed.). Göttingen, Germany: Hogrefe; 2012.
47. Margraf J. [Mini-DIPS: Diagnostic short interview on mental disorders]. Berlin, Germany: Springer; 2013:1-2.
48. Margraf J, Ehlers A, DiNardo P, Barlow D. [DIPS Diagnostic Interview for Mental Disorders: Interview Guide]. Berlin, Germany: Springer; 2013:1-2.
49. Kanter JW, Baruch DE, Gaynor ST. Acceptance and commitment therapy and behavioral activation for the treatment of depression: description and comparison. Behav Anal 2006;29(2):161-185 [FREE Full text] [Medline: 22478462]

50. Mazzucchelli T, Kane R, Rees C. Behavioral activation treatments for depression in adults: a meta-analysis and review. Clin Psychology-Sci Pr 2009;16(4):383-411. [doi: 10.1111/j.1468-2850.2009.01178.x]
51. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki A, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
52. Harris R. [Difficult Situations in Acceptance and Commitment Therapy (ACT)]. Germany: Belz; 2014.
53. Zettle R. ACT for Depression: A Clinician's Guide to Using Acceptance and Commitment Therapy in Treating Depression. Oakland,CA: New Harbinger Publications; 2007:1-2.
54. Lehr D, Hillert A, Schmitz E, Sosnowsky N. [Screening of Depressive Disorders Using the General Depression Scale (ADS-K) and State-Trait Depression Scales (STDS-T)]. Diagnostica 2008 Apr;54(2):61-70. [doi: 10.1026/0012-1924.54.2.61]
55. Goldberg D, Williams PA. A users's guide to the General Health Questionnaire. Windsor,UK: NFER-Nelson; 1988:1-2.
56. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan;27(1):191-197. [Medline: 9122299]
57. Schmitz N, Kruse J, Tress W. Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. Acta Psychiatr Scand 1999 Dec;100(6):462-468. [Medline: 10626926]
58. Hoyer J, Gloster AT. [Measuring psychological flexibility: the questionnaire on acceptance and action II]. Verhaltenstherapie 2013 Feb 22;23(1):42-44. [doi: 10.1159/000347040]
59. Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, et al. Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. Behav Ther 2011 Dec;42(4):676-688. [doi: 10.1016/j.beth.2011.03.007] [Medline: 22035996]
60. Wells A. A multi-dimensional measure of worry: development and preliminary validation of the anxious thoughts inventory. Anxiety Stress Coping 1994 Jan;6(4):289-299. [doi: 10.1080/10615809408248803]
61. Hoyer J. In: Rebstock U, editor. AnTI-Anxious Thoughts Inventory-[German version]. Frankfurt, Germany: Universität Frankfurt; 1996.
62. Berle D, Starcevic V, Moses K, Hannan A, Milicevic D, Sammut P. Preliminary validation of an ultra-brief version of the Penn State Worry Questionnaire. Clin Psychol Psychother 2011;18(4):339-346. [doi: 10.1002/cpp.724] [Medline: 20806422]
63. Meyer T, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [Medline: 2076086]
64. Brooke J. Redhatch Consulting Ltd. 1996. SUS: A quickdirty usability scale URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2019-03-04] [WebCite Cache ID 76cDc4Xep]
65. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
66. Schmidt J, Lamprecht F, Wittmann WW. [Satisfaction with inpatient care. Development of a questionnaire and initial validity studies]. Psychother Psychosom Med Psychol 1989;39:248 [FREE Full text]
67. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [Medline: 10259963]
68. Hannöver W, Dogs C, Kordy H. [Patient satisfaction - a measure of treatment success?]. Psychotherapeut 2000 Sep 25;45(5):292-300. [doi: 10.1007/s002780000094]
69. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127]
70. Lambert MJ, Ogles BM. Using clinical significance in psychotherapy outcome research: the need for a common procedure and validity data. Psychother Res 2009 Jul;19(4-5):493-501. [doi: 10.1080/10503300902849483] [Medline: 20183403]
71. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Hillsdale, NJ: Lawrence Erlbaum Associates; 1988:37-42.
72. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [Medline: 17695343]
73. Mayring P. Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. Berlin, Germany: Leibniz Insitutute; 2018.
74. Marmarosh CL. Introduction to special issue: feedback in group psychotherapy. Psychotherapy (Chic) 2018 Dec;55(2):101-104. [doi: 10.1037/pst0000178] [Medline: 29863390]
75. Burlingame GM, Fuhriman A, Mosier J. The differential effectiveness of group psychotherapy: a meta-analytic perspective. Group Dyn 2003;7(1):3-12. [doi: 10.1037/1089-2699.7.1.3]
76. Burlingame GM, Seebeck JD, Janis RA, Whitcomb KE, Barkowski S, Rosendahl J, et al. Outcome differences between individual and group formats when identical and nonidentical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. Psychotherapy (Chic) 2016 Dec;53(4):446-461. [doi: 10.1037/pst0000090] [Medline: 27918191]
77. McDermut W, Miller I, Brown RA. The efficacy of group psychotherapy for depression: a meta analysis and review of the empirical research. Clin Psychol-Sci Pr 2001;8(1):98-116. [doi: 10.1093/clipsy/8.1.98]

78. Delgadillo J, Kellett S, Ali S, McMillan D, Barkham M, Saxon D, et al. A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy. Behav Res Ther 2016 Dec;87:155-161. [doi: 10.1016/j.brat.2016.09.010] [Medline: 27685938]
79. Thimm JC, Antonsen L. Effectiveness of cognitive behavioral group therapy for depression in routine practice. BMC Psychiatry 2014 Oct 21;14:292 [FREE Full text] [doi: 10.1186/s12888-014-0292-x] [Medline: 25330912]
80. Bauer S, Wolf M, Haug S, Kordy H. The effectiveness of internet chat groups in relapse prevention after inpatient psychotherapy. Psychother Res 2011 Mar;21(2):219-226. [doi: 10.1080/10503307.2010.547530] [Medline: 21347978]
81. Nordh M, Vigerland S, Öst LG, Ljótsson B, Mataix-Cols D, Serlachius E, et al. Therapist-guided internet-delivered cognitive-behavioural therapy supplemented with group exposure sessions for adolescents with social anxiety disorder: a feasibility trial. BMJ Open 2017 Dec 14;7(12):e018345 [FREE Full text] [doi: 10.1136/bmjopen-2017-018345] [Medline: 29247101]
82. Månsson KN, Klintmalm H, Nordqvist R, Andersson G. Conventional cognitive behavioral therapy facilitated by an internet-based support system: feasibility study at a psychiatric outpatient clinic. JMIR Res Protoc 2017 Aug 24;6(8):e158 [FREE Full text] [doi: 10.2196/resprot.6035] [Medline: 28838884]
83. Kaiser T, Schmutzhart L, Laireiter AR. Attitudes of Austrian psychotherapists towards process and outcome monitoring. Adm Policy Ment Health 2018 Dec;45(5):765-779 [FREE Full text] [doi: 10.1007/s10488-018-0862-1] [Medline: 29520535]
84. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]
85. Kenter R, van de Ven P, Cuijpers P, Koole G, Niamat S, Gerrits R, et al. Costs and effects of internet cognitive behavioral treatment blended with face-to-face treatment: results from a naturalistic study. Internet Interv 2015 Mar;2(1):77-83. [doi: 10.1016/j.invent.2015.01.001]
86. Ruwaard J, Kok R. European Health Psychologist. 2015. Wild West eHealth: Time to hold our horses? URL: https://www.ehps.net/ehp/index.php/contents/article/view/765 [accessed 2019-03-04] [WebCite Cache ID 76cF6BN5C]
87. Ivanov V, Enander J, Mataix-Cols D, Serlachius E, Månsson KN, Andersson G, et al. Enhancing group cognitive-behavioral therapy for hoarding disorder with between-session internet-based clinician support: A feasibility study. J Clin Psychol 2018 Jul;74(7):1092-1105. [doi: 10.1002/jclp.22589] [Medline: 29411356]
88. 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]
89. Mackie C, Dunn N, MacLean S, Testa V, Heisel M, Hatcher S. A qualitative study of a blended therapy using problem solving therapy with a customised smartphone app in men who present to hospital with intentional self-harm. Evid Based Ment Health 2017 Dec;20(4):118-122. [doi: 10.1136/eb-2017-102764] [Medline: 29030503]
90. 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]
91. 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]
92. Klein JP, Gamon C, Späth C, Berger T, Meyer B, Hohagen F, et al. Does recruitment source moderate treatment effectiveness? A subgroup analysis from the EVIDENT study, a randomised controlled trial of an internet intervention for depressive symptoms. BMJ Open 2017 Jul 13;7(7):e015391 [FREE Full text] [doi: 10.1136/bmjopen-2016-015391] [Medline: 28710212]
93. Greene CJ, Morland LA, Macdonald A, Frueh BC, Grubbs KM, Rosen CS. How does tele-mental health affect group therapy process? Secondary analysis of a noninferiority trial. J Consult Clin Psychol 2010 Oct;78(5):746-750. [doi: 10.1037/a0020158] [Medline: 20873910]
94. Marton K, Kanas N. Telehealth modalities for group therapy: comparisons to in-person group therapy. Int J Group Psychother 2015 Dec 14;66(1):145-150. [doi: 10.1080/00207284.2015.1096109]

Abbreviations

AAQ-II: acceptance and action questionnaire-II
ACT: Acceptance and Commitment Therapy
AnTi: Anxious Thoughts Inventory
BA: Behavioral Activation
bGT: Blended Group Therapy
CBT: Cognitive Behavioral Therapy
CES-D: Center of Epidemiologic Studies Depression
DIPS: Diagnostic Interview for Psychological Disorders
GAD: Generalized Anxiety Disorder
GHQ: General Health Questionnaire
ITT: intention-to-treat

PSWQ: Penn State Worry Questionnaire
RCI: Reliable Change Index
SUS: System Usability Scale
ZUF-8: Fragebogen zur Patientenzufriedenheit

Edited by G Eysenbach; submitted 07.08.18; peer-reviewed by N Topooco, EM Rathner; comments to author 04.10.18; revised version received 10.10.18; accepted 27.10.18; published 06.05.19.

Please cite as:
Schuster R, Kalthoff I, Walther A, Köhldorfer L, Partinger E, Berger T, Laireiter AR
Effects, Adherence, and Therapists' Perceptions of Web- and Mobile-Supported Group Therapy for Depression: Mixed-Methods Study
J Med Internet Res 2019;21(5):e11860
URL: https://www.jmir.org/2019/5/e11860/
doi:10.2196/11860
PMID:31066700

Review

Web-Based Interventions to Improve Mental Health in Home Caregivers of People With Dementia: Meta-Analysis

Yinan Zhao[1*], MA; Hui Feng[1,2*], PhD; Mingyue Hu[1*], MD; Hengyu Hu[1*], MD; Hui Li[1*], MA; Hongting Ning[1*], MA; Huijing Chen[1*], MA; Lulu Liao[1*], MA; Linlin Peng[3,4*], MD

[1]Xiangya School of Nursing, Central South University, Changsha, China

[2]Xiangya-Oceanwide Health Management Research Institute, Central South University, Changsha, China

[3]Department of Geriatrics, Xiangya Hospital, Changsha, China

[4]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

[*]all authors contributed equally

Corresponding Author:
Hui Feng, PhD
Xiangya School of Nursing
Central South University
No 172 Tongzipo Road
Changsha,
China
Phone: 86 82650292
Email: feng.hui@csu.edu.cn

Abstract

Background: Dementia is a major cause of disability and dependency in older adults worldwide. It is often accompanied by general psychological distress, such as depression and anxiety symptoms, among caregivers of people with dementia (PwD). The physical and mental health of the caregiver is a prerequisite and a promise to help PwD continue to live as long and as well as possible. Web-based interventions can provide convenient and efficient support and an education tool to potentially reduce the negative outcomes associated with providing care.

Objective: The aim of this study was to examine the effect of internet-based interventions on the mental health outcomes of family caregivers of PwD and to explore which components of the Web-based interventions play an important role.

Methods: A comprehensive literature search was conducted in PubMed, Excerpta Medica dataBASE, PsycINFO, Cochrane Database, and the Cumulative Index to Nursing and Allied Health Literature using relevant terms such as Web-based and caregiver as keywords, covering all studies published before June 2018. A total of 2 reviewers independently reviewed all published abstracts, according to established inclusion and exclusion criteria. We extracted information about the participants, interventions, and results and reviewed article quality in terms of the randomized trial methods, using the approach recommended by the Cochrane Handbook for Systematic Reviews of Interventions.

Results: A total of 815 caregivers participated in 6 studies, with 4 of the studies using depression as an outcome. The analysis found that depression scores dropped an average of 0.23 (95% CI −0.38 to −0.07; *P*<.01) after Web-based interventions. In 2 studies of caregivers who were experiencing anxiety symptoms, the average score for anxiety dropped by 0.32 points (95% CI −0.50 to −0.14; *P*<.01). However, in terms of coping, pain, and stress, the Web-based interventions showed a poor effect. On the whole, the addition of professional psychological support on the basis of education can improve caregivers' mental health.

Conclusions: Internet-based interventions were generally effective at reducing anxiety and depression in dementia caregivers, although negative results were found in some studies. As for burden and stress, further research is required.

(J Med Internet Res 2019;21(5):e13415) doi:10.2196/13415

KEYWORDS

internet; education; mental health; caregivers; dementia

Introduction

Background

Dementia is a syndrome, usually of a chronic or progressive nature, caused by a variety of brain illnesses that affect memory, thinking, behavior, and ability to perform everyday activities [1]. It appears that at least 2 cognitive changes result in significant social and occupational restriction [2]. Dementia currently affects approximately 50 million people worldwide, a number that is projected to grow to 82 million by 2030 and 152 million by 2050 [3,4]. With this growth, dementia is predicted to become the single greatest challenge facing health care and medical systems worldwide [5]. It is the second highest cause of disability in individuals aged 70 years and older and the seventh leading cause of death [4,6]. In terms of care costs, the total cost of dementia globally was US $1 trillion in 2018. By 2030, it is expected that the number will rise to US $2 trillion [7].

Dementia is one of the major causes of disability and dependency among the elderly worldwide [8]. Alzheimer's Disease International (ADI) estimates that, globally, approximately 84% of elderly dementia patients live at home [9] and are cared for by nonprofessional family caregivers. Nonprofessional caregivers (NCs) assume responsibility for providing care in a nonprofessional, unpaid manner, based on family or affective ties [10]. However, because of a lack of knowledge and understanding of dementia, NCs do not understand why these elderly behave as they do or what the future may hold for them. In recent years, multiple studies have shown that, in caregivers of people with dementia (PwD), poor mental health is widespread, with anxiety and depression being the most common symptoms, reported by more than 50% of caregivers [11,12]. Mental health is a state of complete happiness, which refers to our ability to enjoy life and cope with challenges [13]. ADI reported higher levels of depression, emotional distress, and physical strain in caregivers of PwD than in caregivers of older adults with physical impairments [7]. Owing to the severity of this mental health situation, recent research also suggests that caregivers may be a high-risk group for suicide [14]. In Joling's study, researchers found that over the course of 2 years, a number of people caring for a relative with dementia repeatedly considered hurting themselves, felt suicidal, or wished they were dead [15].

Providing training and support for the Alzheimer's family and other family caregivers is an important public health priority. The World Health Organization (WHO) points out that we need to make a special effort to ensure that caregivers can obtain the information needed to do their work or training [16]. It is recognized that training programs for family caregivers of PwD can bring about positive results for caregivers [17], and researchers in this field are particularly interested in the training program methods that have been implemented. Ducharme et al [18] used keynote speeches to train family caregivers; increase their knowledge, practical skills, and caring skills; and improve their sense of self-efficacy. In a previous study, Hepburn et al [19] provided problem-solving skills and coping strategies for caregivers by issuing books and manuals to help alleviate caregivers' anxiety and depression symptoms.

Objectives

In recent years, because of the continuous development of the internet, internet-based intervention measures have gradually been applied by researchers, with Web-based interventions providing a more convenient and efficient support and education method for home care providers, as they allow caregivers to learn anytime, anywhere. Eysenbach [20] revealed that the efficiency with which medical services are delivered through internet interventions could reduce health care costs. In addition, particularly in remote and rural areas, caregivers may have easier access to the internet and electronic health care than to health care resources in the real world, thereby increasing equity in access to health care. The WHO also encourages primary care providers around the world to participate in e-learning training courses to improve their knowledge, skills, and confidence. Increased internet or mobile phone usage has opened up encouraging and novel outlets for mental health care dissemination and delivery efforts [21].

Currently, there are many systematic reviews of education and support for home caregivers based on the internet. Most focus on chronic disease caregivers and have shown positive results in terms of mitigating caregivers' psychological issues [17,22-25]. Furthermore, these systematic reviews have summarized various education support measures and investigated their influence on the results of intervention measures for all kinds of care, such as mental health (anxiety, depression, and stress), general nursing results (self-efficacy, emergency coping styles, and social support), general health outcomes (quality of life [QoL], health, and satisfaction), and so on. Comprehensive results show that a Web-based intervention has resulted in positive, ineffective, or negative effects for caregivers. Overall, the results have been mixed, mainly because of methodological limitations, such as too few samples and too high dropout rates. During the search, we found there were only narrative reviews [26,27], and few meta-analyses were conducted on the mental health of caregivers of PwD in this field.

However, we believe that for different diseases, caregivers will have a different mental burden [28,29] and the focus and frequency of interventions will vary. Therefore, we designed and conducted a meta-analysis based on network intervention support for NCs of PwD, which is an extension of systematic evaluation and can be used to analyze comprehensive data using standardized statistical techniques [30]. We strictly evaluated the quality of the articles and critically evaluated the heterogeneity of the results, so that public health agencies can clearly understand current effective interventions in mental health training for Alzheimer's care. The next step is to determine what measures should be taken to achieve widespread implementation of training.

Methods

We conducted this systematic review and meta-analysis by following the Preferred Reporting Items for Systematic Review and Meta-Analysis guidelines [31].

Study Design

A systematic review design was conceptualized using both data synthesis and descriptive analysis of randomized controlled trials (RCTs).

Data Source and Retrieval Strategy

A comprehensive literature search was conducted in PubMed, Excerpta Medica dataBASE, Ovid MEDLINE, and Cochrane databases, covering all studies published before June 2018. We used the Cochrane systematic review method, internet-based terms related to care providers of the elderly and mental health as keywords. The retrieval strategy included correlative subject headings, keywords, and free words. A total of 2 reviewers developed search strategies in advance. Multimedia Appendix 1 lists the search terms of the system evaluation and the search strategies and search results of each database.

The 2 reviewers independently reviewed all published abstracts; extracted data on participants, interventions, and outcomes; and used the Cochrane risk of bias framework [32] to review article quality.

Inclusion and Exclusion Criteria

To achieve high levels of evidence, we only chose RCTs for peer review. In addition, the inclusion criteria were as follows: (1) family caregivers of older adults with Alzheimer's disease or other related types of dementia; (2) using the technology of the internet to provide intervention or support to the caregiver; (3) interventions aimed at improving the mental health of caregivers; and (4) RCTs with or without a blind method published in an English-language peer-reviewed journal. The exclusion criteria included the following: (1) a lack of real internet-based or browser-based intervention, such as video phone, or only adopting a non-internet–based video conference technology intervention; (2) lack of measurements of the outcome indicators of mental health (such as depression and anxiety symptoms) of caregivers; and (3) only providing Web-based education, without any interaction (such as discussion, Bulletin Board System, and feedback on problems).

The studies included a comparison group, defined as receiving no internet-based intervention, which could include blank controls, general controls, the use of paper materials, minimal guidance, or electronic communications for information resources (eg, e-bulletins).

Quality Assessment

We used the Cochrane bias risk assessment tool [33] to assess the risk of bias of the included studies, which included the following 7 items: (1) generation of random sequences (selection bias); (2) allocation concealment (selection bias); (3) blinding of participants and personnel (performance bias); (4) blinding of outcome assessment (detection bias); (5) incomplete outcome data (attrition bias); (6) selective outcome reporting (reporting bias); and (7) other biases. A total of 2 reviewers independently rated the risk of bias in the above categories as *low risk*, *high risk*, and *unclear*.

Data Extraction and Analysis

In total, 2 independent reviewers extracted and tabulated the following data for subsequent analysis: (1) participants (number of participants in the intervention and control groups, average age, and gender); (2) the intervention description (intervention, control group, and study duration); and (3) measurement of outcomes and measurement time. In the process of data extraction, some articles did not have a clear description of an intervention, and the duration of the intervention is sometimes difficult to determine. The problem was identified by contacting the corresponding authors of the published data.

In this study, we used standard mean difference (SMD) and 95% CIs to show the summary result, in which case it is necessary to mark the results of the study as a unified measure unit. For each study result, the difference between baseline, final value, and SD value was used for analysis. The analysis based on baseline change would be more effective and powerful than the comparison based on final value [34], as it removes a component of person variability from the analysis. In articles that did not report SD, we calculated SD from the reported mean, SE, 95% CI, and other information [35].

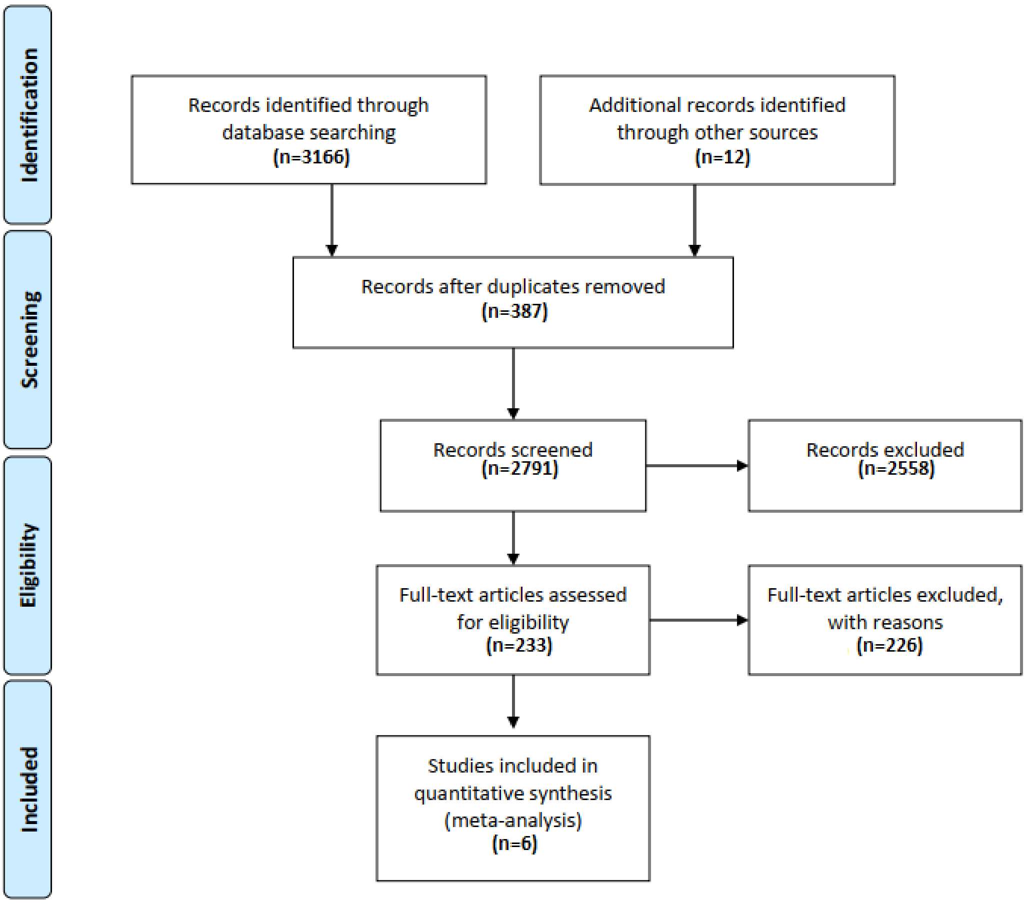
I^2 was used to represent statistical heterogeneity of combinatorial studies [36]. I^2 was used for quantitative statistics of the heterogeneity between the included studies. Under 30% indicated low heterogeneity, 30% to 60% moderate heterogeneity, and greater than 60% high heterogeneity. Review Manager (RevMan 5.3 version, Cochrane Library) was used for all analyses.

Results

Study Selection

A total of 3166 studies were retrieved from various databases, with another 12 from other sources. Of the 3166 studies, 387 reported that the same study was excluded and 2558 did not meet the inclusion criteria. Next, a full-text review of the remaining 233 studies was conducted, including 17 systematic evaluations, 26 non-RCT studies, 73 intervention non-Web–based studies, 6 noncontrol articles, and 3 articles mainly targeted at professional caregivers. Finally, 6 articles were included in this study (Figure 1).

Figure 1. Flow diagram.



Participants

A total of 815 caregivers of PwD or potential memory loss were included in the study. The mean age of the participants in the 6 studies in the report was 58.66 years. A total of 4 articles reported participant gender, with 73.5% of participants who were female. The participants, intervention, comparison, outcome measures, and results details of the included studies are shown in Multimedia Appendix 2.

Types of Intervention

With the classification of Sherifali [25], we divided Web-based interventions for caregivers into single-component interventions and multicomponent interventions.

On the basis of the results of the literature retrieval, we found 2 single-component interventions using information or education and 4 multicomponent interventions, including information or education and peer support (3/4) and information or professional education and psychological support (1/4).

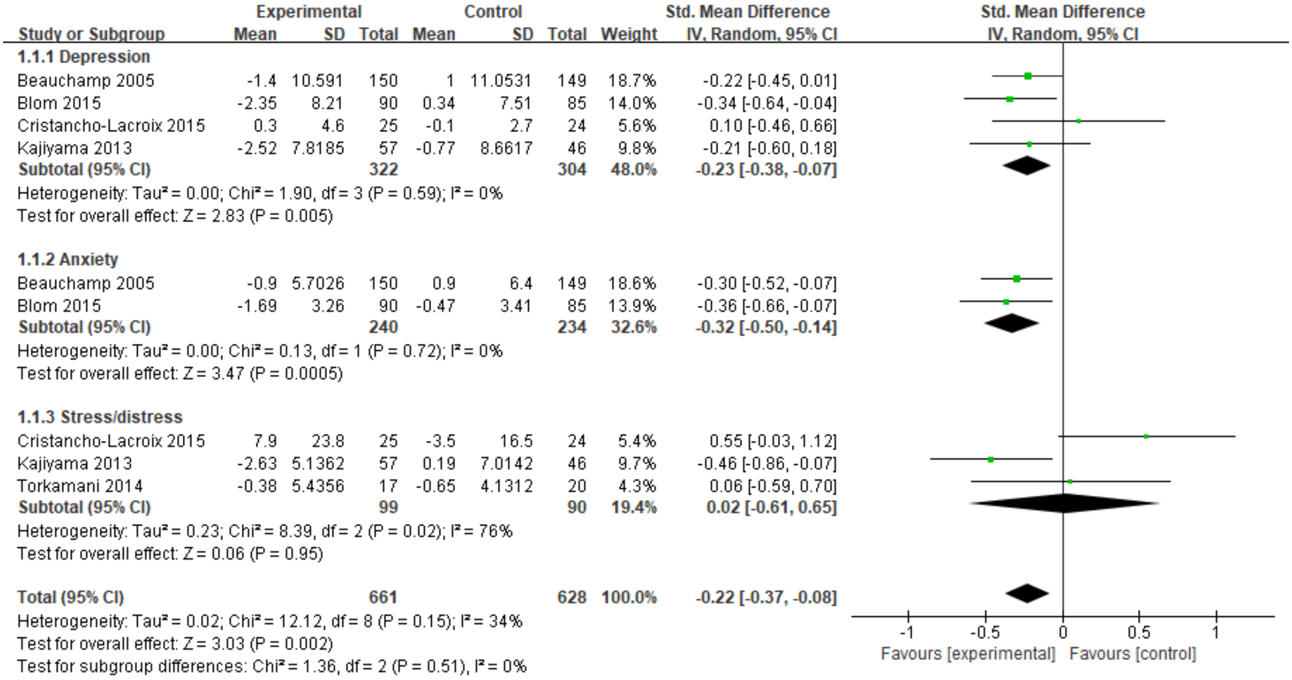
Outcome Measures

According to Isabelle Dor's [13] classification of mental health, the components of mental health include emotional well-being/QoL and psychological and social well-being. We used depression (4/6), anxiety (2/6), and stress/distress (3/6) as primary outcomes and coping (2/6), burden (4/6), and QoL (3/6) as secondary outcomes. The results of the analysis of each variable are as follows.

Depression

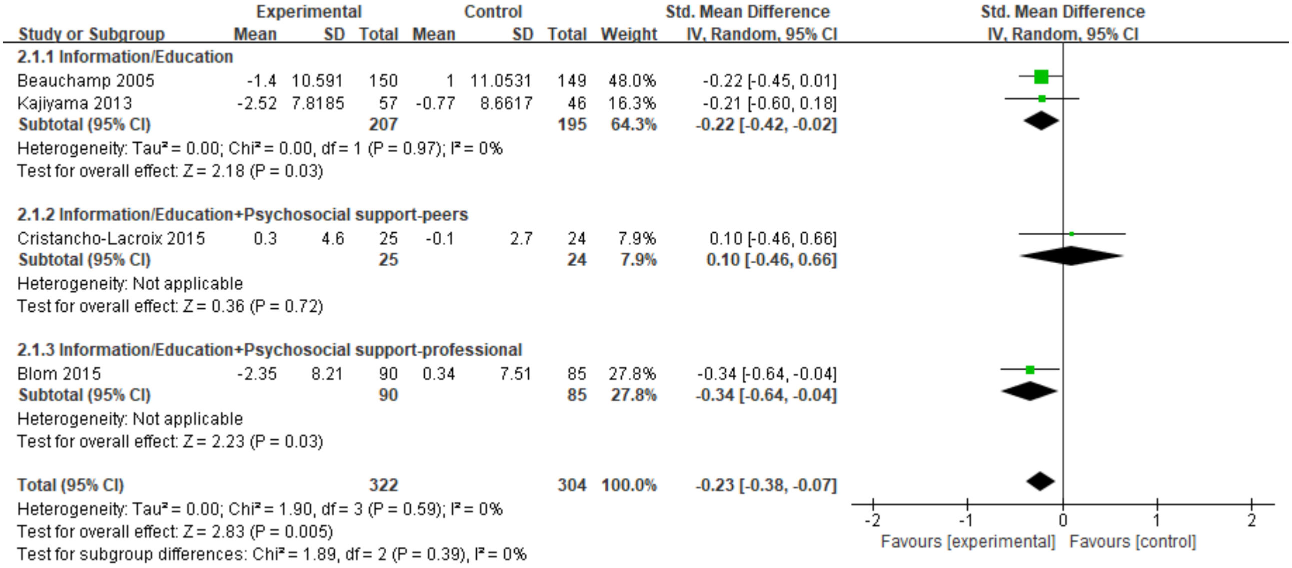
In the 4 studies [37-40], depression was used as the outcome variable, and Center for Epidemiologic Studies Depression Scale (n=3) and Beck Depression Inventory-II scale (n=1) were used to measure depression status. The sample size ranged from 49 to 299. For a sample of 626 caregivers, the overall combined effect of Web-based interventions on depression was statistically significant at −0.23 SMD (95% CI −0.38 to −0.07; *P*=.005). I^2 was equal to 0%, showing low heterogeneity (Figures 2 and 3).

Figure 2. Forest plot for Web-based intervention on mental health. Std: standard.



Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.1.1 Depression								
Beauchamp 2005	-1.4	10.591	150	1	11.0531	149	18.7%	-0.22 [-0.45, 0.01]
Blom 2015	-2.35	8.21	90	0.34	7.51	85	14.0%	-0.34 [-0.64, -0.04]
Cristancho-Lacroix 2015	0.3	4.6	25	-0.1	2.7	24	5.6%	0.10 [-0.46, 0.66]
Kajiyama 2013	-2.52	7.8185	57	-0.77	8.6617	46	9.8%	-0.21 [-0.60, 0.18]
Subtotal (95% CI)			**322**			**304**	**48.0%**	**-0.23 [-0.38, -0.07]**
Heterogeneity: Tau² = 0.00; Chi² = 1.90, df = 3 (P = 0.59); I² = 0%								
Test for overall effect: Z = 2.83 (P = 0.005)								
1.1.2 Anxiety								
Beauchamp 2005	-0.9	5.7026	150	0.9	6.4	149	18.6%	-0.30 [-0.52, -0.07]
Blom 2015	-1.69	3.26	90	-0.47	3.41	85	13.9%	-0.36 [-0.66, -0.07]
Subtotal (95% CI)			**240**			**234**	**32.6%**	**-0.32 [-0.50, -0.14]**
Heterogeneity: Tau² = 0.00; Chi² = 0.13, df = 1 (P = 0.72); I² = 0%								
Test for overall effect: Z = 3.47 (P = 0.0005)								
1.1.3 Stress/distress								
Cristancho-Lacroix 2015	7.9	23.8	25	-3.5	16.5	24	5.4%	0.55 [-0.03, 1.12]
Kajiyama 2013	-2.63	5.1362	57	0.19	7.0142	46	9.7%	-0.46 [-0.86, -0.07]
Torkamani 2014	-0.38	5.4356	17	-0.65	4.1312	20	4.3%	0.06 [-0.59, 0.70]
Subtotal (95% CI)			**99**			**90**	**19.4%**	**0.02 [-0.61, 0.65]**
Heterogeneity: Tau² = 0.23; Chi² = 8.39, df = 2 (P = 0.02); I² = 76%								
Test for overall effect: Z = 0.06 (P = 0.95)								
Total (95% CI)			**661**			**628**	**100.0%**	**-0.22 [-0.37, -0.08]**
Heterogeneity: Tau² = 0.02; Chi² = 12.12, df = 8 (P = 0.15); I² = 34%								
Test for overall effect: Z = 3.03 (P = 0.002)								
Test for subgroup differences: Chi² = 1.36, df = 2 (P = 0.51), I² = 0%								

Figure 3. Forest plot for Web-based intervention on depression. Std: standard.

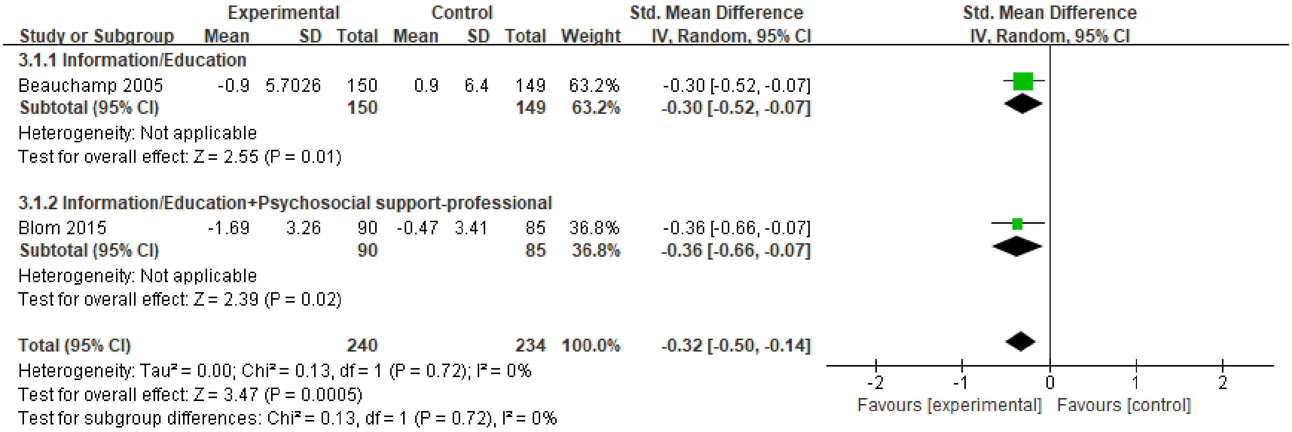


Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Information/Education								
Beauchamp 2005	-1.4	10.591	150	1	11.0531	149	48.0%	-0.22 [-0.45, 0.01]
Kajiyama 2013	-2.52	7.8185	57	-0.77	8.6617	46	16.3%	-0.21 [-0.60, 0.18]
Subtotal (95% CI)			**207**			**195**	**64.3%**	**-0.22 [-0.42, -0.02]**
Heterogeneity: Tau² = 0.00; Chi² = 0.00, df = 1 (P = 0.97); I² = 0%								
Test for overall effect: Z = 2.18 (P = 0.03)								
2.1.2 Information/Education+Psychosocial support-peers								
Cristancho-Lacroix 2015	0.3	4.6	25	-0.1	2.7	24	7.9%	0.10 [-0.46, 0.66]
Subtotal (95% CI)			**25**			**24**	**7.9%**	**0.10 [-0.46, 0.66]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.36 (P = 0.72)								
2.1.3 Information/Education+Psychosocial support-professional								
Blom 2015	-2.35	8.21	90	0.34	7.51	85	27.8%	-0.34 [-0.64, -0.04]
Subtotal (95% CI)			**90**			**85**	**27.8%**	**-0.34 [-0.64, -0.04]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.23 (P = 0.03)								
Total (95% CI)			**322**			**304**	**100.0%**	**-0.23 [-0.38, -0.07]**
Heterogeneity: Tau² = 0.00; Chi² = 1.90, df = 3 (P = 0.59); I² = 0%								
Test for overall effect: Z = 2.83 (P = 0.005)								
Test for subgroup differences: Chi² = 1.89, df = 2 (P = 0.39), I² = 0%								

Anxiety

A total of 2 studies [37,38] included anxiety in the results, and State–Trait Anxiety Inventory (n=1) and Hospital Anxiety and Depression Scale (n=1) were used to measure caregiver anxiety status. The results showed that intervention in the experimental group significantly improved the anxiety status of caregivers with SMD of −0.32 (95% CI −0.50 to −0.14; *P*=.0005). I^2 was equal to 0%, showing low heterogeneity (Figures 2 and 4).

Figure 4. Forest plot for Web-based intervention on anxiety. Std: standard.



	Experimental			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
3.1.1 Information/Education								
Beauchamp 2005	-0.9	5.7026	150	0.9	6.4	149	63.2%	-0.30 [-0.52, -0.07]
Subtotal (95% CI)			**150**			**149**	**63.2%**	**-0.30 [-0.52, -0.07]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.55 (P = 0.01)								
3.1.2 Information/Education+Psychosocial support-professional								
Blom 2015	-1.69	3.26	90	-0.47	3.41	85	36.8%	-0.36 [-0.66, -0.07]
Subtotal (95% CI)			**90**			**85**	**36.8%**	**-0.36 [-0.66, -0.07]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.39 (P = 0.02)								
Total (95% CI)			**240**			**234**	**100.0%**	**-0.32 [-0.50, -0.14]**

Heterogeneity: Tau² = 0.00; Chi² = 0.13, df = 1 (P = 0.72); I² = 0%
Test for overall effect: Z = 3.47 (P = 0.0005)
Test for subgroup differences: Chi² = 0.13, df = 1 (P = 0.72), I² = 0%

Stress/Distress

In the 3 studies [39-41], we used caregiver stress as the outcome variable and used the Perceived Stress Scale (PSS; n=2). In another study, the researcher used the Neuropsychiatric Inventory (NPI) scale to sum up the frequency and severity scores of the caregivers' problem behaviors to obtain the caregivers' total distress score. The sample size ranged from 37 to 103 people. For a sample of 189 caregivers, the overall combined effect of the Web-based intervention on stress was not statistically significant (*P*>.05). I^2 was equal to 32%, showing moderate heterogeneity.

Coping

Coping was used as the outcome variable in 2 studies [37,39]. One of the articles used the Revised Ways of Coping scale and the other used a self-made single item score to evaluate caregiver coping. For the sample of 209 caregivers, the overall combined effect corresponding to the network intervention method was not statistically significant (*P*>.05). I^2 was equal to 0%, showing low heterogeneity.

Burden

Care burden was reported as an outcome in 3 studies [39,41,42]. Of them, 2 studies used the Zarit Burden Scale and the third set up an entry to evaluate care burden. For the sample of 184 caregivers, the overall combined effect corresponding to network intervention was not statistically significant (*P*>.05). I^2 was equal to 0%, showing low heterogeneity.

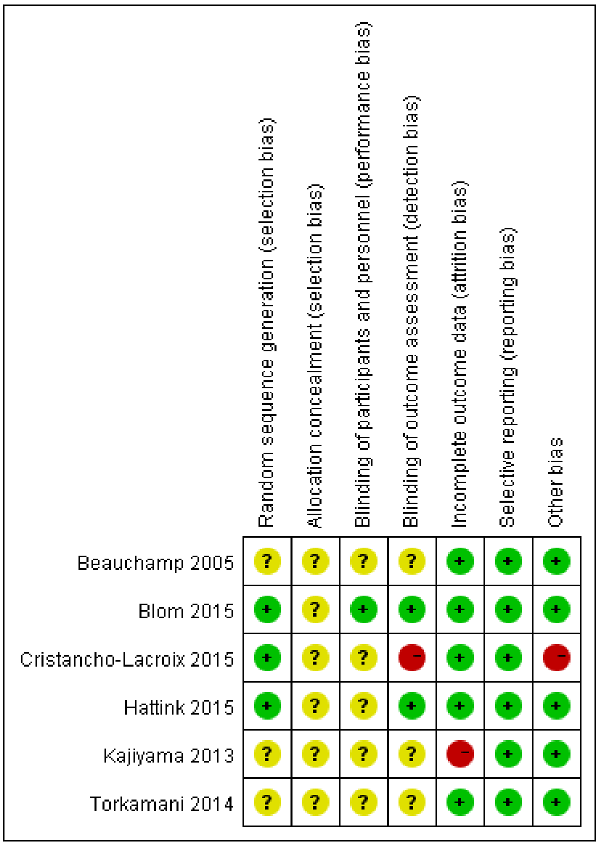
Quality of Life

A total of 3 studies used QoL as outcome measures, using EuroQoL, QoL scale, and perceived QoL to assess changes in participants' QoL. No statistical significance has been found for the intervention to improve caregiver QoL (*P*>.05). I^2 was equal to 39%, showing moderate heterogeneity.

Risk Bias

The quality of the studies included was assessed by the Cochrane risk of bias summary. The overall risk of bias in the published literature is different, and owing to the lack of detailed descriptions reported in the literature, the risk of bias in many projects remains unclear. In all studies, the quality of performance bias and detection bias is relatively inconsistent, whereas the quality of publication bias is relatively low (Figure 5).

Figure 5. Risk of bias summary: review authors' judgments about each risk of bias item for each included study.



	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Beauchamp 2005	?	?	?	?	+	+	+
Blom 2015	+	?	+	+	+	+	+
Cristancho-Lacroix 2015	+	?	?	−	+	+	−
Hattink 2015	+	?	?	+	+	+	+
Kajiyama 2013	?	?	?	?	−	+	+
Torkamani 2014	?	?	?	?	+	+	+

Discussion

Principal Findings

This is the first meta-analysis to only include RCT studies aimed at improving the mental health status of caregivers of older adults with dementia using a Web-based intervention. Many systematic reviews of the mental health of caregivers of people with chronic disease have previously been conducted [17,22-25]; however, the meta-analysis shows high heterogeneity owing to differences in study design, sample size, and intervention measures [25]. Our meta-analysis only included the RCT to constitute the most rigorous design and provide the highest quality evidence. We screened the samples strictly according to the inclusion and exclusion criteria and conducted a subgroup analysis to explore interventions that are the most beneficial for caregivers' mental health, showing that the variable has low heterogeneity. However, the lack of research details and quality reports was a common problem in the published literature.

In this meta-analysis, although our analysis counted the aggregate effect sizes of different degrees of different categories of mental problems, in the 2 dimensions of anxiety and depression, the results were significant, which means that an internet-based intervention is indeed beneficial in improving anxiety and depression in caregivers of PwD. We calculated the effect size as Cohen's *d* [43] where positive effect size represents *symptom improvement* and negative effect size represents *symptom deterioration*. The depression (d=−0.21) and anxiety (d=−0.32) of caregivers in the support group showed low impact and medium impact, consistent with the previous results of Ling [44]. However, there has been no significant improvement in the stress, burden, coping, and life quality of the caregivers in the support group. In terms of burden and pain, only Kajiyama [40] and his colleagues revealed the effects of an intermediate (d=0.46) intervention. However, another study showed that stress symptoms decreased at 3 months but increased at 6 months. This may be caused by a small sample size, low test efficiency, high rate of loss to follow-up, and the evolution of the patient's illness over time. A high rate of loss to follow-up is a common problem in psychotherapy research [45]. Moreover, owing to the lengthy and time-consuming nature of internet courses, the rate may be higher [46]. In the studies we included, the average follow-up rate loss was 28.3%, ranging from 0% to 41.5%. In addition, confusion about the stress source is another important reason, which Statistics Canada noted in 2009 [47]. Caregivers have multiple responsibilities in addition to their caregiver role, with 60% working at a paid job and 28% caring for children under the age of 18 years. These pressures are difficult to eliminate using a simple intervention. At the same time, in this review, we found that the heterogeneity of caregiver stress is high. Alba pointed out that continuous outcomes showed substantially higher I^2 than meta-analyses of binary outcomes [48]. In this review, the source of high heterogeneity may be the difference in the evaluation tools of outcome. In their study, Cristancho-Lacroix [39] and Kajiyama [40] used a sophisticated stress assessment tool, the PSS; however, the caregiver stress data used by Torkamani came from the NPI.

All 6 studies used mature Web-based training platforms, such as STAR, iCare, and Diapason Program, but showed different results. We split the components, shown in Figures 3 and 4, and found that (1) for the depression and anxiety status of dementia caregivers, the effect of increasing professional support on the basis of information/education (SMD=−0.23, 95% CI −0.39 to −0.07) is better than the effect of increasing peer support (SMD=−0.16, 95% CI −0.34 to −0.02). Peeters et al [49] pointed out in a cross-sectional study of caregivers of 984 PwD in the Netherlands that most informal caregivers report they need additional information and advice, for example, about how to cope with their relative's behavioral problems, about the illness trajectory progression, and emotional support and coordination of dementia care. Gaugler et al's [50] empirical findings emphasize that the source of professional information can influence support recommendations for dementia caregiving families in need. On the contrary, in interviews conducted in 1 study [51], some caregivers who participated in peer support activities said that there were always 1 or 2 negative, unhappy people on the team who did not want others to be happy. In terms of pain, only information/education was given, which showed good results. Peer support and professional support were added, but the results were not statistically significant; (2) the improved and comprehensive stress management program showed better outcomes than the modified multicomponent integration program. In the research by Kajiyama [40] and Blom [38], a stress coping model was applied. The course mainly aims to teach nurses a set of core skills aimed at stress management, including relaxation training, learning to increase skills every day, cognitive restructuring, and asking peers/experts for advice. In other training platforms, information on Alzheimer's disease, management of one's emotions, and common processing problems are added on the basis of adjusting coping skills. Therefore, future studies should consider maintaining the systematic integrity of stress coping courses when revising the model; and (3) in his study, Beauchamp personalized the interventions based on the particular concerns of the viewer and the degree of dementia of the person receiving care. It showed good results for depression (d=−0.22), anxiety (d=−0.30), and stress (d=−0.39). This tailored intervention offers a promising and relatively inexpensive method of individualizing content to maximize relevance and impact.

According to the assessment of literature quality, all results of the examination were rated as low or medium, which may reflect the lack of consistency in this emerging field and the resulting research, such as inconsistency in theoretical model selection, inconsistency in intervention course content, and inconsistency in evaluation tools. In addition, the lack of information in some studies leads to an inability to assess the risk of bias in many areas, which is another reason for not showing consistent results. Although there is growing evidence that Web-based multicomponent interventions are feasible, there is no evidence as to which intervention components or platforms are the most useful and which could reduce the cost of training for NCs. Future research can be started from the following aspects: (1) pay attention to NCs of PwD in terms of providing psychological support; (2) when making revisions to the systematic course of stress and coping, pay attention to its integrity and systematisms; (3) for different types of caregivers, develop personalized intervention content that can be based on their needs; and (4) cost-effective interventions can be added in future research.

Strengths and Limitations

This meta-analysis summarizes the effects of internet-based interventions that can be used to evaluate the mental health outcomes of home caregivers of older adults with dementia. Although there have been systematic evaluations in this field previously, we proposed rigorous screening of articles, only included RCTs, developed a comprehensive retrieval strategy, searched 5 databases to ensure we found the most comprehensive literature, and summarized the outcomes of different types of intervention studies by a subgroup analysis. However, after a comprehensive analysis, we found a number of limitations. In all of the studies that were included, the overall quality of evidence was poor, the sample size of some studies was insufficient, the intervention measures varied greatly, and the measurement tools of the outcome variables were different. This review also has some limitations. We divided Web-based interventions into 4 subgroups to reduce heterogeneity but, in some aspects, some heterogeneity remains.

Conclusions

This is the first meta-analysis that only included RCT studies aimed at improving the mental health status of caregivers of older adults with dementia based on a Web-based intervention. Our research results show that in recent years, the literature on Web-based intervention measures has been emerging continuously. Although negative results were found in some studies, internet-based interventions were generally effective at decreasing anxiety and depression in caregivers. Next, the multicomponent internet interventions can be analyzed in further detail using a standardized assessment tool to analyze the outcome indicators to encourage researchers to adopt a more rigorous method in future.

Future researchers are encouraged to take a more rigorous approach and to continue to report on other mental health outcomes, such as stress and burden in dementia caregivers. Ongoing large-scale, high-quality studies are required.

Acknowledgments

The authors sincerely thank the library of the Central South University for its help in the retrieval process for this research and the National Social Science Foundation for its support of their research funds, grant number: 16BSH55.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms.

[PDF File (Adobe PDF File), 171KB - jmir_v21i5e13415_app1.pdf]

Multimedia Appendix 2

Characteristics of included studies.

[PDF File (Adobe PDF File), 184KB - jmir_v21i5e13415_app2.pdf]

References

1. World Health Organization. 2018. 10 Facts on Dementia URL: https://www.who.int/features/factfiles/dementia/en/ [accessed 2019-03-31] [WebCite Cache ID 77HEzY112]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental disorders DSM-5, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
3. World Health Organization. Dementia URL: https://www.who.int/en/news-room/fact-sheets/detail/dementia [accessed 2019-03-31] [WebCite Cache ID 77HF37VbK]
4. World Health Organization. 2017. Dementia: number of people affected to triple in next 30 years URL: https://www.who.int/en/news-room/detail/07-12-2017-dementia-number-of-people-affected-to-triple-in-next-30-years [accessed 2019-03-31] [WebCite Cache ID 77HF9ziz8]
5. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec 16;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6] [Medline: 28735855]
6. World Health Organization. 2018. Mental health URL: https://www.who.int/mental_health/neurology/dementia/en/ [accessed 2019-03-31] [WebCite Cache ID 77HFHjKyw]
7. Patterson C. World Alzheimer Report 2018: The State of the Art of Dementia Research: New Frontiers. London: Alzheimer's Disease International; 2018:027-039.
8. Dagani J, Iozzino L, Ferriari C, Giobbio GM. What is the course of behavioural symptoms and functional conditions in hospitalised older people with dementia? A multicentre cohort study in Italy. Eur Geriatr Med 2015;6(6):554-560. [doi: 10.1016/j.eurger.2015.06.003]
9. Wimo A. Establishing and Funding Long Term Care Infrastructure. 2018 Presented at: 33rd International Conference of Alzheimer's Disease International; July 26-29, 2018; Chicago, USA.
10. Ferrara M, Langiano E, Di Brango T, De Vito E, Di Cioccio L, Bauco C. Prevalence of stress, anxiety and depression in with Alzheimer caregivers. Health Qual Life Outcomes 2008:6. [doi: 10.1186/1477-7525-6-93] [Medline: 18990207]
11. Thompson CA, Spilsbury K, Hall J, Birks Y, Barnes C, Adamson J. Systematic review of information and support interventions for caregivers of people with dementia. BMC Geriatr 2007;7:18 [FREE Full text] [doi: 10.1186/1471-2318-7-18] [Medline: 17662119]
12. Garcia-Alberca J, Lara J, Berthier M. Anxiety and depression in caregivers are associated with patient and caregiver characteristics in Alzheimer's disease. Int J Psychiatry Med 2011;41(1):57-69. [doi: 10.2190/PM.41.1.f] [Medline: 21495522]
13. Doré I, Caron J. [Mental health: concepts, measures and determinants]. Sante Ment Que 2017;42(1):125. [doi: 10.7202/1040247ar]
14. O'Dwyer S, Moyle W, Zimmer-Gembeck M, de Leo D. Suicidal ideation in family carers of people with dementia: a pilot study. Int J Geriatr Psychiatry 2013 Nov;28(11):1182-1188. [doi: 10.1002/gps.3941] [Medline: 23456660]
15. Joling K, O'Dwyer S, Hertogh C, van Hout HP. The occurrence and persistence of thoughts of suicide, self-harm and death in family caregivers of people with dementia: a longitudinal data analysis over 2 years. Int J Geriatr Psychiatry 2018 Dec;33(2):263-270 [FREE Full text] [doi: 10.1002/gps.4708] [Medline: 28379646]
16. World Health Organization. Cagegiver support. In: Integrated Care for Older People: Guidelines on Community-Level Interventions to Manage Declines in Intrinsic Capacity. Geneva: World Health Organization; 2017:21-24.
17. Sousa L, Sequeira C, Ferré-Grau C, Neves P, Lleixà-Fortuño M. Training programmes for family caregivers of people with dementia living at home: integrative review. J Clin Nurs 2016 Oct;25(19-20):2757-2767. [doi: 10.1111/jocn.13266] [Medline: 27273093]
18. Ducharme FC, Lévesque LL, Lachance LM, Kergoat M, Legault AJ, Beaudet LM, et al. "Learning to become a family caregiver" efficacy of an intervention program for caregivers following diagnosis of dementia in a relative. Gerontologist 2011 Aug;51(4):484-494. [doi: 10.1093/geront/gnr014] [Medline: 21383112]
19. Hepburn K, Lewis M, Tornatore J, Sherman CW, Bremer KL. The Savvy Caregiver program: the demonstrated effectiveness of a transportable dementia caregiver psychoeducation program. J Gerontol Nurs 2007 Mar;33(3):30-36. [Medline: 17378189]
20. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
21. Bunnell B, Davidson T, Dewey D, Price M, Ruggiero K. Rural and urban/suburban families' use of a web-based mental health intervention. Telemed J E Health May 2017;23(5):390-396. [doi: 10.1089/tmj.2016.0153] [Medline: 27753542]

22. Dyer E, Kansagara D, Mclnnes D, Freeman M, Woods S. Mobile applications and internet-based approaches for supporting non-professional caregivers: a systematic review. In: VA Evidence-based Synthesis Program Reports. Washington, DC: Department of Veterans A airs Veterans Health Administration Quality Enhancement Research Initiative Health Services Research & Development Service; 2012.
23. Oliver DP, Patil S, Benson J, Gage A, Washington K, Kruse R, et al. The effect of internet group support for caregivers on social support, self-efficacy, and caregiver burden: a meta-analysis. Telemed J E Health 2017 Dec;23(8):621-629. [doi: 10.1089/tmj.2016.0183] [Medline: 28328392]
24. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]
25. Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of Internet-based interventions on caregiver mental health: systematic review and meta-analysis. J Med Internet Res 2018 Jul 3;20(7):e10668 [FREE Full text] [doi: 10.2196/10668] [Medline: 29970358]
26. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. J Med Internet Res 2018 Jun 12;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
27. Egan KJ, Pinto-Bruno Á, Bighelli I, Berg-Weger M, van Straten A, Albanese E, et al. Online training and support programs designed to improve mental health and reduce burden among caregivers of people with dementia: a systematic review. J Am Med Dir Assoc 2018 Mar;19(3):200-6.e1. [doi: 10.1016/j.jamda.2017.10.023] [Medline: 29306605]
28. Brinda E, Rajkumar A, Enemark U, Attermann J, Jacob KS. Cost and burden of informal caregiving of dependent older people in a rural Indian community. BMC Health Serv Res 2014 May 7;14:207 [FREE Full text] [doi: 10.1186/1472-6963-14-207] [Medline: 24886051]
29. Gratao AC, do Vale FA, Roriz-Cruz M, Haas VJ, Lange C, Talmelli LF, et al. The demands of family caregivers of elderly individuals with dementia. Rev Esc Enferm USP 2010 Dec;44(4):873-880 [FREE Full text] [Medline: 21329110]
30. Gregory AT, Denniss AR. An introduction to writing narrative and systematic reviews - tasks, tips and traps for aspiring authors. Heart Lung Circ 2018 Jul;27(7):893-898. [doi: 10.1016/j.hlc.2018.03.027] [Medline: 29857977]
31. Moher D, Liberati A, Tetzlaff J, Altman DG. 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]
32. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011;343:d5928 [FREE Full text] [Medline: 22008217]
33. Deeks JH, Altman D. Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions. USA: Cochrane Collaboration; 2008:211-242.
34. Deeks JH, Altman D. Types of data and effect measures. In: Cochrane Handbook for Systematic Reviews of Interventions. USA: Cochrane Collaboration; 2008:280-281.
35. Deeks JH, Altman D. Summarizing effects across studies. In: Cochrane Handbook for Systematic Reviews of Interventions. USA: Cochrane Collaboration; 2008:263-275.
36. Deeks JH, Altman D. Heterogeneity. In: Cochrane Handbook for Systematic Reviews of Interventions. USA: Cochrane Collaboration; 2008:276-281.
37. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec;45(6):793-801. [doi: 10.1093/geront/45.6.793] [Medline: 16326661]
38. Blom MM, Zarit SH, Groot ZR, Cuijpers P, Pot AM. Effectiveness of an internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
39. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. J Med Internet Res 2015;17(5):e117 [FREE Full text] [doi: 10.2196/jmir.3717] [Medline: 25967983]
40. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. Aging Ment Health 2013;17(5):544-554 [FREE Full text] [doi: 10.1080/13607863.2013.775641] [Medline: 23461355]
41. Torkamani M, McDonald L, Saez AI, Kanios C, Katsanou M, Madeley L, ALADDIN Collaborative Group. A randomized controlled pilot study to evaluate a technology platform for the assisted living of people with dementia and their carers. J Alzheimers Dis 2014;41(2):515-523. [doi: 10.3233/JAD-132156] [Medline: 24643137]
42. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J, et al. Web-based STAR E-learning course increases empathy and understanding in dementia caregivers: results from a randomized controlled trial in the Netherlands and the United Kingdom. J Med Internet Res 2015;17(10):e241 [FREE Full text] [doi: 10.2196/jmir.4025] [Medline: 26519106]
43. Muller K. Statistical power analysis for the behavioral sciences. Technometrics 1988;31(4):499-500. [doi: 10.1080/00401706.1989.10488618]

44. Chien L, Chu H, Guo J, Liao Y, Chang L, Chen C, et al. Caregiver support groups in patients with dementia: a meta-analysis. Int J Geriatr Psychiatry 2011 Oct;26(10):1089-1098. [doi: 10.1002/gps.2660] [Medline: 21308785]
45. 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]
46. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
47. Sinha M. Statistics Canada. 2012. Spotlight on Canadians: results from the General Social Survey: portrait of caregivers URL: https://www150.statcan.gc.ca/n1/en/pub/89-652-x/89-652-x2013001-eng.pdf?st=iFd1MaEI [accessed 2019-04-01] [WebCite Cache ID 77Iswx2Jr]
48. Alba A, Alexander P, Chang J, MacIsaac J, DeFry S, Guyatt G. High statistical heterogeneity is more frequent in meta-analysis of continuous than binary outcomes. J Clin Epidemiol 2016 Feb;70:129-135. [doi: 10.1016/j.jclinepi.2015.09.005] [Medline: 26386323]
49. Peeters J, Van Beek AP, Meerveld JH, Spreeuwenberg PM, Francke AL. Informal caregivers of persons with dementia, their use of and needs for specific professional support: a survey of the National Dementia Programme. BMC Nurs 2010 Jun 7;9:9 [FREE Full text] [doi: 10.1186/1472-6955-9-9] [Medline: 20529271]
50. Gaugler J, Westra B, Kane R. Professional discipline and support recommendations for family caregivers of persons with dementia. Int Psychogeriatr Jun 2016;28(6):1029-1040. [doi: 10.1017/S1041610215002318] [Medline: 26739379]
51. Vaughan C, Trail T, Mahmud A, Dellva S, Tanielian T, Friedman E. Informal caregivers' experiences and perceptions of a web-based peer support network: mixed-methods study. J Med Internet Res 2018 Aug 28;20(8):e257 [FREE Full text] [doi: 10.2196/jmir.9895] [Medline: 30154074]

Abbreviations

ADI: Alzheimer’s Disease International
NC: nonprofessional caregiver
NPI: Neuropsychiatric Inventory
PSS: Perceived Stress Scale
PwD: people with dementia
QoL: Quality of Life
RCT: randomized controlled trial
SMD: standard mean difference
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.01.19; peer-reviewed by B Davis, MS Aslam; comments to author 07.02.19; revised version received 15.03.19; accepted 24.03.19; published 06.05.19.

Please cite as:
Zhao Y, Feng H, Hu M, Hu H, Li H, Ning H, Chen H, Liao L, Peng L
Web-Based Interventions to Improve Mental Health in Home Caregivers of People With Dementia: Meta-Analysis
J Med Internet Res 2019;21(5):e13415
URL: http://www.jmir.org/2019/5/e13415/
doi:10.2196/13415
PMID:31066680

Original Paper

Reducing Stress and Preventing Depression (RESPOND): Randomized Controlled Trial of Web-Based Rumination-Focused Cognitive Behavioral Therapy for High-Ruminating University Students

Lorna Cook[1*], MSc; Mohammod Mostazir[2*], MSc; Edward Watkins[1*], PhD

[1]SMART Lab, Mood Disorders Centre, School of Psychology, University of Exeter, Exeter, United Kingdom
[2]College of Life and Environmental Sciences (CLES), School of Psychology, University of Exeter, Exeter, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Edward Watkins, PhD
SMART Lab, Mood Disorders Centre
School of Psychology
University of Exeter
Sir Henry Wellcome Building for Mood Disorders Research
Queens Drive
Exeter, EX4 4QG
United Kingdom
Phone: 44 (0) 139272 ext 4692
Email: E.R.Watkins@exeter.ac.uk

Abstract

Background: Prevention of depression is a priority to reduce its global disease burden. Targeting specific risk factors, such as rumination, may improve prevention. Rumination-focused Cognitive Behavioral Therapy (RFCBT) was developed to specifically target depressive rumination.

Objective: The primary objective of this study was to test whether guided Web-based RFCBT (i-RFCBT) would prevent the incidence of major depression relative to usual care in UK university students. The secondary objective was to test the feasibility and estimated effect sizes of unguided i-RFCBT.

Methods: To address the primary objective, a phase III randomized controlled trial was designed and powered to compare high risk university students (N=235), selected with elevated worry/rumination, recruited via an open access website in response to circulars within universities and internet advertisements, randomized to receive either guided i-RFCBT (interactive Web-based RFCBT, supported by asynchronous written Web-based support from qualified therapists) or usual care control. To address the secondary objective, participants were also randomized to an adjunct arm of unguided (self-administered) i-RFCBT. The primary outcome was the onset of a major depressive episode over 15 months, assessed with structured diagnostic interviews at 3 (postintervention), 6, and 15 months post randomization, conducted by telephone, blind to the condition. Secondary outcomes of symptoms of depression and anxiety and levels of worry and rumination were self-assessed through questionnaires at baseline and the same follow-up intervals.

Results: Participants were randomized to guided i-RFCBT (n=82), unguided i-RFCBT (n=76), or usual care (n=77). Guided i-RFCBT reduced the risk of depression by 34% relative to usual care (hazard ratio [HR] 0.66, 95% CI 0.35 to 1.25; *P*=.20). Participants with higher levels of baseline stress benefited most from the intervention (HR 0.43, 95% CI 0.21 to 0.87; *P*=.02). Significant improvements in rumination, worry, and depressive symptoms were found in the short-to-medium term. Of the 6 modules, guided participants completed a mean of 3.46 modules (SD 2.25), with 46% (38/82) being compliant (completing ≥4 modules). Similar effect sizes and compliance rates were found for unguided i-RFCBT.

Conclusions: Guided i-RFCBT can reduce the onset of depression in high-risk young people reporting high levels of worry/rumination and stress. The feasibility study argues for formally testing unguided i-RFCBT for prevention: if the observed effect sizes are robustly replicated in a phase III trial, it has potential as a scalable prevention intervention.

Trial Registration: ISRCTN Registry ISRCTN12683436; https://www.isrctn.com/ISRCTN12683436 (Archived by WebCite at http://www.webcitation.org/77fqycyBX)

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-015-1128-9

(J Med Internet Res 2019;21(5):e11349) doi:10.2196/11349

KEYWORDS

cognitive behavioral therapy; depression; prevention; rumination, cognitive; stress, psychological; student health services

Introduction

Background

Depression is the leading cause of disease burden worldwide, accounting for 7.5% of all years lived with disability in 2015 [1], with considerable individual, societal, and economic consequences. Although there are effective evidence-based acute treatments, their impact is limited because of poor access to treatments [2], high rates of nonresponse [3], and the recurrent nature of depression, with 50% to 80% of patients experiencing 2 or more episodes [4]. It is estimated that even with optimal acute treatment at full coverage, only 34% of the disease burden would be averted [5]. As a consequence, a strong case has been made that prevention is needed to reduce the global burden of depression [6].

Preventive interventions, largely using cognitive behavioral therapy (CBT) approaches, can reduce symptoms of depression and prevent the incidence of depression [7-9], with an average reduction of 21% in incidence rates [10]. These meta-analyses suggest that targeted interventions (selective interventions aimed at subgroups with known risk factors and indicated interventions aimed at those with subclinical symptoms) produce larger and longer-lasting effects than universal interventions aimed at entire populations. A meta-analysis [11] of 21 preventive interventions (15 using CBT approaches) found that selective interventions and indicated interventions had lower incidence rate ratios (IRR; 0.72 and 0.76, respectively) relative to controls than universal interventions (IRR=0.90). Merry et al [8] also found that both universal and targeted interventions reduced incidence relative to no intervention in the short-to-medium term (3 to 9 months postintervention) but only targeted interventions reduced incidence at 12 months. Thus, targeting at-risk groups may improve the efficacy of preventive interventions for depression [6], in part because the base rate is higher in targeted samples, so it is easier to detect a significant effect with smaller sample sizes [12].

The incidence of depression rises steeply from the age of 14 years through young adulthood, with increased rates in females (2:1, female:male) emerging at around the age of 12 years and continuing into young adulthood [13]. The UK Adult Psychiatric Morbidity Survey found increasing rates of common mental health disorders (CMDs; incorporating depression and anxiety disorders) among young women (aged 16 to 24 years), rising from 22.2% in 2007 to 28.2% in 2014 [14], with rates in young women almost three times those of young men (10.0%) in 2014. As early onset is linked to greater chronicity [15] and other negative long-term outcomes, such as poor academic and occupational performance [1,16], prevention may be particularly effective and impactful for this age group.

Within this age range, university students are a particularly high-risk group, with a weighted mean prevalence for depression of 30.6% (range 10% to 85%) across 24 studies [17] relative to estimates of 10.8% to 22% in nonstudents of the same age range [18,19]. This increased prevalence may be due to the specific pressures of university and associated lifestyle changes, such as leaving the family home for the first time, forming new friendships, more self-directed learning, and irregular sleep patterns [20]. Students who experience mental health difficulties during their studies are at greater risk of poor academic outcomes [16] and dropout [21].

Despite these challenges, students often do not seek help from relevant services [22,23]. Alternative delivery modes, such as Web-based interventions, offer advantages that may be attractive to students, including availability at any time and place, anonymity which may reduce the stigma of seeking help, and more time to reflect on the treatment material [12,24,25]. A recent systematic review and meta-analysis of 17 Web-based and computer-delivered interventions for higher education students found reductions in depression, anxiety, and stress when compared with inactive controls [26]. However, sample sizes were generally small, and the authors recommend further larger-scale trials to assess the effectiveness of Web-based interventions in university students.

One such relatively large-scale trial by Topper et al [27] tested a guided Web-based targeted preventive intervention for 251 high school and university students aged 15 to 22 years with high levels of self-reported worry and rumination. Participants were randomized to face-to-face group rumination-focused cognitive-behavioral therapy (RFCBT), guided Web-based RFCBT (i-RFCBT), or a no-intervention control group. This preventive intervention is based on RFCBT, previously shown to be effective in treating residual depression [28]. There is considerable evidence that rumination plays a causal role in the onset and duration of major depressive episodes (MDEs) [29,30]. Within a student population, rumination predicts change in depression over 6 months [31]. Rumination interacts with other risk factors to both maintain depression (the combination of rumination, low self-esteem, and stressful life events predicts maintenance of depressive symptoms over 6 weeks [32]) and predict the onset of depressive symptoms (engaging in rumination in response to stress prospectively predicted an increase in subsequent depressive symptoms [33]). These studies suggest that specifically targeting ruminative responses to stressful events could reduce depression.

RFCBT specifically targets repetitive negative thought (RNT), incorporating both rumination and worry, defined as a thinking style that: (1) is repetitive, intrusive, and difficult to disengage from; (2) is perceived as unproductive; and (3) captures mental capacity [34]. RNT is a transdiagnostic process, involved in the onset and maintenance of a range of emotional disorders including depression and anxiety as well as physical health issues [29,35], including in children and adolescents [36-39]. Targeting transdiagnostic risk factors has the potential to improve the efficacy of prevention by impacting on multiple disorders with a single intervention [27].

In the Topper et al [27] trial, both Web-based and group-delivered RFCBT reduced symptoms of depression and anxiety (d=0.36 to 0.72), relative to controls. Cumulative incidence rates at the final 12-month follow-up were significantly lower in both RFCBT intervention conditions for depression (14.7% Web-based; 15.3% group) relative to the usual care control condition (32.4% depression), with no difference between i-RFCBT versus group RFCBT. In support of the hypothesized mechanism of change, reductions in worry and rumination were found to mediate the effects of the interventions on prevalence of depression and Generalized Anxiety Disorder (GAD). These findings suggest that targeting rumination may have preventive effects for depression and are consistent with evidence that targeted prevention can be effective in adolescents and young adults.

Objectives

Topper et al [27] included both secondary and university students to form a heterogeneous sample. Given the evidence that undergraduates may form a distinct at-risk subgroup for depression, the primary aim of this phase III efficacy trial was to test whether these beneficial effects of guided i-RFCBT on onset of depression relative to usual care [27] could be extended to a selective UK high-risk undergraduate population.

We also aimed to address several key limitations of the Topper et al [27] trial: (1) there was no diagnostic interview to assess depression, and self-report measures were only able to estimate point prevalence caseness and (2) as history of depression was not assessed, participants' previous history of depression was not known, and therefore it was not possible to discriminate whether the intervention prevented first onset or relapse/recurrence of depression. To address these methodological limitations, we included a well-validated diagnostic interview (Structured Clinical Interview for DSM-IV [SCID-I] [40]) to increase accuracy of the current diagnostic status and measure retrospective incidence.

We hypothesized that in high-ruminating undergraduates, guided i-RFCBT, relative to usual care, would significantly reduce the onset of MDEs over the course of the 15 months post randomization follow-up (primary outcome). Rumination has also been found to increase the negative effect of stressful life events on depressive symptoms in young people and students [32,33]. This observed interaction is consistent with the evidence that rumination contributes to depression by exacerbating existing negative mood and negative cognitions and by repetitively dwelling on difficulties [29,30], such as results from stressful events. As such, a tendency to ruminate would be expected to have less impact when things are going well and there is less to dwell on, relative to when things are difficult. We therefore hypothesized that i-RFCBT would be particularly beneficial for high-ruminating undergraduates who were also experiencing stressful life events, as this would be the group for whom rumination would be most detrimental.

As a secondary aim, we explored the feasibility and acceptability of an unguided version of i-RFCBT to prevent depression. Topper et al [27] used i-RFCBT that was guided and supported by a therapist because past evidence suggested that, at least for acute treatment for depression, guided Web-based cognitive behavioral therapy (i-CBT) is significantly more effective than unguided (ie, self-help) i-CBT [41-43], and only guided i-CBT produces similar treatment effects to face-to-face therapy in patients with acute depressive symptoms [24]. A key rationale for Web-based therapy is to increase the coverage, availability, and accessibility of treatment, by potentially reaching large numbers of people through the internet and by overcoming hurdles such as geographical distance, poor mobility, and scheduling appointments during standard office hours. However, any form of guided i-CBT (including i-RFCBT) is necessarily limited in its scalability because coverage is determined by the number and availability of therapists. In contrast, an unguided form of Web-based therapy has nearly limitless scalability as there are no such constraints and, thus, even with smaller effect sizes than guided interventions, has significant potential to reduce the disease burden of depression [44]. Such an intervention would be particularly beneficial for preventing depression because effective prevention requires an intervention to be highly scalable and able to reach very high numbers of people. As a secondary question, we therefore explored the feasibility and acceptability of an unguided version of i-RFCBT in a quasi-phase II pilot arm and estimated its effect sizes to inform a fully powered trial, with regard to incidence rates and symptom levels (descriptives and CIs).

Methods

Trial Design

Phase III Efficacy Study

The phase III study consisted of a single (researcher) blind parallel-group randomized controlled trial (RCT), comparing guided i-RFCBT versus a usual care control group. For full details, see the trial protocol paper [45] and Current Controlled Trials ISRCTN12683436.

Quasi-Phase II Pilot Arm

To assess the feasibility of unguided i-RFCBT, a separate adjunct arm of unguided i-RFCBT was included as a quasi-phase II pilot arm. For efficiency, participants were randomized to this arm within the overall trial design, but there was no direct comparison between the unguided and guided arms. The unguided arm was compared with the control group to estimate the effect sizes of an unguided version of i-RFCBT for the planning of future efficacy trials.

Participants

Participants were university students resident in the United Kingdom, aged 18 to 24 years, with elevated RNT, defined as scoring above the 75th percentile on at least one measure of worry/rumination: ≥50 on the Penn State Worry Questionnaire (PSWQ; [46]); ≥40 on the Ruminative Response Scale (RRS; [47]), using the same criteria as Topper et al [27]. As a prevention study, participants were excluded if they met the diagnostic criteria for a current (within the past month) MDE. In addition, potential participants were excluded if they reported any of the following: current and significant substance abuse or dependence; current symptoms/diagnosis of psychosis or bipolar disorder; and current psychological therapy or active suicide risk. In line with standard practice, receipt of antidepressant medication was not an exclusion criterion, providing the dose had been stable for at least 1 month.

Sample Size Calculation and Recruitment

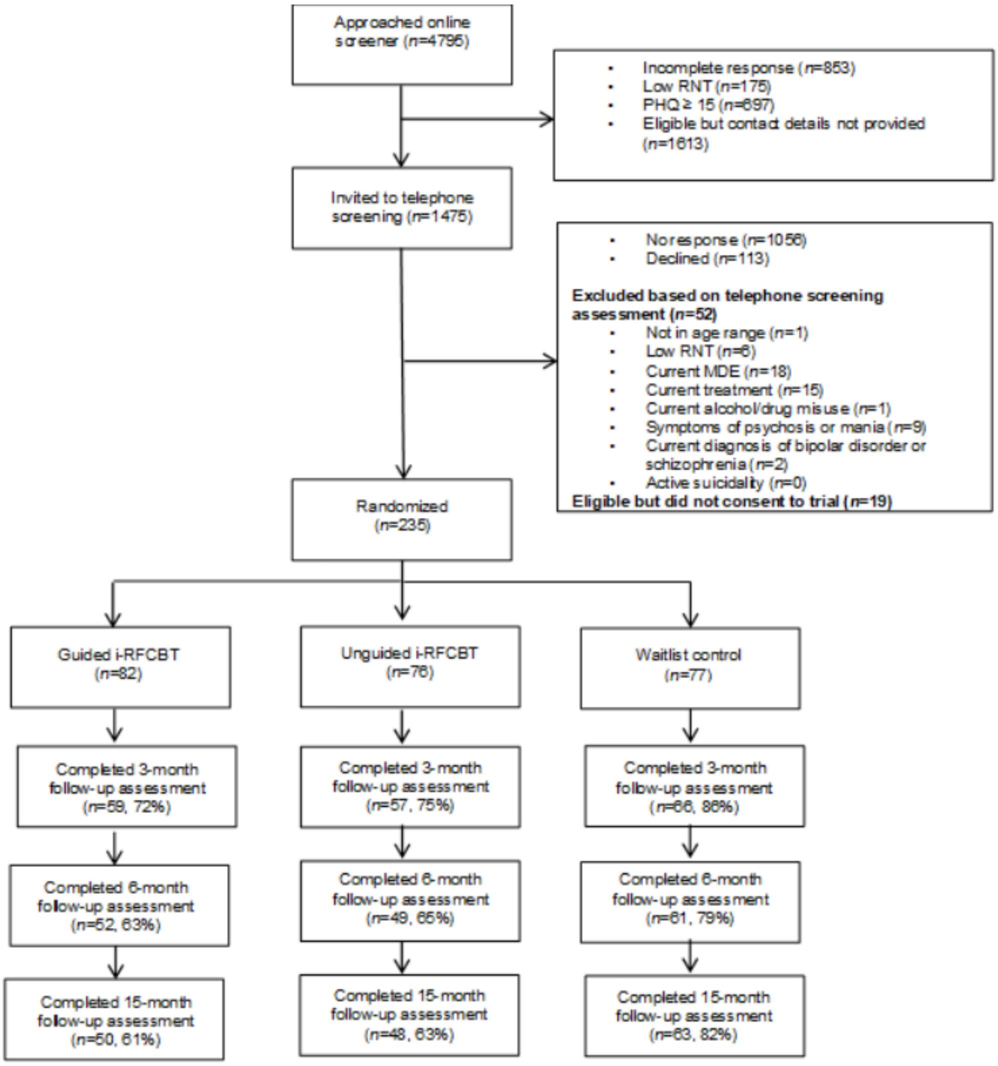
For the primary question comparing guided i-RFCBT to usual care control, assuming a similar hazard ratio of 0.41 for the guided i-RFCBT versus usual care control [27], 75 participants per arm would provide 0.86 power (2-tailed 5% alpha level) to detect this effect. For change in depressive symptoms, the observed effect size was *d*=0.51 [27]. A total of 78 participants per arm would be needed for 80% power to detect a similar effect at the 2-tailed 5% alpha level, allowing for 20% follow-up dropout attrition. With no previous evidence for unguided i-RFCBT, the comparison of unguided i-RFCBT to usual care was conducted as a feasibility study as a first step to conducting a fully powered trial of unguided i-RFCBT. As such, no power or sample size calculations were conducted for the unguided arm. We aimed to recruit the same number (n=78) as the other 2 arms.

The full recruitment procedure is outlined in Cook and Watkins [45]. Briefly, 1834 university departments in the United Kingdom were contacted between November 14, 2013, and December 10, 2014 (1527 contacted twice) and asked to advertise the study. In total, 336 departments confirmed the study was advertised either by email or as a poster. This advertisement contained a link to an open-access screening website. Twitter and Facebook were also used to circulate the advertisement to young people who expressed an interest in the following terms: stress; worry; rumination; mental health; self-esteem; well-being; research; psychology; CBT; and online therapy. In addition, 3 organizations working with young people or in the field of mental health agreed to advertise the study.

A 2-step procedure identified eligible participants. In the first step, an open-access screening website with conditional automated feedback identified potential participants for further screening by screening in those with elevated RNT (>75th percentile) using shortened versions of the PSWQ (4 items, range 4 to 20, cut-off ≥12) and RRS (5 items, range 5 to 20, cut-off ≥10), as developed by Topper et al [48]. A conservative cut-off of 15 on PHQ-8 [49] excluded individuals likely to be experiencing a current MDE. Eligible participants provided contact details as consent to be contacted for further telephone screening.

In the second step, a telephone interview consisted of brief screening questions for alcohol and drug use, symptoms of bipolar disorder and psychosis (Psychosis Screening Questionnaire [PSQ]; [50]), assessment of any relevant past or current treatments, and the SCID-I [40] sections on current and past depressive episodes, dysthymia, and any relevant anxiety disorders and eating disorders. As the primary objective was to investigate the prevention of depression, diagnoses of anxiety disorders and eating disorders were recorded but participants meeting the criteria for any of these disorders were not excluded from the study. The consent to interview was obtained verbally and included providing their general practitioner's (GP) contact details so that appropriate clinical support could be obtained in the event of disclosure of suicidal risk. The interview was audio-recorded, with consent, so that the diagnostic status could be independently checked. The same researcher conducted the baseline and follow-up telephone interview assessments, ensuring continuity of contact between the research team and participants. A total of 254 participants were eligible, of whom 235 returned written informed consent and were randomized to guided i-RFCBT (n=82), unguided i-RFCBT (n=76), or usual care control (n=77). The Consolidated Standards of Reporting Trials diagram (Figure 1) indicates the numbers excluded at baseline for each of the exclusion criteria.

Figure 1. Consolidated Standards of Reporting Trials flowchart. MDE: major depressive episode; PHQ: Patient Health Questionnaire; RNT: repetitive negative thought; i-RFCBT: Web-based rumination-focused cognitive behavioral therapy.



Interventions

Guided Web-Based Rumination-Focused Cognitive Behavioral Therapy

The guided intervention was an English version of i-RFCBT (called MindReSolve), translated and adapted from the version used by Topper et al [27] to include case examples relevant to university students (see Multimedia Appendix 1). RFCBT differs from standard CBT by seeking to change the process of thinking rather than the content of individual thoughts [51]. RFCBT [52] was developed from theoretical models and experimental findings indicating distinct types of repetitive thought (RT) with different consequences [29]: unconstructive RT involves an abstract, evaluative processing mode focused on the meaning and evaluation of events and difficulties, leading to a range of negative consequences such as poorer problem-solving and greater emotional reactivity, relative to constructive RT, which involves concrete, specific, and action-oriented processing [53]. RFCBT therefore aims to shift participants from an abstract and evaluative style to a concrete, specific, and action-oriented style [29], consistent with evidence that concreteness training reduces depression [54].

RNT is also theoretically conceptualized as a mental habit acting as a form of avoidance and maintained by negative reinforcement [55]. RFCBT therefore involves counterconditioning the avoidant ruminative response with more helpful coping strategies and approach behaviors [52]. In practice, this involves the functional analysis of rumination to help users spot triggers for rumination, to distinguish between helpful and unhelpful RT, and to counter condition unhelpful RT with more functional responses through the formulation of contingency *If-Then* plans [52].

The internet treatment was delivered on the internet platform and software owned, programmed, and hosted by Minddistrict [56], accessed by a research licence purchased from Minddistrict by the research team. The specific content of the i-RFCBT intervention was developed and entered into the platform using its content management system by the research team led by Edward Watkins, using the same key intervention principles and techniques as face-to-face RFCBT as described by Watkins [52], adapted for the internet. i-RFCBT contains the same key components as face-to-face RFCBT [52], split into 6 1-hour modules, each in turn split into 3 or 4 sessions consisting of a single Web page, with 1 to 2 weeks recommended per module for practice of the techniques. The content includes psychoeducation, mood diaries, experiential audio exercises, pictures, and video vignettes of university students talking about their own experiences of the intervention. Modules follow the same basic structure: reflection on the previous module; introduction of a new technique; experiential in-session exercises; and plans for implementation. The specific behavior-change techniques are drawn from the following groups in the Behavior Change Technique (BCT) Taxonomy (v1) [57]: goals and planning (goal setting, action planning, review behavior, and behavioral contract), feedback and monitoring (self-monitoring of behavior and outcomes), shaping knowledge (information about antecedents), natural consequences (information about social and environmental consequences and monitoring of emotional consequences), associations (prompts/cues and associative learning), repetition and substitution (behavioral practice/rehearsal, behavior substitution, and habit formation), antecedents (restructuring the physical and social environment and avoiding/reducing cues for the behavior), and self-belief (mental rehearsal of successful performance, focus on past success, and self-talk). The key strategies include coaching participants to spot warning signs for rumination and worry, and then to make If-Then plans in which an alternative strategy is repeatedly practised in response to the warning signs. These strategies include being more active, slowing things down, breaking tasks down, opposite action, relaxation, concrete thinking, becoming absorbed, self-compassion, and assertiveness.

The intervention was accessed individually, for free, on a secure, password-protected website. Access was granted using an email link, inviting the participant to set up a personal account and password. The intervention was supported by qualified clinicians who had received additional specific training in RFCBT. This support consisted of asynchronous written feedback provided by the clinician at the end of each module. Feedback served to highlight positive steps and identified areas to focus on in the following module. Feedback was constrained by template responses for each module, faithful to the RFCBT model, which could be adapted to individual participants' responses. All the content and module order were identical across participants, ensuring treatment fidelity. Each module was self-paced, but the participants were advised to spend 1 to 2 weeks on each and could only access the next module once feedback from the clinician was received, typically within 2 working days. Clinicians monitored log-ons and sent personalized reminder emails if there was no log-on for over a week. The platform also sent an automatic weekly reminder if the platform had not been accessed for a week. Suicidal risk was also monitored using a well-established departmental protocol to determine the level of risk and seek clinical support as appropriate.

Therapists were provided with regular supervision with the developer of RFCBT (EW) to further encourage treatment fidelity. All (100%) feedback reviewed by EW were faithful to the intervention model (over 10% of therapist feedback sampled—a minimum of the 3 initial feedback for each therapist, plus a random subset of later feedback).

Unguided Web-Based Rumination-Focused Cognitive Behavioral Therapy

The content of unguided i-RFCBT was almost identical to guided i-RFCBT with minor adaptations for self-help to include some automatic Web-based conditional feedback addressing common challenges with the exercises. Access was granted via an email link to set up a personal account and password. Participants could then access all modules without restriction but were advised to spend 1 to 2 weeks on each to allow time for practice. Responses were not monitored except for weekly checks of questionnaires to identify and follow up suicidal risk as necessary.

Usual Care Control Condition

Participants in the usual care control condition were permitted to access any other treatments during the study, as necessary. They were also offered access to unguided i-RFCBT at the end of the follow-up period.

Measures

All measures were completed at baseline, 3 months, 6 months, and 15 months unless otherwise stated. Diagnostic interviews for the primary outcome were conducted by telephone, with the option to complete self-report questionnaire measures for secondary outcomes during the telephone interview or request for them to be returned by email/post.

TheSCID-I [40] is a semistructured diagnostic interview for Axis I DSM-IV diagnoses. The SCID-I was used to assess MDE (current and past), anxiety disorders, and eating disorders. The interrater reliability for Axis I diagnoses is fair to excellent, with a mean Kappa of 0.71 [58]. In the event of disclosure of suicidal risk during the diagnostic interview, the researcher followed a well-established departmental protocol to assess risk and obtain clinical support as needed.

The Episodic Life Event Interview, part of the University of California Los Angeles Life Stress Interview [59], assessed the number and impact of stressful events since the previous assessment (for the previous 3 months at baseline). Participants provided a list of events experienced and a subjective rating of stress experienced as a result of the worst event. The original scale ranges from 1 *none* to 5 *severe*. Participants scored 0 if no events were experienced. To aid analysis and interpretation, stress was recoded to collapse 0 *no event* and 1 *event experienced but no stress* into a single *no stress* category. The recoded stress scale therefore ranges from 0 *no stress* to 4 *severe stress*.

The PSWQ [46] is a 16-item self-report questionnaire assessing frequency, intensity, and automaticity of worry (eg, "My worries overwhelm me" and "I know I shouldn't worry about things, but I just can't help it"). It is scored from *1* (not at all typical of me) to *5* (very typical of me), with higher scores indicating higher levels of worry. The internal consistency is high with good test-retest reliability [46]. The PSWQ has also been shown to have good predictive validity for symptoms of anxiety and depression [60].

The RRS [47] is a self-report measure of frequency of ruminative responses to depressed mood, with items relating to the self (eg, “Think about all your shortcomings, failings, faults and mistakes”), one’s symptoms (eg, “Think about how hard it is to concentrate”) and possible causes and consequences of one’s mood (eg, “Go away by yourself and think about why you feel this way”). Items are scored from 1 (almost never) to 4 (almost always). Higher scores indicate higher levels of rumination. The RRS has good internal consistency, moderate test-retest reliability, acceptable convergent validity, and good predictive validity [47,61,62].

The Patient Health Questionnaire (PHQ-9; [63]) is a 9-symptom measure of depressive symptoms. Scores range from 0 to 27, with higher scores indicating greater severity. The PHQ-9 is a reliable and valid measure of severity of depressive symptoms [63].

The GAD Screener (GAD-7; [64]) is a standardized self-report measure of symptoms of anxiety. Scores range from 0 to 21, and higher scores indicate more severe symptoms. Spitzer et al [64] demonstrated good validity and reliability of the GAD-7.

Demographics and Treatment

At baseline, participants were asked if they had any family history of depression (including whom and how recently) and whether they had experienced any physical, sexual, or emotional abuse before the age of 16 years (yes/no questions with no further details requested). Participants were asked to report whether they had received any mental health treatments (medication, therapies, and use of self-help materials) before or during the trial. Timing, duration, and dosage (for medication) were recorded.

Randomization, Allocation Concealment, and Blinding

Independent computer-generated block randomization (block size of 3), stratified by sex (male vs female) and a history of depression (presence or absence of past depressive episodes), was used to allocate participants to the guided i-RFCBT, unguided i-RFCBT, or usual care control in a 1:1:1 ratio. Varying block sizes were not used as the 2 levels of stratification ensured it would be difficult for the researcher to anticipate or determine allocation. A third party not involved in assessing or treating the participants implemented the random allocation sequence and informed the therapist of the condition for each participant. The researcher responsible for recruitment and screening was blind to allocation and unable to influence the order of consents. As a single blind trial, the researcher conducting outcome assessments was blind to allocation. The researcher was not involved with any element of treatment delivery. To preserve researcher blinding, participants were notified of their treatment allocation by a trial therapist. Owing to the nature of the intervention, participants and therapists could not be blinded.

Statistical Analysis for the Phase III Efficacy Trial

Data cleaning followed the protocol set out by Tabachnick and Fidell [65]. Unplanned missing data were handled via multiple imputation (MI). A sensitivity analysis, assuming a variety of MI models (Missing at Random and Missing Not at Random), verified the likely impact of missing data. Auxiliary variables were used to improve the estimation of missing data. Primary analyses were conducted on the intention-to-treat (ITT) sample. Additional analyses assessed the effect of compliance using the Complier Average Causal Effect (CACE) analysis [66]. The CACE analysis provides an unbiased estimate of the benefits of compliance by comparing the compliers in the intervention group to a comparable subgroup of the control group who would have complied had they been offered the intervention. Compliance was defined in the protocol as fully completing at least 4 of 6 modules, that is, accessing all of the sections in each of those modules [45]. Analyses were carried out using statistical software, Stata (StataCorp; version 15.1 [67]).

As a prevention study, the primary outcome was the occurrence and time to onset of any depressive episode by 15-month follow-up. To investigate this, Cox proportional hazard models were fitted to the depression event data, with the diagnosis of an episode of major depression at any point during the follow-up period as the outcome and time to onset measured in weeks from the randomization date. Participants were censored upon measurement dropout or end of study. The Cox proportional hazard model was initially adjusted for both stratification variables: past depression and gender, as they have previously been found to influence likelihood of depression. In addition, as baseline stress was expected to increase the risk of depression and Topper et al [27] controlled for stressful life events, severity of baseline stress was included in the model. To examine the hypothesis that i-RFCBT would be especially beneficial in high ruminators experiencing high stress, we further tested the potential interactions between intervention condition and baseline stress, and intervention condition and history of depression within the Cox proportional hazard analysis.

Secondary outcomes of symptom severity and levels of rumination/worry were examined using mixed model analyses of covariance (ANCOVAs): between group (ITT/CACE) and repeated measures (3- to 15-month follow-ups), controlling for baseline symptom levels.

Feasibility and Acceptability (Quasi-Phase II Pilot Arm)

Feasibility of data collection procedures was assessed by measuring the missing items on clinical outcome measures, number and timing of dropouts, and whether these varied across arms. The acceptability of the intervention was assessed using a behavioral index, measuring the number of Web-based modules completed.

Ethical Approval and Informed Consent

Ethical and professional guidelines were followed at all times, in line with Good Clinical Practice guidelines. Ethical approval was obtained from the Ethics Committee of the School of Psychology, University of Exeter (Ref: 2012/554). Participants returned written informed consent including permission to contact their GP if significant risk was disclosed (see Multimedia Appendix 2).

Results

Demographics

For brevity, baseline demographics for the 3 arms are included in Table 1. As noted earlier, the primary comparison is guided i-RFCBT versus usual care control, with a separate analysis of the feasibility and acceptability of the adjunct unguided i-RFCBT arm.

Survival Analysis: Guided Web-Based Rumination-Focused Cognitive Behavioral Therapy Versus Usual Care Control

A total of 27 participants in the primary comparison of guided i-RFCBT versus the control completed no follow-ups, and no minimum survival time could be estimated, so the ITT survival analyses were conducted on n=132 (guided n=63 and control n=69). Participants with a family history of depression were more likely to be lost to follow-up than those without: $\chi^2_1=3.89$; *P*=.049. No other baseline variables were linked to loss to follow-up: all *t* values on continuous measures were <1.70; all chi-square test values on categorical variables were <1.74; all *P* values were >.09.

There was no overall difference in incidence of depression (*P*=.64): 29% (n=18) of participants receiving guided i-RFCBT and 33% (n=23) of participants receiving usual care experienced an MDE during the follow-up period. A Cox proportional hazard model was conducted, including past depression, gender, and baseline stress as potential predictors of incidence of depression. As the majority (83%) of participants were female and there was no significant effect of gender in predicting depression, this variable was removed from the model, such that the final model controlled for past depression and baseline stress. As expected, history of depression significantly increased risk, with participants with a history of depression over two and a half times more likely to experience an MDE than participants without: hazard ratio (HR) 2.62, 95% CI 1.37 to 5.01; *P*=.004. Baseline stress marginally increased the risk of MDEs: HR 1.40, 95% CI 0.99 to 1.99; *P*=.06. When controlling for both past depression and baseline stress, there was a 34% reduced risk of depression in the guided i-RFCBT condition relative to usual care, although this difference was not significant: HR 0.66, 95% CI 0.35 to 1.25; *P*=.20 (see Figure 2).

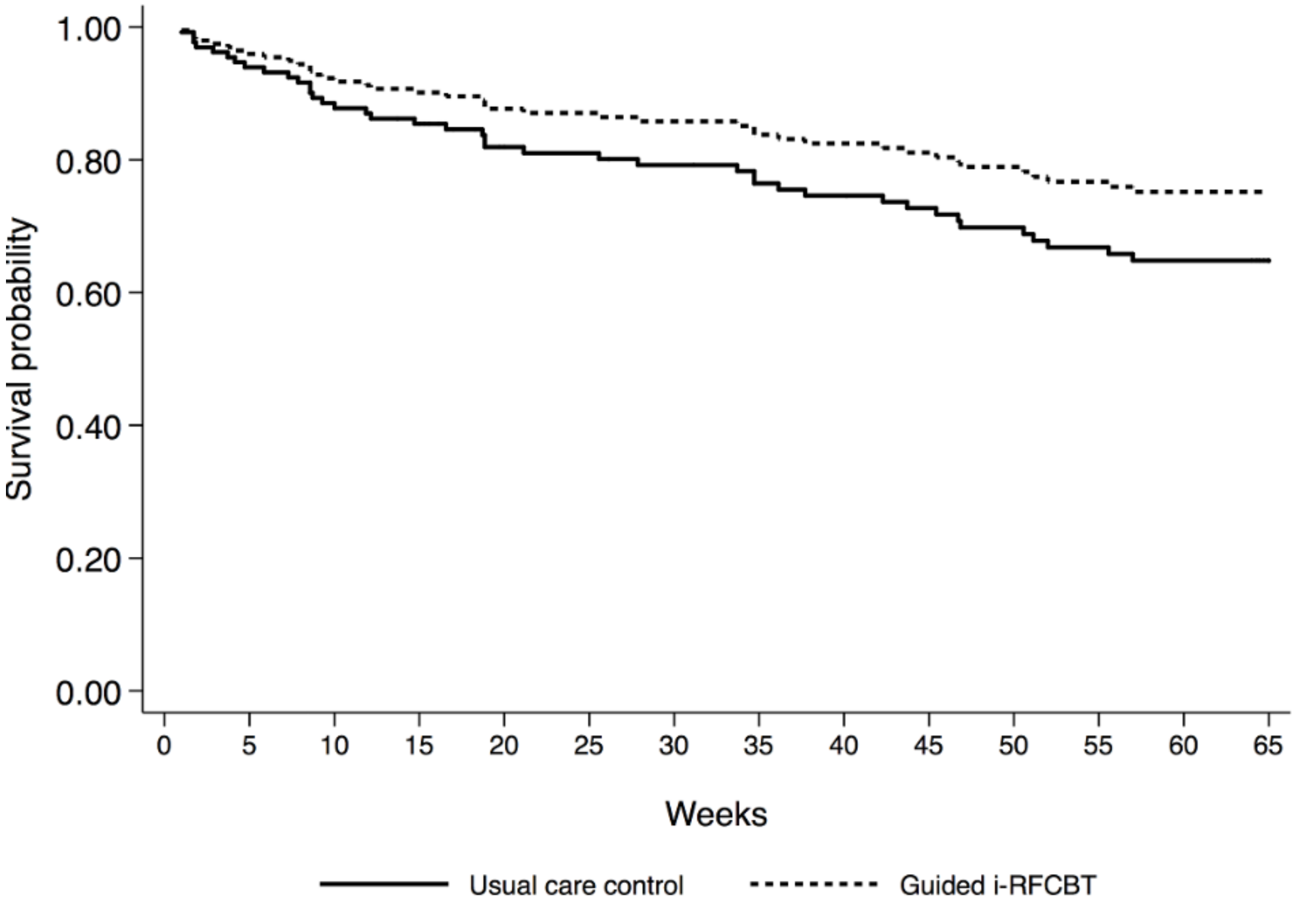
For Cox proportional hazard models including the interactions between intervention condition and baseline stress, and intervention condition and history of depression, there was no differential effect of intervention between first onset (ie, no history of depression) or relapse/recurrence (ie, past history of depression) for incidence of major depression, and this interaction was removed from the final model (HR 0.54, 95% CI 0.15 to 1.94; *P*=.34; guided i-RFCBT: 38.9% first onset; 61.1% relapse vs usual care: 36.4% first onset; 63.6% relapse). Both the effects of past depression (HR 2.52, 95% CI 1.32 to 4.81; *P*=.005) and baseline stress (HR 1.99, 95% CI 1.22 to 3.24; *P*=.006) remained significant. As hypothesized, there was a significant interaction of the intervention condition by baseline stress (HR 0.43, 95% CI 0.21 to 0.87; *P*=.02), indicating a greater benefit of guided i-RFCBT relative to usual care (risk of MDEs decreased by 57%) for undergraduates with higher baseline stress.

Table 1. Baseline characteristics of usual care, guided Web-based rumination-focused cognitive behavioral therapy (i-RFCBT), and unguided i-RFCBT intention-to-treat samples.

Baseline characteristics	Usual care (n=77)	Guided i-RFCBT[a] (n=82)	Unguided i-RFCBT (n=76)
Sex (female), n (%)	64 (83)	68 (83)	64 (84)
Age (years), mean (SD)	20.27 (1.55)	20.43 (1.65)	20.53 (1.30)
Ethnicity (white), n (%)	70 (91)	77 (94)	67 (88)
English mother tongue, n (%)	71 (92)	75 (91)	64 (84)
Previous major depressive episode (yes), n (%)	29 (38)	34 (41)	29 (38)
Received previous mental health treatment (yes), n (%)	38 (49)	38 (46)	31 (41)
Family history of depression (yes), n (%)	39 (51)	42 (51)	33 (43)
Parent with history of depression (yes), n (%)	34 (44)	34 (41)	29 (38)
Reported history of sexual abuse (yes), n (%)	7 (9)	5 (6)	5 (7)
Reported history of physical abuse (yes), n (%)	7 (9)	1 (1)	7 (9)
Reported history of emotional abuse (yes), n (%)	17 (22)	10 (12)	11 (14)
Patient Health Questionnaire-9, mean (SD)	5.6 (4.1)	5.6 (3.2)	5.4 (3.6)
Generalized Anxiety Disorder Screener-7, mean (SD)	6.6 (4.3)	7.3 (4.2)	7.1(4.0)
Penn State Worry Questionnaire, mean (SD)	61.9 (9.0)	62.0 (9.5)	60.3 (10.5)
Ruminative Response Scale, mean (SD)	47.9 (11.1)	49.8 (10.6)	47.2 (10.7)
Stressful events in the past 3 months, mean (SD)	3.6 (2.3)	3.8 (2.4)	3.4 (1.8)
Subjective rating of worst event, mean (SD)	2.20 (1.11)	2.57 (0.96)	2.53 (0.92)

[a] i-RFCBT: Web-based rumination-focused cognitive behavioral therapy.

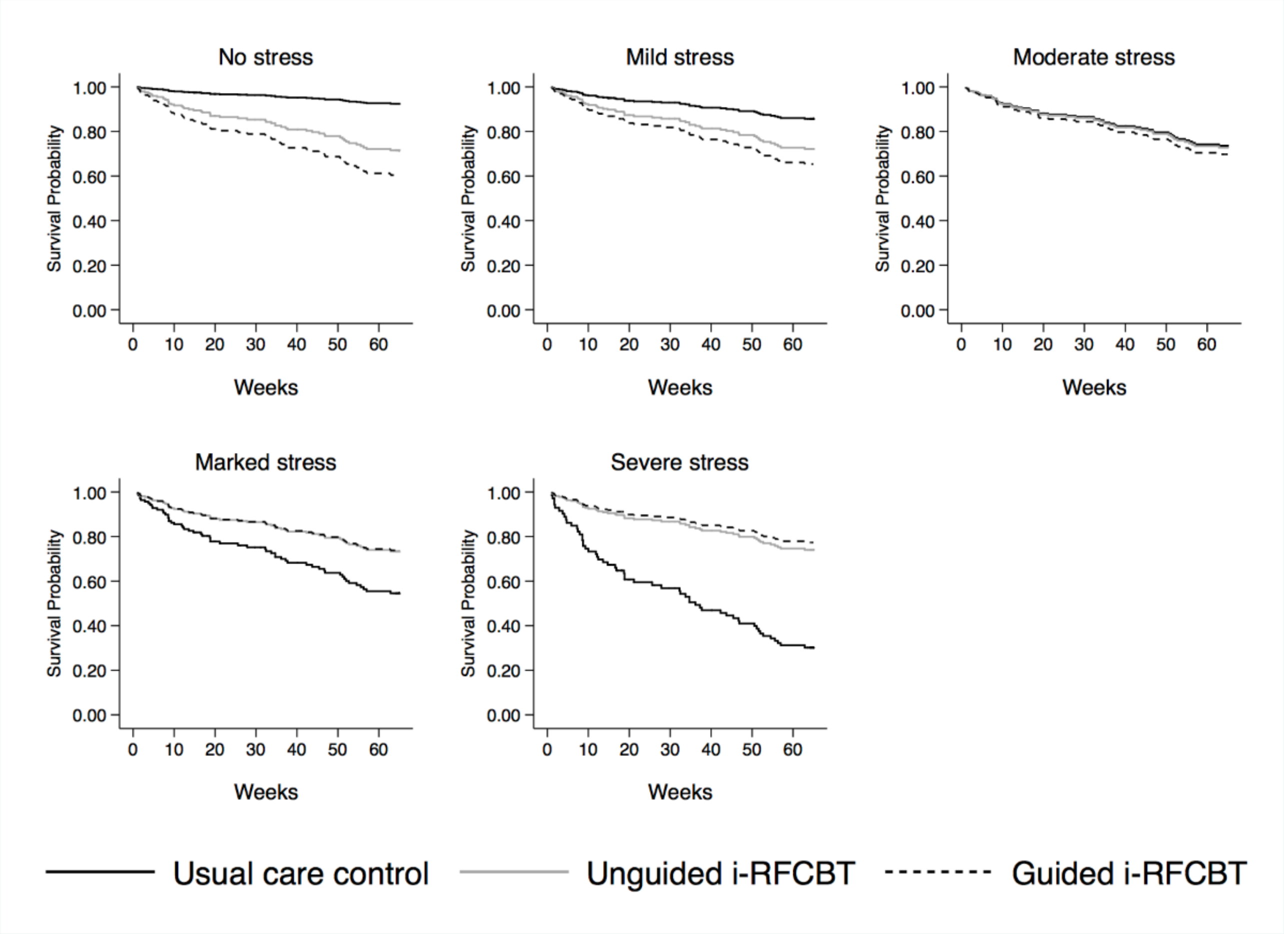
Figure 2. Survival curves for guided Web-based rumination-focused cognitive behavioral therapy (i-RFCBT) and usual care controls, adjusted for past depression and baseline stress.



Plotting the interaction between the intervention group and baseline stress (see Figure 3) suggests that at higher levels of stress, guided i-RFCBT markedly reduces the risk of a depressive episode relative to usual care, with this effect reversing at low levels of stress (albeit in a small number of participants, only 13.1% scoring either 0 or 1).

As further sensitivity analyses, to investigate the effect of compliance on outcomes, we conducted a CACE analysis, using the Loeys and Goetghebur [66] method, which only allows for inclusion of the randomization variable in the model, and using regression-based adjustments to include past depression and baseline stress, which compares compliers in the intervention group to all other participants [68]. The mean completion for guided i-RFCBT was 3.46 (*SD* 2.25) for the full ITT sample (n=159), with 46% (38/82) being compliant by completing at least 4 of the 6 modules. The rates of compliance were higher among those with follow-up outcome data (n=132) as used for the CACE analysis at 60% (38/63). The results of the CACE analyses (see Multimedia Appendix 3) were equivalent to the ITT analysis. We therefore have only reported the primary ITT analysis.

Figure 3. Survival curves for guided Web-based rumination-focused cognitive behavioral therapy (i-RFCBT), unguided i-RFCBT, and usual care controls at each level of baseline stress.



Secondary Analyses on Patient Health Questionnaire, Generalized Anxiety Disorder, Ruminative Response Scale, and Penn State Worry Questionnaire

Baseline adjusted ANCOVAs were conducted for each of the symptom measures, at each of the 3 follow-ups. Estimated means, between-group differences, and CIs are displayed in Table 2 for the case completers and following MIs (50 imputations). For the complete cases, at 3 months, rumination scores were significantly lower for guided i-RFCBT relative to usual care; at 6 months, both worry and depression scores were significantly lower for guided i-RFCBT relative to usual care, and there was no evidence of significant between-group differences at 15 months. Similar patterns were found when using MI to account for differing levels of missing data across the follow-ups.

Table 2. Baseline adjusted symptom measures at 3, 6, and 15 months: guided Web-based rumination-focused cognitive behavioral therapy versus usual care controls.

Timepoint and measure	Guided Web-based rumination-focused cognitive behavioral therapy, mean (95% CI)	Usual care, mean (95% CI)	Between-group difference, difference (95% CI)	Estimated between-group difference after multiple imputations, multiple imputation difference (95% CI)
Follow-up 1 (3 months; n=114)				
PHQ-9[a]	4.75 (3.74 to 5.76)	5.40 (4.48 to 6.33)	−0.65[b] (−2.02 to 0.72)	−0.55[b] (−2.05 to 0.96)
GAD-7[c]	5.58 (4.51 to 6.66)	6.27 (5.28 to 7.25)	−0.69[b] (−2.15 to 0.78)	−0.53[b] (−2.14 to 1.09)
PSWQ[d]	57.27 (54.85 to 59.69)	58.45 (56.23 to 60.67)	−1.18[b] (−4.46 to 2.11)	−0.75[b] (−4.35 to 2.86)
RRS[e]	44.34 (41.66 to 47.02)	48.21 (45.73 to 50.68)	−3.87[f] (−7.53 to −0.21)	−3.69[b] (−8.01 to 0.63)
Follow-up 2 (6 months; n=105)				
PHQ-9	3.70 (2.48 to 4.92)	5.52 (4.42 to 6.62)	−1.82[f] (−3.46 to −0.18)	−1.97[f] (−3.87 to −.063)
GAD-7	4.72 (3.44 to 5.99)	6.06 (4.91 to 7.20)	−1.34[b] (−3.05 to 0.38)	−1.15[b] (−3.16 to 0.85)
PSWQ	54.83 (52.19 to 57.48)	58.41 (56.03 to 60.79)	−3.58[f] (−7.14 to −0.02)	−2.71[b] (−6.68 to 1.25)
RRS	41.74 (38.15 to 45.34)	46.35 (43.12 to 49.58)	−4.60[b] (−9.47 to 0.26)	−3.98[b] (−9.48 to 1.52)
Follow-up 3 (15 months; n=108)				
PHQ-9	4.47 (3.23 to 5.71)	4.82 (3.73 to 5.91)	−0.35[b] (−2.00 to 1.30)	−0.38[b] (−2.30 to 1.55)
GAD-7	4.42 (3.16 to 5.68)	5.73 (4.62 to 6.83)	−1.31[b] (−2.99 to 0.38)	−1.10[b] (−2.10 to 0.80)
PSWQ	54.81 (51.71 to 57.91)	58.11 (55.39 to 60.84)	−3.30[b] (−7.43 to 0.82)	−1.74[b] (−6.53 to 3.06)
RRS	46.15 (42.59 to 49.72)	44.65 (41.53 to 47.78)	1.50[b] (−3.28 to 6.28)	1.16[b] (−3.99 to 6.31)

[a] PHQ-9: Patient Health Questionnaire.

[b] Not significant.

[c] GAD-7: Generalized Anxiety Disorder Screener.

[d] PSWQ: Penn State Worry Questionnaire.

[e] n=115 for Ruminative Response Scale (RRS) at 3 months owing to one partially completed questionnaire set.

[f] *P*<.05.

Retention, Acceptability and Effect Sizes of Unguided Web-Based Rumination-Focused Cognitive Behavioral Therapy

A total of 19 (25%) unguided participants did not complete any follow-up assessments. Participants were significantly more likely to be lost to follow-up in unguided i-RFCBT than in usual care: χ^2_1=4.53; *P*=.03.

Owing to the exploratory nature of the unguided version of i-RFCBT, no formal CACE analysis of compliance was undertaken for the unguided intervention. In the full ITT sample, unguided participants completed an average of 2.66 modules (*SD* 2.35). The rates of compliance (38% unguided) were not significantly different from guided i-RFCBT (χ^2_1=1.08; *P*=.30). For the unguided intervention, participants logged in an average of 6.25 times (*SD* 5.21) and accessed the intervention over an average period of 114.92 days (*SD* 105.51). The median (interquartile range) was 87 days (22-173). Guided participants logged in an average of 7.97 times (*SD* 5.65) and accessed the intervention over an average period of 110.13 days (*SD* 108.44). The median (interquartile range) was 67 days (38-156).

Estimates of Hazard Ratios for Unguided Web-Based Rumination-Focused Cognitive Behavioral Therapy Versus Usual Care

No formal significance analyses were undertaken, but hazard ratios and CIs were estimated relative to usual care. Using a Cox proportional hazard model including past depression and baseline stress, unguided i-RFCBT showed a 36% reduced risk of developing a depressive episode relative to controls: HR 0.64, 95% CI 0.33 to 1.24. A similar interaction between the intervention and baseline stress was found as for guided i-RFCBT (HR 0.48, 95% CI 0.23 to 1.00), such that unguided i-RFCBT had larger effect sizes for undergraduates with moderate-to-severe levels of baseline stress (see Figure 3).

Table 3. Baseline adjusted symptom measures at 3, 6, and 15 months: unguided Web-based rumination-focused cognitive behavioral therapy versus usual care controls.

Timepoint and measure	Unguided Web-based rumination-focused cognitive behavioral therapy, mean (95% CI)	Usual care, mean (95% CI)	Between-group difference, difference (95% CI)	Estimated between-group difference after multiple imputations, multiple imputation difference (95% CI)
Follow-up 1 (3 months; n=116)				
PHQ-9[a]	4.02 (3.07 to 4.96)	5.21 (4.33 to 6.10)	−1.20 (−2.49 to 0.10)	−1.18 (−2.65 to 0.28)
GAD-7[b]	4.94 (3.90 to 5.99)	5.98 (5.01 to 6.96)	−1.04 (−2.47 to 0.39)	−1.06 (−2.60 to 0.49)
PSWQ[c]	55.77 (53.26 to 58.28)	57.60 (55.26 to 59.95)	−1.84 (−5.28 to 1.61)	−1.35 (−4.87 to 2.17)
RRS[d]	44.47 (41.89 to 47.06)	47.01 (44.60 to 49.42)	−2.54 (−6.08 to 1.01)	−2.42 (−6.19 to 1.34)
Follow-up 2 (6 months; n=104)				
PHQ-9	4.38 (3.20 to 5.56)	5.35 (4.30 to 6.40)	−0.97 (−2.56 to 0.61)	−1.04 (−2.89 to −0.81)
GAD-7	4.20 (2.96 to 5.44)	5.93 (4.83 to 7.03)	−1.73 (−3.38 to −0.07)	−2.09 (−3.92 to −0.28)
PSWQ	54.51 (51.60 to 57.42)	58.06 (55.47 to 60.65)	−3.55 (−7.46 to 0.36)	−3.35 (−7.36 to 0.67)
RRS	41.27 (38.07 to 44.47)	45.20 (42.35 to 48.04)	−3.93 (−8.22 to 0.37)	−4.12 (−8.94 to 0.69)
Follow-up 3 (15 months; n=107)				
PHQ-9	4.20 (3.00 to 5.40)	4.69 (3.64 to 5.73)	−0.49 (−2.08 to 1.11)	−0.92 (−2.61 to 0.77)
GAD-7	4.49 (3.28 to 5.70)	5.52 (4.46 to 6.57)	−1.03 (−2.63 to 0.58)	−1.36 (−3.23 to 0.52)
PSWQ	53.78 (50.75 to 56.81)	57.56 (54.93 to 60.19)	−3.78 (−7.79 to 0.24)	−4.34 (−8.57 to −0.09)
RRS	42.07 (38.59 to 45.54)	43.85 (40.84 to 46.86)	−1.78 (−6.40 to 2.83)	−2.61 (−7.93 to 2.71)

[a] PHQ-9: Patient Health Questionnaire.

[b] GAD-7: Generalized Anxiety Disorder Screener.

[c] PSWQ: Penn State Worry Questionnaire.

[d] RRS: Ruminative Response Scale.

Between-group differences for unguided i-RFCBT versus usual care were estimated with baseline adjusted ANCOVAs for both case completers and using multiple imputations (50 imputations). Estimated means and CIs are displayed in Table 3. Owing to the exploratory nature of this comparison, significance testing was not conducted. Patterns of change and CIs indicate similar symptom changes to those found in the guided i-RFCBT versus usual care control ANCOVAs.

Discussion

Principal Findings and Comparison With Previous Work

The main aim was to test if guided i-RFCBT could be effective in preventing depression in undergraduate students in the United Kingdom over 1-year follow-up. When controlling for both past depression and baseline stress, guided i-RFCBT reduced the risk of experiencing an MDE by 34% relative to usual care. Although this effect size was not significant and smaller than that found by Topper et al [27], it is consistent with the wider prevention literature, which reports an average reduction in incidence of 21% [10] and a 28% (IRR=.72) reduction in incidence relative to controls for selective, predominantly CBT, interventions [11]. It may be that this study was underpowered to detect a main preventive effect of i-RFCBT as it used a larger effect size estimate derived from Topper et al [27].

As hypothesized, guided i-RFCBT was significant at preventing the onset of MDEs in high-risk undergraduates relative to usual care when they experienced moderate or above levels of baseline stress, with a hazard ratio of 0.43 when moderated by baseline stress. This is consistent with theoretical models of rumination and the RFCBT treatment approach. The tendency to ruminate about difficulties or low mood is more likely to increase the risk for depression in the context of stressful events, which activates that habitual tendency and provides subject matter to ruminate about. Even someone with a habitual tendency to ruminate is less likely to have frequent rumination in the absence of any difficulties. Furthermore, one key mechanism by which rumination is proposed to increase vulnerability to depression is by exacerbating and prolonging negative affect and distress [29,30]: rumination does not have deleterious effects in the absence of negative mood, and it is thus the confluence of stressful events that lower mood and the tendency to ruminate that particularly confer the risk for depression [32]. This pattern of results suggests a partial replication of Topper et al [27], by indicating that guided i-RFCBT may be a helpful preventive intervention for university students with high levels of rumination and worry, who also experience at least moderate levels of stress.

We note that the observed interaction between i-RFCBT and baseline stress could also be interpreted as indicating that

RFCBT is unhelpful compared with usual care, for users who are experiencing little to no current stress. However, given the small number of participants who reported low levels of baseline stress, this reversal is based on low power and needs to be treated with caution.

The findings on the symptom measures suggest that guided i-RFCBT was effective in the short-to-medium term, by reducing rumination, worry, and symptoms of depression at 3 and 6 months relative to usual care, but that these improvements were not sustained over the longer term. Watkins and Nolen-Hoeksema [55] hypothesized that rumination could be conceptualized as a learnt habit, triggered by particular cues such as low mood. Within this analysis, successful long-term reduction of the ruminative habit requires extensive repetition and rehearsal of alternative more adaptive responses to the triggers for rumination. It may be that i-RFCBT was too brief or that participants did not practise enough to produce long-term change in the ruminative habit. It may also be that further engagement and booster sessions some months after the initial intervention phase would enhance the longer-term effects of the intervention [69]. These could take the form of explicit reminders to practice techniques (flashcards and text/email reminders; [70]) or increasing the generalizability of the new, more helpful techniques across a broader range of contexts [70].

One possible reason for the difference in findings between this study and Topper et al [27] is the means of assessing onset of depression: Reducing Stress and Preventing Depression (RESPOND) used structured clinical interviews, whereas Topper et al [27] used cut-offs on self-report questionnaires, which may overestimate incidence. Another potential explanation is the different samples. Although their sample included university students, the average age was 17.5 years, compared to 20.4 years in RESPOND. Cases of depression begin to rise steeply from the age of 14 [13], so it may be that the developmental risks during mid to late adolescence differ from those in university students and either that i-RFCBT was more efficacious in younger participants or the base rate was higher in the younger sample, increasing the power of the trial.

Compliance rates and the pattern of findings and preliminary effect sizes and CIs for unguided i-RFCBT were similar to those for guided i-RFCBT. These findings are in contrast to the literature on Web-based acute treatment for depression, which generally demonstrates larger effect sizes for guided interventions relative to unguided [41-43]. This benefit of therapist guidance has also previously been found for indicated preventive interventions in university students [71]. As a preliminary finding from a feasibility pilot, further large-scale trials are needed to confirm whether this potential equivalence between unguided and guided i-RFCBT is robust. We speculate that only selecting high-ruminating participants for the trial meant that i-RFCBT was highly relevant and engaging to participants, thus ameliorating the relative benefits of guidance on treatment motivation and completion.

Given the need for widespread dissemination of preventive interventions, an efficacious unguided intervention would be valuable, even if it had somewhat reduced effect sizes relative to the guided version, because it would not be constrained by therapist numbers or availability unlike a guided treatment and could be enormously scaled up to increase accessibility [44]. In addition, unguided interventions may benefit a previously unreached population as many university students do not seek professional help for mental health difficulties [22,23] and could therefore be more attracted to self-help interventions. In support of this, an unguided preventive intervention for students with elevated distress ratings reduced depressive symptoms relative to usual care at 2 months follow-up [72], with two-thirds of the trial completers reporting an unmet need (needing help but not seeking it) in the previous year. As i-RFCBT targets worry and rumination, rather than focusing on depression, this may further attract those who prefer self-help to manage their symptoms as worry is a common experience without the perceived stigma of mental illness [27]. These initial findings on the acceptability and effect sizes of the unguided version provide some promise in terms of potential benefits and suggests the value of further studies to formally test unguided i-RFCBT as a preventive intervention.

Despite the need for larger-scale trials to test the robustness of these findings, several strengths of RESPOND are identified. First, the RESPOND trial addressed some of the methodological limitations of the Topper et al [27] trial by including diagnostic interviewing. This allowed for retrospective diagnoses, capturing any episodes occurring between follow-up interviews, as well as baseline history of depression to assess the effect of previous history on risk of a further MDE.

The use of Web-based and telephone-based measures allowed for recruitment throughout the United Kingdom, with participants from a wide range of university departments and geographical locations, increasing the generalizability of the findings within this demographic. The target sample size was achieved through this recruitment strategy and this would therefore be a suitable approach for a larger scale trial of i-RFCBT.

Limitations

There were several limitations to the study. First, the sample was disproportionately female, limiting the generalizability of the findings. However, females consistently report higher levels of rumination [73] and higher levels of depression, so a trial selecting on this basis will necessarily attract more female participants.

Second, despite a successful recruitment strategy, there was a considerable proportion of missing data at follow-up, particularly in the intervention conditions (albeit in the context of planning the sample size for 20% drop-out attrition). In addition, follow-up assessments were sometimes incomplete as participants did not always return the questionnaires after the follow-up interview, despite reminders being sent. Future trials should therefore further emphasize to participants the importance of follow-up data during the baseline assessment and ensure all measures are completed during the interview. Finally, common to many electronic mental health trials, the participants were not blind to the treatment condition, and, as such, the results could have been influenced by response bias and expectancy effects.

Conclusions

Despite these limitations, taken together, the findings from the Topper et al [27] trial and from this RESPOND trial suggest that i-RFCBT is an effective and acceptable intervention for preventing depression in adolescents and undergraduates experiencing high levels of rumination and worry. This demonstrates the value in targeting a preventive intervention at identified risk factors. This intervention may be particularly effective in individuals experiencing high levels of stress. The initial findings relating to unguided i-RFCBT suggest this may be efficacious in preventing depression, which, if shown to be robust in a fully powered trial, would have significant implications for the scalability of i-RFCBT.

Acknowledgments

This work was supported by the University of Exeter, United Kingdom, through a matched funded PhD studentship to a successful Wellcome Trust Capital Bid. The funders had no involvement in the research or the submission of this manuscript for publication.

The authors are grateful to Jenny Cadman, Dr Yusuke Umegaki (therapists for guided therapy and assisting with randomization), and Rosario Melero (assisting with recruitment and therapist for guided therapy).

Conflicts of Interest

EW developed the original RFCBT intervention. The other authors declare they have no competing interests.

Multimedia Appendix 1

Screenshots of Web-based rumination-focused cognitive behavioral therapy.

[PDF File (Adobe PDF File), 766KB - jmir_v21i5e11349_app1.pdf]

Multimedia Appendix 2

Information sheet and consent form.

[PDF File (Adobe PDF File), 79KB - jmir_v21i5e11349_app2.pdf]

Multimedia Appendix 3

Complier average causal effect analysis.

[PDF File (Adobe PDF File), 41KB - jmir_v21i5e11349_app3.pdf]

Multimedia Appendix 4

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 524KB - jmir_v21i5e11349_app4.pdf]

References

1. World Health Organization. Geneva, Switzerland: WHO; 2017. Depression and other common mental disorders: global health estimates URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1 [accessed 2018-04-23] [WebCite Cache ID 6ytV1RlIN]
2. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
3. Griffiths CA, Griffiths LJ. Recovery and reliable change rates for patients scoring severe on depression, anxiety or impaired functioning in a psychological therapies service: IAPT. Ment Health Rev J 2015 Mar 9;20(1):28-35. [doi: 10.1108/MHRJ-06-2014-0022]
4. Judd LL. The clinical course of unipolar major depressive disorders. Arch Gen Psychiatry 1997 Nov;54(11):989-991. [Medline: 9366654]
5. Andrews G, Issakidis C, Sanderson K, Corry J, Lapsley H. Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. Br J Psychiatry 2004 Jun;184:526-533 [FREE Full text] [Medline: 15172947]
6. Topper M, Emmelkamp P, Ehring T. Improving prevention of depression and anxiety disorders: Repetitive negative thinking as a promising target. Appl Prev Psychol 2010 Jun;14(1-4):57-71. [doi: 10.1016/j.appsy.2012.03.001]
7. Horowitz JL, Garber J. The prevention of depressive symptoms in children and adolescents: a meta-analytic review. J Consult Clin Psychol 2006 Jun;74(3):401-415. [doi: 10.1037/0022-006X.74.3.401] [Medline: 16822098]

8. Merry SN, Hetrick SE, Cox GR, Brudevold-Iversen T, Bir JJ, McDowell H. Psychological and educational interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2011 Dec 7(12):CD003380. [doi: 10.1002/14651858.CD003380.pub3] [Medline: 22161377]
9. Stice E, Shaw H, Bohon C, Marti CN, Rohde P. A meta-analytic review of depression prevention programs for children and adolescents: factors that predict magnitude of intervention effects. J Consult Clin Psychol 2009 Jun;77(3):486-503 [FREE Full text] [doi: 10.1037/a0015168] [Medline: 19485590]
10. van Zoonen K, Buntrock C, Ebert DD, Smit F, Reynolds CF, Beekman AT, et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. Int J Epidemiol 2014 Apr;43(2):318-329 [FREE Full text] [doi: 10.1093/ije/dyt175] [Medline: 24760873]
11. Cuijpers P, van Straten A, Smit F, Mihalopoulos C, Beekman A. Preventing the onset of depressive disorders: a meta-analytic review of psychological interventions. Am J Psychiatry 2008 Oct;165(10):1272-1280. [doi: 10.1176/appi.ajp.2008.07091422] [Medline: 18765483]
12. Muñoz RF, Cuijpers P, Smit F, Barrera AZ, Leykin Y. Prevention of major depression. Annu Rev Clin Psychol 2010;6:181-212. [doi: 10.1146/annurev-clinpsy-033109-132040] [Medline: 20192789]
13. Hankin BL, Young JF, Abela JR, Smolen A, Jenness JL, Gulley LD, et al. Depression from childhood into late adolescence: influence of gender, development, genetic susceptibility, and peer stress. J Abnorm Psychol 2015 Nov;124(4):803-816 [FREE Full text] [doi: 10.1037/abn0000089] [Medline: 26595469]
14. McManus S, Bebbington P, Jenkins R, Brugha T. National Health Service. Leeds, UK: NHS Digital; 2016. Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014 URL: https://files.digital.nhs.uk/pdf/q/3/mental_health_and_wellbeing_in_england_full_report.pdf [accessed 2018-04-23] [WebCite Cache ID 6ytaQR4yA]
15. Pine DS, Cohen P, Gurley D, Brook J, Ma Y. The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. Arch Gen Psychiatry 1998 Jan;55(1):56-64. [Medline: 9435761]
16. Hysenbegasi A, Hass SL, Rowland CR. The impact of depression on the academic productivity of university students. J Ment Health Policy Econ 2005 Sep;8(3):145-151. [Medline: 16278502]
17. Ibrahim AK, Kelly SJ, Adams CE, Glazebrook C. A systematic review of studies of depression prevalence in university students. J Psychiatr Res 2013 Mar;47(3):391-400. [doi: 10.1016/j.jpsychires.2012.11.015] [Medline: 23260171]
18. Aalto-Setälä T, Marttunen M, Tuulio-Henriksson A, Poikolainen K, Lönnqvist J. One-month prevalence of depression and other DSM-IV disorders among young adults. Psychol Med 2001 Jul;31(5):791-801. [Medline: 11459377]
19. Newman DL, Moffitt TE, Caspi A, Magdol L, Silva PA, Stanton WR. Psychiatric disorder in a birth cohort of young adults: prevalence, comorbidity, clinical significance, and new case incidence from ages 11 to 21. J Consult Clin Psychol 1996 Jun;64(3):552-562. [Medline: 8698949]
20. National Institute of Mental Health. Bethesda, MD: NIMH; 2003. Depression and college students URL: https://www.nimh.nih.gov/health/publications/depression-and-college-students/index.shtml [accessed 2018-04-23] [WebCite Cache ID 6ytaW7SOz]
21. Eisenberg D, Golberstein E, Hunt JB. Mental health and academic success in college. B E J Econom Anal Policy 2009;9(1):40. [doi: 10.2202/1935-1682.2191]
22. Eisenberg D, Gollust SE, Golberstein E, Hefner JL. Prevalence and correlates of depression, anxiety, and suicidality among university students. Am J Orthopsychiatry 2007 Oct;77(4):534-542. [doi: 10.1037/0002-9432.77.4.534] [Medline: 18194033]
23. National Union of Students. Lost in transition? – provision of mental health support for 16-21 year olds moving to further and higher education. 2015 Presented at: All Party Parliamentary Group on Students; December 14, 2015; London, UK URL: https://www.nus.org.uk/PageFiles/2161132/APPG%20on%20Students%20-%20December%20-%20Speaker%20briefing%20-%20Mental%20Health%20(002).pdf
24. Andersson G. The promise and pitfalls of the internet for cognitive behavioral therapy. BMC Med 2010;8:82 [FREE Full text] [doi: 10.1186/1741-7015-8-82] [Medline: 21138574]
25. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
26. 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;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
27. Topper M, Emmelkamp PM, Watkins E, Ehring T. Prevention of anxiety disorders and depression by targeting excessive worry and rumination in adolescents and young adults: a randomized controlled trial. Behav Res Ther 2017 Dec;90:123-136. [doi: 10.1016/j.brat.2016.12.015] [Medline: 28049069]
28. Watkins ER, Mullan E, Wingrove J, Rimes K, Steiner H, Bathurst N, et al. Rumination-focused cognitive-behavioural therapy for residual depression: phase II randomised controlled trial. Br J Psychiatry 2011 Oct;199(4):317-322 [FREE Full text] [doi: 10.1192/bjp.bp.110.090282] [Medline: 21778171]
29. Watkins ER. Constructive and unconstructive repetitive thought. Psychol Bull 2008 Mar;134(2):163-206 [FREE Full text] [doi: 10.1037/0033-2909.134.2.163] [Medline: 18298268]

30. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci 2008 Sep;3(5):400-424. [doi: 10.1111/j.1745-6924.2008.00088.x] [Medline: 26158958]
31. Morrison R, O'Connor RC. Predicting psychological distress in college students: the role of rumination and stress. J Clin Psychol 2005 Apr;61(4):447-460. [doi: 10.1002/jclp.20021] [Medline: 15468342]
32. Ciesla JA, Felton JW, Roberts JE. Testing the cognitive catalyst model of depression: does rumination amplify the impact of cognitive diatheses in response to stress? Cogn Emot 2011 Dec;25(8):1349-1357 [FREE Full text] [doi: 10.1080/02699931.2010.543330] [Medline: 21432645]
33. Connolly SL, Alloy LB. Rumination interacts with life stress to predict depressive symptoms: an ecological momentary assessment study. Behav Res Ther 2017 Oct;97:86-95. [doi: 10.1016/j.brat.2017.07.006] [Medline: 28734979]
34. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry 2011 Jun;42(2):225-232 [FREE Full text] [doi: 10.1016/j.jbtep.2010.12.003] [Medline: 21315886]
35. Nolen-Hoeksema S, Watkins ER. A heuristic for developing transdiagnostic models of psychopathology: explaining multifinality and divergent trajectories. Perspect Psychol Sci 2011 Nov;6(6):589-609. [doi: 10.1177/1745691611419672] [Medline: 26168379]
36. Abela JRZ, Brozina K, Haigh EP. An examination of the response styles theory of depression in third- and seventh-grade children: a short-term longitudinal study. J Abnorm Child Psychol 2002 Oct;30(5):515-527. [Medline: 12403154]
37. McLaughlin KA, Hatzenbuehler ML. Mechanisms linking stressful life events and mental health problems in a prospective, community-based sample of adolescents. J Adolesc Health 2009 Feb;44(2):153-160 [FREE Full text] [doi: 10.1016/j.jadohealth.2008.06.019] [Medline: 19167664]
38. Nolen-Hoeksema S, Stice E, Wade E, Bohon C. Reciprocal relations between rumination and bulimic, substance abuse, and depressive symptoms in female adolescents. J Abnorm Psychol 2007 Feb;116(1):198-207. [doi: 10.1037/0021-843X.116.1.198] [Medline: 17324030]
39. Schwartz JA, Koenig LJ. Response styles and negative affect among adolescents. Cogn Ther Res 1996 Feb;20(1):13-36. [doi: 10.1007/BF02229241]
40. First M, Spitzer R, Gibbon M, Williams J. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN). New York: Biometrics Research, New York State Psychiatric Institute; 2002.
41. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
42. 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 Sep;37(9):1217-1228. [doi: 10.1017/S0033291707000062] [Medline: 17306044]
43. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
44. Cuijpers P, Riper H, Andersson G. Internet-based treatment of depression. Curr Opin Psychol 2015 Aug;4:131-135. [doi: 10.1016/j.copsyc.2014.12.026]
45. Cook L, Watkins E. Guided, internet-based, rumination-focused cognitive behavioural therapy (i-RFCBT) versus a no-intervention control to prevent depression in high-ruminating young adults, along with an adjunct assessment of the feasibility of unguided i-RFCBT, in the REducing Stress and Preventing Depression trial (RESPOND): study protocol for a phase III randomised controlled trial. Trials 2016 Jan 04;17:1 [FREE Full text] [doi: 10.1186/s13063-015-1128-9] [Medline: 26725476]
46. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [Medline: 2076086]
47. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. J Pers Soc Psychol 1991 Jul;61(1):115-121. [Medline: 1890582]
48. Topper M, Emmelkamp PM, Watkins E, Ehring T. Development and assessment of brief versions of the Penn State Worry Questionnaire and the Ruminative Response Scale. Br J Clin Psychol 2014 Nov;53(4):402-421. [doi: 10.1111/bjc.12052] [Medline: 24799256]
49. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
50. Bebbington PE, Nyani T. The Psychosis Screening Questionnaire. Int J Methods Psychiatr Res 1995;5:11-19 [FREE Full text]
51. Watkins E, Scott J, Wingrove J, Rimes K, Bathurst N, Steiner H, et al. Rumination-focused cognitive behaviour therapy for residual depression: a case series. Behav Res Ther 2007 Sep;45(9):2144-2154. [doi: 10.1016/j.brat.2006.09.018] [Medline: 17367751]
52. Watkins E. Rumination-Focused Cognitive-Behavioral Therapy for Depression. New York: The Guilford Press; 2016.

53. Watkins E, Moberly NJ, Moulds ML. Processing mode causally influences emotional reactivity: distinct effects of abstract versus concrete construal on emotional response. Emotion 2008 Jun;8(3):364-378 [FREE Full text] [doi: 10.1037/1528-3542.8.3.364] [Medline: 18540752]
54. Watkins ER, Taylor RS, Byng R, Baeyens C, Read R, Pearson K, et al. Guided self-help concreteness training as an intervention for major depression in primary care: a Phase II randomized controlled trial. Psychol Med 2012 Jul;42(7):1359-1371 [FREE Full text] [doi: 10.1017/S0033291711002480] [Medline: 22085757]
55. Watkins ER, Nolen-Hoeksema S. A habit-goal framework of depressive rumination. J Abnorm Psychol 2014 Feb;123(1):24-34. [doi: 10.1037/a0035540] [Medline: 24661156]
56. Minddistrict. URL: https://www.minddistrict.com/ [accessed 2018-06-19] [WebCite Cache ID 70IFG4ZML]
57. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
58. Lobbestael J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). Clin Psychol Psychother 2011;18(1):75-79. [doi: 10.1002/cpp.693] [Medline: 20309842]
59. Hammen C. Episodic Life Event Interview: UCLA Life Stress Interview. Los Angeles, CA: UCLA Department of Psychology; 2004.
60. Hong RY. Worry and rumination: differential associations with anxious and depressive symptoms and coping behavior. Behav Res Ther 2007 Feb;45(2):277-290. [doi: 10.1016/j.brat.2006.03.006] [Medline: 16635479]
61. Butler L, Nolen-Hoeksema S. Gender differences in responses to depressed mood in a college sample. Sex Roles 1994 Mar;30(5-6):331-346. [doi: 10.1007/BF01420597]
62. Nolen-Hoeksema S. The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. J Abnorm Psychol 2000 Aug;109(3):504-511. [Medline: 11016119]
63. 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] [Medline: 11556941]
64. 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]
65. Tabachnick B, Fidell L. Using Multivariate Statistics (4th Edition). Needham Heights, MA: Allyn & Bacon; 2001.
66. Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics 2003 Mar;59(1):100-105. [Medline: 12762446]
67. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
68. Odondi L, McNamee R. Performance of statistical methods for analysing survival data in the presence of non-random compliance. Stat Med 2010 Dec 20;29(29):2994-3003. [doi: 10.1002/sim.4070] [Medline: 20963732]
69. van Voorhees BW, Mahoney N, Mazo R, Barrera AZ, Siemer CP, Gladstone TR, et al. Internet-based depression prevention over the life course: a call for behavioral vaccines. Psychiatr Clin North Am 2011 Mar;34(1):167-183 [FREE Full text] [doi: 10.1016/j.psc.2010.11.002] [Medline: 21333846]
70. Hertel PT, Mathews A. Cognitive bias modification: past perspectives, current findings, and future applications. Perspect Psychol Sci 2011 Nov;6(6):521-536. [doi: 10.1177/1745691611421205] [Medline: 26168375]
71. Conley CS, Durlak JA, Shapiro JB, Kirsch AC, Zahniser E. A meta-analysis of the impact of universal and indicated preventive technology-delivered interventions for higher education students. Prev Sci 2016 Dec;17(6):659-678. [doi: 10.1007/s11121-016-0662-3] [Medline: 27225631]
72. Lintvedt OK, Griffiths KM, Sørensen K, Østvik AR, Wang CE, Eisemann M, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial. Clin Psychol Psychother 2013;20(1):10-27. [doi: 10.1002/cpp.770] [Medline: 21887811]
73. Robichaud M, Dugas MJ, Conway M. Gender differences in worry and associated cognitive-behavioral variables. J Anxiety Disord 2003;17(5):501-516. [Medline: 12941362]

Abbreviations

ANCOVA: analysis of covariance
CACE: complier average causal effect
CBT: cognitive behavioral therapy
CMD: common mental health disorder
GAD: generalized anxiety disorder
GP: general practitioner
HR: hazard ratio
i-RFCBT: Web-based rumination-focused cognitive behavioral therapy
IRR: incidence rate ratio
ITT: intention-to-treat

MDE: major depressive episode
MI: multiple imputation
PHQ: Patient Health Questionnaire
PSQ: Psychosis Screening Questionnaire
PSWQ: Penn State Worry Questionnaire
RCT: randomized controlled trial
RESPOND: Reducing Stress and Preventing Depression
RFCBT: rumination-focused cognitive behavioral therapy
RNT: repetitive negative thought
RRS: Ruminative Response Scale of the Response Styles Questionnaire
RT: repetitive thought
SCID-I: Structured Clinical Interview for DSM-IV

Edited by G Eysenbach; submitted 22.06.18; peer-reviewed by M Deady, L Mclellan, J Apolinário-Hagen; comments to author 06.09.18; revised version received 15.03.19; accepted 24.03.19; published 13.05.19.

Please cite as:
Cook L, Mostazir M, Watkins E
Reducing Stress and Preventing Depression (RESPOND): Randomized Controlled Trial of Web-Based Rumination-Focused Cognitive Behavioral Therapy for High-Ruminating University Students
J Med Internet Res 2019;21(5):e11349
URL: https://www.jmir.org/2019/5/e11349/
doi:10.2196/11349
PMID:31094331

Original Paper

Remediating Reduced Autobiographical Memory in Healthy Older Adults With Computerized Memory Specificity Training (c-MeST): An Observational Before-After Study

Kris Martens[1], MSc; Keisuke Takano[2], PhD; Tom J Barry[3,4], PhD; Jolien Goedleven[1], MSc; Louise Van den Meutter[1], MSc; Filip Raes[1], PhD

[1]Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium

[2]Department of Psychology, Ludwig-Maximilians-University of Munich, Munich, Germany

[3]Department of Psychology, The University of Hong Kong, Hong Kong, China

[4]Department of Psychology, The Institute of Psychiatry, King's College London, London, United Kingdom

Corresponding Author:
Tom J Barry, PhD
Department of Psychology
The University of Hong Kong
Jockey Club Tower
Pokfulam Road
Hong Kong,
China
Phone: 86 3917 8927
Email: tom.j.barry@icloud.com

Abstract

Background: The ability to retrieve specific autobiographical memories decreases with cognitive aging. This decline is clinically relevant due to its association with impairments in problem solving, daily functioning, and depression. A therapist-delivered group training protocol, Memory Specificity Training (MeST), has been shown to enhance the retrieval of specific memories while ameliorating the impairments and negative outcomes associated with reduced specificity. The therapist-delivered nature of this intervention means it is relatively expensive to deliver and difficult for people with mobility impairments, such as older people, to receive.

Objective: The objective of this study was to test if a novel, Web-based computerized version of a group training protocol called Memory Specificity Training, has the potential to increase autobiographical memory specificity and impact associated secondary psychological processes.

Methods: A total of 21 participants (13 female; mean age 67.05, SD 6.55) who experienced a deficit in retrieving specific autobiographical memory were trained with c-MeST. We assessed memory specificity at preintervention and postintervention, as well as secondary processes such as depressive symptoms, rumination, and problem-solving skills.

Results: Memory specificity increased significantly after participants completed c-MeST (r=.57). Session-to-session scores indicated that autobiographical memory specificity improved most from the online baseline assessment to the first Web-based session. Symptoms or secondary processes such as problem-solving skills did not change significantly.

Conclusions: A Web-based automated individual version of MeST is a feasible, low-cost intervention for reduced memory specificity in healthy older adults. Future studies should clarify the preventive impact of c-MeST in other at-risk sample populations with longer follow-up times.

(J Med Internet Res 2019;21(5):e13333) doi:10.2196/13333

KEYWORDS

memory specificity training; autobiographical memory; cognitive aging; online; depression; memory; telemedicine; rumination, cognitive

Introduction

Background

The world's rapidly aging population [1] poses several challenges for societies regarding whether they can develop scalable interventions for maintaining quality of life and independence among an increasingly older population. One important cognitive factor associated with cognitive aging is a decrease in the ability to retrieve specific, personal memories [2]. This factor, referred to as reduced autobiographical memory specificity (rAMS) or overgeneral autobiographical memory [3], is associated with depression [2], impaired problem solving [4], and difficulty maintaining independence [5]. The link between these processes can be explained by the constructive episodic simulation hypothesis, which states that similar episodic processes are central to retrieval of past memories and to construction and simulation of hypothetical events [6]. Consequently, people who can retrieve more specific memories are better able to simulate possible events; they are also better able to formulate solutions to problems that might emerge in their future and plan for how to implement these solutions.

rAMS was first studied in depression (see [3] for a review) and trauma [7], and is now considered a trait marker for depression [8]. The first attempt to remediate rAMS [9] involved a 4-session group training program called Memory Specificity Training (MeST). This intervention improved memory specificity and associated cognitive processes (problem solving, rumination, and hopelessness) in depressed female inpatients. Subsequent investigations showed similar effects of MeST on rAMS in other patient groups [10-12]. In a cluster-randomized controlled platform pilot trial among people with depression, Werner-Seidler and colleagues [13] found that MeST was associated with improvements in memory specificity compared with a group receiving psychoeducation and supportive counselling.

The core component of MeST resembles the Autobiographical Memory Test (AMT) [14] used to assess rAMS. In the AMT, participants are presented with cue words and instructed to retrieve specific memories that these cue words remind them of. In MeST, participants receive similar instructions with the exception that they also receive feedback on the specificity of their responses and instructions for how they might be more specific and more detailed. Exercises are completed during the sessions and as homework assignments. In addition to exercises with cue words, in a second kind of specificity exercise participants are instructed each evening to write down one or two memories of that day (with no cue words given). After retrieving a specific memory, participants are encouraged to retrieve details of this specific moment.

rAMS is also an age-related phenomenon in *healthy* older adults [2], and aging is shown to contribute more than depressive symptoms to rAMS in people older than 50 years [15]. As the ability to retrieve specific memories is considered to be a protective factor for mental health [16], Leahy et al [17] examined whether improving memory specificity was possible among healthy adults over 70 years of age. In their study, they compared 3 groups: a MeST intervention group; a life review group, which also emphasized the recall of specific life events but placed them within the broader context of a person's life narrative; and a control group, which was asked to complete a workbook of cognitively stimulating activities not directly related to autobiographical memory (ie, crossword and Sudoku puzzles). Each intervention took 4 weeks, with a posttraining assessment in the fifth week and a follow-up 3 months later. Leahy and colleagues reported significant improvements in autobiographical memory specificity in the MeST and life review groups at posttraining relative to the control group. However, this effect was not found at 3 months' follow-up. Neither intervention had an effect on depression symptoms, functional independence, or executive functioning, but improvements in memory specificity were significantly related to improvements in social problem solving in both intervention groups.

Remediating rAMS has been found to be beneficial for older adults [17,18]. However, as societies age dramatically, making in-person training accessible to this growing and diverse population, who may not have contact with health care providers or who may have mobility or independence problems, would be challenging. Translating MeST into a computerized individualized platform could offer promise as a solution to these challenges [19]. A recently designed computerized algorithm for scoring the specificity of written autobiographical memories [20] offers new possibilities given that memory specificity training might now be delivered in the absence of a therapist and at home. This scoring algorithm, which has demonstrated good agreement with human-expert scorings [20], was incorporated into a Web-based platform for memory specificity training such that memories are coded and feedback can be given [21]. In a proof-of-concept study with participants with rAMS (operationalized as scoring <50% on the AMT), this Web-based computerized version of MeST (c-MeST) improved rAMS after 2 weeks of training (consisting of 7 sessions of 5 to 8 trials each) and the effect was maintained at 2-week follow-up, compared with a no-training control group.

Objective

In this study, we aimed to examine a Web-based, individually delivered c-MeST that exclusively consisted of specificity trials. In this version of c-MeST, we standardized sessions, as each session contained the same amount of neutral, negative, and positive valence cue words, and cue words were equivalent in valence/pleasantness, activity/arousal, power/dominance [22], and concreteness [23] among the sessions. As a result, we could obtain session-by-session specificity scores and observe each participant's progress. This standardization of sessions is in contrast to the study by Takano and colleagues [21], which followed the standard in-group version of MeST that increased the difficulty in exercises as the session progressed (eg, retrieving two memories in response to a single emotional cue). In addition, we aimed to assess depressive symptoms, rumination, and problem-solving skills online at preintervention and postintervention.

We aimed to test whether c-MeST would remediate rAMS among older adults in terms of change from preintervention to postintervention and the trajectory of change from session to session. We also aimed to test the extent to which c-MeST was associated with change in secondary outcomes and, in particular, a decrease in depressive symptoms and ruminative brooding, and an increase in problem-solving skills. Additionally, we aimed to test the feasibility of c-MeST for older adults in terms of whether, and to what extent, participants completed the intervention, and to gather reports of their experiences with c-MeST.

Methods

Participants

We recruited participants between October 2017 and April 2018 via (1) a network of university-related organizations for older alumni, (2) the website of a public advisory body for older adults, and (3) an online forum of a commercial website targeting older adults. We allowed people to participate regardless of location, but in practice all participants lived in Belgium. The study was described to them as the evaluation of a Web-based training program for a memory problem associated with cognitive aging and that is known to be a general vulnerability factor for associated processes such as impaired social problem solving and depression. The only inclusion criterion mentioned in the description of the study was a minimum age of 50 years. After completing the survey, participants were entered into a lottery to win a shopping coupon (€20). Participants showing rAMS were contacted and invited to participate in the preintervention measurement, until 20 participants completed c-MeST. An extra exclusion criterion at preintervention measurement was not having Dutch as their native language. The study received institutional ethical approval from the Social and Societal Ethics Committee of the KU Leuven (approval number G201709932).

Measures

Autobiographical Memory Test

We measured autobiographical memory specificity before and after training using an online version of the AMT [14]. Participants were instructed to retrieve a specific memory for each of 10 cue words (5 positive, 5 negative; presented in Multimedia Appendix 1). The instructions stated that the memory needed to be specific—that is, the event recalled must have happened once and lasted less than a day but did not have to be an important event. One example of a correct answer and two examples of incorrect answers were provided. Because the assessment was online, in contrast with earlier studies using an in-person verbal version of the AMT (eg, [9]) no practice trials and no feedback during the test could be given and no time limit was applied. The AMT was scored by the online classifier and manually by the fourth author (JG). When scores contradicted each other (382/2010, 18.91% of the entries), the first author (KM) checked the answers and made the final decision. We used 2 sets of cues, and although we matched both sets for imageability, familiarity, and emotional extremity [8], we administered them in counterbalanced order across the 2 test moments to avoid an effect of the cue words. For this study, we operationalized rAMS as a score lower than 70%, which we considered as a deficit in memory specificity to be remediated via training. Published studies have some variability in the inclusion criterion, from no inclusion [9-11,17] to scoring lower than 50% [21] or lower than 70% [13].

Depressive Symptomatology

We used the Patient Health Questionnaire-9 (PHQ-9) [24] to measure depressive symptomatology. The PHQ-9 is a 9-item self-report measure of depressive symptoms, scoring the 9 *Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)* Major Depressive Episode criteria based on the frequency with which they have been experienced in the past 2 weeks, from 0 (“not at all”) to 3 (“nearly every day”). Scores can range from 0 to 27. PHQ-9 showed good internal consistency with a Cronbach alpha=.76 at the preintervention measurement.

Rumination

The Ruminative Response Scale-Brooding subscale (RRS-Brooding) [25,26] is a self-report questionnaire consisting of 5 items measuring brooding from the 22-item Ruminative Response Scale [27]. The items on the brooding factor are considered to measure the maladaptive coping of passively comparing one’s situation with some unachieved standard. For example, participants are asked to rate how frequently they tend to think “Why do I always react this way?” or “Why do I have problems other people do not have?” on a scale from 1 (“almost never”) to 4 (“always”). Scores range from 5 to 20. Cronbach alpha at preintervention was good (alpha=.81).

Problem Solving

We measured participants’ problem-solving skills with an online Dutch version of the Stress Anxiety Depression version of the Means-Ends Problem-Solving Procedure (SAD-MEPS) [28]. the original Stress Anxiety Depression version of the Means-Ends Problem-Solving Procedure (MEPS) [29] consists of a series of short stories or interpersonal problem situations faced by a hypothetical protagonist. Each story starts with the protagonist facing a specific problem, which is immediately followed by a successful ending. Participants are asked to provide the middle part of each story by typing in strategies or means for solving the particular problem. We used an adapted format [28], consisting of 2 versions of each 3 scenarios: 1 depression-related, 1 stress-related, and 1 anxiety-related story. We used 2 sets of stories and administered them in counterbalanced order to avoid an effect of the difficulty of the stories. Answers were scored manually by one of the authors (JG) on 2 dimensions. First, in line with the original manual [29], we scored stories for the number of relevant means (ie, discrete sequenced steps that enable the protagonist to get closer to the stated goal). The more relevant means a participant mentions, the better. Second, in line with Marx et al [30], we also scored stories for their effectiveness from 1 (“totally ineffective”) to 7 (“very effective”). Total scores result from a mean of the scores on the 3 stories.

The c-MeST Intervention

The Web-based c-MeST consisted of 9 sessions of 11 specificity trials, which is similar in dose to the original in-person MeST (99 specificity trials vs 104 specificity exercises; [9]). The original in-person MeST [9] consisted of 1 session each week for 4 weeks, with homework assignments for every day in between sessions. For this study, we instructed participants to train on 1 session every other day, resulting in 17 days of training. The 11 trials of each session, 9 with cue words of different valences, can be categorized into 4 types: 3 positive, 3 negative, 3 neutral, and 2 memories of the day (1 about a memory of yesterday and 1 about today, without cue words). In this version of c-MeST, we standardized the sessions, as each session contained the same amount of each type of trial, and cue words were equivalent in valence/pleasantness, activity/arousal, power/dominance [22], and concreteness [23] among the sessions. Multimedia Appendix 1 lists the cue words. The 9 sets of cue words were presented in a fixed order, but the order of the cue words was randomized within each session.

Participants completed each session on a Web-based platform that contained instructions and tips about autobiographical specificity, similar to the instructions of the AMT but providing more examples. In each of the 11 specificity trials, participants were asked to retrieve a specific memory. The website used the computerized scoring algorithm for the AMT [20] to score entries and to automatically give feedback on whether the entry was specific. The scoring algorithm showed good performance against expert-rated scores in discriminating specific versus nonspecific memories (area under the receiver operating characteristic curve >.90; [31]). If the entry was scored as not specific, participants received feedback stating that their answer was not specific enough, were reminded that they needed to provide a specific memory that occurred on as specific day and that occurred only once, and were encouraged to reenter the memory or another memory with greater specificity. If, despite the feedback, participants could not generate a specific memory within three attempts, the next cue word was presented automatically. If participants succeeded in providing a specific memory, positive feedback was provided and participants were invited to provide more spatiotemporal and contextual details on the next page (ie, “Where did it happen? When did it happen? How long did it take? Who else was there? What can you see, hear, smell or taste? What kind of day was it?”). Participants were instructed to fill out these details only if they had not already provided them in their initial memory entry. Participants could skip a cue word if they wished to do so. There was no time limit per question.

Measures of Training Experiences

After each session, participants were asked 3 closed and 2 open questions regarding (1) to what extent they found that the offered words were helpful or easy for retrieving a specific memory (0 = “not easy at all, words are very difficult to retrieve memories for” to 10 = “very easy, words are very easy to retrieve memories for”), (2) to what extent they experienced the feedback provided by the software as correct (0 = “not at all, a lot of mistakes” to 10 = “very correct, no mistakes”), (3) to what extent they experienced the session to be acceptable in length (1 = “way too short” to 5 = “way too long”), (4) how they experienced the training, and (5) whether they had any other remarks.

Procedure

In the first online assessment, participants received an informed consent form including a question asking whether they wanted to provide contact details to be invited for a follow-up study, in case their results made them eligible. After completing an online AMT, participants who showed rAMS (operationalized as a score <70% on the AMT) were contacted by telephone and invited to participate in this study. We explained to participants that they were selected on their score on the online AMT. We asked participants whether they (1) recognized rAMS in their daily functioning and (2) were interested in participating in a study exploring the possibility of remediating this phenomenon. Because of concerns about feasibility and dropout, we offered participants 2 options: (1) if they wished to receive instructions for face-to-face c-MeST, we invited them to visit the first author (KM) for an in-person conversation (n=9); if not, then (2) instructions were given by telephone or email (n=12). All instructions were provided by the first author, a clinical psychologist, who could potentially refer participants to the appropriate care in case they were worried about cognitive problems. In either case, participants received an email with a link to a preintervention measurement of secondary measures (SAD-MEPS, RRS-Brooding, and PHQ-9), a second informed consent form, and a link to c-MeST. Participants were instructed to complete 1 session every other day, which would result in a training period of 17 days. Each Web-based session contained questions on feasibility. After participants completed c-MeST, another email was sent with an invitation to an online postintervention measurement of memory specificity (AMT) and secondary measures (SAD-MEPS, RRS-Brooding, and PHQ-9). When all data were gathered, participants were provided feedback about their scores and were invited to provide extra feedback on feasibility.

Analysis of Data

We scored c-MeST sessions as the number of trials for which the patient’s first answer was classified as a specific autobiographical memory, in accordance with the logic of the AMT, resulting in a maximum of 11 points per session. We tested memory specificity and secondary outcomes (depressive symptoms, rumination, problem-solving skills) for deviation from the normal distribution using the Kolmogorov-Smirnov test. Results suggested that at both time points (preintervention and postintervention), there were significant deviations from normality (AMT postintervention, *P*=.02; RRS-Brooding preintervention, *P*=.04; SAD-MEPS means postintervention, *P*=.02). For memory specificity per type of trial, the assumption of normality was not satisfied either (all *P*<.001). Therefore, we used nonparametric statistics for all analyses.

To analyze the impact of c-MeST on memory specificity and secondary measures, we used a Wilcoxon signed rank test. We assessed relations between variables and change in variables with a Kendall tau rank correlation. To compare scores on different types of trials, we used a Kruskal-Wallis test with post hoc Mann-Whitney *U* tests. We used IBM SPSS Statistics for Windows version 25.0 (IBM Corporation) for all analyses. Multimedia Appendix 2 shows raw data of memory specificity and secondary measures.

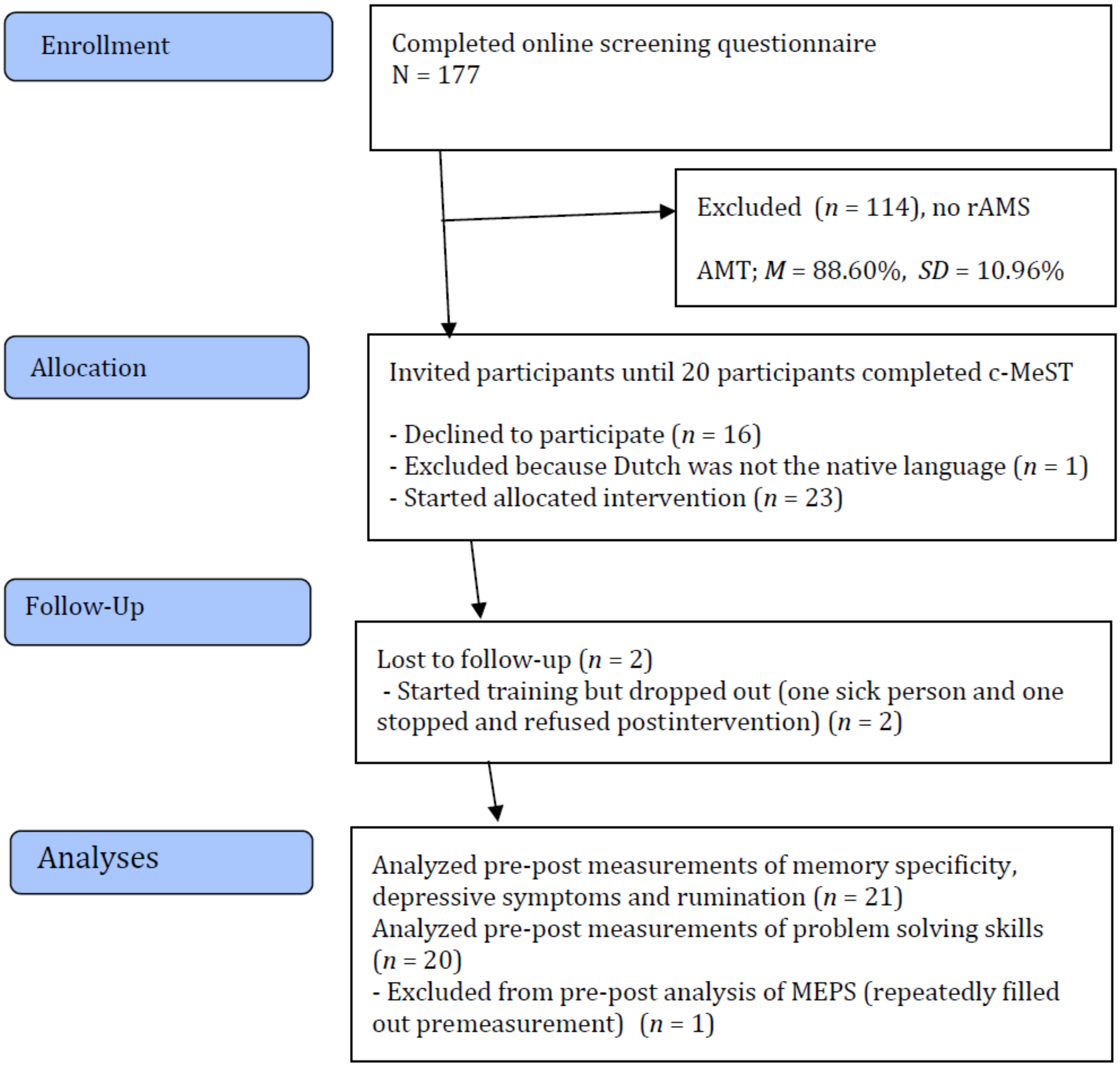
Results

Sample Characteristics

In total, 177 participants aged 50 years and over (121 female; mean age 68.97, SD 6.60 years) filled out an online version of the AMT. This screening assessment identified 63 participants with rAMS, operationalized as a score of less than 70% on the AMT (mean 37.46%, SD 18.58%). Among them, we contacted 40 people to participate in this study. However, 16 people declined to participate and we excluded 1 person because Dutch was not their native language. The remaining 23 participants started c-MeST. During or after the training, 2 participants dropped out (1 person was sick and 1 person stopped during the training without a postintervention measurement). Finally, 21 participants (13 female) completed the postintervention measurements. Figure 1 shows a flow diagram of the selection and inclusion process.

Participants in c-MeST (n=21) were aged between 55 and 77 years (mean 67.05, SD 6.55 years). Participants' age did not significantly correlate with memory specificity or any of the secondary measures (depressive symptoms, brooding, and problem solving) at preintervention measurement (with the biggest correlation being a Kendall tau correlation of –.25, *P*=.14 for brooding). At the preintervention measurement, 4 participants reported mild depressive symptoms (operationalized as a score of >5 on the PHQ-9) and 1 participant showed moderate depressive symptoms (score >10 on PHQ-9).

Figure 1. Flow diagram of the selection and inclusion process. AMT: Autobiographical Memory Test; c-MeST: computerized version of Memory Specificity Training; SAD-MEPS: Stress Anxiety Depression version of the Means-Ends Problem-Solving Procedure; rAMS: reduced autobiographical memory specificity.



Treatment Characteristics

A total of 2 participants did not complete c-MeST but did provide a postintervention measurement: 1 participant stopped after 4 sessions and 1 stopped after 5 sessions. We excluded 1 other participant from analyses of problem-solving skills, as they inadvertently filled out the preintervention assessment several times and thus also completed both versions of the SAD-MEPS task, which made a valid postintervention measurement impossible.

During c-MeST, participants needed to click the OK button after entering their memory, so that the memory was automatically scored, before filling out the details tab. Sometimes participants did not click OK and switched immediately to the details tab, which led to missing values in 1.67% of all memories scored. Participants were also allowed to skip a trial if they found it too difficult; they did so in 11.31% of the provided trials at a first attempt.

Participants were instructed to train every other day, but they were free to complete the sessions at another pace if they wished to. For the 19 participants who completed all sessions, the duration varied from 13 to 29 days (mean 18.37, SD 3.34 days). The number of days between the last session and the time of the postintervention measurement varied as well, between 0 and 16 (mean 3, SD 3.76 days).

Check on Parallel Versions

There were no differences between the sets used for the AMT and SAD-MEPS, counterbalanced between participants across time points, and so subsequent analyses did not use counterbalance as a between-participants factor (see Multimedia Appendix 3).

Memory Specificity

A Wilcoxon signed rank test showed that participants' memory specificity increased significantly (*Z*=–3.70, *P*<.001) between preintervention (mean 30.00%) and postintervention (median 80.00%) as measured by the AMT, which can be regarded as a large effect size (*r*=.57). Comparing the group that received instructions on MeST in an in-person conversation (n=9) versus the group that received instructions by telephone or email (n=12), the groups did not differ significantly in terms of change in memory specificity between preintervention and postintervention measurements, assessed with a Mann-Whitney *U* test (*U*=47.00; *P*=.61).

Session-to-session analyses, based on participants' first attempts to retrieve a specific memory, showed that the proportion of specific answers given by participants improved significantly from the preintervention assessment (median 30.00%) to the end of the first session (median 81.82%; *Z*=3.95, *P*<.001, *r*=.61). We observed no further enhancement of specificity throughout the remaining sessions (varying from a median of 72.73% for sessions 3 and 8 to a median of 81.82% for sessions 1, 2, 4, and 9), as Figure 2 illustrates and Multimedia Appendix 4 shows.

As participants could give a correct answer in a second or even third attempt if they did not do so on a first attempt, we examined whether participants were successfully able to respond to the feedback given to them after their failed first attempt and to report a specific memory in their second or third attempt. Comparing the mean proportion of specific memories given on *first* attempts with the mean across *all* attempts, this increased from 73.24% to 78.36%. A Wilcoxon signed rank test revealed that this increase in memory specificity was statistically significant (*Z*=6.29, *P*<.001), which can be regarded as a large effect size (*r*=.54). Feedback helped participants to retrieve more specific memories.

Figure 2. Median scores with interquartile ranges (25%-75%) of the Autobiographical Memory Test (AMT; pre- and postintervention measurements) and in between session-to-session scores on computerized version of Memory Specificity Training (c-MeST).

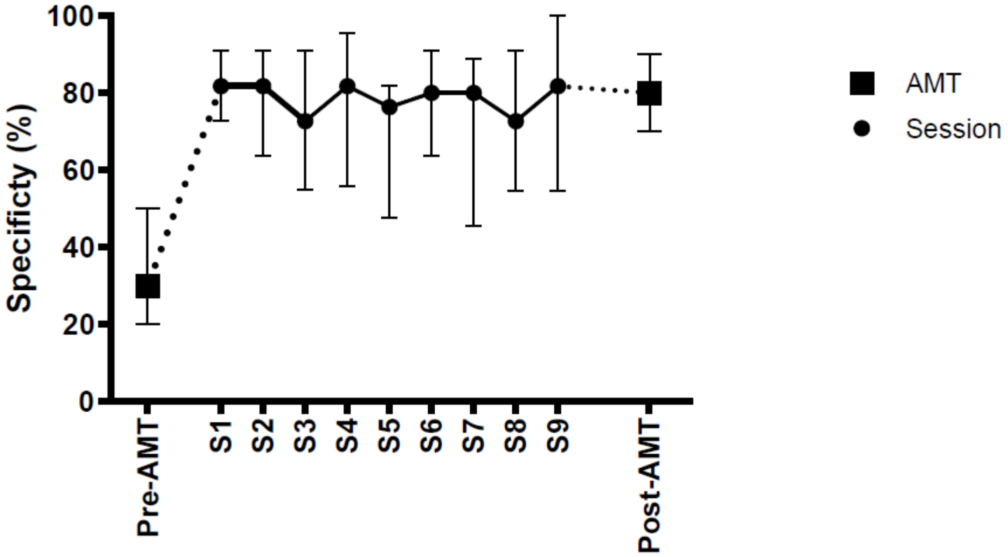


Table 1. Median, range, and effect size using a Wilcoxon signed rank test for all variables at preintervention and postintervention assessments.

Variable	Preintervention		Postintervention		Effect (*r*)
	Median	Range (IQR[a] 25%-75%)	Median	Range (IQR 25%-75%)	
AMT[b]	30.00	20.00-50.00	80.00	70.00-90.00	–.57
PHQ-9[c]	3.00	0.50-4.50	3.00	0.00-4.50	–.09
RRS-Brooding[d]	7.00	6.00-8.50	8.00	6.00-9.00	–.04
SAD-MEPS[e]-Means	2.00	1.42-2.92	1.67	1.67-2.92	–.07
SAD-MEPS-Effectiveness	4.50	3.33-5.33	4.83	3.67-5.33	–.03

[a] IQR: interquartile range.

[b] AMT: Autobiographical Memory Test.

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

[d] RRS-Brooding: Ruminative Response Scale-Brooding subscale.

[e] SAD-MEPS: Stress Anxiety Depression version of the Means-Ends Problem-Solving Procedure.

To check whether certain trials were particularly hard to complete for participants, we compared scores (%) of participants for the 4 different trial types: trials with (1) neutral, (2) positive, and (3) negative cue words and (4) memories of the day. A Kruskal-Wallis test showed a significant difference in scores between different trials (χ^2_3=19.7, *P*<.001, with mean rank scores for neutral cues of 362.81; for positive cues, of 349.44; for negative cues, of 320.68; and for memories of the day, of 409.07). Post hoc Mann-Whitney *U* tests showed a statistically significant difference between scores on the category memories of the day (median 100%) in comparison with 3 other categories of exercises: neutral cues (median 66.67%; *U*=14038.50, *P*=.01), positive cues (median 66.67%; *U*=13509.50, *P*=.002), and negative cues (median 66.67%; *U*=12308.50, *P*<.001). Results indicated no significant differences in scores between types of cue words, but memories of the day can be regarded as the easiest type of trial.

In addition, analyses did not reveal that the number of days it took participants to fulfill the training (τ_b=.10, *P*=.59) or the number of days between the last session and the postintervention measurement (τ_b=.16, *P*=.38) significantly influenced the difference between preintervention and postintervention measurements of memory specificity.

Changes in Secondary Outcomes

Participants reported low levels of depressive symptoms (median 3.00) and brooding (median 7.00) at preintervention measurement. As Table 1 shows, no significant change in reported depressive symptoms and brooding was evident by postintervention. In addition, we found no significant change in problem-solving skills (the number of means or the overall effectiveness of the solutions generated) between preintervention measurement and postintervention measurement (Table 1). Multimedia Appendix 3 shows exploratory analyses, in which we found no relevant association between change in memory specificity and change in secondary measures.

Feasibility: Training Experiences

Overall, participants found the cue words used in each session to be of moderate difficulty (mean score 6.16, SD 2.18), and they experienced the classifier as correct more often than not (mean score 7.29, SD 1.89). The length of the sessions was experienced on average as “just right” and “a bit too long” (mean score 3.52, SD 0.80). Multimedia Appendix 5 shows mean scores on the 3 questions. For the open questions, 5 participants stated throughout the training that the rationale of the training was not clear, 4 participants experienced some technical problems, and 4 participants reported that they got better at retrieving memories more quickly.

Discussion

Principal Findings

This study examined the impact of Web-based memory specificity training (c-MeST) on difficulty retrieving specific autobiographical memories among healthy older adults. This proof-of-concept study showed that translating MeST to a Web-based application resulted in significant improvements in specificity.

Translating MeST to a Web-based application dismantled MeST to its core mechanism. In comparison with in-person, group MeST as used by Leahy and colleagues [17], in c-MeST the introductory session and therapist-plus-group interaction are absent. Other study protocols [13] included psychoeducation on memory problems in depression (session 1) and psychoeducation and exercises on how to notice when one is thinking on an overgeneral level in everyday contexts and how to tackle that (session 4). The results of this study support the idea that mere memory specificity trials are sufficient to improve AMS, which is in line with previous examinations of c-MeST in the context of depression ([21] and K Martens, MSc, TJ Barry, PhD, K Takano, PhD, P Onghena, PhD, F Raes, PhD, unpublished data, 2018). Session-by-session scores revealed an increase in specificity between the online preintervention measurement and the end of the first c-MeST session. A similar finding emerged in the only other MeST or c-MeST investigation to quantify change in specificity on a session-to-session basis (K Martens, MSc, TJ Barry, PhD, K Takano, PhD, P Onghena, PhD, F Raes, PhD, unpublished data, 2018). Critically, this previous investigation used a face-to-face assessment (using a version of the AMT that included feedback) in their preintervention assessment and then an online assessment at the end of their first session. The authors concluded that the rapid improvement in specificity may have been due to a change in modality between measurements. The fact that this sudden increase in memory specificity was observed again, but now with an online preintervention assessment of specificity (without feedback), refutes this suggestion. Instead, it seems that the effects of c-MeST on specificity are realized rapidly. In this study, the addition of automated feedback during the session in comparison with the premeasurement, might have contributed to the sudden increase in memory specificity. However, it remains unclear what dosage of c-MeST (how many sessions) is required for these effects to endure once the intervention ends.

Some discrepancies between specificity measured by the AMT and by c-MeST are also of note. The difference in cue words between AMT and c-MeST might explain why c-MeST evoked more specific memories. First, the addition of neutral cue words and memories of the day to assessments of specificity in c-MeST may have made it easier for participants to retrieve specific memories. Also, including participants with specificity scores lower than 70% at preintervention measurement may have caused the increase in scores at a second measurement to be due to regression to the mean [32]. Future investigations should test these possibilities by comparing c-MeST with a control intervention and by testing differences in specificity across different cue types within the AMT *and* c-MeST. Another interesting route for future investigations is to include a measure of speed (or response time) for each memory retrieval. A decrease in the response time to retrieve a specific memory over the training period may reflect an improvement in memory functions, which could better capture the training effect (or improvement trajectory) rather than the binary score of a specific memory.

Our hypothesis that c-MeST would lead to a decrease in depressive symptoms and rumination was not supported, but this may be due to floor effects for both variables. Participants' scores at the preintervention measurement of depressive symptoms (PHQ-9 mean score 3.19, SD 2.96) fell in the range of scores found in the general population in this age range (from age 45 to >75: mean score 2.8, SD 3.5 to mean 4.4, SD 3.9; [33]). Scores on the rumination brooding scale were also in line with those found in the general population (mean score 7.62, SD 2.27 vs mean 8.6, SD 2.8; [34]). Leahy and colleagues [17] reported similar findings. It might, therefore, be unrealistic to predict further improvements from these low levels. It is of note that, among older adults who are vulnerable to subsequent increases in depression and impairments to quality of life and independence, the potential for c-MeST in preventing increases in these variables is worth further investigation.

We observed no increase in problem-solving skills. This might indicate that c-MeST does not influence problem-solving skills in healthy older adults with rAMS. This might also be explained by the use of an online version of SAD-MEPS, which is a test designed to be conducted face-to-face. After SAD-MEPS was used as a face-to-face measurement among people with depression [9], it was used as an online measurement among healthy students [35]. Both studies found no statistically significant effects from preintervention to postintervention in problem-solving skills. Future studies could assess problem-solving skills using measures that are more appropriate for online delivery or else the test should be conducted in person. The use of an adapted version of MEPS, SAD-MEPS, may not have been optimal for a group of healthy older adults with rAMS, and future research might use the standard MEPS.

The results of this study suggest that Web-based remediation of rAMS is feasible for older adults. Participants perceived the words to be moderately difficult, the feedback from the classifier as correct, and the length of the sessions as tolerable. However, participants varied in their preferences for session length and frequency. Given the nature of this research trial, we instructed participants to train in 9 sessions of 11 trials in 17 days. However, outside of a research context, participants should be able to train at their own pace. The software developed and tested here can enable participants to choose their own dosage and the frequency of training, which could further improve uptake and adherence. People also varied in scores for the 4 different kinds of trials. Future c-MeST could be personalized with an adaptive design, for example, by offering participants with low scores on one sort of cue words more of those similar trials. The software could also be combined with other instructions, such as those used in a life review, where specific memories are retrieved for particular life periods [18].

Limitations

A limitation of this study is that we did not know the participants' educational levels. We can assume that the average educational level was above average, as many participants were members of a university alumni group. Although internet use among older adults is generally high (in Belgium, 79% of older adults between 55 and 64 years of age have been reported to use the internet daily [36]), education and income levels are also positively correlated with internet skills [37]. Future investigations should examine the feasibility of c-MeST among a more diverse socioeconomic range of participants than we used. Another limitation is that we did not control for cognitive functioning. As previous research has indicated that specificity performance is associated with cognitive functioning such as executive functioning [38,39], future research should control for cognitive functioning. However, for this proof-of-concept study, the feasibility of c-MeST is promising.

Conclusions

Web-based memory specificity training can effectively improve rAMS among healthy older adults. Translating the in-group training to a computerized version resulted in a feasible, scalable alternative, but we found no impact of this training on depressive symptoms, rumination, or problem-solving skills. Future investigations require follow-up assessments and control groups to assess the utility of c-MeST as an intervention for rAMS, and in the prevention of other negative outcomes such as increases in depression symptoms, among older adults.

Acknowledgments

This work was supported by KU Leuven, Leuven, Belgium, Program Funding Grant PF/10/005 awarded to FR.

Conflicts of Interest

FR is one of the developers of the original in-group face-to-face MeST. KT, KM, and FR are the developers of the Web-based computerized MeST (c-MeST). KM and FR additionally receive payments for training workshops and presentations related to MeST.

Multimedia Appendix 1

Cue words used in assessments and training.

[PDF File (Adobe PDF File), 18KB - jmir_v21i4e13333_app1.pdf]

Multimedia Appendix 2

Raw data.

[XLSX File (Microsoft Excel File), 48KB - jmir_v21i5e13333_app2.xlsx]

Multimedia Appendix 3

Exploratory analyses.

[PDF File (Adobe PDF File), 128KB - jmir_v21i5e13333_app3.pdf]

Multimedia Appendix 4

Specificity scores for each of the 9 sessions of computerized memory specificity training.

[PDF File (Adobe PDF File), 76KB - jmir_v21i5e13333_app4.pdf]

Multimedia Appendix 5

Results for 3 feasibility questions.

[PDF File (Adobe PDF File), 72KB - jmir_v21i5e13333_app5.pdf]

References

1. World Health Organization. Mental health of older adults. Geneva, Switzerland: WHO; 2017 Dec 12. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults [accessed 2019-03-12] [WebCite Cache ID 76p5SaTpK]
2. Wilson FCL, Gregory JD. Overgeneral autobiographical memory and depression in older adults: a systematic review. Aging Ment Health 2018 May;22(5):575-586. [doi: 10.1080/13607863.2017.1326461] [Medline: 28541754]
3. Williams JMG, Barnhofer T, Crane C, Herman D, Raes F, Watkins E, et al. Autobiographical memory specificity and emotional disorder. Psychol Bull 2007 Jan;133(1):122-148 [FREE Full text] [doi: 10.1037/0033-2909.133.1.122] [Medline: 17201573]
4. Beaman A, Pushkar D, Etezadi S, Bye D, Conway M. Autobiographical memory specificity predicts social problem-solving ability in old and young adults. Q J Exp Psychol (Hove) 2007 Sep;60(9):1275-1288. [doi: 10.1080/17470210600943450] [Medline: 17676558]
5. Holland C, Boukouvalas A, Wallis S, Clarkesmith D, Cooke R, Liddell L, et al. Transition from community dwelling to retirement village in older adults: cognitive functioning and psychological health outcomes. Ageing Soc 2016 May 26;37(07):1499-1526. [doi: 10.1017/S0144686X16000477]
6. Schacter DL, Addis DR. The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philos Trans R Soc Lond B Biol Sci 2007 May 29;362(1481):773-786 [FREE Full text] [doi: 10.1098/rstb.2007.2087] [Medline: 17395575]
7. Moore SA, Zoellner LA. Overgeneral autobiographical memory and traumatic events: an evaluative review. Psychol Bull 2007 May;133(3):419-437 [FREE Full text] [doi: 10.1037/0033-2909.133.3.419] [Medline: 17469985]
8. Brittlebank AD, Scott J, Williams JM, Ferrier IN. Autobiographical memory in depression: state or trait marker? Br J Psychiatry 1993 Jan;162:118-121. [Medline: 8425125]
9. Raes F, Williams JMG, Hermans D. Reducing cognitive vulnerability to depression: a preliminary investigation of MEmory Specificity Training (MEST) in inpatients with depressive symptomatology. J Behav Ther Exp Psychiatry 2009 Mar;40(1):24-38. [doi: 10.1016/j.jbtep.2008.03.001] [Medline: 18407245]

10. Eigenhuis E, Seldenrijk A, van Schaik A, Raes F, van Oppen P. Feasibility and effectiveness of memory specificity training in depressed outpatients: a pilot study. Clin Psychol Psychother 2017 Jan;24(1):269-277. [doi: 10.1002/cpp.1995] [Medline: 26620851]
11. Moradi AR, Moshirpanahi S, Parhon H, Mirzaei J, Dalgleish T, Jobson L. A pilot randomized controlled trial investigating the efficacy of MEmory Specificity Training in improving symptoms of posttraumatic stress disorder. Behav Res Ther 2014 May;56:68-74. [doi: 10.1016/j.brat.2014.03.002] [Medline: 24705337]
12. Maxwell K, Callahan JL, Holtz P, Janis BM, Gerber MM, Connor DR. Comparative study of group treatments for posttraumatic stress disorder. Psychotherapy (Chic) 2016 Dec;53(4):433-445. [doi: 10.1037/pst0000032] [Medline: 26390014]
13. Werner-Seidler A, Hitchcock C, Bevan A, McKinnon A, Gillard J, Dahm T, et al. A cluster randomized controlled platform trial comparing group memory specificity training (MEST) to group psychoeducation and supportive counselling (PSC) in the treatment of recurrent depression. Behav Res Ther 2018 Jun;105:1-9 [FREE Full text] [doi: 10.1016/j.brat.2018.03.004] [Medline: 29587159]
14. Williams JM, Broadbent K. Autobiographical memory in suicide attempters. J Abnorm Psychol 1986 May;95(2):144-149. [Medline: 3711438]
15. Takano K, Mori M, Nishiguchi Y, Moriya J, Raes F. Psychometric properties of the written version of the autobiographical memory test in a japanese community sample. Psychiatry Res 2017 Dec;248:56-63. [doi: 10.1016/j.psychres.2016.12.019] [Medline: 28013087]
16. Latorre JM, Ricarte JJ, Serrano JP, Ros L, Navarro B, Aguilar MJ. Performance in autobiographical memory of older adults with depression symptoms. Appl Cogn Psychol 2012 Nov 04;27(2):167-172. [doi: 10.1002/acp.2891]
17. Leahy F, Ridout N, Mushtaq F, Holland C. Improving specific autobiographical memory in older adults: impacts on mood, social problem solving, and functional limitations. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2018 Dec;25(5):695-723. [doi: 10.1080/13825585.2017.1365815] [Medline: 28825508]
18. Serrano JP, Latorre JM, Gatz M, Montanes J. Life review therapy using autobiographical retrieval practice for older adults with depressive symptomatology. Psychol Aging 2004 Jun;19(2):270-277. [doi: 10.1037/0882-7974.19.2.270] [Medline: 15222820]
19. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
20. Takano K, Ueno M, Moriya J, Mori M, Nishiguchi Y, Raes F. Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm. Behav Res Methods 2017 Dec;49(3):835-852. [doi: 10.3758/s13428-016-0753-x] [Medline: 27338931]
21. Takano K, Moriya J, Raes F. Lost in distractors: reduced autobiographical memory specificity and dispersed activation spreading over distractors in working memory. Behav Res Ther 2017 Jul;94:19-35. [doi: 10.1016/j.brat.2017.04.005] [Medline: 28441521]
22. Moors A, De Houwer J, Hermans D, Wanmaker S, van Schie K, Van Harmelen A, et al. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. Behav Res Methods 2013 Mar;45(1):169-177. [doi: 10.3758/s13428-012-0243-8] [Medline: 22956359]
23. Brysbaert M, Stevens M, De Deyne S, Voorspoels W, Storms G. Norms of age of acquisition and concreteness for 30,000 Dutch words. Acta Psychol (Amst) 2014 Jul;150:80-84. [doi: 10.1016/j.actpsy.2014.04.010] [Medline: 24831463]
24. 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. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
25. Raes F, Schoofs H, Hoes D, Hermans D, Van Den Eede F, Frank E. 'Reflection' en 'brooding' als subtypes van rumineren: herziening van de Ruminative Response Scale. Gedragstherapie 2009;42:205-214.
26. Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: a psychometric analysis. Cogn Ther Res 2003;27(3):247-259. [doi: 10.1023/A:1023910315561]
27. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta earthquake. J Pers Soc Psychol 1991 Jul;61(1):115-121. [Medline: 1890582]
28. Hermans D, Ruys K, Vuerstaeck S, Van Daele T, Raes F. Autobiographical memory specificity and the effects of a psycho-educational programme for stress and anxiety. 2007 Presented at: 41st Annual Convention of the Association for Behavioral and Cognitive Therapies; Nov 15-18, 2007; Philadelphia, PA, USA.
29. Platt J, Spivack G. Manual for the Means-Ends Problem-Solving procedure (MEPS): A Measure of Interpersonal Problem-Solving Skill. Philadelphia, PA: Hahnemann Medical College and Hospital; 1975.
30. Marx EM, Williams JM, Claridge GC. Depression and social problem solving. J Abnorm Psychol 1992 Feb;101(1):78-86. [Medline: 1537977]
31. Takano K, Gutenbrunner C, Martens K, Salmon K, Raes F. Computerized scoring algorithms for the Autobiographical Memory Test. Psychol Assess 2018 Dec;30(2):259-273. [doi: 10.1037/pas0000472] [Medline: 28368170]
32. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005 Feb;34(1):215-220. [doi: 10.1093/ije/dyh299] [Medline: 15333621]

33. Kocalevent R, Hinz A, Brähler E. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2013;35(5):551-555. [doi: 10.1016/j.genhosppsych.2013.04.006] [Medline: 23664569]
34. Raes F, Bijttebier P. Ruminative Response Scale (RRS). Tijdschr Gedragsther 2012;45:81-84 [FREE Full text]
35. Van Daele T, Van den Bergh O, Van Audenhove C, Raes F, Hermans D. Reduced memory specificity predicts the acquisition of problem solving skills in psychoeducation. J Behav Ther Exp Psychiatry 2013 Mar;44(1):135-140. [doi: 10.1016/j.jbtep.2011.12.005] [Medline: 22222179]
36. Statbel (Algemene Directie Statistiek - Statistics Belgium). Enquête naar het ICT-gebruik bij huishoudens. Brussels, Belgium: Statbel; 2017 Dec 22. URL: https://bestat.statbel.fgov.be/bestat/crosstable.xhtml?view=9cbae079-22b2-44d1-aac6-cace3063d18d [accessed 2019-03-18] [WebCite Cache ID 76xxCQ4JV]
37. Hargittai E, Piper A, Morris M. From internet access to internet skills: digital inequality among older adults. Universal Access Infn Soc 2018:1-10. [doi: 10.1007/s10209-018-0617-5]
38. Ros L, Latorre JM, Serrano JP. Working memory capacity and overgeneral autobiographical memory in young and older adults. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2010 Jan;17(1):89-107. [doi: 10.1080/13825580903042650] [Medline: 19626477]
39. Holland CA, Ridout N, Walford E, Geraghty J. Executive function and emotional focus in autobiographical memory specificity in older adults. Memory 2012;20(8):779-793. [doi: 10.1080/09658211.2012.703210] [Medline: 22873516]

Abbreviations

AMT: Autobiographical Memory Test
c-MeST: computerized version of Memory Specificity Training
MEPS: Means-Ends Problem-Solving Procedure
MeST: Memory Specificity Training
PHQ-9: Patient Health Questionnaire-9
rAMS: reduced autobiographical memory specificity
RRS-Brooding: Ruminative Response Scale-Brooding subscale
SAD-MEPS: Stress Anxiety Depression version of the Means-Ends Problem-Solving Procedure

Edited by G Eysenbach; submitted 10.01.19; peer-reviewed by C Holland, F Leahy; comments to author 20.02.19; revised version received 05.03.19; accepted 07.03.19; published 14.05.19.

Please cite as:
Martens K, Takano K, Barry TJ, Goedleven J, Van den Meutter L, Raes F
Remediating Reduced Autobiographical Memory in Healthy Older Adults With Computerized Memory Specificity Training (c-MeST): An Observational Before-After Study
J Med Internet Res 2019;21(5):e13333
URL: https://www.jmir.org/2019/5/e13333/
doi:10.2196/13333
PMID:31094362

Original Paper

An Internet-Based Self-Testing Model (Easy Test): Cross-Sectional Survey Targeting Men Who Have Sex With Men Who Never Tested for HIV in 14 Provinces of China

Xia Jin[1], MS; Junjie Xu[2], MD, PhD; M Kumi Smith[3], MD, PhD; Dong Xiao[4], MA; Erica R Rapheal[3], MD, MPH; Xiangfei Xiu[1], MS; Zhengwei Ding[1], MPH; Yang Zhang[1], MSc; Yang Jie[1], BM; Ying Liao[1], MA; Ningxiao Cao[5], MD; Hao Wu[6], MD, PhD; Yugang Bao[1], MD, PhD

[1]AIDS Healthcare Foundation China Program, Beijing, China

[2]Department of Laboratory Medicine, Key Laboratory of AIDS Immunology of National Health and Family Planning Commission, The First Affiliated Hospital, China Medical University, Shenyang, China

[3]School of Public Health, University of Minnesota Twin Cities, Minneapolis, MN, United States

[4]Tongzhi Welfare, Beijing, China

[5]Chinese Academy of Medical Sciences, Hospital for Skin Diseases and STI Control, Nanjing, China

[6]Beijing Youan hospital, Beijing, China

Corresponding Author:
Yugang Bao, MD, PhD
AIDS Healthcare Foundation China Program
No.13, South Road of the 3rd Ring East, Chaoyang District
Beijing, 100021
China
Phone: 86 10 67735682
Fax: 86 10 67735682
Email: Yugang.Bao@aidshealth.org

Abstract

Background: With China’s explosive internet growth, activities such as socializing and partner seeking among men who have sex with men (MSM) has also become Web based through popular services such as Blued. This creates a new mode of health promotion with the potential to instantly reach large numbers of MSM, including those who rarely access traditional offline testing facilities.

Objective: This study aimed to assess the feasibility of the Easy Test in increasing access and uptake of HIV testing and treatment services among MSM and to identify demographic and behavioral predictors of program uptake to inform future implementation.

Methods: A feasibility study of the Easy Test model was conducted from October 2017 to December 2017 in 14 Chinese provinces. Applicants who provided informed consent completed a self-administered questionnaire and submitted a US $5 deposit to have the free test kit delivered to their homes. Orders were then received, processed, and posted by volunteers from local community-based organizations. Once applicants submitted images of their test results, the deposit was refunded to the applicant. Those whose test results were deemed to be HIV-positive were then connected to a peer navigator to accompany the individual to follow-up medical services. A chi-squared trend test was used to assess the relationship between lifetime HIV testing volume and HIV prevalence. Logistic regression models were used to identify independent risk factors associated with two outcomes: (1) never having tested for HIV and (2) receiving an HIV-positive result.

Results: A total of 879 individuals submitted Web-based requests for test kits. Their median age was 28 (interquartile range 24-34 years); 69.3% (609/879) had at least a college education, and 51.5% (453/879) had a monthly income between US $450 to $750; 77.7% (683/879) of the applicants submitted images of their test results, among whom 14.3% (98/683) had an HIV-positive result. Among the 42.9% (293/683) who were first-time testers, the HIV prevalence was 18.8% (55/293). Nearly three-quarters (71/98, 72.4%) of those with a positive test result were connected with a peer navigator and enrolled in treatment. Among the first-time testers, having multiple sexual partners (2-3 sexual partners: adjusted odds ratio [aOR] 2.44, 95% CI 1.08-5.50; 4 or above sexual partners: aOR 3.55, 95% CI 1.18-10.68) and reporting inconsistent condom use in the previous 3 months (aOR 7.95, 95% CI 3.66-17.26) were both associated with an HIV-positive result. An inverse dose response relationship between lifetime HIV testing volume and HIV prevalence was also observed in this study (χ^2_3=55.0; P<.001).

Conclusions: The Easy Test model reached a larger portion of first-time testers, many who reported higher risk sexual behaviors. This highlights the potential for an internet-based self-test model to increase access to HIV treatment services for HIV-positive MSM in China.

(J Med Internet Res 2019;21(5):e11854) doi:10.2196/11854

KEYWORDS

internet; men who have sex with men (MSM); HIV; China

Introduction

Background

Men who have sex with men (MSM) represent an increasing proportion of newly reported HIV/AIDS cases. In 2017, MSM accounted for 25.54% (34,358/134,512) of all newly reported HIV/AIDS cases in China, compared with 14.70% (10,954/74,517) in 2011 [1,2]. In addition, national sentinel surveillance data indicate an increase in HIV prevalence among MSM from 5.72% (1948/34,009) to 7.98% (3312/41,503) between 2010 and 2015 [3]. Although the Chinese government has provided free HIV testing and counseling since 2003, and health authorities have affected a massive scale-up of facility-based HIV testing services, experts estimate around half of Chinese MSM remain untested for HIV [4].

Facility-based HIV testing is defined as HIV testing conducted in health care facilities, including the Centers for Disease Control and Prevention (CDC), hospitals and clinics, and community-based organizations (CBOs) that conduct HIV antibody screening (often using rapid tests) with linkage services to health care facilities for those who screen positive. The public health utility of internet-based interventions is of growing interest in China owing to the expansion of internet access—internet and mobile internet users currently exceed 800 million people [5]—and the growing use of social networking apps for socializing and sexual partner seeking [6-9]. To date, internet-based interventions have been used to promote HIV testing in young MSM, promote HIV/hepatitis C virus testing among people who inject drugs, reduce HIV transmission among male sex workers, reduce pretreatment loss to follow-up in HIV-diagnosed individuals, and improve the adherence of antiretroviral therapy in people living with HIV/AIDS [10-14]. Although 73.49% (34,712/47,231) of respondents in a nationally representative survey of MSM reported that they had found recent sex partners on the Web, only 26.36% (3890/14,757) of these respondents indicated that they had been tested for HIV and received test results in the last year [15,16]. The high proportion of Chinese MSM who use the internet to find sexual partners and the low rate of HIV testing among internet-using MSM underscore the potential of internet-based interventions to increase HIV testing uptake in this population.

The World Health Organization has recently recommended HIV self-testing to increase HIV testing uptake in light of its convenience, acceptability, timeliness, accuracy, and its provision of privacy to testers [17]. The Chinese Center for Disease Control and Prevention (China CDC) has also cited self-testing as a priority area. However, public health concerns about self-testing, including the risk of the clients misinterpreting results and missed opportunities for safe sex counseling and linkage to care, have slowed its implementation beyond small-scale studies in urban areas [6,18]. To address these shortcomings, the AIDS Healthcare Foundation (AHF) China developed an internet-based self-testing model, Easy Test, in which clients were invited to place Web-based orders for home delivery of free HIV self-test kits. For those testing positive, access to other services (eg, test result confirmation by trained staff or linkage to an in-person peer navigator) would be provided as well. The internet-based approach of Easy Test was designed with the explicit goal of increasing test coverage among MSM, particularly those who may be less likely to access facility-based HIV testing and treatment services.

Objectives

The objectives of this study were (1) to assess the capability of the Easy Test model in increasing access to HIV testing and treatment services for MSM who have never tested for HIV and (2) to identify demographic and behavioral predictors of program uptake to inform future implementation.

Methods

Study Setting and Sampling Strategy

This study was conducted by AHF China from October 2017 to December 2017 in 14 Chinese provinces. The Easy Test model was born out of a collaborative effort between AHF China, regional health facilities, and MSM CBOs local to each of the study cities (Beijing, Tianjin, Shanxi, Chongqing, Liaoning, Guangxi, Yunnan, Hunan, Hainan, Sichuan, Henan, Heilongjiang, Zhejiang, and Xinjiang). Promotion of the Easy Test program was carried out using social media tools (WeChat, QQ, and Blued) as well as offline events organized by local CBOs. Printed Easy Test promotion cards were also included in the self-testing kit packages to encourage participants to share information about the service with spouses, sexual partners, or friends. Eligible participants were at least aged 16 years, born biologically male, had engaged in anal sex with a man, and were of negative or unknown HIV status.

Study Procedures

Participants interested in ordering a self-test kit were instructed to visit the Easy Test website either by typing in the URL or scanning an exclusive Quick Response code. Following eligibility screening, those meeting the requirements were asked to provide electronic informed consent and to participate in a self-administered survey where they provided information on sociodemographic characteristics, sexual risk behaviors, and HIV test history. Participants were also asked for their mobile phone number, shipping address, a US $5 deposit, and a nonrefundable shipping fee of US $2 to $3. All payments were processed using the secure and widely used Web-based payment platforms of Alipay or WeChat. Peer educators at the MSM CBOs processed all incoming orders and dispatched all express mail deliveries with China's most widely used express delivery service. Packages were delivered within 1 to 3 days in an unmarked package with no visible indications of the contents or sender. Self-testing kits were limited to 1 per registered mobile phone number to minimize potential bias from repeat testers.

The self-testing package included a rapid test reagent, a single-use safety lancet, a buffer, a dropper, an alcohol pad, a printed Easy Test promotion card, and printed kit instructions. The rapid test reagent used in this study is Alere Determine TM HIV-1/2 (Alere Medical). The needle of the single use safety spring lancet is designed to permanently and automatically retract after the first use. Participants were also offered to be put in touch with a peer educator, who could provide over-the-phone guidance on self-administration of the test. In addition, an instructional video was provided on the Easy Test website and WeChat account. Participants were refunded their US $5 deposit upon successful upload of a photograph of the results of their completed self-test kit, which was then interpreted by trained staff at each CBO and health facility.

Volunteers from each local CBO contacted participants whose submitted photographs indicated an HIV-positive result to arrange linkage services. Those interested in linkage services were paired with a peer navigator to accompany them to receive confirmatory testing, as well as initial visits for treatment and care following formal diagnosis. Treatment and care for HIV/AIDS are available free through the Chinese national treatment program [19]. The observation period to assess the rate of HIV treatment initiation among diagnosed individuals in this study extended a month beyond the study endpoint to maximize inclusion of successful referrals. Peer navigators were trained volunteers from local CBOs who paired up with local health departments to help ensure successful linkage of participants to confirmatory testing, care, and treatment.

Data Analysis

The information on sociodemographic characteristics, sexual behaviors, and HIV testing history collected as part of the self-testing application was used for descriptive analyses. A chi-squared trend test was used to assess the relationship between lifetime HIV testing volume and HIV prevalence. Independent risk factors for never having tested for HIV and a positive self-test result were assessed using bivariable logistic regression analysis. Variables significant at $P \leq .05$ or determined to be clinically relevant were selected for inclusion in the multivariable logistic model. All statistical analyses were performed using SAS 9.4 (SAS Institute Inc).

Ethical Statement

The study received approval from the Hospital for Skin Diseases and STI Control of Chinese Academy of Medical Sciences.

Results

Study Participants

From October 2016 to December 2016, a total of 1015 participants applied for Web-based self-testing kits among whom 879 (86.60%) met the eligibility criteria and provided informed consent. Those who did not meet the requirements included 11 individuals who reported their sex at birth as female, 8 who were already aware of their HIV-positive status, and 117 who had never engaged in anal sex with a man (Figure 1).

Figure 1. Flowchart of the Easy Test model.

Sociodemographics and Behaviors

Among the 879 eligible participants, the median age was 28 years (interquartile range [IQR] 24-34 years), 91.9% (808/879) were of Han ethnicity, 69.3% (609/879) had a college education or higher, 74.3% (653/879) had never been married, 12.7% (112/879) were currently students, and half (453/879, 51.5%) had a monthly income between 3000 and 5000 Chinese yuan (US $450 to $750).

Overall, 90.0% (791/879) of participants self-identified as homosexual and the median age of sexual debut among the participants was 20 years (IQR 18-22 years). Among the 87.1% (766/879) who had engaged in anal sex with another man in the last 3 months, 53.3% (408/766) had had 2 or more partners and 28.9% (221/766) had not used condoms during anal intercourse. Among all participants, 81.5% (716/879) reported condom use during their last anal intercourse, although only 60.5% (433/716) reported correct condom use; 10.0% (88/879) of participants self-identified as bisexual or unsure, 96.3% (70/88) reported vaginal sex with a female partner, and only 31.4% (22/70) of these reported condom use during vaginal sex in the last 3 months; 29.4% (258/879) of participants reported that they were aware of the HIV status of regular sexual partners.

Participants who self-reported incorrect condom use (adjusted odds ratio [aOR] 1.63, 95% CI 1.17-2.27) and lack of knowledge about their regular sexual partners’ HIV status (aOR 2.45, 95% CI 1.67-3.60) were associated with having never tested for HIV before participation in this study compared with participants reporting correct condom use and awareness of the HIV status of sexual partners. Further details are provided in Table 1.

Testing History, Results, and Linkage to Care

Upon enrollment in this study, 40.0% (352/879) of the participants reported having never tested for HIV before. Over three-quarters (683/879, 77.7%) of participants submitted a photograph of the completed self-test result to the study staff. The overall HIV prevalence among these participants was 14.3% (98/683), among whom 100% were eventually confirmed as HIV infection following confirmatory testing at a local health facility. Participants with confirmed HIV were all peer-navigated to local health facilities and 72% (71/98) of them enrolled in treatment with a 1-month extension beyond the study endpoint.

Nearly half of the participants (293/683, 42.9%) who submitted photographs of their self-test results were first-time testers for HIV, and among whom HIV prevalence was 18.8% (55/293). Among those who had tested before, we observed an inverse dose response relationship between lifetime HIV testing volume and HIV prevalence. In the subgroups of participants who had tested once or twice previously versus 3 to 4 times versus 5 or more times, HIV prevalence was 16.2% (33/204), 7.0% (8/114), and 2.8% (2/72), respectively (χ^2_3=55.0; *P*<.001; Figure 2).

Table 1. Demographics and behaviors of the Easy Test participants and factors associated with those who had never tested for HIV.

Variable	Participants, n (%)	Never tested for HIV, n (%)	Ever tested for HIV, n (%)	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	*P* value for adjusted OR
Age (years)						
<25	234 (26.6)	118 (33.5)	116 (22.0)	2.26 (1.53-3.35)	2.08 (1.25-3.44)	.005
25-34	442 (50.3)	171 (48.6)	271 (51.4)	1.40 (0.98-2.00)	1.78 (1.15-2.75)	.01
≥35	203 (23.1)	63 (17.9)	140 (26.6)	1.0	1.0	N/A[b]
Ethnicity						
Han	808 (91.9)	336 (95.5)	472 (89.6)	2.45 (1.38-4.34)	3.49 (1.62-7.52)	.001
Other	71 (8.1)	16 (4.5)	55 (10.4)	1.0	1.0	N/A
Education						
<12 years	270 (30.7)	122 (34.7)	148 (28.1)	1.36 (1.02-1.82)	N/A	N/A
≥12 years	609 (69.3)	230 (65.3)	379 (71.9)	1.0	N/A	N/A
Marital status						
Never married	653 (74.3)	269 (76.4)	384 (72.9)	1.21 (0.88-1.65)	N/A	N/A
Other[c]	226 (25.7)	83 (23.6)	143 (27.1)	1.0	N/A	N/A
Occupation						
Students	112 (12.7)	60 (17.0)	52 (9.9)	1.88 (1.26-2.80)	N/A	N/A
Other	767 (87.3)	292 (83.0)	475 (90.1)	1.0	N/A	N/A
Monthly income (US $)						
<450	194 (22.1)	107 (30.4)	87 (16.5)	3.30 (2.20-4.95)	2.29 (1.37-3.84)	.002
450-749	453 (51.5)	182 (51.7)	271 (51.4)	1.80 (1.28-2.54)	1.31 (0.86-1.98)	>.05
≥750	232 (26.4)	63 (17.9)	169 (32.1)	1.0	1.0	N/A
Age of sexual debut (years)						
<18	142 (16.2)	54 (15.3)	88 (16.7)	0.90 (0.63-1.31)	N/A	N/A
≥18	737 (83.8)	298 (84.7)	439 (83.3)	1.0	N/A	N/A
Sexual orientation[d]						
Homosexual	791 (90.0)	325 (92.3)	466 (88.4)	1.58 (0.98-2.53)	N/A	N/A
Bisexual/Unsure	88 (10.0)	27 (7.7)	61 (11.6)	1.0	N/A	N/A
Sexual role						
Insertive partner	295 (33.6)	107 (30.4)	188 (35.7)	1.0	N/A	N/A
Both	332 (37.8)	128 (36.4)	204 (38.7)	1.10 (0.80-1.53)	N/A	N/A
Receptive partner	252 (28.6)	117 (33.2)	135 (25.6)	1.52 (1.08-2.15)	N/A	N/A
Number of anal sexual partners in the last 3 months						
1	358 (46.7)	150 (50.8)	208 (44.2)	1.0	N/A	N/A
2-3	335 (43.7)	113 (38.4)	222 (47.1)	0.71 (0.52-0.96)	N/A	N/A
≥4	73 (9.6)	32 (10.8)	41 (8.7)	1.08 (0.65-1.80)	N/A	N/A
Always used condoms during anal sex in the last 3 months						
Yes	545 (71.1)	187 (63.4)	358 (76.0)	1.0	N/A	N/A
No	221 (28.9)	108 (36.6)	113 (24.0)	1.83 (1.33-2.51)	N/A	N/A
Correct condom use during last anal sex[e]						
Yes	433 (60.5)	139 (52.3)	294 (65.3)	1.0	1.0	N/A
No	283 (39.5)	127 (47.7)	156 (34.7)	1.72 (1.26-2.35)	1.63 (1.17-2.27)	.004

Variable	Participants, n (%)	Never tested for HIV, n (%)	Ever tested for HIV, n (%)	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	*P* value for adjusted OR
Know the HIV status of regular sexual partners						
Yes	258 (29.4)	62 (17.6)	196 (37.2)	1.0	1.0	N/A
No	621 (70.6)	290 (82.4)	331 (62.8)	2.77 (2.00-3.84)	2.45 (1.67-3.60)	<.001

[a] OR: odds ratio

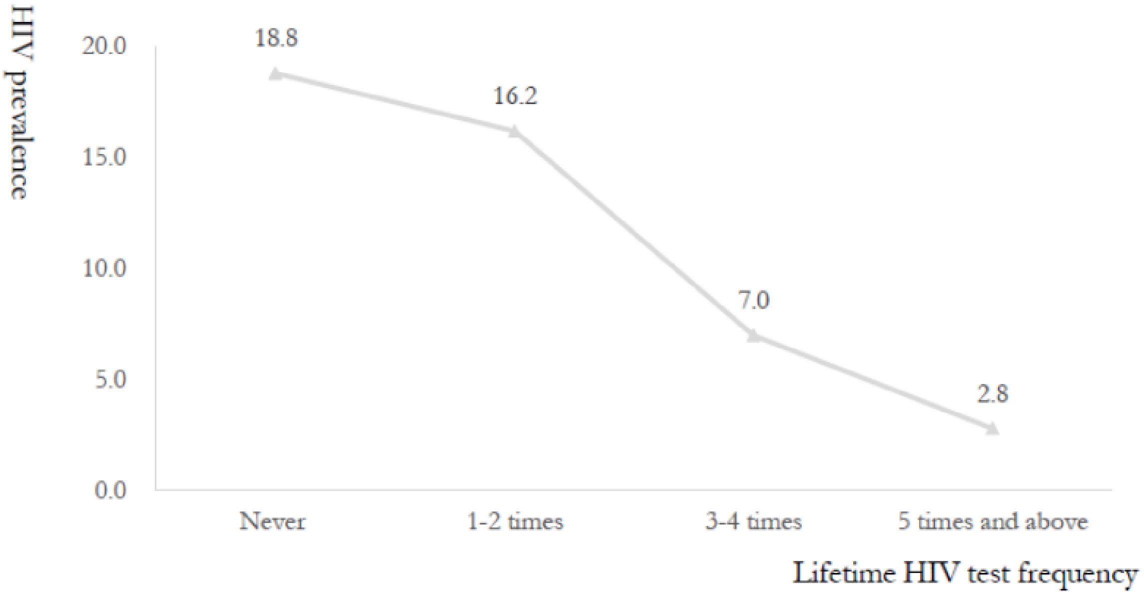
[b] N/A: not applicable.

[c] Marital status: “other” included those who were married, cohabitating, divorced, or widowed.

[d] Sexual orientation is not a behavior but an identity.

[e] Correct condom use was defined as proper use of a condom during the entire process of sexual intercourse (from the beginning of anal sex to the end) without breakage or falling off.

Figure 2. HIV prevalence with lifetime frequency.



Factors Associated With HIV Infection

Results of multivariate models indicated that overall, participants who reported usually being the receptive anal sex partner (aOR 4.58, 95% CI 1.52-13.79) or who assumed both insertive and receptive roles (aOR 4.10, 95% CI 1.39-12.12) had far greater odds of HIV infection relative to those who mainly assumed the insertive role. Participants reporting greater numbers of sexual partners in the last 3 months also had higher odds of HIV infection, with an aOR of 2.74 (95% CI 1.23-6.11) for those with 2 to 3 sexual partners and 3.98 (95% CI 1.27-12.49) for 4 or more sexual partners, compared with those reporting only 1 partner in the same time period. The odds of HIV infection were also higher for those reporting inconsistent condom use during anal intercourse in the last 3 months (aOR 4.64, 95% CI 2.23-9.63) and those reporting incorrect condom use during their most recent anal sex episodes had higher odds of HIV infection (aOR 2.19, 95% CI 1.05-4.58) compared with those reporting consistent and correct usage. Other characteristics associated with higher odds of HIV infection included fewer than 3 lifetime HIV testing episodes (aOR 2.58, 95% CI 1.02-6.49), no knowledge of sexual partners’ HIV infection status (aOR 3.26, 95% CI 1.06-10.00), and never having been married (aOR 4.10, 95% CI 1.31-12.82).

Among the subset of participants who had never tested for HIV, those who reported being a receptive anal partner or who assumed both insertive and receptive roles had far higher odds of HIV infection (aOR 3.78, 95% CI 1.38-10.35 for receptive partners and aOR 3.00, 95% CI 1.09-8.26 for both positions) relative to those who mainly assumed the insertive role. Participants reporting greater numbers of sexual partners in the last 3 months also had higher odds of HIV infection, with an aOR of 2.44 (95% CI 1.08-5.50) for those with 2 to 3 sexual partners and 3.55 (95% CI 1.18-10.68) for 4 or more sexual partners, compared with those reporting only 1 partner in the same period. The likelihood of HIV infection was higher among participants who did not always use condoms during anal intercourse in the last 3 months (aOR 7.95, 95% CI 3.66-17.26) relative to participants who had used condoms during anal sex in the last 3 months. The details are provided in Tables 2 and 3.

Table 2. Factors correlated with HIV infection among the Easy Test participants in China.

Characteristics	Participants (N=879), n	HIV cases, n	Prevalence, %	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	Adjusted *P* value
Age (years)						
<25	193	35	18.1	1.76 (0.95-3.25)	N/A[b]	N/A
25-34	329	45	13.7	1.26 (0.70-2.25)	N/A	N/A
≥35	161	18	11.2	1.0	N/A	N/A
Ethnicity						
Han	624	93	14.9	1.89 (0.74-4.85)	N/A	N/A
Other	59	5	8.5	1.0	N/A	N/A
Education						
<12 years	227	29	12.8	0.82 (0.52-1.31)	N/A	N/A
≥12 years	456	69	15.1	1.0	N/A	N/A
Marital status						
Never married	499	83	16.6	2.25 (1.26-4.01)	4.10 (1.31-12.82)	.015
Other[c]	184	15	8.2	1.0	1.0	N/A
Occupation						
Students	91	17	18.7	1.45 (0.81-2.58)	N/A	N/A
Other	592	81	13.7	1.0	N/A	N/A
Monthly income (US $)						
<450	155	25	16.1	2.40 (1.13-5.06)	N/A	N/A
450-749	380	62	16.3	2.43 (1.24-4.75)	N/A	N/A
≥750	148	11	7.4	1.0	N/A	N/A
Age of sexual debut (years)						
<18	128	21	16.4	1.22 (0.72-2.06)	N/A	N/A
≥18	555	77	13.9	1.0	N/A	N/A
Sexual orientation[d]						
Homosexual	615	93	15.1	2.25 (0.88-5.73)	N/A	N/A
Bisexual/Unsure	68	5	7.4	1.0	N/A	N/A
Sex role						
Insertive partner	230	14	6.1	1.0	1.0	N/A
Both	246	43	17.5	3.27 (1.74-6.15)	4.10 (1.39-12.12)	.011
Receptive partner	207	41	19.8	3.81 (2.01-7.22)	4.58(1.52-13.79)	.007
Number of anal sexual partners in the last 3 months						
1	272	27	9.9	1.0	1.0	N/A
2-3	246	52	21.1	2.43 (1.47-4.02)	2.74 (1.23-6.11)	.014
≥4	64	13	20.3	2.31 (1.12-4.79)	3.98 (1.27-12.49)	.018
Always used condoms during anal sex in the last 3 months						
Yes	393	26	6.6	1.0	1.0	N/A
No	189	66	34.9	7.57 (4.61-12.46)	4.64 (2.23-9.63)	<.001
Correct condom use during last anal sex[e]						
Yes	298	17	5.7	1.0	1.0	N/A
No	248	39	15.7	3.08 (1.70-5.60)	2.19 (1.05-4.58)	.037

Characteristics	Participants (N=879), n	HIV cases, n	Prevalence, %	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	Adjusted *P* value
Lifetime HIV testing frequency						
<3	497	88	16.2	3.79 (1.92-7.46)	2.58 (1.02-6.49)	.045
≥3	10	186	2.8	1.0	1.0	N/A
Know the HIV status of regular sexual partners						
Yes	157	10	6.4	1.0	1.0	N/A
No	526	88	16.7	2.95 (1.50-5.83)	3.26 (1.06-10.00)	.04

[a] OR: odds ratio.

[b] N/A: not applicable.

[c] Marital status: “other” included those who were married, cohabitating, divorced, or widowed.

[d] Sexual orientation is not a behavior but an identity.

[e] Correct condom use was defined as proper use of a condom during the entire process of sexual intercourse (from the beginning of anal sex to the end) without breakage and falling off.

Table 3. Factors correlated with HIV infection among first-time HIV testers in China.

Characteristics	Participants (N=293), n	HIV cases, n	Prevalence, %	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	Adjusted *P* value
Age (years)						
<25	98	22	22.4	1.30 (0.57-2.30)	N/A[b]	N/A
25-34	140	23	16.4	0.89 (0.39-2.01)	N/A	N/A
≥35	55	10	18.2	1.0	N/A	N/A
Ethnicity						
Han	281	53	18.9	1.16 (0.25-5.46)	N/A	N/A
Other	12	2	16.7	1.0	N/A	N/A
Education						
<12 years	109	18	16.5	0.79 (0.42-1.46)	0.41 (0.19-0.89)	.025
≥12 years	184	37	20.1	1.0	1.0	N/A
Marital status						
Never married	224	48	21.4	2.42 (1.04-5.62)	N/A	N/A
Other[a]	69	7	10.1	1.0	N/A	N/A
Occupation						
Students	49	11	22.4	1.32 (0.62-2.78)	N/A	N/A
Other[c]	244	44	18	1.0	N/A	N/A
Monthly income (US $)						
<450	88	15	17	1.73 (0.59-5.09)	N/A	N/A
450-749	158	35	22.2	2.39 (0.88-6.50)	N/A	N/A
≥750	47	5	10.6	1.0	N/A	N/A
Age of sexual debut (years)						
<18	47	13	27.7	1.86 (0.90-3.82)	N/A	N/A
≥18	246	42	17.1	1.0	N/A	N/A
Sexual orientation[d]						
Homosexual	273	53	19.4	2.17 (0.49-9.63)	N/A	N/A
Bisexual/unsure	20	2	10	1.0	N/A	N/A
Sex role						
Insertive partner	94	7	7.4	1.0	1.0	N/A
Both	97	21	21.6	3.43 (1.38-8.52)	3.00 (1.09-8.26)	.034
Receptive partner	102	27	26.5	4.47 (1.84-10.86)	3.78 (1.38-10.35)	.01
Number of anal sexual partners in the last 3 months						
1	127	15	11.8	1.0	1.0	N/A
2-3	87	27	31	3.36 (1.66-6.80)	2.44 (1.08-5.50)	.033
≥4	26	10	38.5	4.67 (1.79-12.14)	3.55 (1.18-10.68)	.024
Always used condoms during anal sex in the last 3 months						
Yes	148	12	8.1	1.0	1.0	N/A
No	92	40	43.5	8.72 (4.24-17.91)	7.95 (3.66-17.26)	<.001
Correct condom use during last anal sex[e]						
Yes	104	9	8.7	1.0	N/A	N/A
No	117	17	14.5	1.79 (0.76-4.22)	N/A	N/A

Characteristics	Participants (N=293), n	HIV cases, n	Prevalence, %	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	Adjusted *P* value
Know the HIV status of regular sexual partners						
Yes	46	4	8.7	1.0	N/A	N/A
No	247	51	20.6	2.73 (0.94-7.97)	N/A	N/A

[a] OR: odds ratio.

[b] N/A: not applicable.

[c] Marital status: “other” included those who were married, cohabitating, divorced, or widowed.

[d] Sexual orientation is not a behavior but an identity.

[e] Correct condom use was defined as proper use of a condom during the entire process of sexual intercourse (from the beginning of anal sex to the end) without breakage and falling off.

Discussion

Principal Findings

The large sample size, high participation among first-time testers (293/683, 42.9%), and the high rate of photograph-confirmed test completion (683/879, 77.7%) all emphasize the feasibility and efficacy of our Easy Test model for promoting internet-based HIV self-testing and linkage to care among MSM in China. Our study assessed the utility of the Easy Test model as a supplement to existing facility-based HIV testing services. The proportion of those who never tested for HIV in this study (293/683, 42.9%) was higher than in a previous Web-based distribution of self-testing kits in MSM in Guangdong (118/380, 31.1%) [18], though the vastly different regional scope (one province in Zhong et al vs 14 provinces in this study), sample size (N=198 in Zhong et al vs N=1015 in this study), and HIV prevalence (4.5% in Zhong et al vs 18.8% in this study) suggest substantial differences in sample makeup that may impede direct comparison.

All participants who shared their positive test results with study staff were eventually confirmed as HIV infected, an encouraging finding for which we believe there are possible explanations. First, inclusion of a one-to-one peer navigation service for those testing positive may have helped assuage known concerns in the study population regarding what to do in the event of a positive test result. The other reason that contributed to this was the rapid test substitute strategy issued by CDC China in 2015 to reduce the time from screening positive to diagnosis as HIV infected within 1 to 3 days [20]. A similar 100% confirmatory testing result of the screened HIV-positive cases was found in a smaller size study (N=198; 8 positives) in Guangdong [18].

Treatment enrollment rates among those who received confirmatory testing were high (71/98, 72%) but still well below the Joint United Nations Programme on HIV/AIDS target of 90% of diagnosed individuals on therapy. As this proportion captures treatment initiated within a month after study close out, this rate may underestimate the ultimate proportion. However, China’s universal and immediate treatment policies suggest that full linkage may be attainable, and future rollout may do well to consider additional strategies such as recruitment of test coordinators from health facilities to conduct community outreach for diagnosed individuals.

Participants in this study reported more sexual risk behaviors than in comparable Web-based samples of MSM in China, including more multiple sexual partnerships, infrequent or incorrect condom use, more frequent receptive anal sex, and less awareness of sexual partners’ HIV status [6,21]. The overall HIV prevalence (98/683, 14.3%) was also far higher than in a similar sample (24/341, 7.0%) [21]. However, the identified risk factors for HIV infection were similar to these studies and included characteristics such as being the receptive anal sex partner (whether primarily or occasionally so), higher numbers of anal sexual partners, and not always using condoms during anal intercourse [6,22-25]. Even though most of the internet-based MSM in the study reported using condoms during anal sex, there is still a significant proportion who reported inconsistent and incorrect condom use, particularly among those who had previously never tested for HIV (283/716, 39.5%). Our findings suggest that public health officials should pay attention to relevant contextual factors that may increase the risk of HIV infection, including greater numbers of anal sexual partners and unprotected anal sexual activities. In addition, HIV/AIDS health education and condoms should be provided with an emphasis on MSM who prefer the receptive anal role to increase condom use and negotiation skills with their sexual partners [6]. Furthermore, the proportion of HIV-positive results decreased with the increased lifetime HIV testing frequency in this study. This information further highlights the importance of routine HIV testing and its impact on better health outcomes.

Limitations

The study findings should be considered in light of several limitations. First, our online recruitment methods may have resulted in a sample which over-represents better-educated, literate, and higher income MSM, a well-documented phenomenon of Web-based samples [26-28]. Second, to understand the potential bias introduced by the nonrandom subset of participants who submitted photographs of their self-testing results, we compared reported behaviors across groups among those who did and did not submit photographs of their completed self-test results. This analysis found that condom use during the last anal sex (68.8% for correct condom use vs 87.3% for not) and HIV status of sexual partners (60.9% for knowing partner status vs 84.7% for not) were independent risk factors associated with not providing feedback regarding their results. It is possible that those with lower HIV risk were more reluctant to submit their test results, which could lead to overestimation of HIV prevalence in this population. Finally, we do not know the proportion of respondents with HIV-positive test results who may have sought care and treatment on their own without the help of peer-navigator services provided as a part of this study.

Conclusions

Our study demonstrated that internet-based self-testing may be an effective approach for increasing HIV test uptake, particularly among those who have never previously tested for HIV. It also shows promise as a strategy for improving linkage to care and treatment initiation among those diagnosed with HIV. Internet-based MSM who engage in self-testing and who have never previously tested for HIV are less likely to use condoms during sex with both male and female partners, suggesting that internet-based testing interventions may be effective in reaching those in greatest need of these services.

Acknowledgments

The authors would like to thank all the study participants and peer educators who took part in this study. This study was supported by AHF China Program (2017-P-001); National Science and Technology Major Project Research of the 13th Five-Year (2018ZX10101001-001-003); and National Natural Science Foundation of China (81872674).

Conflicts of Interest

None declared.

References

1. National Center for AIDS/STD Control and Prevention, China CDC. Update on the AIDS/STD epidemic in China and main response in control and prevention in 2011. Chinese Journal of AIDS & STD 2012;18(02):64 [FREE Full text]
2. NCAIDS, NCSTD, China CDC. Update on the AIDS/STD epidemic in China in December 2017. Chinese Journal of AIDS & STD 2018;173(02):111 [FREE Full text]
3. Ge L, Li D, Li P, Guo W, Cui Y. Population specific sentinel surveillance for HIV infection, syphilis and HCV infection in China, during 2010-2015. Disease Surveillance 2017;2:111-117 [FREE Full text]
4. Chow EPF, Wilson DP, Zhang L. The rate of HIV testing is increasing among men who have sex with men in China. HIV Med 2012 May;13(5):255-263 [FREE Full text] [doi: 10.1111/j.1468-1293.2011.00974.x] [Medline: 22252151]
5. China Internet Network Information Center (CNNIC). 2018. The 42th China Internet Development Statistics Report URL: http://www.cnnic.net.cn/hlwfzyj/ [accessed 2019-03-30] [WebCite Cache ID 77FeztkKR]
6. Pan S, Xu J, Han X, Zhang J, Hu Q, Chu Z, et al. nternet-based sex-seeking behavior promotes HIV infection risk: a 6-year serial cross-sectional survey to MSM in Shenyang, China. Biomed Res Int 2016;2016:2860346 [FREE Full text] [doi: 10.1155/2016/2860346] [Medline: 28105415]
7. Tang W, Best J, Zhang Y, Liu F, Tso LS, Huang S, et al. Gay mobile apps and the evolving virtual risk environment: a cross-sectional online survey among men who have sex with men in China. Sex Transm Infect 2016 Nov;92(7):508-514 [FREE Full text] [doi: 10.1136/sextrans-2015-052469] [Medline: 27288414]
8. Zou H, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
9. Grosskopf NA, LeVasseur MT, Glaser DB. Use of the Internet and mobile-based "apps" for sex-seeking among men who have sex with men in New York City. Am J Mens Health 2014 Nov;8(6):510-520. [doi: 10.1177/1557988314527311] [Medline: 24658284]
10. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874 [FREE Full text] [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
11. Aronson ID, Bennett A, Marsch LA, Bania TC. Mobile technology to increase HIV/HCV testing and overdose prevention/response among people who inject drugs. Front Public Health 2017;5:217 [FREE Full text] [doi: 10.3389/fpubh.2017.00217] [Medline: 28879174]
12. Mimiaga MJ, Thomas B, Biello K, Johnson BE, Swaminathan S, Navakodi P, et al. A pilot randomized controlled trial of an integrated in-person and mobile phone delivered counseling and text messaging intervention to reduce HIV transmission risk among male sex workers in Chennai, India. AIDS Behav 2017 Nov;21(11):3172-3181 [FREE Full text] [doi: 10.1007/s10461-017-1884-5] [Medline: 28831618]
13. Mehta K, Kumar AM, Chawla S, Chavda P, Selvaraj K, Shringarpure KS, et al. 'M-TRACK' (mobile phone reminders and electronic tracking tool) cuts the risk of pre-treatment loss to follow-up by 80% among people living with HIV under programme settings: a mixed-methods study from Gujarat, India. Glob Health Action 2018;11(1):1438239 [FREE Full text] [doi: 10.1080/16549716.2018.1438239] [Medline: 29482468]

14. Kanters S, Park JJ, Chan K, Socias ME, Ford N, Forrest JI, et al. Interventions to improve adherence to antiretroviral therapy: a systematic review and network meta-analysis. Lancet HIV 2017 Dec;4(1):e31-e40. [doi: 10.1016/S2352-3018(16)30206-5] [Medline: 27863996]
15. Wu Z, Xu J, Liu E, Mao Y, Xiao Y, Sun X, National MSM Survey Group. HIV and syphilis prevalence among men who have sex with men: a cross-sectional survey of 61 cities in China. Clin Infect Dis 2013 Jul;57(2):298-309 [FREE Full text] [doi: 10.1093/cid/cit210] [Medline: 23580732]
16. Liu M, Ma B, Mi G. A comparison of web-based survey national sentinel surveillance among MSM in China. Chin J AIDS STD 2016;22(6):441-443 [FREE Full text] [doi: 10.13419/j.cnki.aids.2016.06.12]
17. World Health Organization. 2016. Guidelines on HIV self-testing and partner notification URL: https://www.who.int/hiv/pub/vct/hiv-self-testing-guidelines/en/ [accessed 2019-03-30] [WebCite Cache ID 77Ffgr9SB]
18. Zhong F, Tang W, Cheng W, Lin P, Wu Q, Cai Y, et al. Acceptability and feasibility of a social entrepreneurship testing model to promote HIV self-testing and linkage to care among men who have sex with men. HIV Med 2017 Dec;18(5):376-382 [FREE Full text] [doi: 10.1111/hiv.12437] [Medline: 27601301]
19. Hao Y, Sun X, Xia G, Jiao Y, Song L, Shen J, et al. Program in HIV/AIDS prevention and treatment since implementing the Four Frees and One Care AIDS policy in China. Chin J AIDS STD 2014;20(4):228-232 [FREE Full text]
20. National Guideline for Detection of HIV/AIDS. Beijing: China CDC; Dec 2015:38-40.
21. Qin Y, Liu F, Tang W, Tang S, Liu C, Mao J, et al. HIV self-testing among high-risk men who have sex with men in China: a cross-sectional study. Lancet 2016 Oct;388:S76 [FREE Full text] [doi: 10.1016/S0140-6736(16)32003-7]
22. Xu J, An M, Han X, Jia M, Ma Y, Zhang M, et al. Prospective cohort study of HIV incidence and molecular characteristics of HIV among men who have sex with men(MSM) in Yunnan Province, China. BMC Infect Dis 2013 Jan 04;13:3 [FREE Full text] [doi: 10.1186/1471-2334-13-3] [Medline: 23286213]
23. Zhang W, Xu J, Zou H, Zhang J, Wang N, Shang H. HIV incidence and associated risk factors in men who have sex with men in Mainland China: an updated systematic review and meta-analysis. Sex Health 2016 May 26;13(4):373-382. [doi: 10.1071/SH16001] [Medline: 27294689]
24. Xu J, Zhang C, Hu Q, Chu Z, Zhang J, Li Y, et al. Recreational drug use and risks of HIV and sexually transmitted infections among Chinese men who have sex with men: Mediation through multiple sexual partnerships. BMC Infect Dis 2014 Dec 02;14:642 [FREE Full text] [doi: 10.1186/s12879-014-0642-9] [Medline: 25443542]
25. Xu J, Zhang M, Brown K, Reilly K, Wang H, Hu Q, et al. Syphilis and HIV seroconversion among a 12-month prospective cohort of men who have sex with men in Shenyang, China. Sex Transm Dis 2010 Jul;37(7):432-439 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181d13eed] [Medline: 20375928]
26. Ross MW, Tikkanen R, Månsson SA. Differences between internet samples and conventional samples of men who have sex with men: implications for research and HIV interventions. Soc Sci Med 2000 Sep;51(5):749-758. [Medline: 10975234]
27. Ross MW, Månsson SA, Daneback K, Cooper A, Tikkanen R. Biases in internet sexual health samples: comparison of an internet sexuality survey and a national sexual health survey in Sweden. Soc Sci Med 2005 Jul;61(1):245-252. [doi: 10.1016/j.socscimed.2005.01.019] [Medline: 15847976]
28. Wang C, Mollan KR, Hudgens MG, Tucker JD, Zheng H, Tang W, et al. Generalisability of an online randomised controlled trial: an empirical analysis. J Epidemiol Community Health 2018 Dec;72(2):173-178 [FREE Full text] [doi: 10.1136/jech-2017-209976] [Medline: 29183956]

Abbreviations

AHF: AIDS Healthcare Foundation
aOR: adjusted odds ratio
CBO: community-based organization
CDC: Centers for Disease Control and Prevention
IQR: interquartile range
MSM: men who have sex with men

Edited by G Eysenbach; submitted 07.08.18; peer-reviewed by Z Wu, D Katz; comments to author 12.10.18; revised version received 06.11.18; accepted 24.03.19; published 15.05.19.

Please cite as:
Jin X, Xu J, Smith MK, Xiao D, Rapheal ER, Xiu X, Ding Z, Zhang Y, Jie Y, Liao Y, Cao N, Wu H, Bao Y
An Internet-Based Self-Testing Model (Easy Test): Cross-Sectional Survey Targeting Men Who Have Sex With Men Who Never Tested for HIV in 14 Provinces of China
J Med Internet Res 2019;21(5):e11854
URL: https://www.jmir.org/2019/5/e11854/
doi:10.2196/11854
PMID:31094339

Original Paper

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

Janine Clarke[1,2], MPsychol (Clin), PhD; Samineh Sanatkar[2,3], PhD; Peter Andrew Baldwin[1,2], PhD; Susan Fletcher[4], PhD; Jane Gunn[4], MD, PhD; Kay Wilhelm[2], MBBS, MD, FRANZCP; Lesley Campbell[5], MBBS, FRCP, FRACP AM; Nicholas Zwar[6], MBBS, PhD; Mark Harris[7], MBBS, PhD; Helen Lapsley[2], PhD; Dusan Hadzi-Pavlovic[2], BSc, MPsychol; Helen Christensen[1], PhD; Judy Proudfoot[1,2], PhD

[1]Black Dog Institute, Sydney, Australia
[2]School of Psychiatry, UNSW Sydney, Sydney, Australia
[3]The Black Dog Institute, Sydney, Australia
[4]Department of General Practice, University of Melbourne, Melbourne, Australia
[5]Diabetes and Metabolism Division, Garvan Institute of Medical Research, Sydney, Australia
[6]School of Medicine, University of Wollongong, Wollongong, Australia
[7]Centre for Primary Health Care and Equity, UNSW Sydney, Sydney, Australia

Corresponding Author:
Peter Andrew Baldwin, PhD
Black Dog Institute
Hospital Road
Randwick
Sydney,
Australia
Phone: 61 0293824521
Email: p.baldwin@unsw.edu.au

Abstract

Background: Depressive symptoms are common in people with type 2 diabetes mellitus (T2DM). Effective depression treatments exist; however, access to psychological support is characteristically low. Web-based cognitive behavioral therapy (CBT) is accessible, nonstigmatizing, and may help address substantial personal and public health impact of comorbid T2DM and depression.

Objective: The aim of this study was to evaluate the Web-based CBT program, myCompass, for improving social and occupational functioning in adults with T2DM and mild-to-moderate depressive symptoms. myCompass is a fully automated, self-guided public health treatment program for common mental health problems. The impact of treatment on depressive symptoms, diabetes-related distress, anxiety symptoms, and self-care behavior was also examined.

Methods: Participants with T2DM and mild-to-moderate depressive symptoms (N=780) were recruited online via Google and Facebook advertisements targeting adults with T2DM and via community and general practice settings. Screening, consent, and self-report scales were all self-administered online. Participants were randomized using double-blind computerized block randomization to either myCompass (n=391) for 8 weeks plus a 4-week tailing-off period or an active placebo intervention (n=379). At baseline and postintervention (3 months), participants completed the Work and Social Adjustment Scale, the primary outcome measure. Secondary outcome measures included the Patient Health Questionnaire-9 item, Diabetes Distress Scale, Generalized Anxiety Disorder Questionnaire-7 item, and items from the Self-Management Profile for Type 2 Diabetes.

Results: myCompass users logged in an average of 6 times and completed an average of .29 modules. Healthy Lifestyles users logged in an average of 4 times and completed an average of 1.37 modules. At baseline, mean scores on several outcome measures, including the primary outcome of work and social functioning, were near to the normal range, despite an extensive recruitment process. Approximately 61.6% (473/780) of participants completed the postintervention assessment. Intention-to-treat analyses revealed improvement in functioning, depression, anxiety, diabetes distress, and healthy eating over time in both groups. Except for blood glucose monitoring and medication adherence, there were no specific between-group effects. Follow-up analyses suggested the outcomes did not depend on age, morbidity, or treatment engagement.

Conclusions: Improvement in social and occupational functioning and the secondary outcomes was generally no greater for myCompass users than for users of the control program at 3 months postintervention. These findings should be interpreted in light of near-normal mean baseline scores on several variables, the self-selected study sample, and sample attrition. Further attention to factors influencing uptake and engagement with mental health treatments by people with T2DM, and the impact of illness comorbidity on patient conceptualization and experience of mental health symptoms, is essential to reduce the burden of T2DM.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12615000931572; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=368109&isReview=true (Archived by WebCite at http://www.webcitation.org/7850eg8pi)

(J Med Internet Res 2019;21(5):e12246) doi:10.2196/12246

KEYWORDS

type 2 diabetes; depression; internet

Introduction

Background

Type 2 diabetes mellitus (T2DM) is a global public health problem, affecting more than 370 million people worldwide [1,2]. The disease is the fastest growing chronic condition in Australia, with approximately 1 million people diagnosed and up to 500,000 undiagnosed [3]. In addition to increased risk of adverse health consequences (including macro and microvascular diseases), people diagnosed with T2DM are at greater risk of psychiatric and neurodegenerative disorders, as well as physical disability and functional decline, than the general population [4,5]. With the prediction that global prevalence rates will increase to nearly 600 million by 2035 [3], T2DM presents a substantial escalating challenge for health care delivery and public health infrastructure in Australia and around the globe.

Depression is frequently comorbid with T2DM, affecting up to 40% of people with the disease [6,7]. Comorbid depression is associated with greater disease morbidity, mortality, and health care costs. These relationships are likely because of poor occupational and social functioning, reduced regimen adherence, poor quality of life, and increased need for outpatient and inpatient health services in people with both disorders [6,8,9]. The interrelationship between T2DM and depressive symptoms [10] further complicates the situation, with research showing that each condition increases the negative functional and health impacts of the other [6]. A population-based early intervention program for depression in people with T2DM is urgently needed to reduce both the substantial personal burden and public health impact of these comorbid conditions.

International diabetes treatment guidelines now recommend regular screening for depressive symptoms, with referral to appropriate psychological treatments as part of standard diabetes care [11]. Cognitive behavioral therapy (CBT) is the most widely validated treatment for depression. Recent reviews show positive impacts of face-to-face CBT on depressive symptoms, quality of life, fasting glucose levels [12], and self-care [13] in people with T2DM. However, the majority of people with T2DM are cared for in the primary care setting, and depression screening in primary care is highly variable. Within a standard-length consultation, it is difficult for general practitioners to focus on mental health as well as glycemic control and prevention of diabetes complications; in addition, many patients are reluctant to accept a referral for face-to-face psychotherapeutic interventions [14]. Concerns about confidentiality, stigma, treatment cost, and time and lifestyle constraints are further barriers to patients' help seeking, and psychological services are often scarce in rural and remote areas. In addition, the sizeable base rates of comorbid T2DM and depressive symptoms [15] mean that improved screening and diagnosis will likely place further pressure on face-to-face psychological services, many of which already struggle to meet the needs of their community [16,17]. Therefore, testing more flexible and scalable models of mental health service delivery for people with T2DM is necessary [12].

Internet delivery of evidence-based psychological therapies is now established as apopular, clinically effective, and cost-efficient means of upscaling access to psychological treatments. In people with T2DM, both diabetes-specific [18,19] and generic (ie, not diabetes sensitive but with therapist guidance) [20] depression interventions have yielded positive treatment effects for both depression symptoms and diabetes-related distress. There is also some suggestion that these may provide a potentially cost-effective solution to the substantial disability resulting from depression-diabetes comorbidity [21]. The current literature suggests that such interventions are most effective for people experiencing mild-to-moderate depressive symptoms [22]. As subclinical depression is more prevalent in T2DM than severe depression [23,24], testing of low-intensity Web-based interventions in T2DM patients with mild-to-moderate depressive symptoms is a public health priority.

myCompass is a broadly available Web-based CBT program available free of charge in Australia. A previous trial reported symptom reduction of mild-to-moderate depression in the general population compared with a placebo condition [25]. Despite being self-guided and transdiagnostic (ie, not diabetes specific) in therapeutic content, an uncontrolled feasibility study suggested that myCompass may potentially improve functional and mental health outcomes in people with diabetes [26]. As such, it may provide a more flexible, scalable alternative to disease-specific interventions and may benefit people with T2DM whose depressive symptoms do not warrant more intensive face-to-face intervention and/or who wish to manage their mental health themselves.

Objectives

The primary aim of this randomized controlled trial (RCT) was to evaluate the effectiveness of myCompass in improving work and social functioning—a major contributor to the T2DM burdens [8]—in adults with T2DM and mild-to-moderate depressive symptoms. On the basis of our feasibility data, we hypothesized that participants using the myCompass intervention would show improved scores on a self-report measure of work and social functioning relative to participants using an active placebo program. The secondary aim was to determine the effectiveness of myCompass for improving a range of diabetes-specific outcomes linked with depression and also shown to impact health outcomes in T2DM, including diabetes-related distress, diabetes self-care, and anxiety symptoms. The inclusion of these variables enabled us to examine whether treatment of depression with a public health intervention was capable of impacting diabetes-specific outcomes.

Methods

Design

This study is the primary outcomes evaluation of a 2-arm RCT called *SpringboarD*. The full protocol for the SpringboarD Trial is published elsewhere [27]. In the full SpringboarD trial, outcomes will be assessed at baseline and 3, 6, and 12 months postrandomization. This paper reports data from baseline and 3 months postintervention. Participants in the active and control groups had uninterrupted access to usual treatment for their diabetes throughout the study. This study was approved by Human Research Ethics Committee (HREC) at University of New South Wales (UNSW) Sydney (HREC 15090) and registered with the Australia and New Zealand Clinical Trials Registry (ACTRN12615000931572).

Participants and Setting

The study utilized a broadly-based recruitment strategy, including offline and online recruitment methods, to enroll the required sample size for sufficient statistical power. Recruitment began in September 2015 and continued until November 2017. Offline recruitment occurred via letters from participating general practices in New South Wales (NSW) and Victoria to their patients with T2DM, distribution of promotional materials in general practice settings throughout NSW and Victoria (eg, study flyers and posters), and print advertisements in national diabetes-related publications.

Online recruitment involved a range of techniques targeting health professionals and individuals. Health professionals were targeted through member associations such as the Australian Association of Practice Managers, the Australian Diabetes Educators Association, the Australian Association of Practice Nurses and the Australian Primary Health Care Nurses Association. Contact was via electronic direct mail, informing members of the study and inviting them to refer appropriate candidates to the SpringboarD website for screening. Individuals were recruited via Google and Facebook advertisements targeting an Australia-wide audience aged 18 years and over, with interest in *diabetes mellitus type 2 awareness*, *diabetes type 2 awareness* and/or *diabetes awareness*. Advertisements provided a click-through link to a dedicated page on Black Dog Institute's website from which the SpringboarD website could be accessed for information and screening. Potential candidates were also contacted via email through Black Dog Institute's Volunteer Research Register (VRR) and the Sax Institute's *45 and Up Study*. The VRR distributed 2378 emails to research volunteers aged between 18 and 75 years with a history of depressive symptoms. The 45 and Up Study is a large, longitudinal population-based cohort study of healthy aging in NSW, Australia, described in detail elsewhere [28]. The 267,153 study participants are considered largely representative of the Australian population. From June 2017 to September 2017, 4175 participants from the *45 and Up Study* aged between 45 and 75 years with self-reported diabetes were emailed invitations to participate in the SpringboarD study. All promotional material directed potential candidates to a secure study-specific website [29], which guided interested participants through the consent process and provided instructions regarding completion of the screening questionnaires.

Eligibility Criteria

People were eligible to take part if they were aged 18 to 75 years, diagnosed with T2DM by a physician, screened positive for depression on the self-report 2-item Patient Health Questionnaire (PHQ-2; ie, ≥2) [30], and had access to an internet-enabled device (eg, computer, tablet, and/or mobile phone). People who screened positive for depression completed the 9-item PHQ (PHQ-9) [31] at screening so that the level of symptom severity could be determined. Exclusion criteria included insufficient English literacy, extremely severe depressive symptoms on the full PHQ-9 (score >19), probable psychosis (measured by the psychosis screener developed for the Australian National Mental Health and Wellbeing Survey) [32], currently receiving face-to-face counseling or therapy for depression, change in antidepressant medication within the previous 2 months, high suicide risk (assessed by the PHQ-9 Item-9), and previous use of the myCompass program.

Eligible participants received immediate feedback of their suitability for the study via the website and were provided onscreen instructions for completion of the baseline assessment. Approximately 25 min were required to complete the screening and baseline assessment. All data captured by the study website during the screening, baseline, and postintervention assessments were stored in password-protected files on secure servers that comply with UNSW HREC privacy regulations regarding online data collection.

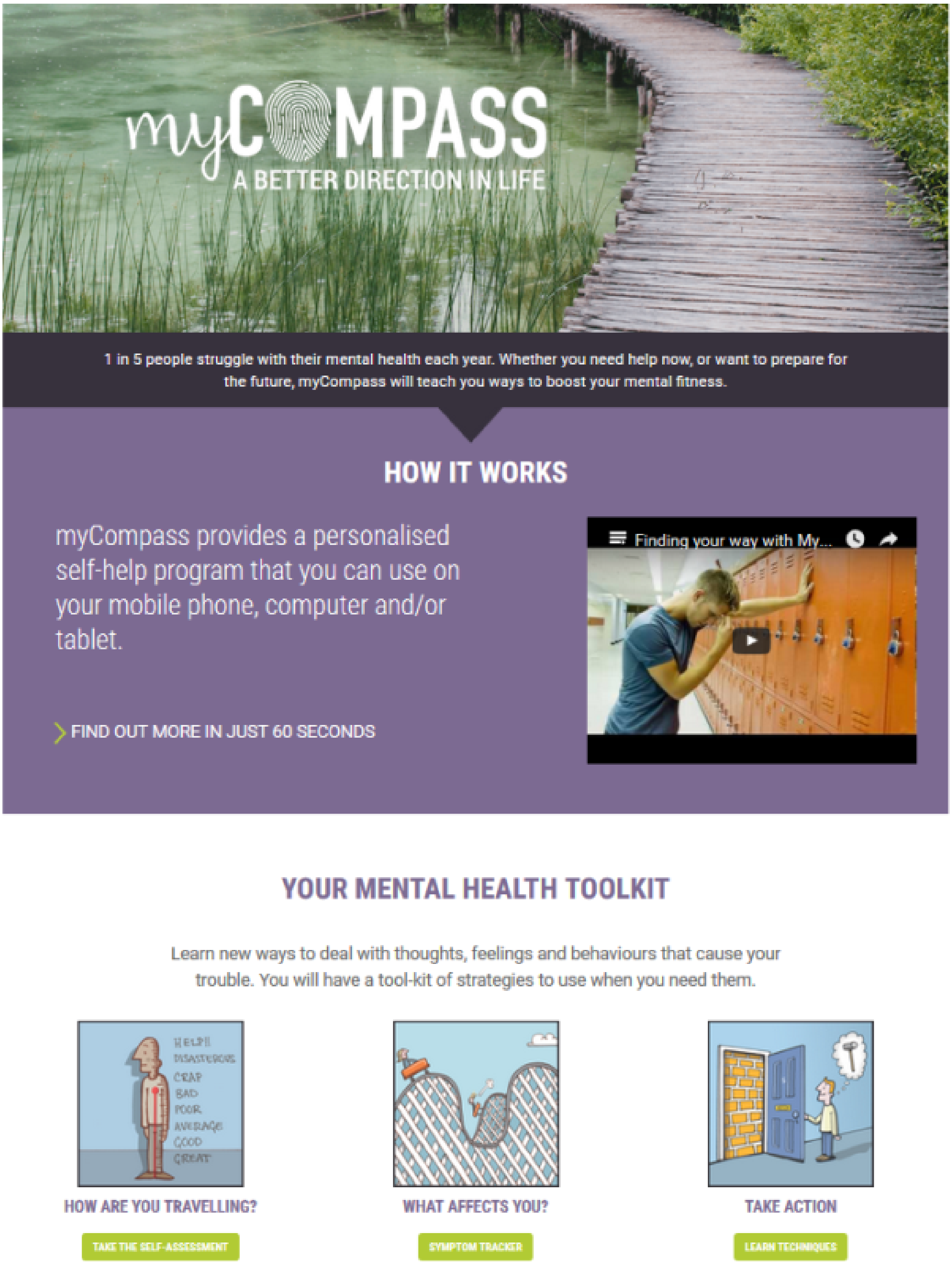
Randomization

Computerized block randomization with blocks of 8 was undertaken to assign participants to the intervention and control condition at a 1:1 allocation ratio. Randomization was initiated immediately after the completion of the baseline assessment by the automated randomization system built into Black Dog Institute's study management software. Allocation was concealed from participants and researchers. Though participants were ineligible if they had previously used myCompass, some participants allocated to the intervention condition may have independently accessed public information about myCompass over the course of the trial and therefore become aware that they were allocated to an existing Web-based therapy program.

Interventions

Active Intervention (myCompass)

myCompass (mycompass.org.au) is a public health fully automated, self-guided CBT intervention that users complete in their own time and at their own pace on their computer and/or mobile phone (see Figures 1 and 2).



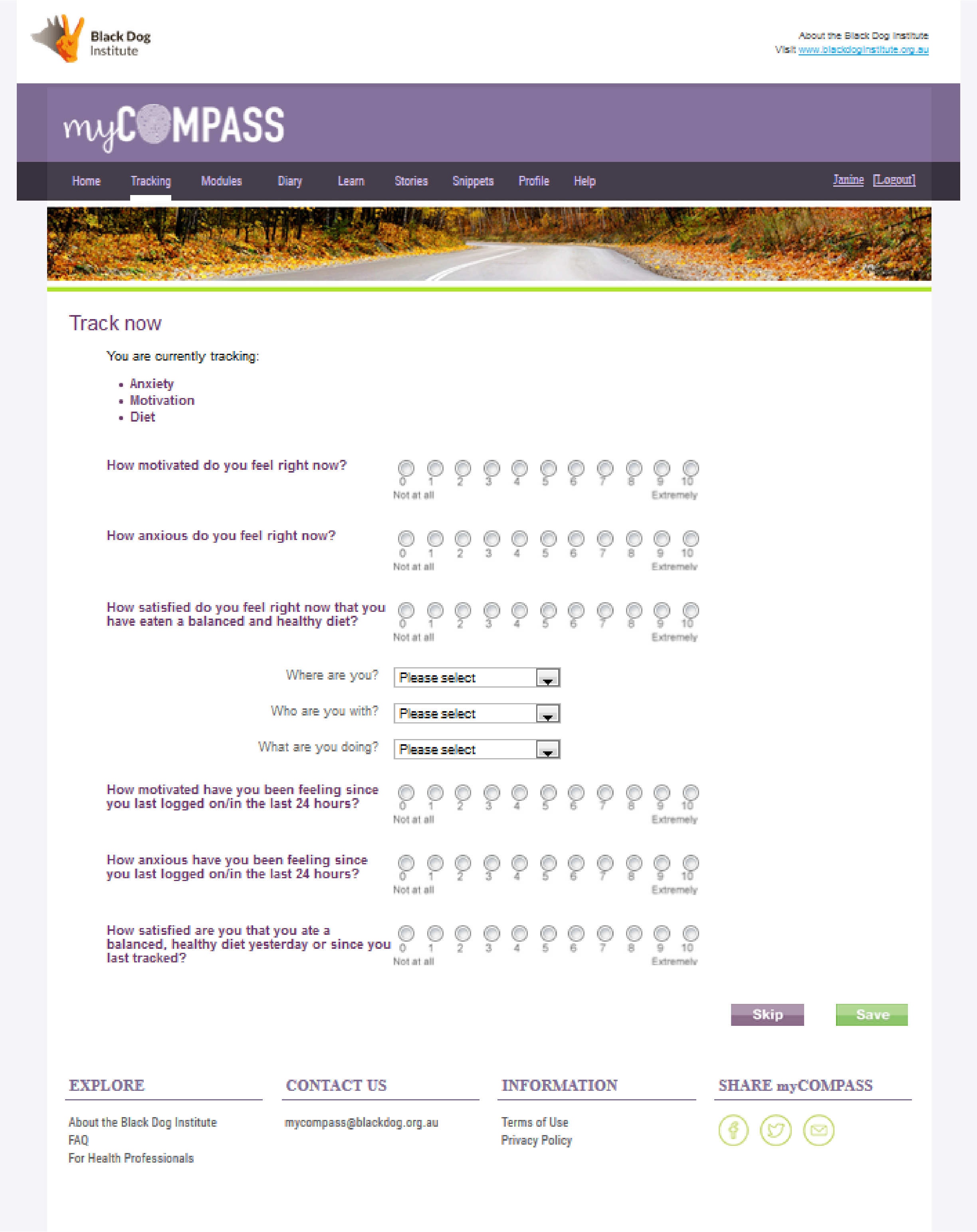
1 in 5 people struggle with their mental health each year. Whether you need help now, or want to prepare for the future, myCompass will teach you ways to boost your mental fitness.

myCompass provides a personalised self-help program that you can use on your mobile phone, computer and/or tablet.

Learn new ways to deal with thoughts, feelings and behaviours that cause your trouble. You will have a tool-kit of strategies to use when you need them.

Figure 1. Screenshot of the myCompass landing page.

Figure 2. Screenshot of the myCompass self-monitoring page.



Track now

You are currently tracking:

- Anxiety
- Motivation
- Diet

How motivated do you feel right now? ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 (Not at all – Extremely)

How anxious do you feel right now? ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 (Not at all – Extremely)

How satisfied do you feel right now that you have eaten a balanced and healthy diet? ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 (Not at all – Extremely)

Where are you? Please select

Who are you with? Please select

What are you doing? Please select

How motivated have you been feeling since you last logged on/in the last 24 hours? ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 (Not at all – Extremely)

How anxious have you been feeling since you last logged on/in the last 24 hours? ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 (Not at all – Extremely)

How satisfied are you that you ate a balanced, healthy diet yesterday or since you last tracked? ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 (Not at all – Extremely)

The program offers a personalized treatment plan based on an assessment of user symptoms, either at registration or at any time during program use [25,27]. The core program consists of 12 interactive CBT modules and 20 cognitive and/or behavioral variables for self-monitoring. For maximum outcomes, users are recommended to complete 3 modules and 3 self-monitoring variables. The myCompass program offers flexibility for users to select their own CBT modules and self-monitoring variables, or, if they prefer, algorithm-based guidance. It provides access to a range of other resources including short message service (SMS) text messaging and/or email self-monitoring reminders, home practice activities to facilitate skill generalization, mental health care tips and motivational statements delivered by email/SMS text messaging, and graphical reporting of self-monitoring data. Registering to use myCompass is free, and users are not billed for the SMSs they receive.

Participants randomized to the myCompass arm were provided access to the full program for 8 weeks. The program recommended that users complete 3 CBT modules and self-monitor up to 3 symptoms or behaviors. A 4-week tailing-off period followed, in which only the symptom monitoring function was accessible. Studies have shown that the use of Web-based interventions is improved when users receive program feedback that is personalized in its content [33]. For this reason, myCompass users receive automated and personalized feedback via email about their use of the program's self-monitoring and module functions in weeks 1, 3, 5, and 7.

myCompass user privacy is managed by a password-protected log-on and by ensuring that user-generated data (eg, self-monitoring ratings) are not stored on the user's device but are instead transferred via the internet using secure sockets layer protocols (which encrypt transmitted data by rendering data unreadable to anyone other than the intended recipient) and by storing the data in secure servers. Participant user data were identified in myCompass using email details provided during study registration. Once extracted, data were deidentified and stored in a password-protected file.

Placebo Intervention (Healthy Lifestyles)

The placebo control program, *Healthy Lifestyles*, was adapted from a control program used in previous studies by members of the research team [34] to replicate the mode of delivery and key functionality of myCompass, but without the therapeutic content. The program contains 12 modules that deliver health and lifestyle information across a range of topics (eg, skin care and mobile phone hygiene), interactive exercises, and the potential for program tailoring via a brief survey completed at registration. Program users received an email at weeks 1, 3, 5, and 7, containing a brief reminder to log into the program but no feedback about their program use. They also received a weekly SMS text message containing a fact relevant to the content of Healthy Lifestyles for the first 4 weeks of the intervention period to match the SMS functionality of myCompass. The Healthy Lifestyles program was designed to have high face validity as a health and lifestyle intervention without any symptomatic benefit [34]. Participants had full access to the Healthy Lifestyles program for 8 weeks. The core features of each program are presented in Table 1.

Outcome Measures

A summary of the measures obtained from participants at baseline and 3 months postintervention is presented in Table 2.

Table 1. A comparison of the core features of the myCompass and Healthy Lifestyles programs.

Feature	MyCompass	Healthy Lifestyles
Modality	Website	Website
Symptom tracking	Yes: track up to 3 symptoms	Not available
Usage reminders	Yes: automated short message service text message or email	Yes: automated email
Number of modules	12: including interactive learning activities	12: including interactive learning activities
Module content	Cognitive and behavioral strategies for mood, anxiety, stress, and mental well-being	Generic health literacy information

Table 2. Measures obtained at each assessment phase.

Measure	Time 1	Time 2
Demographic and disease-related information		
Demographic data (eg, age and gender)	X	—[a]
Disease-relevant data (eg, age at diagnosis and treatment)	X	X
Mental health history (eg, previous diagnoses and help-seeking)	X	—
Primary outcome		
Work and Social Adjustment Scale [31]	X	X
Secondary outcome		
Patient Health Questionnaire-9 [26]	X	X
Diabetes Distress Scale [33]	X	X
Generalized Anxiety Disorder Scale [34]	X	X
Self-management Profile for Type 2 Diabetes Scale (behavior items only) [35]	X	X
Glycosylated hemoglobin	X	—
Days out of role [36]	X	X
Health service utilization	X	X

[a] Not applicable.

Primary Outcome

The primary outcome for the trial was a between-group difference in work and social functioning, measured by the Work and Social Adjustment Scale (WSAS). The WSAS is a psychometrically sound measure of daily functioning across 5 domains, including work, social leisure activities, private leisure activities, home management, and personal relationships [35,36]. Scores range from 0 to 40, with higher scores indicating poorer functioning.

Secondary Outcomes

Symptoms of depression and anxiety were measured by PHQ-9 [31] and the 7-item Generalized Anxiety Disorder scale (GAD-7) [37], respectively. Both scales are well validated, are used widely as screening tools in primary care settings and are frequently included as outcome measures in studies of Web-based interventions [38,39]. Both scales use cut-off scores of 5, 10, and 15 to identify people with mild, moderate, and moderately severe symptoms, respectively.

Emotional adjustment to diabetes—or *diabetes-related distress* —was measured by the 17-item Diabetes Distress Scale (DDS) [40]. The DDS yields a total score plus 4 subscale scores assessing the perceived emotional burden of diabetes, along with physician-related distress, regimen-related distress, and interpersonal distress. DDS total and subscale scores are calculated as the average across all items in the scale/subscale and range from 1 to 6, with higher scores indicating greater distress. A score of >3 indicates clinically relevant distress.

Diabetes self-care was assessed using a subset of items from the Self-Management Profile for Type 2 Diabetes (SMP-T2D). The SMP-T2D was designed for use in clinical trials to assess level and perceived ease of performance of specific T2DM regimen behaviors. Perceived coping and confidence dealing with diabetes and ease of weight management are also assessed, and all constructs demonstrate appropriate internal consistency, validity, and sensitivity [41]. To avoid redundancy (eg, the DDS also asks about difficulties with self-care activities), to reduce participant assessment burden, and as we were primarily interested in learning about participants' behavioral engagement in self-care activities, we only administered items that measured the level of self-care across 4 patient behavior domains: blood glucose monitoring, medication adherence, healthy eating, and physical activity. Scores in each behavior domain are converted to a percentage of the previous week spent engaging in a particular self-care behavior. Higher scores indicate greater time spent on self-care [41].

Additional Measurements

At baseline, we collected disease-related (eg, age of onset and treatment regimen), sociodemographic (eg, age, gender, education, and occupation) and mental health history data (eg, service use and previous diagnoses). We also obtained participants' most recent glycosylated hemoglobin (HbA_{1c}) from their medical records as an indicator of their overall blood glucose management before study commencement. Recent service utilization for physical and mental health problems was assessed at baseline and postintervention, along with days out of role, defined as the number of days in the previous 30 days that a participant was unable to perform work or normal activities because of problems with his/her physical or mental health [42].

At the conclusion of the intervention period, program engagement data were extracted for myCompass and Healthy Lifestyles, including frequency of log-in, number of modules started and completed, and self-monitoring frequency (myCompass only).

Sample Size

Power calculations indicated that a study sample of 600 was needed to detect a minimum difference of .3 standard deviations between groups in mean change in scores on the WSAS at 3 months post intervention, with power of 80%, 2-tailed alpha=.05, and assuming an attrition rate of 40%. Owing to early indications of a higher attrition rate at postintervention than anticipated, a further 180 people were recruited into the study to ensure sufficient power to test the research hypotheses.

Analyses

Primary analyses employed an intention-to-treat (ITT) approach using mixed-model repeated measures analyses (MMRM) computed within the Mixed procedure of SPSS version 23 (IBM Corp.). MMRM makes use of all available data to obtain parameter estimates and is widely recognized as an appropriate strategy for analyzing incomplete datasets [43]. In this study, restricted maximum likelihood estimation was employed to estimate model parameters, and error degrees of freedom were calculated using Satterthwaite approximation. In line with Fairclough and Helms' [44] recommendation that the covariance structure be restricted in situations of high attrition, analyses assumed a compound symmetric structure. Repeated measures (Level 1) were nested within individuals (Level 2), and a random intercept was used at the individual level to account for intraindividual correlations on repeated measures.

In addition to the ITT analyses, we conducted completer analyses to examine the effects of treatment on those individuals who completed the study, defined as having provided complete data at 3-month follow-up. In these analyses, individuals with any missing data at 3-month follow-up were deleted case wise (myCompass n=175; Healthy Lifestyles n=151), and repeated-measures analyses of variance were conducted for each of the primary and secondary outcomes on the remaining participants. As the sample characteristics and treatment effects in the completer analyses did not differ from those of the ITT analyses, only the ITT results are reported.

Results

Sample Characteristics

Overall sample characteristics are presented in Table 3, and participant flow through the study is presented in Figure 3. Of the 6145 visits to the SpringboarD Project website, 3223 consented to Web-based screening, yielding 888 eligible participants who commenced the baseline assessment. The main reasons for ineligibility included the following: did not meet inclusion criteria for the presence of depressive symptoms, that is, a score of <2 on the PHQ-2 (52.07% [1021/1961]), currently receiving face-to-face mental health support (26.31% [516/1961]), and screening results indicating severe depression (9.33% [183/1961]). A total of 780 individuals completed the baseline assessment and were randomized. A total of 57 people subsequently withdrew study consent, leaving a final study sample of 723 individuals.

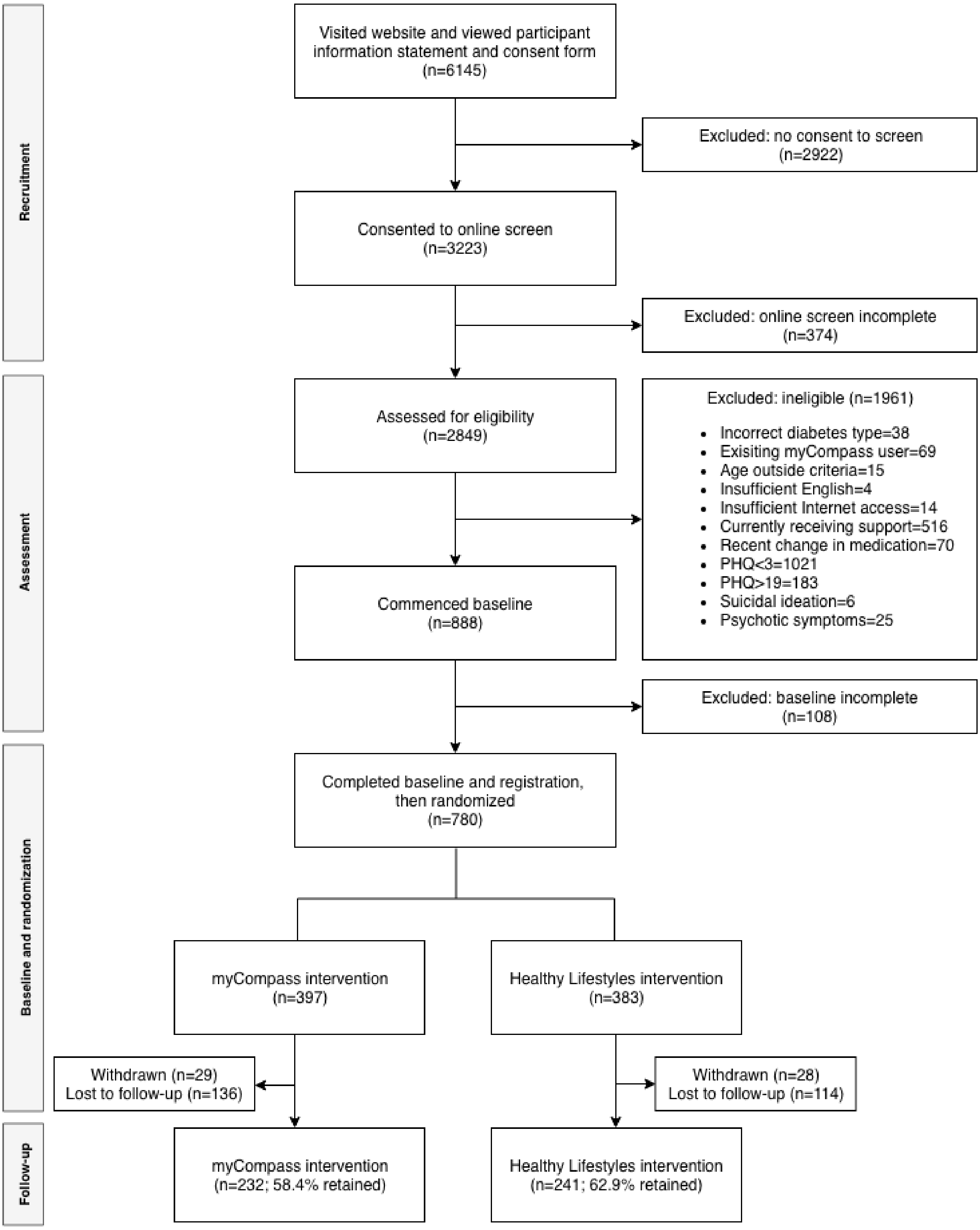
As shown in Table 3, groups were well matched between the trial arms. The myCompass group reported stable antidepressant use of somewhat longer duration than the Healthy Lifestyles group, with no other notable differences in demographics and mental health. The overall sample was predominantly female (68.8% [498/723]), married (55.7% [408/723]), employed at least part time (50.8% [367/723]), university educated (34.02% [246/723]), with an average age of 58 years (SD 10.35). More than 3 quarters of the sample (83.9% [607/723]) had at some time sought professional help for common mental health issues (eg, low mood, anxiety and/or stress), and almost half (43.8% [317/723]) had previously received at least one mental health diagnosis, the most frequent being depression (40.9% [296/723]).

Table 3. Sample characteristics for myCompass and Healthy Lifestyles groups.

Characteristic	myCompass (N=368)	Healthy Lifestyles (N=355)
Demographics		
Age, mean (SD)	57.7 (10.6)	57.7 (10.0)
Female (n=465), n (%)	229 (62)	236 (66)
Married (n=387), n (%)	204 (55)	183 (52)
Employed (n=351), n (%)	173 (47)	178 (50)
Education level, n (%)		
Secondary school or lower (n=220)	112 (30)	108 (30)
Trade certificate or diploma (n=270)	133 (36)	137 (39)
University undergraduate or more (n=233)	123 (33)	110 (31)
Mental health		
Lifetime history, n (%)		
Sought professional support for mental health (n=571)	296 (80)	275 (77)
Received mental health diagnosis (n=300)	155 (42)	145 (41)
Diagnosed with depressive symptoms or major depressive disorder (n=279)	143 (38)	136 (38)
Past 6 weeks, n (%)		
Sought professional support for mental health (n=113)	65 (18)	48 (14)
Current, n (%)		
Taking antidepressant medication (n=241)	125 (34)	116 (33)
Months using antidepressant medication	97.70 (94.72)	73.67 (56.21)[a]
Diabetes		
Age at diagnosis, mean (SD)	46.6 (11.1)	47.2 (10.9)
Diabetes treatment, n (%)		
Healthy eating (n=437)	230 (63)	207 (58)[a]
Physical activity (n=323)	176 (48)	147 (41)
Oral medication (n=583)	295 (80)	288 (81)
Insulin (n=216)	113 (31)	103 (29)
Exenatide (n=32)	21 (6)	11 (3)
Past 6 weeks, n (%)		
Visited general practitioner for diabetes (n=419)	218 (59)	201 (57)
Frequency of general practitioner visit	1.31 (.78)	1.37 (.71)
Hospitalized for diabetes (n=24)	13 (4)	11 (3)
Frequency of hospitalization for diabetes	1.46 (1.5)	1.36 (.9)

[a] Means differ significantly at *P*<.05.

Figure 3. Consolidated Standards for Reporting of Trials diagram of participant flow through the SpringboarD trial. PHQ: Patient Health Questionnaire.



Diabetes-related characteristics were also largely similar between the groups. The average age of onset of T2DM for the sample was 47 years (SD 10.84), and for the 60.0% (434/723) of participants for whom HbA_{1c} data were available from medical records, the mean value was 7.5% (SD 1.6).

Baseline scores for key outcome measures for both groups are presented in Table 4. Again, randomization was largely successful, with 2 exceptions noted for the WSAS ($F_{1,782}$=5.807; P=.02; d=0.17) and SMP-T2D Healthy Eating Domain ($F_{1,782}$=15.925; P=.03; d=0.29). However, when the relevant normative cut-offs were applied, baseline scores on the key outcome variables were near to the normal range. WSAS scores were at the lower end of the *significant impairment* range of 10 to 20 [35] (sample mean 12.89), PHQ-9 scores were slightly above the most common diagnostic cut-off of 10 [31] (sample mean 11.06), scores on the DDS were below the recommended clinical cut-off of ≥3 [40] (sample mean 2.54), and scores on the GAD-7 were also below the recommended clinical cut-off of ≥3 [37] (sample mean 7.45).

Intention-to-Treat Analyses (Mixed-Model Repeated Measures Analyses)

The observed scores for all study outcomes postintervention are presented in Table 4, and the group and time fixed effects from the MMRM analyses are shown in Table 5. Irrespective of the intervention received, all participants reported a significant reduction in WSAS scores across time. Similarly, all participants showed symptomatic improvements on the PHQ-9, DDS, and GAD-7 between baseline and postintervention, with no between-group differences detected posttreatment on any of these measures. The average mean reductions across time were generally small and did not meet any recommended threshold for clinical significance.

For the SMP-T2D, no group or time differences were observed for the Physical Activity Domain, and both groups reported significant improvement in the Healthy Eating Domain. Scores on the Blood Glucose Monitoring Domain and Medication Adherence Domain increased over time for the Healthy Lifestyles group relative to the myCompass group, with Healthy Lifestyles participants increasing time spent each week monitoring blood glucose and taking medication as recommended by approximately .4 to .5 days, compared with their myCompass counterparts.

Posthoc Treatment Effect Moderation Analyses

Level of glycemic control [45] and severity of depression symptoms [46] are potential moderators of treatment effects, and age remained a consistent predictor of overall symptomatic improvement in our sample. We therefore examined the possible moderating role of these variables by repeating our MMRM approach with an additional 3-way interaction term that included the moderator variable of interest (ie, Group×Time×Moderator). Depressive symptoms were dichotomized into “clinical” and “subclinical” groups on the basis of the PHQ-9 diagnostic cut-off of 12 recommended for individuals with diabetes [31]. Level of glycemic control was determined using conventional HbA_{1c} targets for T2DM, with >7% indicative of suboptimal control and ≤7% indicative of well-controlled diabetes [47]. Age was grouped using a median split. No moderator demonstrated a significant impact on treatment effects (all *P*>.05).

Table 4. Baseline and postintervention means (SDs) on key outcome variables for myCompass and Healthy Lifestyles groups.

Measure	Baseline, mean (SD)		Post (3 months), mean (SD)	
	myCompass (n=397)	Healthy Lifestyles (n=387)	myCompass (n=232)	Healthy Lifestyles (n=241)
WSAS[a]	13.64[b] (7.93)	12.23[b] (7.77)	12.24 (9.31)	10.82 (9.19)
PHQ-9[c]	11.30 (4.03)	10.86 (4.11)	8.68 (5.63)	8.24 (5.54)
GAD-7[d]	7.60 (4.10)	7.33 (4.20)	6.66 (4.73)	6.20 (4.58)
DDS[e]	2.56 (0.97)	2.56 (0.95)	2.24 (0.99)	2.18 (0.95)
SMP-BG[f]	49.61 (39.50)	45.55 (38.85)	49.25 (39.33)	55.83 (40.21)
SMP-MA[g]	87.50 (22.47)	87.20 (22.82)	86.34 (24.15)	92.04 (15.12)
SMP-HE[h]	49.96[b] (29.16)	48.26[b] (28.61)	53.52 (28.94)	54.87 (28.25)
SMP-PA[i]	53.44 (33.48)	48.26 (33.61)	55.98 (34.64)	52.44 (34.07)

[b] Means differ significantly at *P*<.05.

[a] WSAS: Work and Social Adjustment Scale.

[c] PHQ-9: Patient Health Questionnaire.

[d] GAD-7: Generalized Anxiety Disorder Scale.

[e] DDS: Diabetes Distress Scale.

[f] SMP-BG: Self-Management Profile for Type 2 Diabetes—Blood Glucose Monitoring.

[g] SMP-MA: Self-Management Profile for Type 2 Diabetes—Medication Adherence.

[h] SMP-HE: Self-Management Profile for Type 2 Diabetes—Healthy Eating.

[i] SMP-PA: Self-Management Profile for Type 2 Diabetes—Physical Activity.

Table 5. Mixed-model repeated measures analyses fixed effects for time, group, and time×group on key outcome variables.

Variable and effect	Beta[a]	SE	*t* test (*df*)	*P* value	95% CI
WSAS[b]					
Time	1.344	0.48	2.800 (544.48)	.005[c]	0.401 to 2.287
Group×Time	0.196	0.684	0.287 (550.76)	.77	–1.147 to 1.540
PHQ-9[d]					
Time	2.395	0.31	7.736 (596.32)	<.001[c]	1.787 to 3.003
Group×Time	0.144	0.44	0.326 (604.61)	.74	–0.720 to 1.008
DDS[e]					
Time	0.302	0.047	6.361 (518.94)	<.001[c]	0.208 to 0.395
Group×Time	–0.062	0.068	–0.919 (523.64)	.36	–0.194 to 0.070
GAD-7[f]					
Time	1.009	0.264	3.825 (553.83)	<.001[c]	0.490 to 1.526
Group×Time	–0.276	0.375	–0.736 (560.75)	.46	–1.013 to 0.460
SMP-BG[g]					
Time	–8.253	2.278	–3.623 (538.46)	<.001[c]	–12.727 to –3.778
Group×Time	8.98	3.248	2.764 (545.46)	.006[c]	2.598 to 15.360
SMP-MA[h]					
Time	–3.126	1.24	–2.520 (490.84)	.01[c]	–5.562 to –0.688
Group×Time	5.117	1.763	2.901 (496.65)	.004[c]	1.651 to 8.582
SMP-HE[i]					
Time	–5.704	1.604	–3.553 (527.04)	<.001[c]	–8.856 to –2.550
Group×Time	3.266	2.282	1.430 (532.61)	.15	–1.217 to 7.750
SMP-PA[j]					
Time	–3.440	2.228	–1.543 (559)	.12	–7.816 to 0.936
Group×Time	0.083	3.165	0.541 (566.75)	.98	–6.301 to 6.134

[a] Beta: unstandardized regression coefficient for the effect holding constant age and sex.

[b] WSAS: Work and Social Adjustment Scale.

[c] Significant at *P*<.05.

[d] PHQ-9: Patient Health Questionnaire-9 item.

[e] DDS: Diabetes Distress Scale.

[f] GAD-7: Generalized Anxiety Disorder Scale-7 item.

[g] SMP-BG: Self-Management Profile for Type 2 Diabetes—Blood Glucose Monitoring.

[h] SMP-MA: Self-Management Profile for Type 2 Diabetes—Medication Adherence.

[i] SMP-HE: Self-Management Profile for Type 2 Diabetes—Healthy Eating.

[j] SMP-PA: Self-Management Profile for Type 2 Diabetes—Physical Activity.

Study Attrition

Of the total trial participants, 59.2% (235/397) of the participants from the myCompass program and 63.3% (245/387) of the participants from the Healthy Lifestyles program provided at least one postintervention measure. A multivariate analysis of variance comparing nonresponders (ie, participants who did not provide any postintervention measures) with responders at baseline revealed differences on several key outcome variables. Nonresponders reported more severe depressive symptoms (*F*=8.362; *P*=.004; *d*=0.27), more severe anxiety symptoms (*F*=3.845; *P*=.05; *d*=0.18), greater diabetes-related distress (*F*=9.095; *P*=.003; *d*=0.28) and poorer medication adherence (*F*=6.564; *P*=.011; *d*=0.19). In a follow-up logistic regression predicting responder status from baseline WSAS, PHQ-9, GAD-7, DDS, and SMP-T2D scores, PHQ-9 (beta=–.058;

P=.021; OR [odds ratio] .943; 95% CI [.898, .991]), DDS (beta=–.221; *P*=.026; OR 1.029; 95% CI [.660, .974]) and WSAS (beta=.029; *P*=.016; OR .802; 95% CI [1.005, 1.054]) scores independently predicted nonresponse. Though effects were generally small, it appears that levels of distress and functioning had some influence on participants remaining in the study.

Program Use and Feedback

Participants in the treatment group logged in to myCompass an average of 6 times (SD 9.01; range 1-71), started a mean of 0.71 modules (SD 1.18; range 0 to 8), fully completed an average of .29 modules (SD .87; range 0-7), and monitored their symptoms an average of 2 times (SD 5.79; range 0-53). Participants in the control group logged in to Healthy Lifestyles an average of 4 times (SD 3.22; range 1 to 17), started a mean of 2.61 (SD 2.78; range 0 to 8), and completed an average of 1.37 modules (SD 2.24; range 0 to 8). Participants who logged in to their assigned Web-based program did not differ significantly from those who did not, except for a slightly higher anxiety score on the GAD-7 among myCompass users who logged in (*F*=10.76; *P*=.001; *d*=0.39). Adherence indices did not correlate with baseline scores on any primary or secondary outcome. Approximately 54.7% (127/232) of myCompass participants and 11.2% (27/241) of Healthy Lifestyles participants reported that, overall, they found their assigned Web-based program both easy and convenient to use.

Discussion

Principal Findings

This trial examined the effectiveness of a self-help, unguided Web-based CBT program (myCompass) for improving work and social functioning in people with T2DM and mild-to-moderate depressive symptoms compared with an active placebo intervention. Our ITT analyses showed that participants' work and social functioning improved significantly postintervention, irrespective of the intervention received. Significant improvements were also observed for both groups in depressive and anxiety symptoms, diabetes distress, and most aspects of diabetes self-management. Our Healthy Lifestyles group showed small but significant improvements in blood glucose monitoring and medication adherence over and above those observed for the myCompass group. As morbidity may influence the outcomes of depression treatments [46] and as age remained a significant outcome predictor in our models, we examined the potential moderator effects on treatment outcomes of age, depressive symptoms, and diabetes control. Analyses revealed that treatment effect estimates were not impacted by these variables.

The absence of differential treatment effects for work and social functioning and depressive symptoms for people with T2DM following treatment with myCompass was surprising in light of previous findings of accelerated symptom gains in myCompass users in the general community [27] and in our pilot study [26]. Although our findings also contrast with previous studies of Web-based depression interventions in people with diabetes [18,20], there are important points of differentiation between this study and previous diabetes trials that need to be considered in interpreting our findings.

Our recruitment methods targeted people with T2DM and mild-to-moderate depressive symptoms to investigate a public health rather than a clinical application of Web-based CBT; however, our final sample unexpectedly comprised a group with minimal symptoms of depression (47% of participants in the *minimal-to-mild* range on the PHQ-9). This contrasts with previous studies of people with diabetes that included only patients with more severe depressive symptoms [18] or those meeting criteria for major depressive disorder (Newby et al, 2017). Moreover, although 58% of our participants reported *impaired* work and social functioning at baseline, scores on the WSAS suggested that the level of impairment experienced by our participant group was again at the milder end of the disability spectrum. In other words, by all measures, our sample was only minimally impaired at baseline.

Even in the absence of true depression, people with chronic illness typically report poorer daily functioning than those without [48]. Consequently, our statistical analyses were based on a participant group with baseline scores on the WSAS and PHQ-9 that were potentially at *floor* for this cohort, and therefore, any improvement was likely to be minor. Also contrasting with previous studies [20], we observed a systematic pattern of attrition, such that increased severity of depressive symptoms was linked to study dropout. Systematic attrition of those in greatest need of intervention, and for whom treatment benefits were likely to be largest, may have further magnified any floor effects in our study [49] and precluded us from finding larger and more significant functional and symptom improvements following treatment with myCompass.

It is important to understand why and how increasing severity of depressive symptoms compromises study involvement in trials of Web-based CBT to inform the take-up of suitable interventions for this patient cohort and to maximize treatment effectiveness. One possibility is that attrition is related to increasing levels of amotivation, concentration difficulties, and behavioral inactivation that are hallmark symptoms of depressive disorders [50]. Alternatively, anhedonia (ie, reduced capacity to experience and anticipate enjoyment from activities) has been shown to compromise reward-seeking behavior and decision making [51] and might interfere with program uptake by rendering user behavior less responsive to reinforcement history (eg, motivational feedback) and anticipatory benefits (eg, information about program effectiveness) [52]. An accumulating body of evidence suggests that analysis of individual or *clusters* of depression symptoms may be necessary for understanding behavioral health outcomes in diabetes patients [53,54]. More precise understanding of relationships between depressive symptoms and indicators of program uptake and treatment benefit in T2DM is a challenge for future research.

Importantly, the myCompass program is a completely self-guided mental health intervention that is generic in content. It is, therefore, lower in treatment intensity than previously studied therapist-guided programs [20], and it lacks the disease specificity of diabetes-themed programs [18,19]. Fisher and colleagues [24] have previously suggested that scores on depression screeners may be less reflective of mood disturbance than general emotional and diabetes-specific distress, that is, distress experienced in response to the daily challenges and demands of living with diabetes. In line with this, we have recently published data showing that mildly depressed adults with diabetes maintain a very nuanced conceptualization of their mental health symptoms in which low mood, anxiety, and stress are generally perceived as features of the diabetes diagnosis (and warranted by contextual stressors) and not separate or comorbid conditions to be managed [55]. Lack of differentiation of mild-to-moderate depression symptoms from adjustment to diabetes has potential to attenuate the personal relevance and clinical effectiveness of interventions targeting broad CBT skills and techniques. At the same time, mild-to-moderate depression covers different levels of symptom severity that may differ in terms of responsiveness to internet-based CBT. Further exploration of the conceptualization of depression in T2DM and associated implications for depression treatment is warranted. Moreover, we recommend that researchers examine the nature of comorbidity between depression and diabetes-related distress. If discrete patterns of depressive symptoms and diabetes-related distress can be distinguished, then different interventions (eg, therapist guided versus unguided, generic versus diabetes specific) may be needed to maximize symptom improvement and increase social and occupational functioning in each.

Of the set of secondary outcomes, the only consistent group effect was for medication adherence, with a small but significant positive effect for the attention control. This finding was isolated and seemingly counterintuitive. However, it is possible that improvement in medication taking was prompted by the focus paid by the attention control program to healthy lifestyle behaviors.

Strengths and Limitations

Recruitment and retention in RCTs targeting comorbid physical and mental illness can be difficult [56]. Despite a lengthy and comprehensive community recruitment strategy, and our adoption of retention strategies that have been used successfully in other studies, participant enrollment and retention for our trial was challenging (and will be discussed in a future publication). Although we were successful in retaining a sample that afforded us sufficient power to test our research hypotheses, attrition in our study was systematic. As a result, near-normal scores on baseline variables may have influenced program engagement (that was generally low) and weakened tests of treatment effects. Moreover, our data are mostly relevant to people with T2DM experiencing mild functional impairment and mild depressive symptoms. Future studies may benefit from broadening eligibility symptom thresholds. Moreover, research designs that include greater program guidance and feedback (eg, module recommendations and homework follow-up) and provide more regular research support (eg, reminders and encouragement) may be more acceptable to participants with higher levels of impairment, for whom motivational factors are likely to impact ongoing study involvement. Importantly, recruitment and retention in future trials may benefit from further investigation of factors influencing individuals' decisions to decline trial participation or drop out post study commencement.

Although previous trials of Web-based CBT in people with diabetes have compared the active treatment with either treatment as usual [20] or waitlist control [18], a key strength of our design was the inclusion of an attention-placebo condition using a health literacy tool that had been validated elsewhere [25,34]. However, it is possible that lifestyle information had unexpected relevance to our chronically ill cohort, who were mostly experiencing only minor psychological distress. This may have resulted in increased engagement with the Healthy Lifestyles program and afforded benefit to our control condition in addition to the intended placebo effect. Further assessment with a waitlist and/or treatment-as-usual control group is required.

Conclusions

In conclusion, this study sought to determine if a broadly available public health Web-based CBT intervention could help individuals with T2DM and mild-to-moderate depressive symptoms improve their daily lives. Functioning and symptom outcomes improved between baseline and postintervention; however, no treatment advantage was observed for the myCompass group. Further research is necessary to identify the factors that impact participant retention in studies of Web-based interventions in T2DM, especially among those in greatest need of psychological support and who stand to benefit most from treatment. At the same time, there is a need to better understand how individuals with diabetes conceptualize mood symptoms in the context of diabetes to ensure that Web-based interventions are both personally and clinically relevant. The personal and societal health burdens posed by comorbid T2DM and depressive symptoms are considerable and will continue to grow. Continued investigation of the potential for Web-based CBT to provide a solution in T2DM is essential.

Acknowledgments

The authors gratefully acknowledge the dedication and input of all study participants. The authors also acknowledge the support of the Australian Government Department of Health in funding the development and dissemination of the myCompass program and the support of the National Health and Medical Research Council in funding this research trial (NHMRC Grant 1083116). JP, JC, SS, and PB are grateful to Mark Donoghoe of UNSW Stats Central for assistance with MMRM analysis procedures.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 7MB - jmir_v21i5e12246_app1.pdf]

References

1. Shaw J, Sicree R, Zimmet P. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract Title 2010 Jan;87(1):4-14. [doi: 10.1016/j.diabres.2009.10.007] [Medline: 19896746]
2. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004 Apr 26;27(5):1047-1053. [doi: 10.2337/diacare.27.5.1047]
3. Nanayakkara N, Pease A, Ranasinha S, Wischer N, Andrikopoulos S, Speight J, et al. Depression and diabetes distress in adults with type 2 diabetes: results from the Australian National Diabetes Audit (ANDA) 2016. Sci Rep 2018 May 18;8(1):7846 [FREE Full text] [doi: 10.1038/s41598-018-26138-5] [Medline: 29777153]
4. Biessels GJ, Staekenborg S, Brunner E, Brayne C, Scheltens P. Risk of dementia in diabetes mellitus: a systematic review. Lancet Neurol 2006 Jan;5(1):64-74. [doi: 10.1016/s1474-4422(05)70284-2]
5. Rubin RR, Peyrot M. Psychological issues and treatments for people with diabetes. J Clin Psychol 2001 Apr;57(4):457-478. [doi: 10.1002/jclp.1041]
6. Semenkovich K, Brown ME, Svrakic DM, Lustman PJ. Depression in type 2 diabetes mellitus: prevalence, impact, and treatment. Drugs 2015 Apr 8;75(6):577-587. [doi: 10.1007/s40265-015-0347-4]
7. Roy T, Lloyd CE. Epidemiology of depression and diabetes: a systematic review. J Affect Disorders 2012 Oct;142:S8-S21. [doi: 10.1016/s0165-0327(12)70004-6]
8. Egede LE, Bishu KG, Walker RJ, Dismuke CE. Impact of diagnosed depression on healthcare costs in adults with and without diabetes: United States, 2004–2011. J Affect Disorders 2016 May;195:119-126. [doi: 10.1016/j.jad.2016.02.011]
9. Holt RI, de Groot M, Golden SH. Diabetes and depression. Curr Diab Rep 2014 Apr 18;14(6):-. [doi: 10.1007/s11892-014-0491-3]
10. Renn BN, Feliciano L, Segal DL. The bidirectional relationship of depression and diabetes: a systematic review. Clin Psychol Rev 2011 Dec;31(8):1239-1246. [doi: 10.1016/j.cpr.2011.08.001] [Medline: 21963669]
11. Hermanns N, Caputo S, Dzida G, Khunti K, Meneghini LF, Snoek F. Screening, evaluation and management of depression in people with diabetes in primary care. Prim Care Diabetes 2013 Apr;7(1):1-10 [FREE Full text] [doi: 10.1016/j.pcd.2012.11.002] [Medline: 23280258]
12. Li C, Xu D, Hu M, Tan Y, Zhang P, Li G, et al. A systematic review and meta-analysis of randomized controlled trials of cognitive behavior therapy for patients with diabetes and depression. J Psychosom Res 2017 Dec;95:44-54. [doi: 10.1016/j.jpsychores.2017.02.006] [Medline: 28314548]
13. Safren SA, Gonzalez JS, Wexler DJ, Psaros C, Delahanty LM, Blashill AJ, et al. A randomized controlled trial of cognitive behavioral therapy for adherence and depression (CBT-AD) in patients with uncontrolled type 2 diabetes. Diabetes Care 2014 Oct;37(3):625-633 [FREE Full text] [doi: 10.2337/dc13-0816] [Medline: 24170758]
14. Fleer J, Tovote KA, Keers JC, Links TP, Sanderman R, Coyne JC, et al. Screening for depression and diabetes-related distress in a diabetes outpatient clinic. Diabet Med 2013 Jan;30(1):88-94. [doi: 10.1111/dme.12001] [Medline: 22924587]
15. Schmitz N, Gariépy G, Smith KJ, Clyde M, Malla A, Boyer R, et al. Recurrent subthreshold depression in type 2 diabetes: an important risk factor for poor health outcomes. Diabetes Care 2014 Apr;37(4):970-978. [doi: 10.2337/dc13-1832] [Medline: 24198303]
16. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. Use of mental health services in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004 Jun(420):47-54. [doi: 10.1111/j.1600-0047.2004.00330.x] [Medline: 15128387]
17. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009 May;28(3):w490-w501. [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
18. 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]
19. 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]
20. 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 Dec 15;19(5):e157 [FREE Full text] [doi: 10.2196/jmir.7274] [Medline: 28506956]
21. Nobis S, Ebert DD, Lehr D, Smit F, Buntrock C, Berking M, et al. Web-based intervention for depressive symptoms in adults with types 1 and 2 diabetes mellitus: a health economic evaluation. Br J Psychiatry 2018 Dec;212(4):199-206. [doi: 10.1192/bjp.2018.10] [Medline: 30071909]

22. 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]
23. Fisher L, Mullan JT, Arean P, Glasgow RE, Hessler D, Masharani U. Diabetes distress but not clinical depression or depressive symptoms is associated with glycemic control in both cross-sectional and longitudinal analyses. Diabetes Care 2010 Jan;33(1):23-28 [FREE Full text] [doi: 10.2337/dc09-1238] [Medline: 19837786]
24. Fisher L, Gonzalez JS, Polonsky WH. The confusing tale of depression and distress in patients with diabetes: a call for greater clarity and precision. Diabet Med 2014 Jul;31(7):764-772 [FREE Full text] [doi: 10.1111/dme.12428] [Medline: 24606397]
25. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
26. 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]
27. 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]
28. Banks E. Cohort profile: the 45 and Up study. Int J Epidemiol 2007 Sep 19;37(5):941-947. [doi: 10.1093/ije/dym184]
29. SpringboarD. URL: https://springboard.blackdoghealth.org.au/ [accessed 2019-05-09] [WebCite Cache ID 78Fkz7rqs]
30. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
31. Kroenke K, Spitzer RL, Williams JB. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
32. Degenhardt L, Hall W, Korten A, Jablensky A. Use of a Brief Screening Instrument for Psychosis: Results of an ROC Analysis. Sydney: National Drug and Alcohol Research Centre; 2005. URL: https://ndarc.med.unsw.edu.au/resource/use-brief-screening-instrument-psychosis-results-roc-analysis [accessed 2019-05-02] [WebCite Cache ID 7854oshja]
33. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
34. Christensen H, Griffiths KM, Mackinnon AJ, Kalia K, Batterham PJ, Kenardy J, et al. Protocol for a randomised controlled trial investigating the effectiveness of an online e health application for the prevention of Generalised Anxiety Disorder. BMC Psychiatry 2010 Mar 21;10:25 [FREE Full text] [doi: 10.1186/1471-244X-10-25] [Medline: 20302678]
35. Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2018 Jan 2;180(05):461-464. [doi: 10.1192/bjp.180.5.461]
36. Pedersen G, Kvarstein EH, Wilberg T. The Work and Social Adjustment Scale: psychometric properties and validity among males and females, and outpatients with and without personality disorders. Personal Ment Health 2017 Dec;11(4):215-228. [doi: 10.1002/pmh.1382] [Medline: 28681505]
37. 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]
38. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010 Jul;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
39. van Steenbergen-Weijenburg KM, de Vroege L, Ploeger RR, Brals JW, Vloedbeld MG, Veneman TF, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. BMC Health Serv Res 2010 Aug 12;10:235 [FREE Full text] [doi: 10.1186/1472-6963-10-235] [Medline: 20704720]
40. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the Diabetes Distress Scale. Diabetes Care 2005 Feb 25;28(3):626-631. [doi: 10.2337/diacare.28.3.626]
41. Peyrot M, Bushnell DM, Best JH, Martin ML, Cameron A, Patrick DL. Development and validation of the self-management profile for type 2 diabetes (SMP-T2D). Health Qual Life Outcomes 2012 Oct 05;10:125 [FREE Full text] [doi: 10.1186/1477-7525-10-125] [Medline: 23039868]
42. Üstün T, Kostanjsek N, Chatterji S, Rehm J. Manual for WHO Disability Assessment Schedule (WHODAS 2.0). Measuring Health and Disability URL: https://www.who.int/sorry/ [accessed 2019-05-02] [WebCite Cache ID 7855KEogc]
43. Graham J, Cumsille P, Elek-Fisk E. Methods for Handling Missing Data. In: Handbook of Psychology. Oxford: Wiley; 2003.
44. Fairclough DL, Helms RW. A mixed linear model with linear covariance structure: a sensitivity analysis of maximum likelihood estimators. J Stat Comput Sim 2007 Mar 30;25(3-4):205-236. [doi: 10.1080/00949658608810933]

45. Markowitz SM, Gonzalez JS, Wilkinson JL, Safren SA. A review of treating depression in diabetes: emerging findings. Psychosomatics 2011;52(1):1-18 [FREE Full text] [doi: 10.1016/j.psym.2010.11.007] [Medline: 21300190]
46. Button KS, Wiles NJ, Lewis G, Peters TJ, Kessler D. Factors associated with differential response to online cognitive behavioural therapy. Soc Psychiatry Psychiatr Epidemiol 2011 May 4;47(5):827-833. [doi: 10.1007/s00127-011-0389-1]
47. American DA. Standards of medical care in diabetes--2013. Diabetes Care 2013 Jan;36(Suppl 1):S11-S66 [FREE Full text] [doi: 10.2337/dc13-S011] [Medline: 23264422]
48. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]
49. Fogg L, Gross D. Threats to validity in randomized clinical trials. Res Nurs Heal 2000;23(1):79-87. [doi: 10.1002/(sici)1098-240x(200002)23:1<79::aid-nur9>3.0.co;2-r]
50. MacPhillamy DJ, Lewinsohn PM. Depression as a function of levels of desired and obtained pleasure. J Abnorm Psychol 1974;83(6):651-657. [doi: 10.1037/h0037467]
51. Der-Avakian A, Markou A. The neurobiology of anhedonia and other reward-related deficits. Trends Neurosci 2012 Jan;35(1):68-77 [FREE Full text] [doi: 10.1016/j.tins.2011.11.005] [Medline: 22177980]
52. Harlé KM, Guo D, Zhang S, Paulus MP, Yu AJ. Anhedonia and anxiety underlying depressive symptomatology have distinct effects on reward-based decision-making. PLoS One 2017 Oct 23;12(10):e0186473. [doi: 10.1371/journal.pone.0186473]
53. Fried EI, Nesse RM, Guille C, Sen S. The differential influence of life stress on individual symptoms of depression. Acta Psychiatr Scand 2015 Feb 04;131(6):465-471. [doi: 10.1111/acps.12395]
54. Nefs G, Pouwer F, Denollet J, Kramer H, Wijnands - van Gent C, Pop V. Suboptimal glycemic control in type 2 diabetes: a key role for anhedonia? J Psychiat Res 2012 Apr;46(4):549-554. [doi: 10.1016/j.jpsychires.2012.01.013]
55. Clarke J, Proudfoot J, Vatiliotis V, Verge C, Holmes-Walker DJ, Campbell L, et al. Attitudes towards mental health, mental health research and digital interventions by young adults with type 1 diabetes: a qualitative analysis. Health Expect 2018 Jan 10;21(3):668-677. [doi: 10.1111/hex.12662]
56. Kanuch S, Cassidy K, Dawson N, Athey M, Fuentes-Casiano E, Sajatovic M. Recruiting and retaining individuals with serious mental illness and diabetes in clinical research: lessons learned from a randomized, controlled trial. J Health Dispar Res Pract 2016;9(3):115-126 [FREE Full text] [Medline: 28533944]

Abbreviations

CBT: cognitive behavioral therapy
DDS: Diabetes Distress Scale
GAD: Generalized Anxiety Disorder scale
HbA_{1c}: glycosylated hemoglobin
HREC: Human Research Ethics Committee
ITT: intention-to-treat
MMRM: mixed-model repeated measures analyses
NSW: New South Wales
OR: odds ratio
PHQ: Patient Health Questionnaire
RCT: randomized controlled trial
SMP-T2D: Self-Management Profile for Type 2 Diabetes
SMS: short message service
T2DM: type 2 diabetes mellitus
UNSW: University of New South Wales Sydney
VRR: Volunteer Research Register
WSAS: Work and Social Adjustment Scale

Edited by G Eysenbach; submitted 19.09.18; peer-reviewed by J Byrne, H Tateno; comments to author 22.11.18; revised version received 11.12.18; accepted 12.12.18; published 21.05.19.

Please cite as:
Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, Campbell L, Zwar N, Harris M, Lapsley H, Hadzi-Pavlovic D, Christensen H, Proudfoot J
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;21(5):e12246
URL: http://www.jmir.org/2019/5/e12246/
doi:10.2196/12246
PMID:31115345

Original Paper

Methodology Used in Ecological Momentary Assessment Studies About Sedentary Behavior in Children, Adolescents, and Adults: Systematic Review Using the Checklist for Reporting Ecological Momentary Assessment Studies

Catiana Leila Possamai Romanzini[1*], PhD; Marcelo Romanzini[1*], PhD; Mariana Biagi Batista[2*], PhD; Cynthia Correa Lopes Barbosa[3*], PhD; Gabriela Blasquez Shigaki[4,5*], PhD; Genevieve Dunton[6*], PhD; Tyler Mason[6*], PhD; Enio Ricardo Vaz Ronque[1*], PhD

[1]Londrina State University, Department of Physical Education, Londrina, Brazil
[2]Federal University of Mato Grosso do Sul, Campus of Pantanal, Corumbá, Brazil
[3]Federal University of Technology of Paraná, Academic Department of Humanities, Apucarana, Brazil
[4]Rio Preto University Center, Department of Physical Education, São José do Rio Preto, Brazil
[5]Paulista University, Department of Physical Education, São José do Rio Preto, Brazil
[6]University of Southern California, Department of Preventive Medicine, Los Angeles, CA, United States
[*]all authors contributed equally

Corresponding Author:
Catiana Leila Possamai Romanzini, PhD
Londrina State University
Department of Physical Education
Rodovia Celso Garcia Cid | Pr 445 Km 380 | Campus Universitário Cx Postal 10.011
Londrina, 86.057-970
Brazil
Phone: 55 4333714238
Email: clpossamai@uel.br

Abstract

Background: The use of ecological momentary assessment (EMA) to measure sedentary behavior (SB) in children, adolescents, and adults can increase the understanding of the role of the context of SB in health outcomes.

Objective: The aim of this study was to systematically review literature to describe EMA methodology used in studies on SB in youth and adults, verify how many studies adhere to the Methods aspect of the Checklist for Reporting EMA Studies (CREMAS), and detail measures used to assess SB and this associated context.

Methods: A systematic literature review was conducted in the PubMed, Scopus, Web of Science, PsycINFO, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and SPORTDiscus databases, covering the entire period of existence of the databases until January 2018.

Results: This review presented information about the characteristics and methodology used in 21 articles that utilized EMA to measure SB in youth and adults. There were more studies conducted among youth compared with adults, and studies of youth included more waves and more participants (n=696) than studies with adults (n=97). Most studies (85.7%) adhered to the Methods aspect of the CREMAS. The main criteria used to measure SB in EMA were self-report (81%) with only 19% measuring SB using objective methods (eg, accelerometer). The main equipment to collect objective SB was the ActiGraph, and the cutoff point to define SB was <100 counts/min. Studies most commonly used a 15-min window to compare EMA and accelerometer data.

Conclusions: The majority of studies in this review met minimum CREMAS criteria for studies conducted with EMA. Most studies measured SB with EMA self-report (n=17; 81.0%), and a few studies also used objective methods (n=4; 19%). The standardization of the 15-min window criteria to compare EMA and accelerometer data would lead to a comparison between these and new studies. New studies using EMA with mobile phones should be conducted as they can be considered an attractive method for capturing information about the specific context of SB activities of young people and adults in real time or very close to it.

(J Med Internet Res 2019;21(5):e11967) doi:10.2196/11967

KEYWORDS

physical activity; accelerometry; health behavior

Introduction

Background

Sedentary behavior (SB) is defined as any activity performed during awake time with low energy expenditure (below or equal to 1.5 metabolic equivalents) in a sitting or reclining position [1-3]. SB can be measured using subjective assessments (eg, questionnaires) [4] or using objective assessments (eg, accelerometers and inclinometers) [4]. Both methods of assessment have a number of strengths and weaknesses. Subjective measures of SB are able to distinguish between various types of SB (eg, watching television and internet use); however, subjective measures are prone to retrospective recall biases given that they depend on one’s ability to accurately recall previous SB [5]. Oppositely, objective measures of SB provide a fine-grained assessment of individuals’ level of SB but do not distinguish between different types of SB [5].

The discussion about the best method to measure SB is relevant because the more accurate information is collected about this behavior, the more precise information we can have about the context and the type of SB that the participant did, and thus better understand how this behavior can occur, distributed in different contexts throughout the day. Although the literature shows that the use of inclinometers [6] is the best method for measuring SB (because they can detect different postures adopted by subjects), more than half of studies (51%) that investigate physical activity (PA) and its domains still use ActiGraph monitors to obtain information about the total amount of time in a specific behavior [7].

Although both subjective and objective assessments are useful as measures of SB [8], they are not able to identify the context of the activities being performed by the subject [4]. Contextual factors include intrapersonal (eg, affect), interpersonal (eg, who one is with), and environmental factors (eg, location) in which SB occurs in daily life and are integral in understanding reasons for SB as well as outcomes of SB. Contextual factors of SB are often time-varying such that they change over the course of minutes, hours, or days. Owing to the time-varying nature of contextual factors, they cannot be adequately assessed using traditional survey measures. In addition, although survey measures of SB are able to capture information on types of SB, these measures are plagued by retrospective recall biases. Thus, despite researchers’ efforts to understand how SB is associated with health outcomes [9-12], there is still a gap in knowledge of the social and environmental context of SB in different populations and how context is associated with health outcomes [13].

The context of SB can be obtained through the use of questionnaires, but its application is burdensome for the participant and labor-intensive for the researcher and, in addition, may be inadequate for long-term monitoring studies [14]. Thus, it is highly likely that, in each of these contexts, there are distinct determinants for the subject to assume this behavior, as they are shaped by the attributes of the environments where they occur and the social structure involved [15].

Ecological momentary assessment (EMA) remedies many of the described limitations including identification of the type, environment, and context in which SB occurs as well as reduced reliance on retrospective recall [16]. EMA is an approach to collecting data in real time in individuals’ natural environment. Participants in EMA studies are instructed to respond to self-report surveys over the course of the day for a short period of time (eg, weeks and months) using an electronic device such as a mobile phone. Several methods may be used to collect EMA surveys including randomly signaling participants to complete a survey (ie, random interval contingent), having participants complete surveys at predefined intervals (ie, fixed interval contingent), or having participants complete surveys when a specific event occurs (ie, event contingent) [16].

To our knowledge, publications about PA using EMA have already been published among children and adolescents [17,18] and with adults [19]. However, considering that the importance of reducing time in SB has recently received significant attention and that the EMA is a relatively new methodology for investigating contexts of SB among children, adolescents, and adults, it is important to increase the understanding of best practices for using EMA to assess SB, mainly in the verification of association between context of SB and many health outcomes.

A primary advantage of using EMA to study SB is increased understanding of the context in which activities are being performed [5]. Thus, in addition to collecting the information that the subject was involved in a certain number of minutes at a given intensity of PA during the week, it is also possible to gain information about where and what type of activity this was, from the information recorded by the subject through EMA, which is very close to the moment when it happened [16,19]. More recently, research suggested the standardization of EMA use and proposed to study the Checklist for Reporting EMA Studies (CREMAS) for enhancing reliability, efficacy, and overall interpretation of the findings for future studies that use EMAs [20].

Objectives

Therefore, the aim of this study was to systematically review the literature on EMA in SB by researching in children, adolescents, and adults, to verify the number of studies that adhere to the Methods aspect of the CREMAS, and to provide recommendations for measuring SB in EMA.

Methods

Information Sources

A systematic review of the literature was conducted in the following databases: PubMed, Scopus, Web of Science, PsycINFO, CINAHL, and SPORTDiscus, seeking to identify studies that used EMA to measure SB in children, adolescents, and adults. We considered an adolescent to be a person aged 10 to 19 years, as proposed by the World Health Organization.

The search comprised the entire period of existence of the databases until January 2018, in the English language. The search strategy used the following structure of keywords and Boolean operators: (“ecological momentary assessment” OR “EMA”) AND (“sedentary behavior” OR “sedentary behaviour”).

Selection Criteria

The eligibility criteria for studies were as follows: (1) involved participants aged >8 years to <60 years; (2) not being a review or systematic review study; (3) used an EMA-based data collection method; and (4) focused on the assessment of SB. After the exclusion of studies according to the eligibility criteria, the remaining studies were analyzed by abstract or full-text reading and were excluded if they did not assess the main outcome (SB) via EMA. The selection of studies was conducted independently by 2 researchers (CLPR and ERVR). In case of inconsistency in the selection of records, a third researcher (MR) was invited to review the selection.

Extraction Criteria

In total, 2 researchers independently extracted information from records about study characteristics including sample size, mean age, outcomes, and measures. Data extraction was also conducted for specific methods used in EMA studies from the CREMAS [20] including the main technology used, prompt approach (such as prompting design), wave duration, monitoring period, and prompt frequency. In addition, compliance (ie, overall response rate to EMA prompts) was extracted from studies. Finally, the authors detailed the measures used to assess SB, the context, and if it is the current or past measure. The authors choose to keep the records of different studies even when the overlapping of participants was found.

Results

Characteristics of the Studies

A total of 115 potential studies were identified. Figure 1 presents the diagram of systematic reviews for analysis of studies proposed by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA).

After the records were located in each of the databases, all were imported into the reference program. Following the PRISMA model, 65 duplicate records were excluded, after which the eligibility criteria were applied, and finally the analysis was carried out by reading the abstracts. Thus, 21 discrete studies were included in the qualitative synthesis of the data.

The characteristics for youth and adult studies are presented in Table 1. Of the 21 studies included in the qualitative synthesis, 16 were with youth [21,22,23-30,31-36] and 5 with adults [37-41].

The publication dates of articles ranged between 2007 and 2017 and 42.8% of these were published in the past 4 years. The majority of studies included both females and males; however, some collected information only with females [29,33,34] or only with males [30]. The mean number of participants per study was higher in studies with youth (696 participants) than in studies with adults (80 participants).

Only 3 studies collected information solely about SB [21,31,35], whereas the majority collected information about SB and other outcomes, such as PA, environmental factors, nutrition information, and depressive symptoms.

Only 1 study with youth [25] and 3 studies with adults [37,38,40] combined EMA with an objective measurement of SB (ie, accelerometers).

The methodological characteristics of the studies on SB in youth and adults are presented in Table 2.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

Table 1. Characteristics of ecological momentary assessment studies on sedentary behavior in children, adolescents, and adults.

Study	N	Sample size		Age (years)		Outcome	Tools
		Girls	Boys	Mean (SD)	Range		
Gorely et al, 2007 [30][a]	1371	850	521	14.7 (0.9)	13-16	PA[b]/SB[c]	EMA[d] Diaries
Gorely et al, 2007 [29][a]	923	923	0	14.7 (0.9)	13-16	PA/SB	EMA Diaries
Biddle et al, 2009 [21][a]	1484	923	561	14.6 (0.9)	13-16	SB	EMA Diaries
Biddle et al, 2009 [22][a]	991	606	385	14.1 (0.8)	13-16	PA/SB	EMA Diaries
Biddle et al, 2009 [23][a]	1493	927	566	14.1 (0.8)	13-16	SB/other	EMA Diaries
Biddle et al, 2009 [24][a]	623	376	247	15.5 (0.9)	13-18	PA/SB	EMA Diaries
Gorely et al, 2009 [27][a]	1171	694	477	14.8 (0.8)	13-16	PA/SB	EMA Diaries
Gorely et al, 2009 [28][a]	561	561	0	14.6 (0.8)	13-16	PA/SB	EMA Diaries
Dunton et al, 2011 [26][a]	121	59	62	40.4 (9.7)	9-13	PA/SB	MP EMA[e] and ACC (no measure SB)
Soos et al, 2012 [35][a]	635	384	251	16.0 (1.0)	13-18	SB	EMA Diaries
Liao et al, 2014 [31][a]	120	58	62	__[f]	9-13	SB	MP EMA
Soos et al, 2014 [36][a]	812	464	348	15.6 (1.0)	9-13	PA/SB/other	EMA Diaries
Dunton et al, 2015 [25][a]	200	200	0	__[f]	8-12	SB/other	MP EMA, questionnaire, and ACC
Raudsepp, 2016 [33][a]	341	341	0	15.3 (3.0)	__[f]	SB/other	EMA Diaries and questionnaire
O'Connor et al, 2017 [32][a]	175	91	84	9.6 (0.9)	8-12	SB/PA/other	MP EMA
Raudsepp and Riso, 2017 [34][a]	122	122	0	11.4 (0.7)	11-12	SB/other	EMA Diaries and questionnaire
Rouse and Biddle, 2010 [41][g]	84	38	46	19.5 (1.1)	__[f]	PA/SB	EMA Diaries
Dunton et al, 2012 [38][g]	110	80	30	40.4 (9.7)	27-73	PA/SB	MP EMA and ACC[h]
Graves et al, 2015 [39][g]	47	37	10	38.6 (9.5)	__[f]	SB/other	EMA Diaries, exams and questionnaires
Liao et al, 2015 [40][g]	110	80	30	__[f]	27-73	PA/SB/other	MP EMA and ACC
Bruening et al, 2016 [37][g]	133	97	36	18.8 (0.6)	__[f]	PA/SB/other	Automated Self-Administered 24-hour, Mobile ecological assessment, and ACC

[a] Studies with children and adolescents.

[b] PA: physical activity.

[c] SB: sedentary behavior.

[d] EMA: ecological momentary assessment.

[e] MP EMA: mobile phone ecological momentary assessment.

[f] Missing data.

[g] Studies with adults.

[h] ACC: accelerometer.

Table 2. Methodological characteristics of the ecological momentary assessment studies on sedentary behavior in children, adolescents, and adults.

Study	Technology; Response rates	Prompt design, Waves	Monitoring period, frequency	SB[a] measures; Context of SB; Current, past SB
Gorely et al, 2007 [30][b]	Paper and pencil diaries; 50.2%	RIC[c], 2 times	4 days (3 vs 1), every 15 min	SR[d] times per 15 min; 23 contexts; Current[e]
Gorely et al, 2007 [29][b]	Paper and pencil diaries; 92%	RIC, 2 times	4 days (3 vs 1), every 15 min	SR times per 15 min; 23 contexts; Current[f]
Biddle et al, 2009 [21][b]	Paper and pencil diaries; not informed	RIC, 2 times	4 days (3 vs 1), every 15 min	SR times per 15 min; 23 contexts by gender; Current[e]
Biddle et al, 2009 [22][b]	Paper and pencil diaries; 95%	RIC, 2 times	4 days (3 vs 1), every 15 min	SR times per 15 min; 18 contexts; Current[f]
Biddle et al, 2009 [23][b]	Paper and pencil diaries; 85%	RIC, 2 times	4 days (3 vs 1), every 15 min	SR times per 15 min; 22 contexts; Current[f]
Biddle et al, 2009 [24][b]	Paper and pencil diaries; Hungary 96%, Romania 78%, Slovakia 86%	RIC	4 days (3 vs 1), every 15 min	SR times per 15 min; 23 contexts; Current[f]
Gorely et al, 2009 [27][b]	Paper and pencil diaries; 50.2%	RIC, 2 times	4 days (3 vs 1), every 15 min	SR times per 15; 23 contexts by gender; Current[e]
Gorely et al, 2009 [28][b]	Paper and pencil diaries; 91.4%	RIC, 2 times	4 days (3 vs 1), every 15 min	SR times per 15 min; 23 contexts; Current[f]
Dunton et al, 2011 [26][b]	HTC Shadow MP[g], MyExperience; 80% with ACC[h]	RIC, 1 time	4 days, (3 vs 7)	SR; 4 contexts; Current[f,i]
Soos et al, 2012 [35][b]	Paper and pencil diaries; 75%	RIC	4 days (3 vs 1), every 15 min	SR times per 15 min; 22 contexts; Current[f]
Liao et al, 2014 [31][b]	HTC Shadow MPMy Experience; 77%	RIC, 1 time	4 days, (3 vs 7)	SR; 2 contexts; Current[f]
Soos et al, 2014 [36][b]	Paper and pencil diaries; 75%	RIC	4 days (3 vs 1), every 15 min	SR times per 15 min; 23 contexts; Current[f]
Dunton et al, 2015 [25][b]	Mobile Phone Android or Motorola Moto G; Mothers 80%, Children 69%	RIC, 6 times	7 days (3 vs 7), 1 add for mothers	SR; 1 context; Current[f,j]
Raudsepp, 2016 [33][b]	Paper and pencil diaries; 95.6%	RIC, 3 times	4 days (3 vs 1), every 15 min	SR times per 15 min; 4 contexts; Current[f]
O'Connor et al, 2017 [32][b]	Mobile Phone Android or Motorola Moto G; not informed	RIC, 1 time	8 days (3 vs 4), 7 vs 8	SR; 1 context; Current[f]
Raudsepp and Riso, 2017 [34][b]	Paper and pencil diaries; 81.8%	RIC, 4 times	4 days (3 vs 1), every 15 min add per day (feedback)	SR times per 15 min; 12 contexts; Current[f]
Rouse and Biddle, 2010 [41][k]	Paper and pencil diaries; 57%	RIC, 1 time	2 days (1 vs 1), every 15 min	SR times per 15 min; 3 contexts; Current[e]
Dunton et al, 2012, 2015 [38][k]	HTC Shadow MPMyExperience; 82%, 85% with ACC	RIC, 1 time	4 days, 8 per day	SR; 3 contexts; Current[f,l]
Graves et al, 2015 [39][k]	Paper based EMA diaries; not informed	RIC, 1 time	5 days, every 15 min	SR times per 15 min; 1 context (sitting); Current[f]
Liao et al, 2015 [40][k]	HTC Shadow MPMyExperience; 82%	RIC, 1 time	4 days, 8 per day	SR; 3 contexts; Current[f,m]
Bruening et al, 2016 [37][k]	Android; iPhone Operational System, MP; Motorola Moto G; Mobile ecological momentary assessment, 84% with ACC	RIC, 1 time	4 days, 7 per day; 1 past per day	SR;10 contexts; Current[f,n] and Past

[a] SB: sedentary behavior.

[b] Studies with children and adolescents.

[c] RIC: random interval contingent.

[d] SR: self-report.

[e] Location and company.

[f] Performing the behavior at this exact moment or very close to it.

[g] MP: mobile phone.

[h] ACC: accelerometer.

[i] Objectively ActiGraph GT2M (no measure SB)

[j] Objectively ActiGraph GT3X+<100 counts per min

[k] Studies with adults.

[l] Objectively; ActiGraph GTM2<100 counts per min/15 min interval

[m] Objectively; ActiGraph GTM2<100 counts per min/15-min window

[n] Objectively/ActiGraph GT3X+<100 counts per min/5 min prior EMA prompt

In relation to the Methods aspect of the CREMAS, 85.7% of the studies adhered to these items (18 studies of 21 met all the criteria; 3 studies did not mention wave duration). The most widely used software for mobile phone studies was My Experience [26,31,38,40]. One study used a version of an application (devilSPARC) created specifically for the study [37] and another 2 studies did not specify the software used [25,32]. The majority of studies included in this review (85.7%) presented information about response rates to EMA prompts, adhering to reporting recommendations. The mean range was between 50.2% and 95.6%.

All of the studies analyzed used the random interval contingent (EMA prompts were set to be randomized throughout the day) [20] to deliver the prompts. In general, studies with youth used multiple measurement waves of monitoring periods (ranging from 1 to 4 waves). All of the studies with adults used only one data collection wave.

Each wave ranged from 2 to 8 days of EMA monitoring. The most common prompting frequency was every 15 min mainly in studies with youth (75%). In studies with adults, 40% prompted at a frequency ≤ 15 min; another 40% prompted at a frequency ≥ 2 hours.

All articles used self-report measures to define different contexts of SB. The contexts utilized ranged from 1 to 23 different contexts. About 81% of the articles included questions related to the behavior of the moment, such as: *What are you doing now?*, *What are you doing and where are you?*, *What were you doing right before you got this text?*, *Which of these things have you done?*, *Who (if anyone) was with you while you were doing this?*, *What were you DOING right before the beep went off?*, and *Have you engaged in any of the following activities during the past two hours?*; 19% sought to identify the location and the company.

Only 4 studies (19%) used an objective method to measure SB, 1 with youth [25] and 3 with adults [37,38,40]. Although 1 study cited the use of an accelerometer, the main outcome (SB) was not measured with this method; only the total number of steps and moderate-to-vigorous PA were measured. In all cases, the accelerometer used was the ActiGraph (models GTM2 or GT3X+), and the cutoff point to define SB was <100 counts/min. A 15-min window for more and less of the registration prompt obtained by EMA was the most common.

Discussion

The aim of this study was to systematically review the literature on EMA studies of SB in youth and adults, verify the number of studies that adhered to the Methods aspect of the CREMAS, and describe the measures used to assess SB in EMA. There was a paucity of EMA studies of SB in adults (n=5), indicating a need for more research in this area. The majority of studies were conducted with youth, and these studies typically collected more waves of data and had larger sample sizes compared with studies in adults. In general, the samples in the reviewed studies often had a greater number of women compared with men.

Principal Findings

Related to the studies conducted with mobile phones, the most commonly used software was MyExperience. This was one of the first software programs developed for EMA and it is an open license software developed especially for Windows mobile devices [42]. However, MyExperience was used in older studies and currently newer technology is available for use in EMA studies. One factor to consider when choosing software is accessibility on various mobile operating systems—some are only available for use on Android and others use Android and iPhone Operational System. Another consideration is cost of the software, which may limit applicability for large-scale studies [4].

This review indicated that EMA monitoring periods lasted from 4 to 8 days, 4 days being the most cited for capturing behavioral information.

Although the use concomitant of EMA in mobile phones and accelerometers can represent the best method to measure SB, we believe that 4 days of concomitant use of EMA and an accelerometer is a good recommendation to obtain more accurate information on the context and pattern of SB, representing a typical week of this behavior.

Comparison With Previous Work

The fact that the subject is monitored several times during the day may induce them not to respond to the activity being performed at the time of prompting; however, this is a limitation not only of this type of record, but also of self-reporting instruments and questionnaires [4]. On the contrary, the use of EMA could also lead to decreased burden on the participant compared with the use of questionnaires, potentially yielding higher rates of compliance and lower rates of missing data [37]. In general, the response rates in studies that use EMA to assess with mobile phones are very high [25,37,38,40], providing evidence of the feasibility of using EMA to collect information about SB. This could be because of the easy access to mobile phones nowadays.

In our systematic review, we only identified 3 studies with adults that combined EMA with mobile phones and accelerometers [37,38,40]. All of these indicated the acceptability of 4 days of use of the EMA protocol in mobile phones to measure SB in adults and that this method can offer an innovative approach to capture PA and the context of SB. For purposes of comparison of the records between EMA and the accelerometer, it might be useful to standardize the 2 pieces of information to the same number of days.

This analysis may provide researchers with a more accurate interpretation of the context in which each sedentary period occurs, as well as identification of the intensity of PA that the subject assumes when he or she interrupts SB [43], for example. Moreover, it could help to eliminate the information bias of subjective instruments, which depend on the subject's capacity to remember what they did at a particular moment of the day, as it has already been found that there is great variation between subjective and objective measures [44].

Identification of the context of the activities is important, as, in addition to the term *sedentary behavior*, there are particular SBs that generally occur in a variety of contexts such as watching television and other recreational activities with a screen in the home environment and occupations that require prolonged sitting in work and transport environments [45]. Thus, it is highly likely that in each of these contexts, there are distinct determinants for the subject to assume this behavior, as they are shaped by the attributes of the environments where they occur and the social structure involved [45]. In addition, a recent study showed that time watching television has a stronger magnitude of effect on all-cause mortality [15]. In this sense, the use of EMA can provide this identification about the context, and combined with information of the accelerometer, we can estimate the time in which the subject can spend in a specific type of SB context. This kind of information has not yet been researched and gains a body of evidence to future research.

Recently, more advantages have been highlighted with regard to the use of mobile phones, such as collecting data more quickly from a large number of people than traditional cross-sectional or other methods. In addition, this paper introduces new concepts that must be explored in the field of research on PA behavior, which relate to synchrony, sequentiality, and instability [46].

In total, 2 systematic reviews, aimed at providing an overview of existing studies on sedentary time in children and adolescents [47] and in adults [48], through a joint programming initiative called DEterminants of DIet and Physical ACtivity [49], made important notes on the need for harmonization and standardization of methods to assess sedentary time in this population, mainly in combination with objective and self-report methods.

In this sense and seeing that the evidence indicates that the use of EMA is still limited in the adult population, further studies should be conducted, as, in this population, life contexts can lead to an even greater amount of SB involvement owing to commitments such as transportation and work. Thus, as SB can be identified in different contexts, new studies can also be conducted that identify which of these may eventually have the greatest impact on health risks. In addition, new research could advance the comparative analysis of SB measurements obtained by EMA and objective measures such as accelerometry, or analyze the posture adopted by participants (inclinometers).

Limitations

This review has strengths and limitations. The strength of this review is that it is the first to systematically review the literature focusing on EMA studies of SB. In addition, this is the first systematic review that used the CREMAS specifically to identify methodological characteristics of the EMA studies [20], demonstrating the importance of tailoring methods to the unique futures of EMA studies, especially those using mobile phone technology. Although an exhaustive literature search was conducted, it is possible that some studies may have been missed and some details about quantitative study characteristics may have been omitted.

Conclusions

This review systematically evaluated information about the characteristics and methodology of empirical articles using EMA to measure SB in youth and adults. The majority of studies adequately presented the minimum criteria necessary for describing studies conducted with EMA, as proposed in the CREMAS. The main assessment used to measure SB with EMA was self-report; only a few studies used objective methods (accelerometer). There was no single standard adopted in the included studies to compare the EMA and accelerometer data, but we believe that the standardization of criteria (eg, use of a 15-min window to compare EMA and accelerometer and the use of the same cutoff point to define SB such as <100 counts/min) lead to higher quality studies. New studies using mobile phones should be carried out as they can be considered an attractive method for capturing information about the specific context of SB activities of youth and adults, in real time or very close to it.

Acknowledgments

The authors would like to express thanks to Coordination of Improvement of Higher Education Personnel (CAPES/Brazil) for the doctorate scholarship conferred to MBB, and the National Council of Technological and Scientific Development (CNPq/Brazil) for the grants conceded to ERVR. This study was partially supported by Researcher Support Office (PROPPG) from Londrina State University.

Conflicts of Interest

None declared.

References

1. Owen N. Sedentary behavior: understanding and influencing adults' prolonged sitting time. Prev Med 2012 Dec;55(6):535-539. [doi: 10.1016/j.ypmed.2012.08.024] [Medline: 22968124]
2. Pate R, O'Neill JR, Lobelo F. The evolving definition of. Exerc Sport Sci Rev 2008 Oct;36(4):173-178. [doi: 10.1097/JES.0b013e3181877d1a] [Medline: 18815485]
3. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Dec 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
4. Atkin AJ, Gorely T, Clemes SA, Yates T, Edwardson C, Brage S, et al. Methods of measurement in epidemiology: sedentary behaviour. Int J Epidemiol 2012 Oct;41(5):1460-1471 [FREE Full text] [doi: 10.1093/ije/dys118] [Medline: 23045206]
5. Healy GN, Clark BK, Winkler EA, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. Am J Prev Med 2011 Aug;41(2):216-227 [FREE Full text] [doi: 10.1016/j.amepre.2011.05.005] [Medline: 21767730]
6. Hart TL, Ainsworth BE, Tudor-Locke C. Objective and subjective measures of sedentary behavior and physical activity. Med Sci Sports Exerc 2011 Mar;43(3):449-456. [doi: 10.1249/MSS.0b013e3181ef5a93] [Medline: 20631642]
7. Wijndaele K, Westgate K, Stephens SK, Blair SN, Bull FC, Chastin SFM, et al. Utilization and harmonization of adult accelerometry data: review and expert consensus. Med Sci Sports Exerc 2015 Oct;47(10):2129-2139 [FREE Full text] [doi: 10.1249/MSS.0000000000000661] [Medline: 25785929]
8. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
9. Evenson KR, Wen F, Metzger JS, Herring AH. Physical activity and sedentary behavior patterns using accelerometry from a national sample of United States adults. Int J Behav Nutr Phys Act 2015 Feb 15;12:20 [FREE Full text] [doi: 10.1186/s12966-015-0183-7] [Medline: 25889192]
10. García-Hermoso A, Martínez-Vizcaíno V, Sánchez-López M, Recio-Rodriguez JI, Gómez-Marcos MA, García-Ortiz L, EVIDENT Group. Moderate-to-vigorous physical activity as a mediator between sedentary behavior and cardiometabolic risk in Spanish healthy adults: a mediation analysis. Int J Behav Nutr Phys Act 2015 Jun 20;12:78 [FREE Full text] [doi: 10.1186/s12966-015-0244-y] [Medline: 26437664]
11. McVeigh JA, Winkler EA, Howie EK, Tremblay MS, Smith A, Abbott RA, et al. Objectively measured patterns of sedentary time and physical activity in young adults of the Raine study cohort. Int J Behav Nutr Phys Act 2016 Mar 24;13:41 [FREE Full text] [doi: 10.1186/s12966-016-0363-0] [Medline: 27009327]
12. van der Berg JD, Stehouwer CD, Bosma H, van der Velde JH, Willems PJ, Savelberg HH, et al. Associations of total amount and patterns of sedentary behaviour with type 2 diabetes and the metabolic syndrome: The Maastricht Study. Diabetologia 2016 Apr;59(4):709-718 [FREE Full text] [doi: 10.1007/s00125-015-3861-8] [Medline: 26831300]
13. Kim Y, Welk GJ. Characterizing the context of sedentary lifestyles in a representative sample of adults: a cross-sectional study from the physical activity measurement study project. BMC Public Health 2015 Dec 9;15:1218 [FREE Full text] [doi: 10.1186/s12889-015-2558-8] [Medline: 26646428]
14. Loveday A, Sherar LB, Sanders JP, Sanderson PW, Esliger DW. Novel technology to help understand the context of physical activity and sedentary behaviour. Physiol Meas 2016 Dec;37(10):1834-1851. [doi: 10.1088/0967-3334/37/10/1834] [Medline: 27654030]
15. Ekelund U, Steene-Johannessen J, Brown W, Fagerland M, Owen N, Powell K, Lancet Physical Activity Series 2 Executive Committe, Lancet Sedentary Behaviour Working Group. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. Lancet 2016 Sep 24;388(10051):1302-1310. [doi: 10.1016/S0140-6736(16)30370-1] [Medline: 27475271]
16. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
17. Dunton GF, Dzubur E, Kawabata K, Yanez B, Bo B, Intille S. Development of a smartphone application to measure physical activity using sensor-assisted self-report. Front Public Health 2014;2:12 [FREE Full text] [doi: 10.3389/fpubh.2014.00012] [Medline: 24616888]
18. Dunton GF, Intille SS, Wolch J, Pentz MA. Investigating the impact of a smart growth community on the contexts of children's physical activity using Ecological Momentary Assessment. Health Place 2012 Jan;18(1):76-84 [FREE Full text] [doi: 10.1016/j.healthplace.2011.07.007] [Medline: 22243909]

19. Marszalek J, Morgulec-Adamowicz N, Rutkowska I, Kosmol A. Using ecological momentary assessment to evaluate current physical activity. Biomed Res Int 2014;2014:915172 [FREE Full text] [doi: 10.1155/2014/915172] [Medline: 25126580]
20. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: An adapted STROBE Checklist for Reporting EMA Studies (CREMAS). J Med Internet Res 2016 Dec 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
21. Biddle SJ, Gorely T, Marshall SJ. Is television viewing a suitable marker of sedentary behavior in young people? Ann Behav Med 2009 Oct;38(2):147-153. [doi: 10.1007/s12160-009-9136-1] [Medline: 19809858]
22. Biddle SJ, Gorely T, Marshall SJ, Cameron N. The prevalence of sedentary behavior and physical activity in leisure time: a study of Scottish adolescents using ecological momentary assessment. Prev Med 2009 Feb;48(2):151-155. [doi: 10.1016/j.ypmed.2008.10.025] [Medline: 19046984]
23. Biddle SJ, Marshall SJ, Gorely T, Cameron N. Temporal and environmental patterns of sedentary and active behaviors during adolescents' leisure time. Int J Behav Med 2009;16(3):278-286. [doi: 10.1007/s12529-008-9028-y] [Medline: 19238558]
24. Biddle SJH, Soos I, Hamar P, Sandor I, Simonek J, Karsai I. Physical activity and sedentary behaviours in youth: data from three Central-Eastern European countries. Eur J Sport Sci 2009 Sep;9(5):295-301. [doi: 10.1080/17461390902829234]
25. Dunton GF, Liao Y, Dzubur E, Leventhal AM, Huh J, Gruenewald T, et al. Investigating within-day and longitudinal effects of maternal stress on children's physical activity, dietary intake, and body composition: protocol for the MATCH study. Contemp Clin Trials 2015 Jul;43:142-154 [FREE Full text] [doi: 10.1016/j.cct.2015.05.007] [Medline: 25987483]
26. Dunton GF, Liao Y, Intille SS, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011 Jun;19(6):1205-1212 [FREE Full text] [doi: 10.1038/oby.2010.302] [Medline: 21164502]
27. Gorely T, Atkin AJ, Biddle SJ, Marshall SJ. Family circumstance, sedentary behaviour and physical activity in adolescents living in England: Project STIL. Int J Behav Nutr Phys Act 2009 Jun 11;6:33 [FREE Full text] [doi: 10.1186/1479-5868-6-33] [Medline: 19519913]
28. Gorely T, Biddle SJ, Marshall SJ, Cameron N. The prevalence of leisure time sedentary behaviour and physical activity in adolescent boys: an ecological momentary assessment approach. Int J Pediatr Obes 2009;4(4):289-298. [doi: 10.3109/17477160902811181] [Medline: 19922044]
29. Gorely T, Marshall SJ, Biddle SJ, Cameron N. The prevalence of leisure time sedentary behaviour and physical activity in adolescent girls: an ecological momentary assessment approach. Int J Pediatr Obes 2007;2(4):227-234. [doi: 10.1080/17477160701408833] [Medline: 17852544]
30. Gorely T, Marshall SJ, Biddle SJ, Cameron N. Patterns of sedentary behaviour and physical activity among adolescents in the United Kingdom: Project STIL. J Behav Med 2007 Dec;30(6):521-531. [doi: 10.1007/s10865-007-9126-3] [Medline: 17712619]
31. Liao Y, Intille S, Wolch J, Pentz M, Dunton G. Understanding the physical and social contexts of children's nonschool sedentary behavior: an ecological momentary assessment study. J Phys Act Health 2014 Mar;11(3):588-595. [doi: 10.1123/jpah.2011-0363] [Medline: 23493261]
32. O'Connor SG, Koprowski C, Dzubur E, Leventhal AM, Huh J, Dunton GF. Differences in mothers' and children's dietary intake during physical and sedentary activities: an ecological momentary assessment study. J Acad Nutr Diet 2017 Aug;117(8):1265-1271 [FREE Full text] [doi: 10.1016/j.jand.2017.02.012] [Medline: 28392348]
33. Raudsepp L. Brief report: longitudinal associations between sedentary behaviours and depressive symptoms in adolescent girls. J Adolesc 2016 Dec;51:76-80. [doi: 10.1016/j.adolescence.2016.06.001] [Medline: 27322892]
34. Raudsepp L, Riso E. Longitudinal associations between sedentary behavior of adolescent girls, their mothers, and best friends. Pediatr Exerc Sci 2017 Dec;29(3):419-426. [doi: 10.1123/pes.2016-0255] [Medline: 28486024]
35. Soos I, Biddle S, Boros-Balint I, Sandor I, Szabo P, Hamar P, et al. Prevalence of sedentary behaviour in young people in Romania and Slovakia. Eur Phy Educ Rev 2012 Feb 23;18(1):19-46. [doi: 10.1177/1356336x11430659]
36. Soos I, Biddle SJ, Ling J, Hamar P, Sandor I, Boros-Balint I, et al. Physical activity, sedentary behaviour, use of electronic media, and snacking among youth: an international study. Kinesiology 2014;46(2):155-163 [FREE Full text]
37. Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G. A mobile ecological momentary assessment tool (devilSPARC) for nutrition and physical activity behaviors in college students: a validation study. J Med Internet Res 2016 Dec 27;18(7):e209 [FREE Full text] [doi: 10.2196/jmir.5969] [Medline: 27465701]
38. Dunton GF, Liao Y, Kawabata K, Intille S. Momentary assessment of adults' physical activity and sedentary behavior: feasibility and validity. Front Psychol 2012;3:260 [FREE Full text] [doi: 10.3389/fpsyg.2012.00260] [Medline: 22866046]
39. Graves LE, Murphy RC, Shepherd SO, Cabot J, Hopkins ND. Evaluation of sit-stand workstations in an office setting: a randomised controlled trial. BMC Public Health 2015 Nov 19;15:1145 [FREE Full text] [doi: 10.1186/s12889-015-2469-8] [Medline: 26584856]
40. Liao Y, Intille SS, Dunton GF. Using ecological momentary assessment to understand where and with whom adults' physical and sedentary activity occur. Int J Behav Med 2015 Feb;22(1):51-61. [doi: 10.1007/s12529-014-9400-z] [Medline: 24639067]
41. Rouse PC, Biddle SJ. An ecological momentary assessment of the physical activity and sedentary behaviour patterns of university students. Health Educ 2010 Mar 26;69(1):116-125. [doi: 10.1177/0017896910363145]

42. MyExperience. URL: http://myexperience.sourceforge.net/ [accessed 2019-03-31] [WebCite Cache ID 77I81vzMm]
43. Altenburg TM, Chinapaw MJ. Bouts and breaks in children's sedentary time: currently used operational definitions and recommendations for future research. Prev Med 2015 Aug;77:1-3. [doi: 10.1016/j.ypmed.2015.04.019] [Medline: 25937587]
44. Dyrstad S, Hansen B, Holme I, Anderssen S. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc 2014 Jan;46(1):99-106. [doi: 10.1249/MSS.0b013e3182a0595f] [Medline: 23793232]
45. Owen N, Sugiyama T, Eakin E, Gardiner P, Tremblay M, Sallis J. Adults' sedentary behavior determinants and interventions. Am J Prev Med 2011 Aug;41(2):189-196. [doi: 10.1016/j.amepre.2011.05.013] [Medline: 21767727]
46. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017 Dec;45(1):48-54 [FREE Full text] [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
47. Verloigne M, Loyen A, van Hecke L, Lakerveld J, Hendriksen I, de Bourdheaudhuij I, et al. Variation in population levels of sedentary time in European children and adolescents according to cross-European studies: a systematic literature review within DEDIPAC. Int J Behav Nutr Phys Act 2016 Jun 28;13:69 [FREE Full text] [doi: 10.1186/s12966-016-0395-5] [Medline: 27350043]
48. Loyen A, Verloigne M, van Hecke L, Hendriksen I, Lakerveld J, Steene-Johannessen J, DEDIPAC consortium. Variation in population levels of sedentary time in European adults according to cross-European studies: a systematic literature review within DEDIPAC. Int J Behav Nutr Phys Act 2016 Jun 28;13(52):71-30 [FREE Full text] [doi: 10.1186/s12966-016-0397-3] [Medline: 27350251]
49. Lakerveld J, van der Ploeg HP, Kroeze W, Ahrens W, Allais O, Andersen LF, DEDIPAC consortium. Towards the integration and development of a cross-European research network and infrastructure: the DEterminants of DIet and Physical ACtivity (DEDIPAC) Knowledge Hub. Int J Behav Nutr Phys Act 2014 Nov 22;11:143 [FREE Full text] [doi: 10.1186/s12966-014-0143-7] [Medline: 25731079]

Abbreviations

CREMAS: Checklist for Reporting EMA Studies
EMA: ecological momentary assessment
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SB: sedentary behavior

Edited by G Eysenbach; submitted 17.08.18; peer-reviewed by M Yamakita, H Calvert; comments to author 22.11.18; revised version received 28.11.18; accepted 09.12.18; published 15.05.19.

Please cite as:
Romanzini CLP, Romanzini M, Batista MB, Barbosa CCL, Shigaki GB, Dunton G, Mason T, Ronque ERV
Methodology Used in Ecological Momentary Assessment Studies About Sedentary Behavior in Children, Adolescents, and Adults: Systematic Review Using the Checklist for Reporting Ecological Momentary Assessment Studies
J Med Internet Res 2019;21(5):e11967
URL: https://www.jmir.org/2019/5/e11967/
doi:10.2196/11967
PMID:31094349

Original Paper

Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care in British Columbia: Phase I

Roopan Gill[1], MD, MPH, FRCSC; Gina Ogilvie[1], MD, MPH, PhD; Wendy V Norman[1], MD, MHSc, CCFP, FCFP; Brian Fitzsimmons[1], MD, FACOG, FRCSC; Ciana Maher[1], BA, MSc; Regina Renner[1], MD, MPH

Department of Obstetrics and Gynaecology, Women's Health Research Institute, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Roopan Gill, MD, MPH, FRCSC
Department of Obstetrics and Gynaecology
Women's Health Research Institute
University of British Columbia
4500 Oak Street
Vancouver, BC,
Canada
Phone: 1 604 369 0774
Email: roopan.gill@gmail.com

Abstract

Background: Over 30% of women in Canada undergo an abortion. Despite the prevalence of the procedure, stigma surrounding abortion in Canada leads to barriers for women to access this service. The vast majority of care is concentrated in urban settings. There is evidence to support utilization of innovative mobile and other technology solutions to empower women to safely and effectively self-manage aspects of the abortion process. This study is part 1 of a 3-phase study that utilizes user-centered design methodology to develop a digital health solution to specifically support follow-up after an induced surgical abortion.

Objective: This study aimed to (1) understand how women at 3 surgical abortion clinics in an urban center of British Columbia utilize their mobile phones to access health care information and (2) understand women’s preferences of content and design of an intervention that will support follow-up care after an induced abortion, including contraceptive use.

Methods: The study design was based on *development-evaluation-implementation* process from Medical Research Council Framework for Complex Medical Interventions. This was a mixed-methods formative study. Women (aged 14-45 years) were recruited from 3 urban abortion facilities in British Columbia who underwent an induced abortion. Adaptation of validated surveys and using the technology acceptance model and theory of reasoned action, a cross-sectional survey was designed. Interview topics included demographic information; type of wireless device used; cell phone usage; acceptable information to include in a mobile intervention to support women’s abortion care; willingness to use a mobile phone to obtain reproductive health information; optimal strategies to use a mobile intervention to support women; understand preferences for health information resources; and design qualities in a mobile intervention important for ease of use, privacy, and security. Responses to questions in the survey were summarized using descriptive statistics. Qualitative analysis was conducted with NVivo using a thematic analysis approach. This study was approved by the local ethics board.

Results: A waiting-room survey was completed by 50 participants, and semistructured interviews were completed with 8 participants. The average age of participants was 26 years. Furthermore, 94% (47/50) owned a smartphone, 85% (41/48) used their personal phones to go online, and 85% would use their cell phone to assist in clinical care. Qualitative analysis demonstrated that women prefer a comprehensive website that included secure email or text notifications to provide tools and resources for emotional well-being, contraceptive decision making, general sexual health, and postprocedure care.

Conclusions: A community-based mixed-methods approach allowed us to understand how women use their cell phones and what women desire in a mobile intervention to support their postabortion care. The findings from this formative phase will assist in the development and testing of a mobile intervention to support follow-up care after an induced surgical abortion.

(J Med Internet Res 2019;21(5):e13387) doi:10.2196/13387

KEYWORDS

mHealth; family planning; abortion, induced; sexual health; digital health; user-centered design

Introduction

Background

Despite abortion being without any criminal restriction in Canada and over 30% of women in Canada undergo an abortion, recent studies suggest that women who have an abortion experience isolation, particularly for those who travel far distances, perceived need for secrecy, feelings of stigma, and other social factors, suggesting that access to the procedure is only one aspect of what it means to deliver quality abortion care [1,2]. Specifically, studies have demonstrated that the dynamics of the political and social climate in Canada can lead to both externalized and internalized stigma for individuals who undergo an abortion [1,3,4]. Despite global guidance that states that follow-up is not necessary after an induced surgical abortion, there is evidence documenting women's desire for postabortion support [5-7]. Opportunities to use innovations using digital solutions for abortion care can support women to self-manage part, if not all, of the abortion process [6,8-11]. More importantly, evidence has pointed to its potential to provide follow-up support, particularly for women who live in rural and remote areas [4].

The intersection of digital health with self-care is moving rapidly and being utilized in various aspects of health care [12-17]. The New England Journal of Medicine published a special report on Telehealth—an example of digital health—in the United States, highlighting its utility and future. In 2016, Kaiser Permanente of Northern California reported that its virtual (email, telephone, and video) communications had exceeded in-person visits [18]. Similarly, research supports the safe and effective use of digital health solutions for provision of medical abortion care globally [19-25]. Evidence supports its safe, effective, and acceptable utility in provision of abortion. Studies utilizing hotlines, text messaging, and mobile apps in Cambodia, Indonesia, and South Africa are just a few examples of how digital health is being utilized to improve the abortion experience for women who face barriers to abortion care globally [14,20,26,27]. Despite the development and implementation of mobile health (mHealth) innovations for family planning, research is limited in understanding the follow-up needs of women who undergo an abortion, how they would perceive a digital health tool to support them, and more importantly, engage them as active participants in the design process.

Objectives

Given the existing evidence in support of mHealth for family planning innovations, we aimed to determine if a mobile technology intervention would be acceptable and feasible to women to support follow-up care after first and second trimester surgical abortion. This study was designed and conducted before the availability of mifepristone for medical abortion in Canada, and thus, women undergoing a surgical abortion were the focus population for this study [28]. We developed a 3-phased study based on user-centered design and the *development-evaluation-implementation* process from Medical Research Council Frameworks for Complex Interventions [29]. Phase I was a mixed-methods formative design to understand how women at 3 surgical abortion clinics utilize their mobile phones to access health care information and to understand their preferences of content and design of a mobile intervention that will support follow-up care after a surgical abortion. Phase II was the design, development, and usability testing aspect of the study, and finally, phase III was a prospective mixed-methods pilot study to determine the feasibility and acceptability of the mobile intervention to support follow-up care after a surgical abortion. This study is the first to utilize mHealth and user-centered design in Canada as a novel approach to support follow-up care for women who undergo surgical abortion. The focus of this manuscript is phase I methodology and results.

Methods

Participants

Participants were recruited from publicly funded abortion clinics, 3 in urban settings and 1 in a rural site within British Columbia, Canada. The eligibility criteria were that the participant should (1) have consented to undergo a first or second trimester surgical abortion procedure, (2) be able to read and write English, (3) be able to participate in study procedures, and (4) be aged ≥14 years. The participants were excluded if they were (1) attending the clinics because of fetal anomaly or miscarriage, (2) undergoing medical abortion, (3) in a situation that may be dangerous to utilize a mobile intervention, and (4) unable to provide consent to participate. To elicit whether a woman was in a dangerous situation, counselors asked patients as part of routine care if they felt safe in their current relationships. In cases where risk is identified, counselors provided resources and would refer to the appropriate provider or service.

Theoretical Framework

The study design and instruments were developed using the technology acceptance model (TAM) and theory of reasoned action (TRA). These theories have been validated in studies for the development of mHealth solutions in low- and middle-income countries [27,30,31]. TAM identifies 2 distinct attitudes that could predict the adoption of a new information system: perceived ease of use and usefulness of the system. TRA states that a person's behavior can be determined by 3 factors: (1) subjective norms, defined as whether or not people important to the individual believe they should perform the behavior or not; (2) individual's attitude toward the behavior; and (3) an individual's intention to engage in a behavior [30-32]. The study instruments for all 3 phases were developed using these theories based on validated survey tools [13,31,33,34]. Specifically, these tools have been utilized for formative research on developing digital health solutions specific to health behaviors.

Study Design

The overall study design is a mixed-methods user-centered design approach with 3 phases based on the *development-evaluation-implementation* process from the Medical Research Council Frameworks for Complex Medical Interventions [35]. Phase I is a community-based mixed-methods study conducted in 3 urban clinics in Vancouver, British Columbia, between May and August 2017 with the goal of understanding how women utilize technology to access health care information and their preferences of content and design of a mobile intervention to support follow-up care after an induced surgical abortion. The quantitative component of the study included an anonymous survey that was adapted from validated surveys from studies conducted for mHealth interventions and contraception use [13,31,33,36]. The survey asked questions about (1) how they use technology and their preferences for its use in clinical care and follow-up after an abortion, (2) mobile device use, privacy, and security, and (3) their past experience with contraception. The qualitative interview guide included topics on the types of wireless device used, technology engagement, acceptable information to include in a mobile intervention to support women's abortion care, willingness to use a mobile phone to obtain reproductive health information, optimal strategies to use the mobile intervention to support women, understanding preferences for health information resources, design qualities of mobile intervention, and privacy and security. Counselors at each of the abortion clinics obtained consent from eligible participants and distributed the survey for participants to complete in the waiting room or counselors' office before going for their procedure. Embedded within the survey, we asked participants whether they consented to being contacted within a week of their visit by our research coordinator to participate in an optional semistructured individual interview for the qualitative component of this study. Participants who consented to a follow-up telephone call received information regarding the qualitative component of the study, and if they were interested in participating, a convenient time was set for the principal investigator or the research coordinator to interview participants over the phone. Verbal consent was obtained at the start of the audio-recorded interviews, which lasted approximately 45 min each. Participants received a Can $25 gift card for participating in the semistructured interview. A convenience sample of approximately 50 participants for the quantitative component of the study was desired and, for the qualitative component, 10 to 20 participants or until saturation of themes was reached. Studies have suggested that thematic saturation could be reached with as few as 12 to 15 participants [37].

This study was approved by the Children's and Women's Research Ethics Board (H16-02823).

Data Analysis

Phase I

Descriptive analysis of each variable from the waiting room survey was reported as a mean (SD) or median for continuous variables and count (%) for categorical variables. The semistructured interview transcripts were uploaded to NVivo 11 (QSR International Pty Ltd) and read by 2 researchers. Inductive analysis was done to identify emerging themes, which were further refined through collaborative analysis with the first author and coinvestigator. Highlighted text was coded into nodes representing similar or repeating ideas. Some text was coded to more than 1 node to reflect the number of ideas presented. Nodes were then grouped together in a respective theme and a thematic map was developed and discussed with the research team and triangulation of data with expert opinion was used to enhance the validity of the findings.

Results

Participant Characteristics

Table 1 provides the demographic profile of participants. A total of 50 participants were recruited, and of these participants, 78% were Canadian, 84% were under the age of 30 years, 69% had an annual income less than Can $35,000, and 36% had high school education, whereas 38% were between high school and a bachelor's degree. There was variation in the distance traveled by participants, with 30% traveling less than 10 km, 22% traveling between 20 and 40 km, and 18% who traveled over 100 km. In addition, 94% (47/50) of participants reported owning a smartphone. Of the 50 participants who completed the survey, 25 consented to being contacted about the interview, and a total of 8 participants were recruited to participate in the semistructured individual phone interviews.

Table 1. Demographic profile of participants (N=50).

Demographic details	n (%)
Birthplace	
Canada	39 (78)
Other	11 (22)
Age (years)	
18-25	27 (54)
26-30	17 (34)
31-37	6 (12)
Annual income	
Can $0-15,000	16 (31)
Can $15,000-35,000	18 (38)
Can $35,000-55,000	8 (17)
Can $55,000+	7 (13)
Education	
High school	17 (36)
Between high school and bachelor’s	19 (38)
Bachelor’s or higher	13 (23)
Distanced traveled to clinic	
<10 km	15 (30)
10-20 km	5 (10)
20-40 km	11 (22)
40-60 km	6 (12)
60-80 km	4 (8)
100+ km	9 (18)
Own cell phone	
Yes (basic)	3 (6)
Yes (smartphone)	47 (94)

Quantitative Results

Results from the waiting room survey are summarized in Table 2. Specifically, 89% (42/47) of participants stated they use a smart phone most often out of a list of five commonly used devices, 92% (45/49) of participants’ phone plans included unlimited SMS plans, 92% (45/49) of participants use the internet on their phone. 88% (42/49) of the participants use their personal phones to go online, 85% (41/48) liked the idea of using a cell phone to assist in their clinical care and follow-up and 92% (45/49) of participants used mobile apps. Participants reported being more comfortable with receiving information about contraception, general postabortion care information, signs and symptoms after an abortion, mental health information and sexual health via email followed by website and mobile apps opposed to other modalities. Results are presented in Figures 1-4.

Table 2. Phase I survey results.

Survey questions (N=50)	n (%)
Which device do you use?	
Basic mobile phone	4 (8)
Smartphone	47 (94)
Tablet	17 (34)
Laptop	35 (70)
Desktop computer	16 (32)
Which device do you use the most? (N=47)	
Basic mobile phone	2 (4)
Smartphone	42 (89)
Tablet	0 (0)
Laptop	2 (4)
Desktop computer	0 (0)
All of the above	1 (2)
When you have question about your health, what do you use? (N=50)	
Internet	46 (92)
Social media	3 (6)
Mobile apps	3 (6)
Friends	20 (40)
Family	20 (40)
Healthcare Provider	34 (68)
What do you use the most often? (N=50)	
Internet	23 (46)
Social media	5 (10)
Mobile app	2 (4)
Friends	2 (4)
Family	2 (4)
HP	6 (12)
Does your cell phone plan include text messaging? (N=49)	
Yes, pay per text	0 (0)
Yes, limited short message service (SMS)	2 (4)
Yes, unlimited SMS	45 (91)
No	1 (2)
Don’t know	1 (2)
Do you use the internet on your phone? (N=49)	
Yes	45 (91)
No	3 (6)
Don’t know	1 (2)
Would like the idea of using cell phone to assist with follow-up care? (N=48)	
Yes	41 (85)
No	3 (6)
I don’t know	4 (8)
How participants go online[a] (N=49)	

Survey questions (N=50)	n (%)
Personal computer	34 (69)
Work	7 (14)
Family/friends' computer	5 (10)
Personal phone	42 (85)
Internet café	1 (2)
Family/friends' phone	1 (2)
All of the above	0 (0)
I don't go online	0 (0)
Mobile app use (N=49)	
Yes	45 (92)
No	3 (6)
I don't know	1 (2)
Features that make a mobile app easy to use[a] (N=49)	
Simple design	44 (90)
Not overly time consuming	31 (63)
Controlled push notifications	18 (37)
Easy scrolling	28 (57)
Link to other websites	4 (8)
Log in by Facebook/Google+	16 (33)
Integration with other apps	8/49 (16)
I don't know	2 (4)
Frequency of messages from clinic (N=48)	
Daily	1 (2)
Once/week	14 (29)
Twice/week	8 (4)
Once/month	25 (52)
Prefer not to see the words abortion, birth control, and contraception in a text message (N=48)	
Yes, prefer if not used	35 (73)
No, I don't care if they are used	13 (27)

[a] Participants to *mark all that applied*.

Figure 1. Number of participants who are comfortable receiving information about contraception via different modalities on their mobile device.

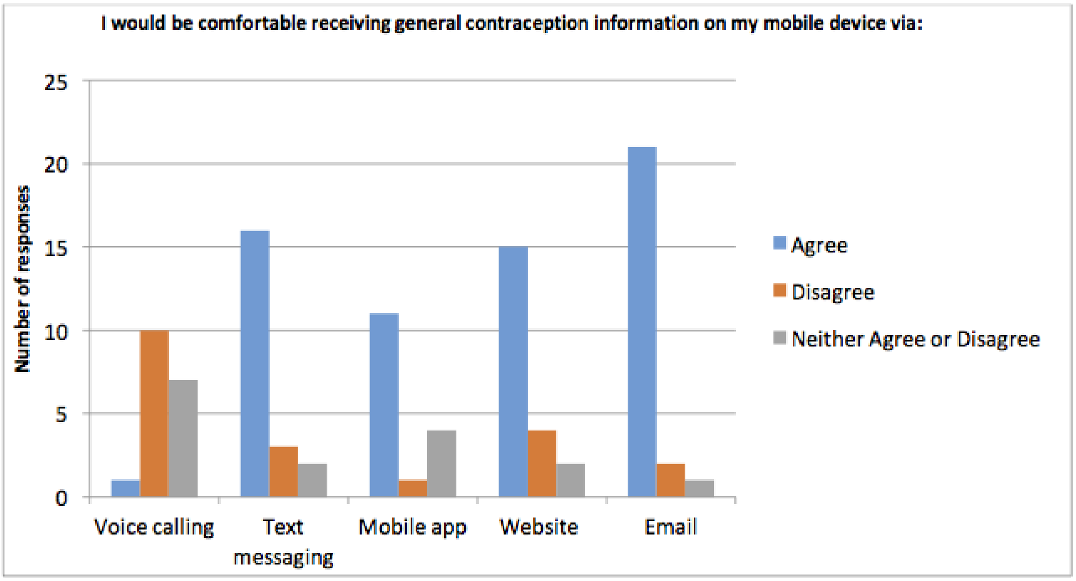


Figure 2. Number of participants who are comfortable receiving information about postabortion care via different modalities on their mobile device.

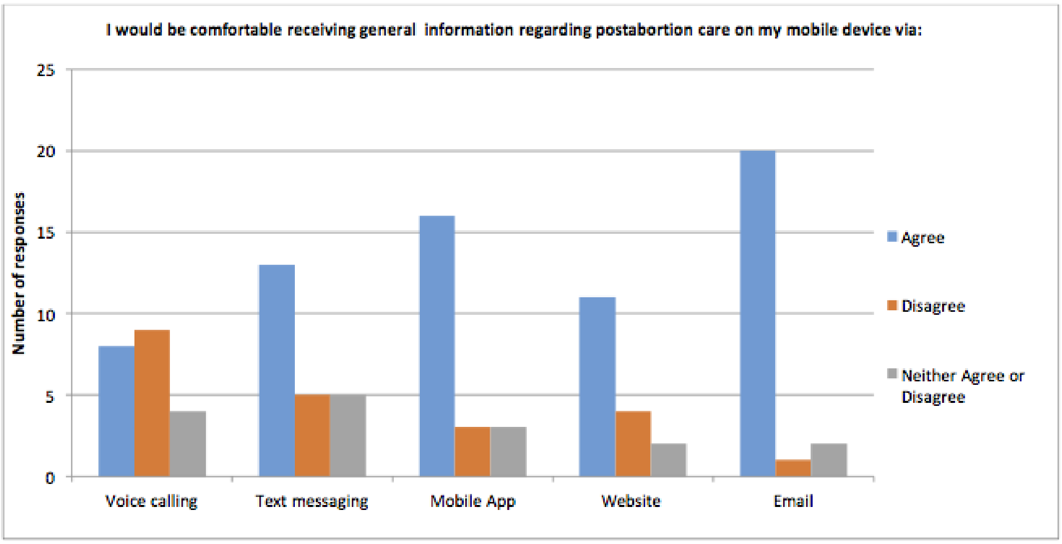


Figure 3. Number of participants who are comfortable receiving information about postoperative signs and symptoms via different modalities on their mobile device.

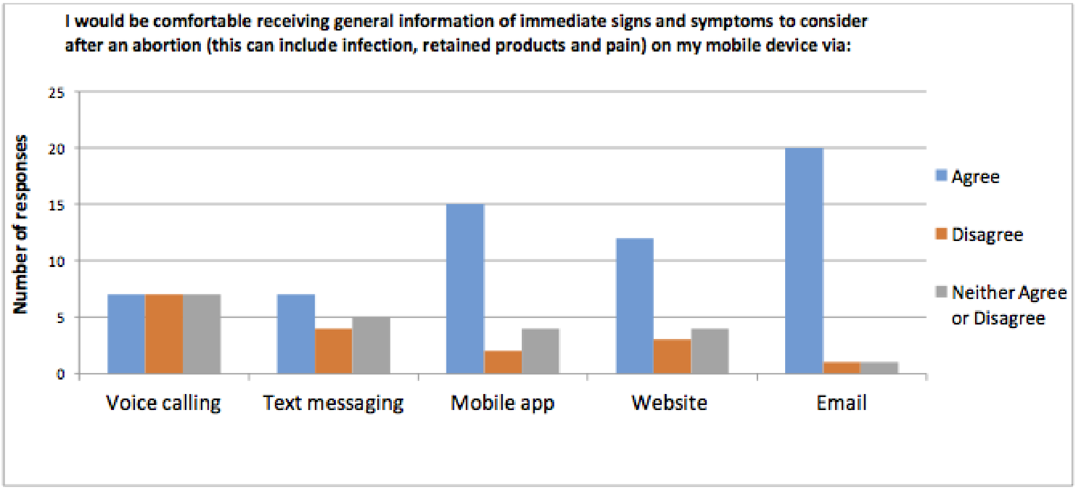


Figure 4. Number of participants who are comfortable receiving information about mental health via different modalities on their mobile device.

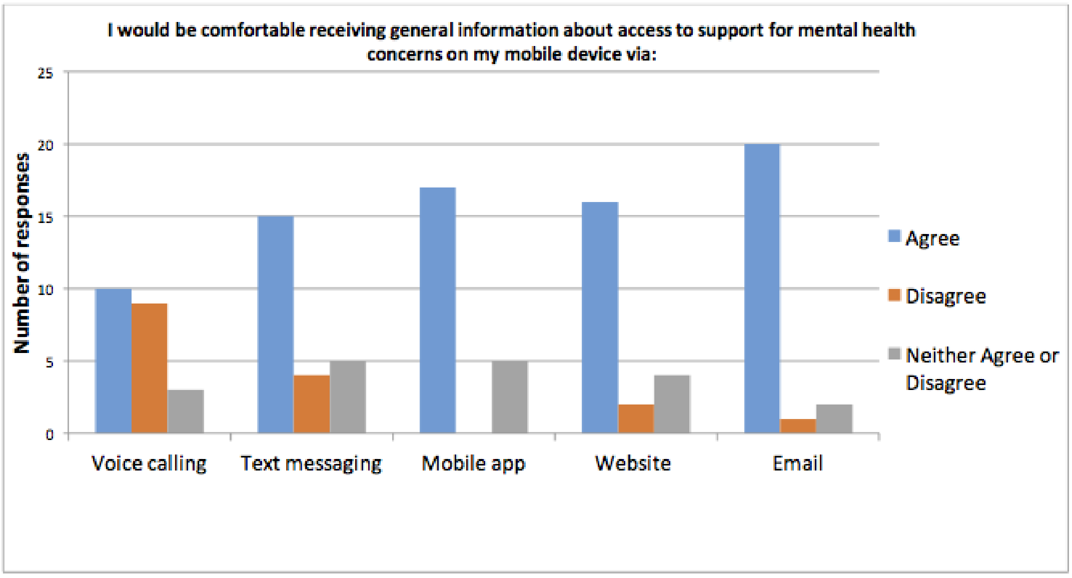


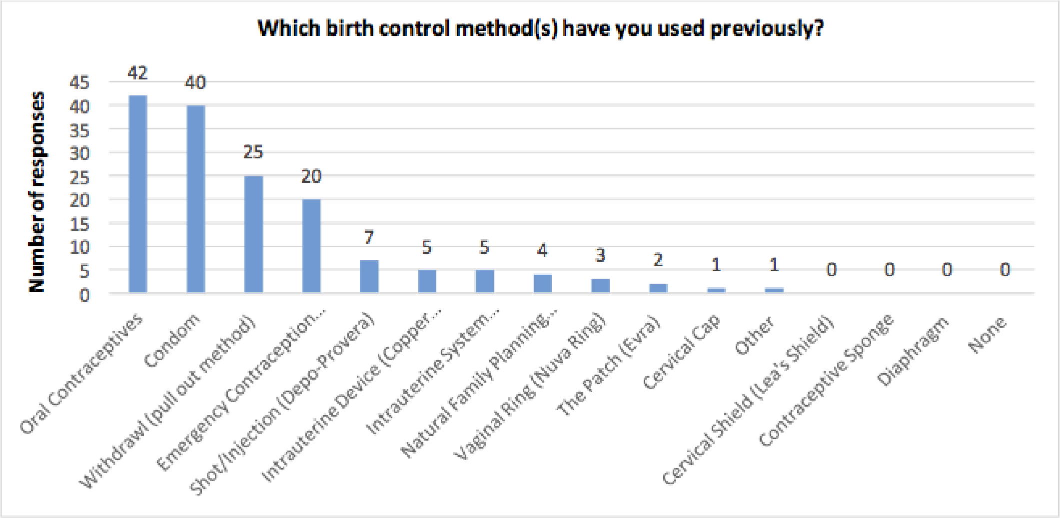
Figure 5 highlights previous contraceptive use. The top 2 methods were oral contraceptives, followed by condoms. Participants were asked if they were using a birth control method just before the index pregnancy and 48% (24/50) stated yes, 28% (14/50) said no, and 24% (12/50) said sometimes.

We further explored questions around contraceptive use and behavior using an adapted version of ORTHO Birth Control Satisfaction Assessment Tool [33]. We analyzed the questions according to lifestyle impact (63), compliance/adherence (51), and assurance/confidence on a scale of 0 to 100 (58). Higher scores indicate better satisfaction; however, the results ranged between 50 and 65, suggesting that participants were neutral with regard to the degree of satisfaction with previous contraceptive use. This information was used to assist with content and design elements for the mobile intervention.

Semistructured Interviews

A thematic analysis approach was used to analyze the interviews. We explored themes around 3 specific categories: (1) participants' overall abortion experience and how that may inform their thoughts on follow-up both for physical and/or emotional well-being, (2) current level of interaction with technology and specifically around health care–based issues, and (3) preferences for a mobile intervention to support follow-up care following a surgical abortion. Overall, 10 key themes were identified. These are listed in Textbox 1 with definitions of each on the basis of the various nodes and subnodes that were established from the analysis. Half of the participants who were interviewed traveled greater than 50 km to obtain care at 1 of the 3 urban abortion clinics. Importantly, they noted that for women such as themselves who travel far distances or are from rural areas, a resource that is comprehensive, secure, simple, and accessible is an important need and would address a gap in postabortion care.

Figure 5. Previous contraceptive methods (numerical representation), n=50.



Textbox 1. Definitions of themes on basis of nodes and subnodes. (Numbers represent the nodes and the letters [a, b, c...] represent the subnodes.)

Talking about abortion

- Decision making
- Abortion experience at the clinic
- Stigma

Accessing health information

- Health professionals
- Internet
- Friends/family
- Other sources

Barriers to health care

- Time
- Appointments/wait times

Contraception

- Access
- Education
- Current and past use
- Motivation
- Side effects

Desire for Postabortion counseling

- Accessing counseling

General Postprocedure needs

- Physical
- Emotional
- Support networks

Understanding follow-up

- Definition
- Needed
- Desired

Features and considerations of a mobile intervention

- Design
- Content
    i. Language
    ii. Topics
- Experience with previous mobile apps

Features and considerations for a text messaging service

- Content
- Convenience
- Frequency
- Preference

Privacy and confidentiality

Talking About Abortion

Participants unanimously described the experience as lonely, particularly as they felt that they were unable to talk to someone about the procedure. When asked if participants would want an intervention that allowed them to connect with other women or to resources where women share their stories, many supported this. One participant stated:

> *...I like the principle of women being able to talk about their experiences and I think that just creates more and more awareness, and awareness is a positive thing*

Some felt that for religious or cultural reasons they may not feel comfortable speaking to their health care provider and that having a mobile intervention that was comprehensive and evidence based would be a good way to access information.

Accessing Health Information

The majority of our participants utilized Google as their first point to access information but did not view this as a panacea to health information. Ultimately, participants preferred to obtain information from a trusted health care provider. One participant stated:

> *...sometimes I would probably go online first and just Google things and try and find things out but I usually try and go to a doctor eventually. When I go online it's just like premature before I decide to go to the doctor*

Participants would also use the internet to elicit opinions of other women on birth control:

> *I Googled it [referring to the intrauterine device] to see what people thought of it. I do that with every kind of birth control. But at the same time you have to take it with a grain of salt because people don't post reviews on contraception if they like it. They only post negative reviews.*

This information was useful for our study team and key stakeholders when considering content development for contraception.

Barriers to Health Care

Participants provided valuable information to highlight that stigma around abortion continues to exist among health care providers and specifically in rural areas. This theme was valuable to consider when we engaged with our rural key stakeholders to share and elicit information from their perspective of the challenges of rural abortion care. One participant stated:

> *I live in a remote community. For sure, and it's a really small town. So I probably wouldn't even feel comfortable going to somebody because it's a small town.*

Another participant commented that there is not enough information on the abortion services provided in British Columbia on the internet, and rather, the first resources that are presented are pregnancy options services that are in fact centers to convince women to keep their pregnancies. This participant stated:

> *...because when I was looking it up on the Internet a bunch of places came up for pregnancy options but none for actual clinics that do these procedures and what not.*

This highlighted that we needed to consider a resource that had a comprehensive and vetted list of resources that women could refer to.

Contraception

We asked participants about their contraceptive behaviors and preferences because this is a component of abortion care and subsequent follow-up. Participants shared their frustration that contraception is not free in Canada. Furthermore, concerns about good access were brought up frequently, as was the need to be well informed that contraception is not just for pregnancy prevention but also for prevention of sexually transmitted infections. Participants were motivated to use contraception to prevent future pregnancies.

Desire for Postabortion Counseling

This was a consistent theme among participants. Many stated that it would be useful to have access in person as well as remotely to postabortion counseling. One participant who had her second procedure had used counseling postabortion after her first termination and stated:

> *when I first got a termination years ago, I was super against it [counselling] because I was an idiot and I was 20. And then I did it anyway and it was the best thing I could have ever done.*

General Postprocedure Needs

We asked participants about their general postprocedure needs. The vast majority felt good physically but found more emotional hardships. One participant stated:

> *Physically I felt better than I expected. I was in bed for maybe a day, but it wasn't bad. It wasn't painful. Like, I have bad cramps on my period and it was at most like that. But that only lasted a short time. But otherwise the pregnancy symptoms left pretty quickly, I guess in about a week, some of them sooner than that.*

On the emotional experience, a participant stated:

> *Emotionally it was harder than I expected it to be. I knew that this was something I needed to do if this was to happen. But I never obviously had to do it before, so it was definitely harder, even though I knew it was the right thing to do. It still brought up a lot of emotional questions and feelings.*

Another participant highlighted the importance for a postprocedure intervention to include a phone call or text message that would be a check-in with the patient. Many stated it was an isolating and lonely experience and to have either a service that connected them to other resources and services or a live chat group would be something valuable.

Understanding Follow-Up

Despite follow-up being an important aspect of health care, there are little data published that examine how individuals perceive follow-up. This is important to consider when developing a postabortion support tool. When asked what follow-up means, a participant stated:

> *...with a healthcare provider, I guess just following up with any physical, medical relevant issues.*

Furthermore, some participants expressed desiring follow-up based on whether or not they had an intrauterine device inserted at the time of abortion. One participant stated:

> *They just told me to come back in six weeks to get an ultrasound of my IUD to make sure it's in the right place...and I'm assuming they'll probably see how I'm feeling emotionally, physically and—that was my understanding of the follow-up.*

When we asked participants whether they felt follow-up was needed and/or desired, a vast majority stated that it would be desired for emotional support, and they did state it was a necessity physically as that was what was told to them at the time of their procedure. However, many had recovered physically a few days after and noted that emotional and social support was desired. Some also commented that it might be good to have a mobile intervention that allowed women who lived in remote communities to have access to. The following are excerpts from individuals who highlighted the desire for emotional support and the benefit of a secure resource for those living in rural areas.

> *I feel like it's a—like, for mental health I feel like maybe it's more desired*
>
> *I think it would probably be beneficial for more people who are living in more rural or conservative communities who might not feel comfortable talking to their doctor or don't have very good access to the aftercare. I'm fine 'cause I'm in downtown Vancouver. But other people might not have such an easy time.*

Features and Considerations of a Mobile Intervention

Participants consistently stated that they wanted an easy-to-navigate, simple, professional mobile intervention that was concise and user-friendly. In addition, 2 of our participants expressed that it should be:

> *Easy to navigate...more simple, straightforward but with options to look at more information if you wanted some*

and:

> *I think just simplicity and user-friendly ease sort of is important. Not having too much information which to be honest I don't know how you do it because there's a lot of information on this topic.*

Participants also stated that the intervention should allow them to unsubscribe, particularly, if it included text or email notifications. There was also encouragement to include partners in the intervention and to have information relevant for partners specifically. One participant commented on the importance to ensure that the app or website is functioning and checked frequently. She stated:

> *Mediocre apps are just an inconvenience, and I feel like this category, it's not—I don't know, I feel like it just wouldn't get the funding that it needed for it to really work properly. I have no patience for apps that don't work well. So if it's not awesome, it's not on my phone and I'm not wasting my time using it.*

Participants suggested that options listed in the form of a dropdown list would allow for ease of navigation. Participants also wanted to be able to book counseling services directly online with the counselors from the respective abortion clinics where they received services. This was a service already provided by the counselors but participants encouraged this be part of a follow-up mobile technology resource. Finally, having more concise, patient-centered information on contraception in an easy-to-access manner was an expressed need:

> *More information on contraception and all the different types that women can use because I know a lot of my girlfriends the only type they use is a condom or birth control...I feel like there should be more information on the IUDs because it would be a really good alternative for some women who are like me and would prefer just to not have to worry about taking oral contraceptives at a certain time if they're busy or something like that. Because my current experience with birth control is after going to the clinic and receiving the information that I got, I realized the IUD was best for me. And there should be more information about that, whether it be on an app or on the Internet.*

Features and Considerations for a Text Messaging Service

Overall, participants encouraged follow-up by either an individual from the clinic texting them at set times or an email or text messaging service that was automatic and timed in appropriate frequencies. It was important that there be an unsubscribe feature. Furthermore, participants stated that weekly messages at first would be useful but then to decrease these messages to monthly up to a maximum of 6 months. The following excerpts highlight this theme:

> *I think it's a really good thing. I would definitely have somebody remind me about appointments, remind me to get a checkup,...I would definitely use that. But I think as well—as long as you consent to it and you can set the frequency yourself, like, I think it's a great idea.*
>
> *I think that would be handy as long as it had the option to stop at any time, an unsubscribe feature*

Privacy and Confidentiality

Privacy and confidentiality was very important for women who were interviewed, which was slightly different from the results of the quantitative survey. Participants stated that they would want something that is discrete and did not mention the words abortion or have any way to link the patient to the clinic where they had their procedure:

> *I think confidentiality is a big part of it, just because—especially with some people, they don't really want everyone to know. I know I don't want everybody in the whole world to know what happened.*
>
> *Privacy is everything for me. That's the most important thing for me. And I think that's maybe because I do live in a small town. It's really, really important to have that here. If you're in a big anonymous city it might be different.*
>
> *Obviously with mobile technology you want to be really considerate of privacy and have secure servers and—especially if there's contact information with a whole list of women...as it relates to mobile, that's definitely something to consider.*

Discussion

This study represents the first phase of a 3-phase study and highlights the importance of formative research that incorporates the voice of end users to assist with development of a mobile intervention for follow-up after an induced surgical abortion. Specifically, this formative mixed-methods study explored women's interactions with technology and important aspects to design and content for a mobile intervention to support follow-up after a surgical abortion. It is important to include a formative aspect to the development of technology-based interventions to improve women's health generally, but abortion care specifically [20,36,38-40]. The theoretical framework for all 3 phases included both the TAM and TRA theories. In this study particularly, we were able to elicit information about what would be important in the design of the tool, what are the contextual factors to consider in designing the tool but also around the stigma women continue to experience around abortion and how that can inform one's attitude and intentions to engage in a tool to support them in care related to an abortion. This information was utilized in the development of the intervention and further testing for phases II and III.

According to the literature, including the voice of end users early in the development of an intervention is a fundamental principle of human-centered design [16,20,38,39]. A study conducted by Smith et al [20] explored women's needs in Cambodia to develop a mobile phone–based intervention to support postabortion family planning, specifically, contraceptive adherence [20]. Similarly, researchers from University of San Francisco's Program in Women-Centered Contraception, developed a tablet-based contraceptive decision support tool for women [41]. This study utilized a multiphase approach that incorporated the end user throughout the entire design of the project. The formative research findings from these studies emphasized that using an iterative process informed by patient and provider input throughout the development and testing led to a more patient-centered innovation [42]. In addition, the formative research in these studies identified main patterns of mobile use of the women from the respective countries, main reasons and expectations for contraception and abortion care, researchers drew on components of existing interventions and behavior change theories to then develop conceptual frameworks. It also highlights that anytime a mobile intervention is to be adapted to a new context, formative research is essential to ensure the tool can be adapted to particular contexts and cultures.

In this study, a theme that resonated was the desire for a tool that incorporated emotional support as part of follow-up abortion care, which has been previously described in the published literature. A qualitative study conducted in Canada explored women's abortion experiences in the Yukon territory, a remote service area, highlighting that "fragmented services left women unsatisfied, stressed and upset about lack of information, multiple appointments and lengthy wait times" [43]. Women further expressed frustration with lack of follow-up counseling and recommended that it be routinely offered as they feel contact with health care providers is cut off after the procedure [43]. In addition to access issues, barriers of cost, knowledge among the general public, health care provider competence, and attitudes have also been highlighted in the literature [44]. Another study explored women's expressed desire for postabortion support services, highlighting the stigma around abortion that exists in political and social contexts, preventing women from sharing their experiences [1]. This study specifically highlighted that though women may not necessarily need mandatory physical follow-up, they desire access to postabortion support for emotional well-being [1]. Furthermore, there is great deal of inconsistency in the type of support and information available to women after an abortion. In addition to emotional support, participants stressed key characteristics of the mobile intervention such as privacy and confidentiality, evidence-based information, some form of interaction with the health care system, and is professional and easy to use.

Finally, it was important to capture the voice of women who traveled from rural and remote areas. According to our demographic data, over one-third of our participants traveled over 50 km to obtain an abortion. Though abortion is legal in Canada, there are access barriers, particularly for women living in remote and rural areas. Studies that have developed strategies in restrictive settings utilizing mHealth could be adapted to rural and remote contexts that have restricted access to abortion services in Canada. Providing safe and effective means, such as telemedicine, text messaging services, or mobile apps, has been proven to be acceptable and satisfactory for women in legally restricted settings [12,21-23,26] and, therefore, could have potential for Canadian women living in rural and remote areas, where abortion is legal but barriers to service delivery exist.

The limitations for this study include overall generalizability to other populations and small convenience sample sizes. Furthermore, abortion continues to be stigmatized, which can contribute to difficulties with recruitment and loss to follow-up. Accordingly, we noted that recruitment for the qualitative interviews took longer than expected and further assumed that lack of participant engagement may be associated with conducting research a few weeks out after the abortion, where participants have moved on and are resuming back in their busy lives. Consideration of recruitment strategies will need to be taken into consideration for future studies, particularly when thinking about diversifying the participants recruited and obtaining robust response rates for analysis.

Balancing these limitations are the strengths of our study, including user engagement early in the research design, a mixed-methods design incorporating both quantitative and qualitative data, engagement with both urban and rural key stakeholders, and robust findings that will inform the design of the mobile intervention.

Details of phases II and III will be published as separate papers. Specifically, phase II incorporated these findings to develop, design, and test an intervention based on findings from this study, and phase III was a pilot, prospective mixed-methods study to determine the feasibility and acceptability of the tool for women who undergo a surgical abortion.

Implications

This study was the first phase of 3 and perhaps the most important phase as it determined crucial findings about women's interactions with technology and their preferences for design and content of an intervention that could support their care after a surgical abortion. Moreover, there is great deal of momentum toward self-managed abortion care and technology has a role to play. This study is the first to utilize mHealth and user-centered design in Canada as a novel approach to provide a means for women to self-manage their follow-up care after a surgical abortion.

Conflicts of Interest

None declared.

References

1. LaRoche KJ, Foster AM. "I kind of feel like sometimes I am shoving it under the carpet": documenting women's experiences with post-abortion support in Ontario. FACETS 2017 Sep;2(2):754-763. [doi: 10.1139/facets-2017-0059]
2. Norman WV. Induced abortion in Canada 1974-2005: trends over the first generation with legal access. Contraception 2012 Feb;85(2):185-191. [doi: 10.1016/j.contraception.2011.06.009] [Medline: 22036474]
3. LaRoche K, Foster AM. Toll free but not judgment free: evaluating postabortion support services in Ontario. Contraception 2015 Nov;92(5):469-474. [doi: 10.1016/j.contraception.2015.08.003] [Medline: 26260687]
4. Sethna C, Doull M. Spatial disparities and travel to freestanding abortion clinics in Canada. Womens Stud Int Forum 2013 May;38:52-62. [doi: 10.1016/j.wsif.2013.02.001]
5. World Health Organization. 2012. Safe abortion: technical and policy guidance for health systems, 2nd Edition URL: https://www.who.int/reproductivehealth/publications/unsafe_abortion/9789241548434/en/ [accessed 2019-05-12] [WebCite Cache ID 78Jysuww5]
6. World Health Organization. 2015. Health worker roles in providing safe abortion care and post-abortion contraception URL: https://www.who.int/reproductivehealth/publications/unsafe_abortion/abortion-task-shifting/en/ [accessed 2019-05-12] [WebCite Cache ID 78Jz9bRsU]
7. National Abortion Federation. 2018. Clinical Policy Guidelines URL: https://prochoice.org/resources/clinical-policy-guidelines/ [accessed 2019-05-16] [WebCite Cache ID 78PzsXe6u]
8. Wainwright M, Colvin CJ, Swartz A, Leon N. Self-management of medical abortion: a qualitative evidence synthesis. Reprod Health Matters 2016 May;24(47):155-167 [FREE Full text] [doi: 10.1016/j.rhm.2016.06.008] [Medline: 27578349]
9. Gerdts C, Tunçalp O, Johnston H, Ganatra B. Measuring abortion-related mortality: challenges and opportunities. Reprod Health 2015 Sep 16;12:87 [FREE Full text] [doi: 10.1186/s12978-015-0064-1] [Medline: 26377189]
10. Murtagh C, Wells E, Raymond EG, Coeytaux F, Winikoff B. Exploring the feasibility of obtaining mifepristone and misoprostol from the internet. Contraception 2018 Dec;97(4):287-291 [FREE Full text] [doi: 10.1016/j.contraception.2017.09.016] [Medline: 29030227]
11. Bracken H, Lohr PA, Taylor J, Morroni C, Winikoff B. RU OK? The acceptability and feasibility of remote technologies for follow-up after early medical abortion. Contraception 2014 Jul;90(1):29-35. [doi: 10.1016/j.contraception.2014.03.016] [Medline: 24815098]
12. Grossman D, Grindlay K. Safety of medical abortion provided through telemedicine compared with in person. Obstet Gynecol 2017 Dec;130(4):778-782. [doi: 10.1097/AOG.0000000000002212] [Medline: 28885427]
13. Lester R, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
14. de Tolly KM, Constant D. Integrating mobile phones into medical abortion provision: intervention development, use, and lessons learned from a randomized controlled trial. JMIR Mhealth Uhealth 2014 Feb 14;2(1):e5 [FREE Full text] [doi: 10.2196/mhealth.3165] [Medline: 25098569]

15. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
16. Gilbert M, Haag D, Hottes TS, Bondyra M, Elliot E, Chabot C, et al. Get Checked… Where? The development of a comprehensive, integrated internet-based testing program for sexually transmitted and blood-borne infections in British Columbia, Canada. JMIR Res Protoc 2016 Sep 20;5(3):e186 [FREE Full text] [doi: 10.2196/resprot.6293] [Medline: 27649716]
17. World Health Organization. 2018. Classification of Digital Health Interventions v1.0 URL: https://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/ [accessed 2019-05-12] [WebCite Cache ID 78K15eTRy]
18. Tuckson R, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
19. Wiebe ER. Use of telemedicine for providing medical abortion. Int J Gynaecol Obstet 2014 Feb;124(2):177-178. [doi: 10.1016/j.ijgo.2013.07.038] [Medline: 24332519]
20. Smith C, Vannak U, Sokhey L, Ngo TD, Gold J, Free C. Mobile Technology for Improved Family Planning (MOTIF): the development of a mobile phone-based (mHealth) intervention to support post-abortion family planning (PAFP) in Cambodia. Reprod Health 2016 Jan 5;13(1):1-015. [doi: 10.1186/s12978-015-0112-x]
21. Grindlay K, Lane K, Grossman D. Women's and providers' experiences with medical abortion provided through telemedicine: a qualitative study. Womens Health Issues 2013;23(2):e117-e122. [doi: 10.1016/j.whi.2012.12.002] [Medline: 23410620]
22. Aiken A, Digol I, Trussell J, Gomperts R. Self reported outcomes and adverse events after medical abortion through online telemedicine: population based study in the Republic of Ireland and Northern Ireland. Br Med J 2017 May 16;357:j2011 [FREE Full text] [doi: 10.1136/bmj.j2011] [Medline: 28512085]
23. Aiken A, Gomperts R, Trussell J. Experiences and characteristics of women seeking and completing at-home medical termination of pregnancy through online telemedicine in Ireland and Northern Ireland: a population-based analysis. BJOG 2017 Jul;124(8):1208-1215 [FREE Full text] [doi: 10.1111/1471-0528.14401] [Medline: 27748001]
24. Smith C, Ngo TD, Gold J, Edwards P, Vannak U, Sokhey L, et al. Effect of a mobile phone-based intervention on post-abortion contraception: a randomized controlled trial in Cambodia. Bull World Health Organ 2015 Dec 01;93(12):842-50A [FREE Full text] [doi: 10.2471/BLT.15.160267] [Medline: 26668436]
25. 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]
26. Gerdts C, Hudaya I. Quality of care in a safe-abortion hotline in Indonesia: beyond harm reduction. Am J Public Health 2016 Dec;106(11):2071-2075. [doi: 10.2105/AJPH.2016.303446] [Medline: 27631756]
27. Mburu S, Franz E, Springer T. A conceptual framework for designing mHealth solutions for developing countries. In: Proceedings of the 3rd ACM MobiHoc workshop on Pervasive wireless healthcare. 2013 Presented at: ACM MobiHoc workshop on Pervasive wireless healthcare; July 29-29, 2013; Bangalore, India p. 1-6. [doi: 10.1145/2491148.2491154]
28. Dunn S, Cook R. Medical abortion in Canada: behind the times. Can Med Assoc J 2014 Jan 07;186(1):13-14 [FREE Full text] [doi: 10.1503/cmaj.131320] [Medline: 24277708]
29. Eckman M, Gorski I, Mehta K. Leveraging design thinking to build sustainable mobile health systems. J Med Eng Technol 2016;40(7-8):422-430. [doi: 10.1080/03091902.2016.1218560] [Medline: 27535325]
30. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
31. Smillie K, Van Borek N, Abaki J, Pick N, Maan EJ, Friesen K, et al. A qualitative study investigating the use of a mobile phone short message service designed to improve HIV adherence and retention in care in Canada (WelTel BC1). J Assoc Nurses AIDS Care 2014;25(6):614-625 [FREE Full text] [doi: 10.1016/j.jana.2014.01.008] [Medline: 24768442]
32. Davis F, Bagozzi R, Warshaw P. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
33. Colwell H, Mathias SD, Cimms TA, Rothman M, Friedman AJ, Patrick DL. The ORTHO BC-SAT--a satisfaction questionnaire for women using hormonal contraceptives. Qual Life Res 2006 Dec;15(10):1621-1631. [doi: 10.1007/s11136-006-0026-8] [Medline: 17031502]
34. Gilbert M, Salway T, Haag D, Fairley CK, Wong J, Grennan T, et al. Use of GetCheckedOnline, a comprehensive web-based testing service for sexually transmitted and blood-borne infections. J Med Internet Res 2017 Dec 20;19(3):e81 [FREE Full text] [doi: 10.2196/jmir.7097] [Medline: 28320690]
35. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
36. van der Kop ML, Memetovic J, Patel A, Marra F, Sadatsafavi M, Hajek J, et al. The effect of weekly text-message communication on treatment completion among patients with latent tuberculosis infection: study protocol for a randomised

controlled trial (WelTel LTBI). BMJ Open 2014 Apr 09;4(4):e004362-e002013 [FREE Full text] [doi: 10.1136/bmjopen-2013-004362] [Medline: 24719431]

37. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
38. Cai R, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, et al. Developing and evaluating JIapp: acceptability and usability of a smartphone app system to improve self-management in young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2017 Aug 15;5(8):e121 [FREE Full text] [doi: 10.2196/mhealth.7229] [Medline: 28811270]
39. Mummah S, King AC, Gardner CD, Sutton S. Iterative development of Vegethon: a theory-based mobile app intervention to increase vegetable consumption. Int J Behav Nutr Phys Act 2016 Aug 08;13:90 [FREE Full text] [doi: 10.1186/s12966-016-0400-z] [Medline: 27501724]
40. Huang E, Williams H, Hocking JS, Lim MS. Safe sex messages within dating and entertainment smartphone apps: a review. JMIR Mhealth Uhealth 2016 Nov 08;4(4):e124 [FREE Full text] [doi: 10.2196/mhealth.5760] [Medline: 27826133]
41. Fitzpatrick J, Anderson N, Kuppermann M, Steinauer J, Wittman A, Dehlendorf C. Pilot study of “My Birth Control,” a contraceptive counseling decision support tool. Contraception 2014 Sep;90(3):320-321. [doi: 10.1016/j.contraception.2014.05.098]
42. Dehlendorf C, Fitzpatrick J, Steinauer J, Swiader L, Grumbach K, Hall C, et al. Development and field testing of a decision support tool to facilitate shared decision making in contraceptive counseling. Patient Educ Couns 2017 Jul;100(7):1374-1381 [FREE Full text] [doi: 10.1016/j.pec.2017.02.009] [Medline: 28237522]
43. Cano J, Foster AM. "They made me go through like weeks of appointments and everything": documenting women's experiences seeking abortion care in Yukon territory, Canada. Contraception 2016 Dec;94(5):489-495. [doi: 10.1016/j.contraception.2016.06.015] [Medline: 27373541]
44. Hulme J, Dunn S, Guilbert E, Soon J, Norman W. Barriers and facilitators to family planning access in Canada. Healthc Policy 2015 Feb;10(3):48-63 [FREE Full text] [Medline: 25947033]

Abbreviations

mHealth: mobile health
TAM: technology acceptance model
TRA: theory of reasoned action

Edited by G Eysenbach; submitted 21.01.19; peer-reviewed by B Payne, K Reiss, C Smith, D Levine, T Kuhn; comments to author 14.02.19; revised version received 10.04.19; accepted 28.04.19; published 29.05.19.

Please cite as:
Gill R, Ogilvie G, Norman WV, Fitzsimmons B, Maher C, Renner R
Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care in British Columbia: Phase I
J Med Internet Res 2019;21(5):e13387
URL: http://www.jmir.org/2019/5/e13387/
doi:10.2196/13387
PMID:31144668

Original Paper

Perceived Patient-Provider Communication Quality and Sociodemographic Factors Associated With Watching Health-Related Videos on YouTube: A Cross-Sectional Analysis

Aisha Langford[1], PhD, MPH; Stacy Loeb[2], MD, MSc

[1]Department of Population Health, NYU Langone Health, New York, NY, United States

[2]Department of Urology and Population Health, NYU Langone Health, New York, NY, United States

Corresponding Author:
Aisha Langford, PhD, MPH
Department of Population Health
NYU Langone Health
227 E 30th St
New York, NY, 10016
United States
Phone: 1 646 501 2914
Email: aisha.langford@nyulangone.org

Abstract

Background: Approximately 73% of US adults use YouTube, making it the most popular social media platform. Misinformation on social media is a growing concern; recent studies show a high proportion of misinformative health-related videos. Several studies on patient-provider communication and general health information seeking have been conducted. However, few studies to date have examined the potential association between patient-provider communication and health information seeking on specific social media platforms such as YouTube. A better understanding of this relationship may inform future health communication interventions.

Objective: The aim was to use nationally representative cross-sectional data to describe the association between perceived patient-provider communication quality and sociodemographic factors on watching YouTube health-related videos.

Methods: Data from the 2018 Health Information National Trends Survey were analyzed (N=3504). The primary outcome was whether participants watched a health-related video on YouTube over the past 12 months. A patient-provider communication composite score was created by summing responses about how often providers did the following: (1) gave you the chance to ask all the health-related questions you had, (2) gave attention to your feelings, (3) involved you in health care decisions as much as you wanted, (4) made sure that you understood the things you needed to do to take care of your health, (5) explained things in a way that you could understand, (6) spent enough time with you, and (7) helped you deal with feelings of uncertainty. Sociodemographic factors included age, gender, race/ethnicity, and education. Descriptive statistics and multivariable logistic regression were conducted.

Results: Approximately 1067 (35% weighted prevalence) participants reported watching a health-related video on YouTube. Higher perceived quality of patient-provider communication on the composite score was significantly associated with lower odds of watching health-related videos on YouTube. Regarding sociodemographic factors, increasing age and being a high school graduate (compared with college graduate) were associated with lower odds of watching health-related videos on YouTube; whereas, Hispanic and non-Hispanic Asians were more likely to have watched a health-related video on YouTube. For individual aspects of patient-physician communication, two of seven patient-provider communication variables were significant. Those who reported that providers "sometimes" spent enough time with them had higher odds of watching a health-related video on YouTube, compared with those who said providers "always" spent enough time with them. Participants reporting that they "never" have a chance to ask all their health-related questions also had higher odds of watching health-related videos on YouTube compared with those who reported "always."

Conclusions: Higher perceived quality of patient-provider communication is associated with lower odds of watching health-related videos on YouTube. When providers do not spend enough time or give an opportunity to ask questions, patients are more likely to pursue health information on social media.

(J Med Internet Res 2019;21(5):e13512) doi:10.2196/13512

KEYWORDS

social media; communication; health communication; ethnic groups; physician-patient relations; emotions; attention; cross-sectional studies; logistic models; HINTS

Introduction

Background

Historically, patient-provider communication has been associated with various health outcomes [1-3] and health information-seeking behaviors [4,5]. For example, one study found that problems with patient-centered communication and clinical care coordination were associated with a higher likelihood of independent eHealth engagement [6], whereas another study found that internet health information seeking could improve the patient-physician relationship depending on the history of the relationship and whether the patient discussed the information with their doctor [4]. However, an underexplored area of research is the association between patient-provider communication and health information seeking on specific social media platforms. As of January 2018, the Pew Research Center reported that nearly 70% of Americans use at least one social media site [7]. By popularity, YouTube ranks first with 73% of Americans using the site, followed by Facebook at 69% and Instagram at 37% [8].

Given the popularity of social media and relative ease in which information can be posted online, health-related misinformation on social media has become a growing public health concern that may affect patient-provider communication [9-12]. A number of studies have explored how and why patients use social media for health-specific purposes [13-15]. For example, Benetoli et al [16] conducted focus groups with Australian consumers and found that blogs helped consumers learn about other people’s experiences with the same condition, Facebook allowed them to follow health-related pages of interest and participate in disease-specific group discussions, Wikipedia was used to help gather information about health conditions and treatments, and YouTube was used to learn about medical procedures including surgery. Other studies have evaluated both the quantity and quality of YouTube video content for various health conditions and behaviors including prostate cancer, infertility, and smoking [11,17-24], as well as, more broadly, how YouTube videos tags are assigned and described by the disseminator [25].

Although many health-related videos on YouTube are deemed educationally useful and are of high quality [22,26-29], some studies show that health-related videos on YouTube are often of poor quality, misleading, and/or have commercial content designed to sell products or services [11,12,30-32]—all which may have serious implications for consumer attitudes and medical decision making. For example, our group recently examined the top 150 YouTube videos on prostate cancer and found that 77% had misinformative and/or biased content in the video or comments, and that there was an inverse relationship between views and thumbs up with expert-rated quality [11]. Another study showed that YouTube videos portraying immunization negatively were more highly rated by users than positive videos, but 45% of the negative videos contained misinformation [33]. In general, viewer engagement with YouTube videos appears to be higher when health-related videos contain personal stories, misinformation, and/or nonrecommended therapies [11,17,34,35].

Theory Considerations and Health Communication

The rapid growth of internet and social media use over the past 20 years also has implications for health communication or “the study and use of communication strategies to inform and influence individual and community decisions that enhance health” [36]. Health communication strategies may include, but are not limited to, the exchange of information between individuals, development of health messages, providing multiple ways for people to access health information, and ensuring that health information meets the needs of people at varying health literacy levels [37]. Although no single theory or model captures all the factors affecting health communication, theories can be combined to better understand these processes; this decision should be based on the health issue or problem, target population, and context [38]. For example, the Basic Communication Model highlights that source and channel, message, and audience are the key elements of communication [39]; whereas, the Chronic Care Model posits that an activated patient and a prepared, proactive health care team are needed to have productive interactions [40]. Street’s [41] Ecological Model of Communication in Medical Encounters helps explain “how communication in these interactions is (or can be) affected by the interpersonal, organizational, media, political-legal, and cultural environments within which they take place.” This model also suggests that patients and providers have predisposing factors such as communication style, verbal and nonverbal behaviors, and cognitive-affective influences that impact the patient-provider communication process [41]. The Technology Acceptance Model, which is an adaption of the Theory of Reasoned Action, was originally developed to assess perceived usefulness and ease of use of technology in workplace settings [42-44] but has also been used to understand consumer acceptance of health technology [45-47].

More broadly, information processing theories such as the Elaboration Likelihood Model can also be informative in the contexts of patient-provider communication and information seeking on social media [48,49]. The Elaboration Likelihood Model suggests that persuasive messages are processed through one of two routes: central or peripheral. Central route processing occurs when people are highly motivated to think about an issue (and have sufficient time and cognitive resources to do so), which should lead to greater “elaboration” or thoughtful, deliberate weighing of the message attributes. In contrast, peripheral route processing occurs in low motivation situations in which people are more prone to rely on peripheral cues such as a celebrity endorsement of a product or behavior, or the esthetic features of the health information product. Taken together, the aforementioned theories and the Health Information National Trends Survey (HINTS) framework [50] might suggest that patient-provider communication may affect how a patient's health information needs are met, which in turn may influence their health information-seeking behaviors, potentially leading them to seek health information on social media platforms.

The purpose of this study was to describe the association between perceived patient-provider communication quality and sociodemographic factors on the likelihood of watching of health-related videos on YouTube. We hypothesized that younger people, minorities, and individuals who do not feel like they had enough time with their provider or did not have their questions answered would be more likely to seek health information on YouTube. This research is important because it may shed light on which specific patient-provider communication aspects are associated with health information seeking on social media and which patient subgroups are more likely to use YouTube to watch health-related videos, both of which may inform future health communication interventions. Social media platforms including YouTube have great potential for the delivery of behavioral interventions, but this area of research is still in its infancy. A better understanding of which health consumers are using these platforms for health information and what perceived gaps they fill could help with this burgeoning field.

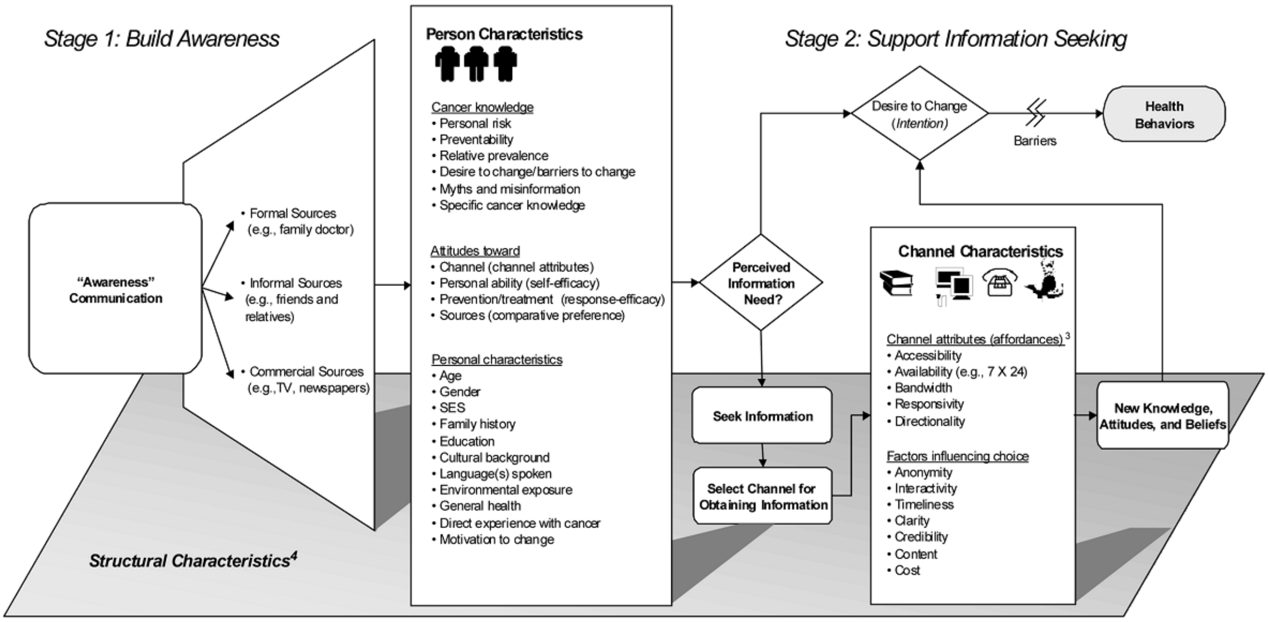
Methods

Brief Overview of the Health Information National Trends Survey

HINTS is a probability-based, nationally representative cross-sectional survey of noninstitutionalized US adults aged 18 years and older that was designed to monitor trends in the American public's use of cancer and other health-related information. The survey was developed by the National Cancer Institute's Health Communication and Informatics Research Branch and has been administered approximately every 2 years since 2003. HINTS is guided by a conceptual framework (see Figure 1) informed by the communication and behavioral science literature. To help ensure participant understanding, HINTS surveys involve at least two rounds of cognitive testing; field testing is also conducted. Criteria for inclusion of survey items are based on scientific validity (ie, use of established measures for assessing constructs of interest), utility of information for key stakeholders, and implementation considerations [50]. For race and ethnicity, blacks and Hispanics are oversampled to help enhance minority representation. Full details about general HINTS methodology are documented elsewhere [50-52].

This study evaluated participant data from HINTS 5, Cycle 2. The sample design consisted of two stages. In the first stage, an equal probability sample of addresses was selected from within each explicit sampling stratum. In the second stage, one adult was selected within each sampled household using the next birthday method. The sampling frame consisted of a database of addresses used by Marketing Systems Group to provide random samples of addresses. Data were collected between January and May 2018 via a mailed survey. A US $2 prepaid monetary incentive was included to encourage participation. Benefits of the sampling approach include enhanced geographic and demographic diversity. Additionally, data collection was anonymous and done via a pen-and-paper survey, two strategies to help reduce social desirability bias. Complete data for HINTS 5, Cycle 2, were collected from 3504 respondents and the overall survey response rate was 32.9%. As this study involved an analysis of a nonidentified publicly available data, institutional review board review approval was not required.

Figure 1. Health Information National Trends Survey (HINTS) framework, which relates psychosocial characteristics to the dynamic process of becoming aware of, and then seeking new information on, public health messages. Source: [50]; reprinted with permission from Taylor & Francis Ltd.



Measures

Health-Related Video on YouTube

The primary outcome was whether participants watched a health-related video on YouTube over the past 12 months (yes/no).

Patient-Provider Communication

A patient-provider communication composite score was created by summing responses to questions about how often doctors, nurses, or other health care professionals did the following: (1) gave you the chance to ask all the health-related questions you had, (2) gave the attention needed for your feelings and emotions, (3) involved you in health care decisions as much as you wanted, (4) made sure that you understood the information needed to take care of your health, (5) explained things in a way that you could understand, (6) spent enough time with you, and (7) helped you deal with feelings of uncertainty. Per Epstein and Street [53], these survey items represent key aspects of patient-centered care and have been used in several studies [6,54-58]. Response options were recoded so that a higher number would indicate more positive patient-provider communication (ie, 1=never, 2=sometimes, 3=usually, and 4=always). The minimum and maximum patient-provider composite scores were 7 and 28, respectively. The Cronbach alpha for the seven patient-provider communication items was .92; this high internal consistency provides support for creating a composite score. Patient-provider communication items were also evaluated individually.

Sociodemographic Factors

Age, gender (male/female), race/ethnicity (white, black, Hispanic, Asian), and education (less than high school, high school graduate, some college, and college graduate) were evaluated.

Statistical Analyses

Per the HINTS analytic recommendations, replicate weights were applied to compute accurate standard errors for statistical testing procedures and to estimate US population-level percentages. Descriptive and inferential statistics including chi-square and *t* tests, and two multivariate logistic regression models were conducted. The first logistic regression model explored the association between watching a health-related video on YouTube and the patient-provider composite score, age, gender, race/ethnicity, and education. The second logistic regression model explored the association between watching a health-related video on YouTube and the seven individual patient-provider communication items, age, race/ethnicity, and education. Gender was removed from the second model because it was not significant in either model. All analyses were conducted using Stata 14.2 [59]. We excluded missing data from the analyses.

Results

As shown in Table 1, the participants included in this analysis were mostly non-Hispanic white, with some college, and a mean age of 48.9 (SE 0.3) years. The overall mean composite score for patient-provider communication was 23.8 (SE 0.1). Approximately 1067 participants (35% weighted prevalence) reported watching a health-related video on YouTube in the last 12 months. Participants who reported watching a health-related video on YouTube were generally younger (eg, 66% were younger than 50 years old) and slightly more were female (54%). Those who watched a health-related video on YouTube had a lower mean composite score for patient-provider communication compared to those who did not report watching a video (mean 23.1, SE 0.2 vs mean 24.2, SE 0.1; $t_{49}=-3.74$, $P<.001$). Additionally, compared with those who did not report watching a health-related video on YouTube, those who watched a health-related video on YouTube had lower percentages for endorsing that health care providers “always” did the seven patient-provider communication behaviors listed (versus never, sometimes, or usually).

In the first multivariable logistic regression model (Table 2), a higher composite score for patient-provider communication (odds ratio [OR] 0.95, 95% confidence interval [CI] 0.92-0.99, $P=.02$) and increasing age (OR 0.96, 95% CI 0.95-0.97, $P<.001$) were associated with lower odds of having watched a health-related video on YouTube. Compared with non-Hispanic whites, Hispanics and non-Hispanic Asians had higher odds of watching a health-related video on YouTube (OR 1.65, 95% CI 1.13-2.42, $P=.01$ and OR 2.40, 95% CI 1.19-4.81, $P=.02$, respectively). Regarding education, high school graduates had lower odds of having watched a health-related video on YouTube compared with those with a college degree or higher (OR 0.57, 95% CI 0.39-0.85, $P=.007$).

In the second multivariable logistic regression model that evaluated the seven patient-provider communication items individually (Table 3), only two items were significant. Compared with those who said providers “always” spent enough time with them, those who reported that providers “sometimes” spent enough time with them had higher odds of watching a health-related video on YouTube (OR 1.92, 95% CI 1.17-3.14, $P=.01$). Additionally, participants who said providers “never” give them to chance to ask all their health-related questions had higher odds of watching a health-related video on YouTube compared with those who said providers “always” did so (OR 4.78, 95% CI 1.16-19.63, $P=.03$). Age, race/ethnicity, and education remained significant.

Table 1. Participant characteristics according to watching health-related videos on YouTube.

Characteristic and category (raw counts)	All (N=3504)	Watched YouTube video (n=1067)	Did not watch YouTube video (n=2361)	*P* value[a]
Age (n=3417), mean (SE)[a]	48.9 (0.3)	42.0 (0.8)	52.6 (0.6)	<.001
Age group (years), weighted %[a]				**<.001**
18-34 (n=406)	24	36	16	
35-49 (n=658)	27	30	25	
50-64 (n=1113)	30	26	32	
65-74 (n=736)	11	6	14	
≥75 (n=504)	8	2	11	
Gender, weighted %[a]				**.11**
Men (n=1394)	49	46	50	
Women (n=2054)	51	54	49	
Race/Ethnicity, weighted %[a]				**<.001**
Non-Hispanic white (n=1983)	67	59	71	
Non-Hispanic black (n=444)	11	11	11	
Hispanic (n=461)	17	20	15	
Non-Hispanic Asian (n=138)	5	9	3	
Education, weighted %[a]				**<.001**
Less than high school (n=275)	9	6	10	
High school graduate (n=631)	22	16	26	
Some college (n=1039)	40	45	38	
College degree (n=1508)	29	34	26	
Patient-Provider communication composite score (n=2871), mean (SE)[a]	23.8 (0.1)	23.1 (0.2)	24.2 (0.1)	<.001
Patient-Provider communication individual items,[b] weighted %				
Chance to ask health-related questions (n=2945)	63	57	65	.001
Attention needed for your feelings and emotions (n=2936)	49	43	51	.05
Involved you in health care decisions (n=2933)	57	51	59	.007
Made sure you understood things needed to do (n=2936)	65	62	68	.07
Explained things in a way that you could understand (n=2933)	66	61	69	.12
Spent enough time with you (n=2923)	48	40	53	<.001
Helped you deal with feelings of uncertainty (n=2917)	47	41	50	.06

[a] Testing for differences in distributions between those who have and have not watched a health-related video on YouTube.

[b] Values for these variables represent the percent of people who answered “always.”

Table 2. Multivariable logistic regression of the association between patient-provider communication composite score, sociodemographic factors, and watching health-related videos on YouTube (N=2408).

Item	OR[a] (95% CI)	*P* value
Patient-Provider communication score (continuous)	0.95 (0.92-0.99)	.02
Age (continuous)	0.96 (0.95-0.97)	<.001
Gender		
Men (ref[b])	1.0	
Women	1.25 (0.91-1.71)	.15
Race/Ethnicity		
Non-Hispanic white (ref)	1.0	
Non-Hispanic black	1.25 (0.80-1.96)	.31
Hispanic	1.65 (1.13-2.42)	.01
Non-Hispanic Asian	2.40 (1.19-4.81)	.02
Education		
≥College degree (ref)	1.0	
Some college	1.14 (0.82-1.58)	.41
High school graduate	0.57 (0.39-0.85)	.007
Less than high school	0.66 (0.29-1.47)	.31

[a] OR: odds ratio.

[b] ref: reference.

Table 3. Multivariable logistic regression of the association between individual patient-provider communication items, sociodemographic factors, and watching health-related videos on YouTube (N=2427).

Item	OR[a] (95% CI)	*P* value
Chance to ask health-related questions		
Always (ref[b])	1.0	
Usually	1.06 (0.72-1.54)	.75
Sometimes	0.49 (0.22-1.08)	.08
Never	4.78 (1.16-19.63)	.03
Attention needed for your feelings and emotions		
Always (ref)	1.0	
Usually	0.85 (0.53-1.36)	.49
Sometimes	0.95 (0.50-1.79)	.88
Never	0.81 (0.31-2.13)	.67
Involved you in health care decisions		
Always (ref)	1.0	
Usually	1.11 (0.70-1.74)	.64
Sometimes	1.50 (0.78-2.88)	.21
Never	3.17 (0.83-12.08)	.09
Made sure you understood the things you needed to do		
Always (ref)	1.0	
Usually	0.97 (0.60-1.58)	.93
Sometimes	0.98 (0.46-2.11)	.97
Never	0.72 (0.19-2.75)	.63
Explained things in a way that you could understand		
Always (ref)	1.0	
Usually	0.94 (0.51-1.73)	.84
Sometimes	0.73 (0.26-2.04)	.55
Never	0.45 (0.06-3.29)	.43
Spent enough time with you		
Always (ref)	1.0	
Usually	1.39 (0.90-2.13)	.12
Sometimes	1.92 (1.17-3.14)	.01
Never	2.62 (0.83-8.22)	.10
Helped you deal with feelings of uncertainty		
Always (ref)	1.0	
Usually	0.89 (0.52-1.51)	.68
Sometimes	0.96 (0.48-1.92)	.93
Never	0.63 (0.26-1.53)	.31
Age (continuous)	0.96 (0.95-0.97)	<.001
Race/Ethnicity		
Non-Hispanic white (ref)	1.0	
Non-Hispanic black	1.29 (0.82-2.02)	.26
Hispanic	1.68 (1.12-2.53)	.01
Non-Hispanic Asian	2.27 (1.12-4.60)	.02

Item	OR[a] (95% CI)	*P* value
Education		
College degree (ref)	1.0	
Some college	1.14 (0.82-1.58)	.42
High school graduate	0.57 (0.38-0.85)	.007
Less than high school	0.66 (0.32-1.39)	.28

[a] OR: odds ratio.

[b] ref: reference.

Discussion

Principal Findings

The purpose of this study was to describe the association between perceived patient-provider communication quality and sociodemographic factors on watching health-related videos on YouTube. In summary, we found that perceived patient-provider communication quality (measured on a composite scale and as individual items), age, race/ethnicity, and education were significantly associated with watching health-related videos on YouTube. Our findings are in line with previous studies showing poor patient-provider communication was associated with higher online health information seeking [6,60] and sociodemographic differences in use of social media platforms [8,61-63]. New findings are that higher perceived patient-provider communication quality is associated with lower odds of watching health-related videos on YouTube.

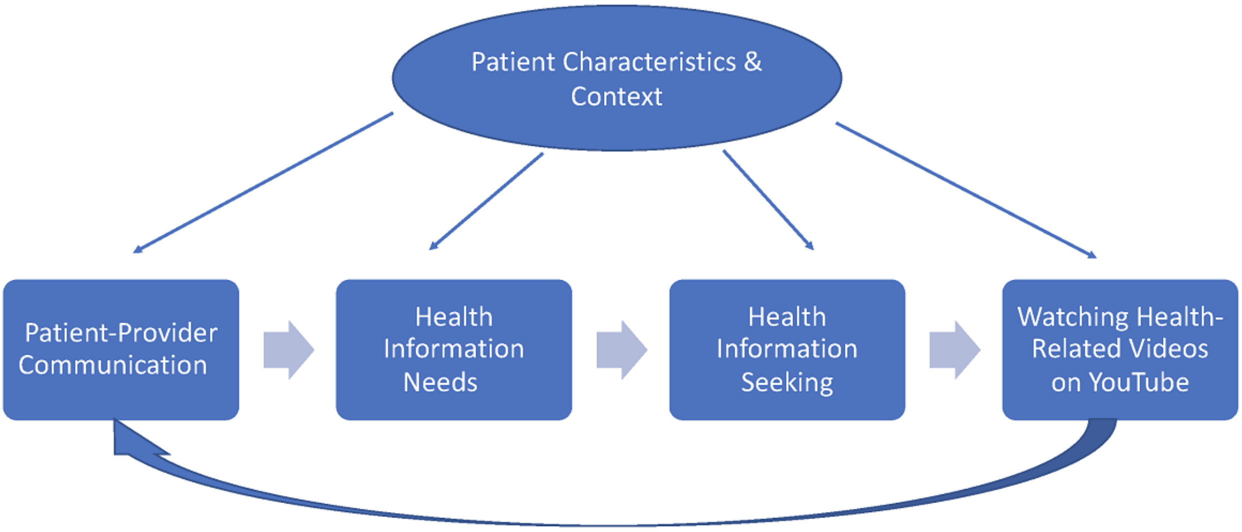
To date, several studies have used HINTS data to explore different aspects of patient-provider communication including disparities in communication [55], the effects of health utilization and sociodemographic factors on patient-provider communication [64], degree of patient-centeredness in communication with cancer survivors [54], and the role of patient-centered communication on different types of eHealth usage [6]. Other studies have explored online health information seeking after medical visits broadly [60,65] and health care information on YouTube specifically [12]. However, ours is the first-known study to examine patient-provider communication in the specific context of watching health-related videos on YouTube. This research is important because there may be unique reasons that some people seek out health information on YouTube as compared with other channels. There may also be opportunities for health care providers to assist patients in finding credible sources of information on YouTube before, during, and after a medical visit. In addition, these data are important for further efforts to use YouTube for delivery of behavioral interventions, which is a burgeoning area of research.

Patient-Provider Communication and Watching Health-Related Videos on YouTube

Overall, those with a higher perceived quality of patient-provider communication had lower odds of watching health-related videos on YouTube. Regarding the seven patient-provider communication items that were evaluated individually, only two were associated with watching health-related videos on YouTube: (1) "never" having a chance to ask all health-related questions and (2) "sometimes" feeling that health care provider spends enough time with you compared with people who reported that providers "always" did these behaviors. Although several studies have been conducted on the link between patient-provider communication and health outcomes, our findings provide a new window into the specific associations between patient-provider communication and health information seeking on YouTube. Our findings may support the idea of a feedback loop regarding patient-provider communication and health information seeking on YouTube (see Figure 2).

Poor patient-provider communication during a medical visit may have implications for how, if at all, a person's health information needs are addressed. This can lead a person to conduct more health information seeking after the medical visit by watching health videos on YouTube because they still have questions or because they are anxious, which may potentially expose the patient to misinformation on YouTube. In turn, exposure to misinformation on YouTube may affect the quality of patient-provider communication in future medical visits because the time is largely spent discussing non-evidenced-based recommendations found on YouTube (eg, injecting herbs into the prostate will cure prostate cancer, immunization is dangerous) [11,33,66]. Consequently, there may be less time to talk about other issues, such as which pharmacological and nonpharmacological treatments may be most efficacious for the patient and the harms, benefits, and quality of life considerations associated with each option. Conversely, watching health-related videos on YouTube before or after a medical encounter may improve a patient's knowledge of a health condition and related treatment options, which may enhance patient-provider communication in future visits because the patient is more informed and potentially clearer about their goals and preferences [4,67]. In both scenarios, patient characteristics and context will impact the patient-provider communication process, health information needs, and health information-seeking behaviors.

Figure 2. Conceptualization of the relationship between patient-provider communication quality and watching health-related videos on YouTube.



One can also argue that discussing *any* information that a patient finds online, regardless of a subjective assessment of whether that information is misinformative, provides additional opportunities to build the patient-provider relationship. For example, a hypertensive patient may seek out YouTube videos on "natural ways" to cure hypertension because they are concerned about overtreatment and potentially falling if their blood pressure is treated too aggressively. Another patient may be drawn to watching certain health-related videos on YouTube because of their illness beliefs [68], philosophies about medical interventions [69], or from a desire to learn more about how other patients such as them manage their condition [70]. If providers are nonjudgmental and invite discussion about *why* certain types of health information resonate with a patient, this may provide an opportunity for the provider to better understand the patient's concerns, values, and preferences.

Sociodemographic Factors and Watching Health-Related Videos on YouTube

We found that Hispanic and non-Hispanic Asian participants had higher odds of watching health-related videos on YouTube compared with white participants. Although we cannot fully explain this finding, a potential explanation may be due to cultural perspectives about the authority of health care professionals, which may lead these groups to be less likely to ask questions during medical visits, potentially leaving unanswered questions and leading them to seek out health videos on YouTube. Trust of health care professionals may have also played a role, which may have affected patient-provider communication. In a prior study using HINTS data from 2014, Singh et al [64] found that Hispanics and Asians reported lower quality patient-centered provider communication compared to whites; whereas, in a separate study on shared decision making, Levine et al [71] found lower scores for patient-perceived shared decision making in Asians compared to white adults.

We also found that as age increased, the odds of watching a health-related video on YouTube decreased. These findings are consistent with general trends in social media consumption by age [8,61,63]. Finally, regarding education, people who were high school graduates were 43% less likely to report watching a health-related video on YouTube compared to those with a college degree. It is likely that those with greater levels of education have greater self-efficacy for finding information online or are possibly exploring alternative knowledge sources beyond their health care providers. Moreover, lower levels of health technology acceptance [72,73] and eHealth literacy [74,75] among high school graduates may also partially explain this finding.

Strengths, Limitations, and Future Directions

Our findings provide an important and nuanced contribution to the literature given that we evaluated two aspects of health communication typically examined separately: patient-provider communication and mass communication via social media. Strengths of this study include the use of HINTS, a large nationally representative survey designed to track changes in health communication and information technology in the United States. The questions asked in HINTS allowed us to examine perceived patient-provider communication quality and health information seeking on YouTube which is the most commonly used social media platform in the United States. Despite its strengths, some limitations should be noted. First, we do not know the reasons why participants were watching health-related videos on YouTube (eg, to learn about a screening test or disease), for whom they were watching (eg, themselves, parent, significant other), or the quality of the videos viewed. Second, the YouTube-related question in HINTS was general and did not provide a specific definition of what constituted a "health-related" video; therefore, we do not know what participants considered as being "health related" (eg, videos about getting a mammography, weight lifting, mindfulness meditation). HINTS 5, Cycle 2, did not include measures of eHealth literacy or the “ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem” [76]. This may have affected how and why people sought out health information on YouTube.

Directions for Future Research

This study raises several potential questions for future inquiry. First, it should be acknowledged that social media has the potential to support patients in several important ways, including meeting their informational and emotional needs [67]. Social media also has the potential to affect attitudes and beliefs in ways that are counter to the goals of many health care professionals (eg, people avoiding vaccines because of antivaxer messages on social media) [12,33]. A better understanding of why and how patients are using different social media platforms for health-related purposes may shed light on which needs that are not being addressed in the routine clinical encounters and how these needs may be partially supported with population-level approaches to health communication. Second, future work should explore the degree to which health care professionals are helping patients find health information on YouTube and which sources of information they are recommending. Third, future research should explore whether strategies to enhance patient-provider communication and access to interprofessional health care teams (eg, doctors, nurses, pharmacists, health coaches, consumer health librarians) affects the likelihood that patients will watch health-related videos on YouTube as team-based care theoretically provides multiple “touch points” and, thus, more opportunities for quality patient-provider communication. Fourth, as patients navigate the vast sea of health information on YouTube, further investigation is needed to understand the process by which patients determine whether health information is useful and credible.

Fifth, the notion of misinformation on social media has garnered a lot of attention in recent years [9,11,32]; however, it is not clear what exactly constitutes misinformation and who gets to determine whether something is misinformative. This raises other questions for health communicators regarding what the “gold standard” of health information should be when equipoise exists or there is disagreement between professional societies about guidelines. Relatedly, different studies have used different approaches to assess the quality of health information. Future work should aim to develop more standardized approaches to assessing and labeling content as misinformative. Sixth, further studies are needed on the impact of illness beliefs and representation on patient-provider communication and subsequent information seeking on YouTube [77-79]. It is possible that people who are watching health-related videos on YouTube are doing so because they (1) want to confirm the information given by their physicians, (2) simply want more information to better understand their health condition, or (3) disagree with the treatment plan recommended by their doctor and are exploring other options and knowledge sources that support their beliefs. The latter option supports the notion of “confirmation bias” or the concept that people seek out information that supports their preexisting beliefs [80,81]. Finally, further evaluation of which theory (or combination of theories and constructs) best explains the relationship between patient-provider communication and watching health-related videos on YouTube is needed. Notably, several concepts from various disciplines are relevant, including technology acceptance, eHealth literacy, information processing, social media self-efficacy, interpersonal communication, mass communication, and message framing.

Conclusions

Lower perceived quality of patient-provider communication is associated with higher odds of watching health-related videos on YouTube. This study could have implications for health professionals (eg, try to give their patients enough time to ask all their questions) and for researchers seeking to use social media for health promotions to understand their potential audience. As social media grows in popularity, more research is needed on the relationship between the impact of patient-provider communication and health-related information seeking on YouTube. In particular, which features of health-related videos are most likely to engage consumers also needs to be examined so that health communication professionals can design population-level health communication interventions that are credible and evidence-based, yet still appealing to the audiences they are trying to reach.

Acknowledgments

AL thanks the National Heart, Lung, and Blood Institute for career development support (1K01HL135467-01), which helped facilitate the development of this manuscript. SL is supported by the Prostate Cancer Foundation and the Edward Blank and Sharon Cosloy-Blank Family Foundation. SL reports reimbursed travel to the Prostate Cancer Foundation Retreat from Sanofi; consulting fees from Lilly, Bayer and Lumenis; and equity in Gilead.

Conflicts of Interest

None declared.

References

1. Street R, Makoul G, Arora N, Epstein R. 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]

2. Okunrintemi V, Spatz ES, Di Capua P, Salami JA, Valero-Elizondo J, Warraich H, et al. Patient-Provider communication and health outcomes among individuals with atherosclerotic cardiovascular disease in the United States: Medical Expenditure Panel Survey 2010 to 2013. Circ Cardiovasc Qual Outcomes 2017 Dec;10(4):1. [doi: 10.1161/CIRCOUTCOMES.117.003635] [Medline: 28373270]
3. Nicoloro-SantaBarbara J, Rosenthal L, Auerbach MV, Kocis C, Busso C, Lobel M. Patient-provider communication, maternal anxiety, and self-care in pregnancy. Soc Sci Med 2017 Dec;190:133-140. [doi: 10.1016/j.socscimed.2017.08.011] [Medline: 28863336]
4. 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]
5. Tanis M, Hartmann T, Te PF. Online health anxiety and consultation satisfaction: a quantitative exploratory study on their relations. Patient Educ Couns 2016 Dec;99(7):1227-1232. [doi: 10.1016/j.pec.2016.01.021] [Medline: 26873545]
6. Senft N, Everson J. eHealth engagement as a response to negative healthcare experiences: cross-sectional survey analysis. J Med Internet Res 2018 Dec 05;20(12):e11034 [FREE Full text] [doi: 10.2196/11034] [Medline: 30518513]
7. Pew Research Center. 2018 Feb 05. Social Media Fact Sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-05-07] [WebCite Cache ID 78C83LnOY]
8. Perrin A, Anderson M. Pew Research Center. 2019 Apr 10. Share of US adults using social media, including Facebook, is mostly unchanged since 2018 URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-05-07] [WebCite Cache ID 78C8HLwUK]
9. Chou WS, Oh A, Klein WM. Addressing health-related misinformation on social media. JAMA 2018 Nov 14:e1. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
10. Merchant RM, Asch DA. Protecting the value of medical science in the age of social media and fake news. JAMA 2018 Nov 19:e1. [doi: 10.1001/jama.2018.18416] [Medline: 30453315]
11. Loeb S, Sengupta S, Butaney M, Macaluso JN, Czarniecki SW, Robbins R, et al. Dissemination of misinformative and biased information about prostate cancer on YouTube. Eur Urol 2019 Dec;75(4):564-567. [doi: 10.1016/j.eururo.2018.10.056] [Medline: 30502104]
12. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
13. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
14. Patel R, Chang T, Greysen SR, Chopra V. Social media use in chronic disease: a systematic review and novel taxonomy. Am J Med 2015 Dec;128(12):1335-1350. [doi: 10.1016/j.amjmed.2015.06.015] [Medline: 26159633]
15. Gage-Bouchard EA, LaValley S, Mollica M, Beaupin LK. Cancer communication on social media: examining how cancer caregivers use Facebook for cancer-related communication. Cancer Nurs 2017;40(4):332-338. [doi: 10.1097/NCC.0000000000000418] [Medline: 27442210]
16. Benetoli A, Chen TF, Aslani P. Consumer health-related activities on social media: exploratory study. J Med Internet Res 2017 Dec 13;19(10):e352 [FREE Full text] [doi: 10.2196/jmir.7656] [Medline: 29030326]
17. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
18. Gonzalez-Estrada A, Cuervo-Pardo L, Ghosh B, Smith M, Pazheri F, Zell K, et al. Popular on YouTube: a critical appraisal of the educational quality of information regarding asthma. Allergy Asthma Proc 2015;36(6):e121-e126. [doi: 10.2500/aap.2015.36.3890] [Medline: 26534743]
19. Huang J, Kornfield R, Emery SL. 100 Million views of electronic cigarette YouTube videos and counting: quantification, content evaluation, and engagement levels of videos. J Med Internet Res 2016 Mar 18;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
20. Ruppert L, Køster B, Siegert AM, Cop C, Boyers L, Karimkhani C, et al. YouTube as a source of health information: analysis of sun protection and skin cancer prevention related issues. Dermatol Online J 2017 Jan 15;23(1):1 [FREE Full text] [Medline: 28329465]
21. Harrison D, Wilding J, Bowman A, Fuller A, Nicholls SG, Pound CM, et al. Using YouTube to disseminate effective vaccination pain treatment for babies. PLoS One 2016;11(10):e0164123 [FREE Full text] [doi: 10.1371/journal.pone.0164123] [Medline: 27695054]
22. Kumar N, Pandey A, Venkatraman A, Garg N. Are video sharing web sites a useful source of information on hypertension? J Am Soc Hypertens 2014 Jul;8(7):481-490. [doi: 10.1016/j.jash.2014.05.001] [Medline: 25064770]
23. Romer D, Jamieson PE, Jamieson KH, Jones C, Sherr S. Counteracting the influence of peer smoking on YouTube. J Health Commun 2017 Dec;22(4):337-345 [FREE Full text] [doi: 10.1080/10810730.2017.1290164] [Medline: 28306481]
24. Kelly-Hedrick M, Grunberg PH, Brochu F, Zelkowitz P. "It's totally okay to be sad, but never lose hope": Content analysis of infertility-related videos on YouTube in relation to viewer preferences. J Med Internet Res 2018 Dec 23;20(5):e10199 [FREE Full text] [doi: 10.2196/10199] [Medline: 29792296]

25. Konstantinidis S, Fernandez-Luque L, Bamidis P, Karlsen R. The role of taxonomies in social media and the semantic web for health education. A study of SNOMED CT terms in YouTube health video tags. Methods Inf Med 2013;52(2):168-179. [doi: 10.3414/ME12-02-0005] [Medline: 23450378]
26. Wong K, Doong J, Trang T, Joo S, Chien AL. YouTube videos on botulinum toxin A for wrinkles: a useful resource for patient education. Dermatol Surg 2017 Dec;43(12):1466-1473. [doi: 10.1097/DSS.0000000000001242] [Medline: 28877151]
27. Lim K, Kilpatrick C, Storr J, Seale H. Exploring the use of entertainment-education YouTube videos focused on infection prevention and control. Am J Infect Control 2018 Nov;46(11):1218-1223. [doi: 10.1016/j.ajic.2018.05.002] [Medline: 29884576]
28. Garg N, Venkatraman A, Pandey A, Kumar N. YouTube as a source of information on dialysis: a content analysis. Nephrology (Carlton) 2015 May;20(5):315-320. [doi: 10.1111/nep.12397] [Medline: 25641264]
29. 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]
30. Staunton PF, Baker JF, Green J, Devitt A. Online curves: a quality analysis of scoliosis videos on YouTube. Spine (Phila Pa 1976) 2015 Dec;40(23):1857-1861. [doi: 10.1097/BRS.0000000000001137] [Medline: 26571065]
31. Rittberg R, Dissanayake T, Katz SJ. A qualitative analysis of methotrexate self-injection education videos on YouTube. Clin Rheumatol 2016 May;35(5):1329-1333. [doi: 10.1007/s10067-015-2910-5] [Medline: 25739847]
32. Albarracin D, Romer D, Jones C, Hall Jamieson K, Jamieson P. Misleading claims about tobacco products in YouTube videos: experimental effects of misinformation on unhealthy attitudes. J Med Internet Res 2018 Jun 29;20(6):e229 [FREE Full text] [doi: 10.2196/jmir.9959] [Medline: 29959113]
33. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007 Dec 5;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
34. Qi J, Trang T, Doong J, Kang S, Chien AL. Misinformation is prevalent in psoriasis-related YouTube videos. Dermatol Online J 2016 Nov 15;22(11):13030. [Medline: 28329562]
35. 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 Jan 04;16(5):572-579. [doi: 10.1513/AnnalsATS.201809-644OC] [Medline: 30608877]
36. Centers for Disease Control and Prevention. 2019 Mar 08. What is health communication? URL: https://www.cdc.gov/healthcommunication/healthbasics/WhatIsHC.html [accessed 2019-05-07] [WebCite Cache ID 78C8d3ofs]
37. Centers for Disease Control and Prevention. 2015 Mar 18. Health communication strategies URL: https://npin.cdc.gov/pages/health-communication-strategies [accessed 2019-05-07] [WebCite Cache ID 78C8p52d4]
38. National Cancer Institute. 2004 Aug 01. Making Health Communication Programs Work: A Planner's Guide, Pink Book URL: https://pubs.cancer.gov/ncipl/detail.aspx?prodid=T068 [accessed 2019-05-07] [WebCite Cache ID 78C91Y6vn]
39. Nelson D, Hesse B, Croyle R. Making Data Talk: Communicating Public Health Data To The Public, Policy Makers, And The Press. New York: Oxford University Press; May 20, 2009.
40. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4 [FREE Full text] [Medline: 10345255]
41. Thompson TL, Miller K, Parrott R, Dorsey A. The Routledge Handbook of Health Communication. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2003.
42. National Library of Medicine: Consumer Health Informatics Research Resource. Technology Acceptance Model (perceived usefulness and perceived ease of use) URL: https://chirr.nlm.nih.gov/tam.php [accessed 2019-05-07] [WebCite Cache ID 78CBEvLLl]
43. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
44. Davis F, Bagozzi R, Warshaw P. Extrinsic and intrinsic motivation to use computers in the workplace. J Appl Social Pyschol 1992 Jul;22(14):1111-1132. [doi: 10.1111/j.1559-1816.1992.tb00945.x]
45. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]
46. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
47. Portz J, Bayliss E, Bull S, Boxer R, Bekelman D, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604. [doi: 10.2196/11604]
48. Glanz K, Rimer B, Viswanath K. Health Behavior: Theory, Research, And Practice. San Francisco, CA: Jossey-Bass; 2015.
49. Petty R, Cacioppo J. The Elaboration Likelihood Model of Persuasion. In: Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York: Springer; 1986:1.
50. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-460; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]

51. Finney RL, Davis T, Beckjord EB, Blake K, Moser RP, Hesse BW. Picking up the pace: changes in method and frame for the health information national trends survey (2011-2014). J Health Commun 2012;17(8):979-989 [FREE Full text] [doi: 10.1080/10810730.2012.700998] [Medline: 23020763]
52. Westat. Health Information National Trends Survey 5 (HINTS 5) Cycle 2 Methodology Report. Bethesda, MD: National Cancer Institute; 2018 Jul. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle_2_Methodology_Report.pdf [accessed 2019-05-07] [WebCite Cache ID 78CHzFkMq]
53. Epstein RM, Street RL. National Cancer Institute Division of Cancer Control & Population Sciences. 2007. Patient-centered communication in cancer care: promoting healing and reducing suffering URL: https://healthcaredelivery.cancer.gov/pcc/pccm_flyer.pdf [accessed 2019-05-07] [WebCite Cache ID 78CIBCKRo]
54. Blanch-Hartigan D, Chawla N, Moser RP, Finney RL, Hesse BW, Arora NK. Trends in cancer survivors' experience of patient-centered communication: results from the Health Information National Trends Survey (HINTS). J Cancer Surviv 2016 Dec;10(6):1067-1077. [doi: 10.1007/s11764-016-0550-7] [Medline: 27193357]
55. Spooner KK, Salemi JL, Salihu HM, Zoorob RJ. Disparities in perceived patient-provider communication quality in the United States: trends and correlates. Patient Educ Couns 2016 May;99(5):844-854. [doi: 10.1016/j.pec.2015.12.007] [Medline: 26725930]
56. Reeve BB, Thissen DM, Bann CM, Mack N, Treiman K, Sanoff HK, et al. Psychometric evaluation and design of patient-centered communication measures for cancer care settings. Patient Educ Couns 2017 Jul;100(7):1322-1328 [FREE Full text] [doi: 10.1016/j.pec.2017.02.011] [Medline: 28214167]
57. Marks R, Ok H, Joung H, Allegrante JP. Perceptions about collaborative decisions: perceived provider effectiveness among 2003 and 2007 Health Information National Trends Survey (HINTS) respondents. J Health Commun 2010;15 Suppl 3:135-146. [doi: 10.1080/10810730.2010.522701] [Medline: 21154089]
58. Halpern MT, Urato MP, Lines LM, Cohen JB, Arora NK, Kent EE. Healthcare experience among older cancer survivors: analysis of the SEER-CAHPS dataset. J Geriatr Oncol 2018 Dec;9(3):194-203. [doi: 10.1016/j.jgo.2017.11.005] [Medline: 29249645]
59. StataCorp. 2015. How do I cite the Stata software, documentation, and FAQs? URL: https://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/ [accessed 2019-05-10] [WebCite Cache ID 78GevVZr7]
60. Li N, Orrange S, Kravitz RL, Bell RA. Reasons for and predictors of patients' online health information seeking following a medical appointment. Fam Pract 2014 Oct;31(5):550-556. [doi: 10.1093/fampra/cmu034] [Medline: 24963151]
61. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
62. 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]
63. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
64. Singh S, Evans N, Williams M, Sezginis N, Baryeh NA. Influences of socio-demographic factors and health utilization factors on patient-centered provider communication. Health Commun 2018 Dec;33(7):917-923. [doi: 10.1080/10410236.2017.1322481] [Medline: 28541816]
65. Bell RA, Hu X, Orrange SE, Kravitz RL. Lingering questions and doubts: online information-seeking of support forum members following their medical visits. Patient Educ Couns 2011 Dec;85(3):525-528. [doi: 10.1016/j.pec.2011.01.015] [Medline: 21315538]
66. Basch CH, Zybert P, Reeves R, Basch CE. What do popular YouTube videos say about vaccines? Child Care Health Dev 2017 Dec;43(4):499-503. [doi: 10.1111/cch.12442] [Medline: 28105642]
67. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
68. Langford AT, Solid CA, Gann LC, Rabinowitz EP, Williams SK, Seixas AA. Beliefs about the causes of hypertension and associations with pro-health behaviors. Health Psychol 2018 Dec;37(12):1092-1101. [doi: 10.1037/hea0000687] [Medline: 30307273]
69. Scherer LD, Caverly TJ, Burke J, Zikmund-Fisher BJ, Kullgren JT, Steinley D, et al. Development of the Medical Maximizer-Minimizer Scale. Health Psychol 2016 Nov;35(11):1276-1287. [doi: 10.1037/hea0000417] [Medline: 27617512]
70. Hale BJ, Gonzales AL, Richardson M. Vlogging cancer: predictors of social support in YouTube cancer vlogs. Cyberpsychol Behav Soc Netw 2018 Sep;21(9):575-581. [doi: 10.1089/cyber.2018.0176] [Medline: 30132690]
71. Levine DM, Landon BE, Linder JA. Trends in patient-perceived shared decision making among adults in the United States, 2002-2014. Ann Fam Med 2017 Nov;15(6):552-556 [FREE Full text] [doi: 10.1370/afm.2132] [Medline: 29133495]
72. Or CK, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]

73. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and acceptance of mHealth in patients with cardiovascular diseases: a cross-sectional study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]
74. Xesfingi S, Vozikis A. eHealth Literacy: in the quest of the contributing factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
75. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
76. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
77. Diefenbach M, Leventhal H. The Common-Sense Model of Illness Representation: theoretical and practical considerations. J Soc Distress Homel 1996 Jan;5(1):11-38. [doi: 10.1007/BF02090456]
78. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res 2006 Jun;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]
79. Phillips LA, Leventhal H, Leventhal EA. Physicians' communication of the common-sense self-regulation model results in greater reported adherence than physicians' use of interpersonal skills. Br J Health Psychol 2012 May;17(2):244-257 [FREE Full text] [doi: 10.1111/j.2044-8287.2011.02035.x] [Medline: 22107169]
80. Klayman J, Ha Y. Confirmation, disconfirmation, and information in hypothesis testing. Psychol Rev 1987;94(2):211-228. [doi: 10.1037/0033-295X.94.2.211]
81. Klayman J. Varieties of confirmation bias. In: Decision Making From A Cognitive Perspective, Volume 32: Advances In Research And Theory (psychology Of Learning & Motivation). Cambridge, MA: Academic Press; 1995:385-418.

Abbreviations

HINTS: Health Information National Trends Survey
OR: odds ratio

Edited by G Eysenbach; submitted 27.01.19; peer-reviewed by J Hazzam, C Weijs, S Singh, P Zelkowitz; comments to author 02.03.19; revised version received 22.04.19; accepted 26.04.19; published 17.05.19.

Please cite as:
Langford A, Loeb S
Perceived Patient-Provider Communication Quality and Sociodemographic Factors Associated With Watching Health-Related Videos on YouTube: A Cross-Sectional Analysis
J Med Internet Res 2019;21(5):e13512
URL: http://www.jmir.org/2019/5/e13512/
doi:10.2196/13512
PMID:31102372

Original Paper

Internet Searches for Medical Symptoms Before Seeking Information on 12-Step Addiction Treatment Programs: A Web-Search Log Analysis

George Nitzburg[1*], PhD; Ingmar Weber[2*], PhD; Elad Yom-Tov[3,4,5*], PhD

[1]Teachers College, Columbia University, New York, NY, United States
[2]Social Computing Department, Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[3]Microsoft Research, Redmond, WA, United States
[4]Microsoft Research, Herzeliya, Israel
[5]Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa, Israel
[*]all authors contributed equally

Corresponding Author:
George Nitzburg, PhD
Teachers College
Columbia University
525 West 120th Street
New York, NY, 10027
United States
Phone: 1 212 678 3267
Email: nitzburg@tc.columbia.edu

Abstract

Background: Brief intervention is a critical method for identifying patients with problematic substance use in primary care settings and for motivating them to consider treatment options. However, despite considerable evidence of delay discounting in patients with substance use disorders, most brief advice by physicians focuses on the long-term negative medical consequences, which may not be the best way to motivate patients to seek treatment information.

Objective: Identification of the specific symptoms that most motivate individuals to seek treatment information may offer insights for further improving brief interventions. To this end, we used anonymized internet search engine data to investigate which medical conditions and symptoms preceded searches for 12-step meeting locators and general 12-step information.

Methods: We extracted all queries made by people in the United States on the Bing search engine from November 2016 to July 2017. These queries were filtered for those who mentioned seeking Alcoholics Anonymous (AA) or Narcotics Anonymous (NA); in addition, queries that contained a medical symptom or condition or a synonym thereof were analyzed. We identified medical symptoms and conditions that predicted searches for seeking treatment at different time lags. Specifically, symptom queries were first determined to be significantly predictive of subsequent 12-step queries if the probability of querying a medical symptom by those who later sought information about the 12-step program exceeded the probability of that same query being made by a comparison group of all other Bing users in the United States. Second, we examined symptom queries preceding queries on the 12-step program at time lags of 0-7 days, 7-14 days, and 14-30 days, where the probability of asking about a medical symptom was greater in the 30-day time window preceding 12-step program information-seeking as compared to all previous times that the symptom was queried.

Results: In our sample of 11,784 persons, we found 10 medical symptoms that predicted AA information seeking and 9 symptoms that predicted NA information seeking. Of these symptoms, a substantial number could be categorized as nonsevere in nature. Moreover, when medical symptom persistence was examined across a 1-month time period, a substantial number of nonsevere, yet persistent, symptoms were identified.

Conclusions: Our results suggest that many common or nonsevere medical symptoms and conditions motivate subsequent interest in AA and NA programs. In addition to highlighting severe long-term consequences, brief interventions could be restructured to highlight how increasing substance misuse can worsen discomfort from common medical symptoms in the short term, as well as how these worsening symptoms could exacerbate social embarrassment or decrease physical attractiveness.

(J Med Internet Res 2019;21(5):e10946) doi:10.2196/10946

KEYWORDS

alcohol use disorder; substance use disorder; 12-step programs; brief intervention; brief physician advice; anonymized internet search log data

Introduction

Background

Alcohol use disorder (AUD) exerts a heavy burden on psychiatric and medical facilities, which has led to early identification and intervention efforts with problem drinkers who are engaging in regular risky drinking patterns but do not yet have physical dependence on alcohol or an immediate medical need for abstinence. For many problem drinkers, the first point of contact often occurs at routine office visits to primary care physicians, where roughly 10%-20% of primary care patients are found to be drinking alcohol at hazardous levels [1-3]. Risky drinking patterns (such as binge or regular heavy drinking) have been linked with numerous adverse medical consequences [4-9] including numerous cancers [5], heart disease and stroke [10-13], pneumonia [14], tuberculosis [15], epilepsy [16], diabetes [17], pancreatitis [18-21], and liver disease [22]. Although the abovementioned conditions are among the more severe long-term medical consequences of alcohol, heavy alcohol consumption has been implicated in the onset or exacerbation of over 100 diseases and conditions [23]. Since primary care office visits can often be prompted by the need to address medical symptoms partially or can fully arise from heavy drinking patterns, such visits offer an opportunity to motivate problem drinkers toward alcohol-reduction treatment options.

To help reduce drinking among their patients, from hazardous levels down to low-risk levels, physicians have made efforts to integrate minimal treatment protocols into their routine care (via face-to-face or electronic delivery), known as brief intervention (BI) [24,25]. The largest venue for BI, which follows after The Substance Abuse and Mental Health Services Administration’s largest-of-its-kind federally funded Screening, Brief Intervention, and Referral to Treatment (SBIRT) initiative, is primary care clinics and emergency rooms. SBIRT has been a vital source of early identification of alcohol and drug use disorders, which has so far screened and offered BI or treatment referral in 29 US states [26]. SBIRT programs have demonstrated effectiveness with drug and alcohol reductions at 6 months postintervention [27], along with clinically and statistically significant improvements in physical and mental health as well as social, legal, housing, and employment outcomes [28]. In addition to expanding the scale of the existing SBIRT program, identifying ways to further maximize BI’s effectiveness is also of critical importance to SBIRT.

One largely unexplored avenue for enhancing BI’s personalized feedback for more severe cases is to evaluate which negative medical consequences most contribute to information seeking for abstinence-based treatment. Twelve-step programs remain the most utilized form of substance use disorder (SUD) treatment [29,30]. Specifically, BI’s health risk information could be improved by understanding which of the >100 alcohol-implicated diseases and conditions are most likely to motivate severe cases to consider 12-step treatment options. Many potential improvements may be possible regarding this health risk information.

In particular, many of the medical consequences emphasized in BI occur over the long term (eg, cancer). However, meta-analytic evidence has shown that individuals who engage in addictive behaviors make more impulsive decisions, have greater difficulty delaying gratification for greater rewards, and tend to discount delayed rewards or other reinforcers more quickly than healthy controls [31,32]. Thus, the actual and perceived speed of reward or reinforcer delivery is particularly critical to motivation in patients with AUD, and as a result, they may be more motivated by less chronic but more immediately discomforting and socially embarrassing symptoms such as gastrointestinal distress. Thus, BI’s health risk information could potentially be improved by focusing on an increased likelihood of experiencing specific sets of immediately discomforting and socially embarrassing symptoms rather than a risk of diagnosis with various medical diseases, some of which occur in the long term (eg, cancer). Reorganizing BI’s health-risk information around the medical symptoms that can be empirically shown to motivate treatment seeking holds the potential to increase BI’s efficacy for the more severe cases often seen in primary care settings.

Prior Work

Analysis of large-scale anonymized Web-search log data offers a promising new method for identifying medical symptoms that are most predictive of subsequent seeking of 12-step treatments. Recent analyses of subtle and temporal changes in medical symptom Web-search queries have been able to predict later medical events, including influenza [33], pancreatic and breast cancers [33,34], and adverse side effects of medications [35]. Such analyses can be readily adapted from the prediction of dichotomous health events (eg, search log evidence of having cancer versus not having cancer) to the prediction of dichotomous behavioral health care choices (eg, search log evidence of interest in 12-step treatment versus no such treatment). Although many researchers investigating online patterns in psychiatric disorders and SUD have opted to examine social media activity rather than search engine queries [36-48], Web-search log data analysis is particularly well suited to the study of alcohol-related medical symptoms. Past work has shown that medical health information seeking most often takes place using search engines, especially when medical symptoms may be perceived as potentially connected to stigmatized conditions (eg, symptoms perceived to be the possible negative medical consequences of addiction) [49]. It is most likely due to the associated stigma that regional variation in the interest in the topics of alcohol and Alcoholics Anonymous (AA) on Facebook show inconsistent patterns with respect to regional variation in the rates of alcohol abuse [50]. Due to its anonymous nature, aggregate and anonymous Web-search volume from Google Trends has also been used to study the link between macroeconomic conditions and problem drinking and has shown that a 5% rise in unemployment is followed by an approximate 15% increase in alcoholism-related searches in the next 12 months [51]. In addition, using data from Google Trends, other researchers have assessed the relationship between different legal statuses of marijuana and the level of search interests for “dabbing,” the vaping of high-potency marijuana concentrates [52].

At the individual level, anonymized Web-search log data offer an ecologically valid and representative sampling of individuals who are likely to show up at primary care clinics and disclose their alcohol-related medical symptoms to a physician when prompted. A recent systematic review found that individuals experiencing acute or conspicuous medical symptoms searched for online health information prior to seeking real-world medical care. These individuals were also subsequently more persuaded to see a physician than those who did not do a Web search, endorsing less embarrassment and concern about bothering their physician with a trivial complaint [53]. Search log data analysis is also well suited to reorganizing BI’s brief advice by a physician, regarding the medical consequences of alcohol misuse, as the same type of analysis has already helped inform how online cancer information may be reorganized to best address the distinct needs of patients and their caregivers when patients are at different levels of disease progression [54]. Taken together, past work suggests that Web-search log data analysis is the best-suited technique to provide unique insights into which negative medical consequences of alcohol misuse are most persuasive for motivating interest in the 12-step treatment.

Goal of the Study

The present study used anonymous internet search log data to better understand how seeking medical symptom information contributed to seeking AA and narcotics anonymous (NA) treatment information, including searches for treatment locators. We examined anonymous Microsoft Bing search data to identify individuals that graduated from using Bing search to investigate their medical complaints (which may be potentially alcohol related) to using Bing search to seeking AA or NA information or finding AA or NA meetings. We then examined which of these medical symptom searches most increased the likelihood of conducting either of these two types of Bing searches. Such data can help inform BI-personalized feedback to enhance acceptance of SBIRT’s treatment referrals for patients spanning the broader range of alcohol and drug use severity levels often seen in primary care physicians’ offices.

Of note, we chose to include NA-related queries in our primarily AUD-focused study since alcohol and drug use disorders are highly comorbid, AA meetings are far more ubiquitous than NA meetings, people with drug use disorders are not typically turned away from AA meetings after revealing their drug of choice is not alcohol, these two types of 12-step programs share many characteristics, and both meeting types can often have mixed attendance [55]. Given such comorbidity and cross-attendance, our online search data analysis was unable to reasonably estimate users’ diagnostic groups, and we instead opted to separately analyze all the various combinations and permutations of AA- and NA-related queries to gauge the convergence/divergence of symptom profiles that preceded them.

Hypotheses

We proposed the following hypotheses:

1. The medical symptom queries motivating subsequent seeking of information on 12-step programs would be nonsevere but immediately discomforting and socially embarrassing.
2. Symptoms or conditions considered severe or medically dangerous would not require persistence over time to induce queries for the 12-step programs, with individuals’ taking near-immediate action to seek out information on the 12-step programs soon after experiencing these symptoms.
3. Symptoms or conditions considered common or nonsevere would require persistence over time to induce queries about the 12-step programs, with individuals only taking action to seek out information on the 12-step programs after a period of enduring discomfort from these symptoms.

Methods

Data

We extracted all queries made by people in the United Stated on the Bing search engine from November 2016 to July 2017 (inclusive). Each query contained an anonymous user-identifier tracked via a browser cookie, the time and date of the query, the text of the query, and the Web address of pages clicked by users in response to the answers shown to them. Although Bing’s US market share is currently estimated at ~24% [56], past work has found it to be representative of the US population [57,58] and has shown that Bing is a viable data source even in the absence of other search engines. This study was approved by the Institutional Review Board of the Technion, Israel Institute of Technology.

Analysis

Identifying Queries of People Seeking Information on Alcoholics Anonymous and Narcotics Anonymous

We then found all queries made by 50 or more people, which resulted in a click to one or more of the 193 pages. There were a total of 100 such queries (eg, “alcoholic anonymous meeting directory”).

Identifying the Target Population

Based on the queries identified, we defined two target groups (ie, two 12-step information-seeking samples). The first larger overall sample comprised all 11,784 people who made an AA- or NA-seeking query during July 2017 and who did not make such queries in any of the preceding months (November 2016 to June 2017). The second sample, a subset of the first, included only those people who specifically queried for AA or NA meeting locators by seeking queries that also contained the term “meeting.” A total of 3820 people were included in this subsample.

Our overall sample consisted of a total of 10,522 people who sought AA information, 1022 people who sought NA information, and 240 people who sought both AA and NA information. Our subsample (whose first 12-step–related search was for AA or NA meeting information) consisted of 2324 people who sought information specifically on AA meetings, 501 people who sought information specifically on NA meetings, and 115 people who sought information on both types of meetings.

Of note, we did not require a medical symptom query to have been conducted in order for the query to be included in our comparison group. In other words, our overall sample also included people who searched for information on the alcohol or drug 12-step program without ever having queried a medical complaint.

Identifying Symptoms Correlated With Seeking Treatment Information

Queries were deemed to contain medical symptoms if their text included a symptom or a synonym thereof, as provided in the list of 195 symptoms and their synonyms in the study by Yom-Tov and Gabrilovich [35]. Colloquial symptoms were mapped to their medical symptom wording, which is why medical terminology descriptors appear in our data (eg, a search query for “sweating” would be mapped onto the more medical descriptor “diaphoresis”). This list included medical symptoms that were indicative of both medical and psychological distress (eg, fever and cough as well as anxiety and depression). The list also included both symptoms that would be immediately discomforting as well as distant or long-term diagnoses (eg, gastrointestinal distress versus cancer).

We compared the symptoms prior to seeking information on the 12-step programs in two ways: First, we compared the entire population of Bing users in the United States. Second, we performed a within-group comparison, taking into account the temporal sequence of searches. When examining these temporal sequences, we categorized “symptoms with persistence in time” as symptoms that were repeatedly queried across any two of our three measurement points (measurement points at 0-7, 7-14, and 14-30 days prior to a 12-step–related search); symptoms were defined as persistent when they were queried during both the 0-7 day and 14-30 day time lags, during both the 7-14 day and 14-30 day time lags, or during both the 0-7 day and 14-30 day time lags. A full schematic of procedures for data collection and filtering as well as the between-group comparison and the time-lagged within-group analysis of medical symptom queries for identified users are provided in Figure 1.

For the first analysis, we calculated the probability that users in the treatment population will ask about each symptom in their queries, divided by the same probability for all Bing users in the United States. For the second analysis, we defined a window of time prior to the first query seeking information on the 12-step programs for each person. We then constructed a 2 × 2 table for each symptom (Table 1) and calculated the Chi-square score of this table. We only retained symptoms where the probability of asking about a symptom was greater in the time window than all previous times and where the Chi-square score was statistically significant at *P*<.05 with Bonferroni correction (*P*=.05/[152*3]=.00011). This approach identified cases where there was a statistically significant increase in the queries for a symptom in the days prior to the first query seeking information about treatment.

Figure 1. Procedure for data collection, filtering, and analysis with a between-group comparison and time-lagged within-group comparison of medical symptom queries. NA: narcotics anonymous; AA: alcoholics anonymous.

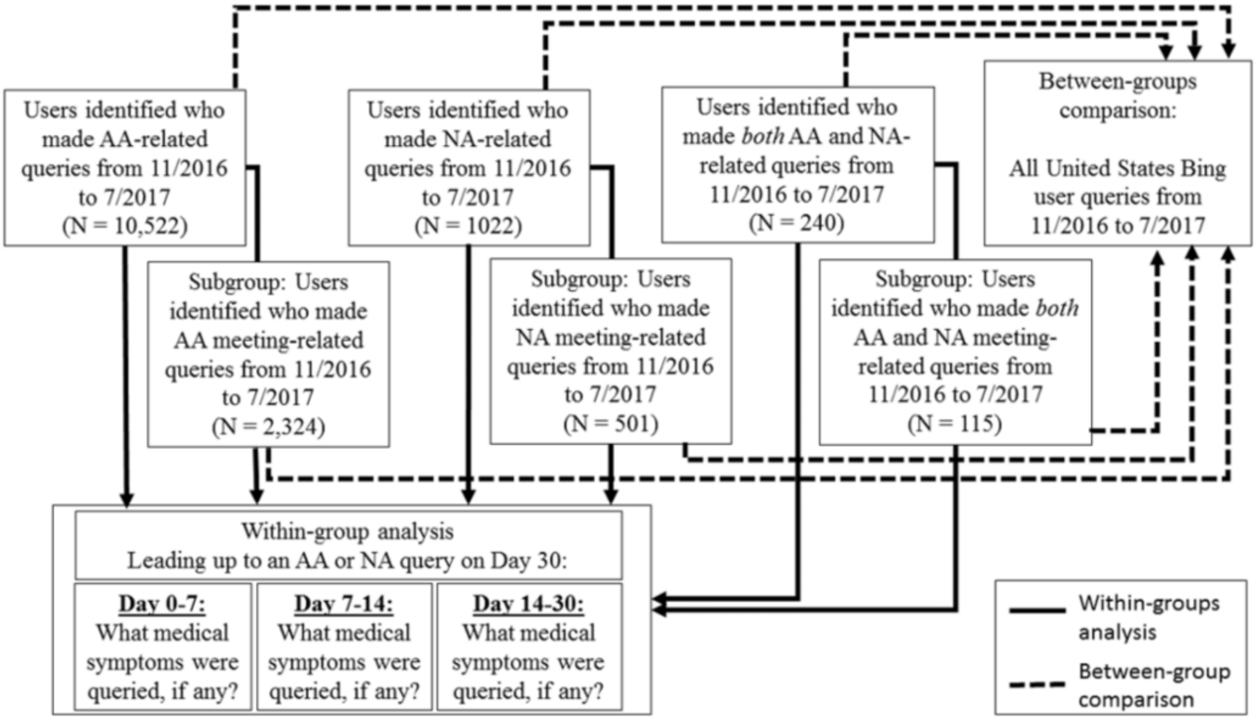


Table 1. A 2 × 2 table for each symptom to evaluate the change in queries for the symptom in each population.

Queries made in T days before seeking information on the 12-step programs	Number of people who queried about the symptom	Number of people who did not query about the symptom
Yes	A	C
No	B	D

Results

When comparing the data of all the people who queried a medical symptom in the United States, our analysis of 195 medical symptoms revealed that 10 medical symptoms predicted AA information seeking and 9 symptoms predicted NA information seeking. Results also identified 10 medical symptoms that predicted subsequent querying for an AA meeting and one symptom (bloating) that predicted subsequent querying for an NA meeting. Of these symptoms, a substantial number could be categorized as nonsevere in nature (Textbox 1). Moreover, when medical symptom persistence was examined across a 1-month time period, a substantial number of nonsevere, yet persistent, symptoms were identified (Table 2).

We first hypothesized that given the evidence of heightened levels of delayed discounting in SUD populations compared to controls [31,32], more immediately discomforting and socially embarrassing medical symptoms and conditions would be most prevalent for those who subsequently search for information on the 12-step treatment program. This first hypothesis was only partially confirmed. As hypothesized, compared to all other Bing search users in the United States, the probability of a user querying about AA or NA, in general, or the meeting information was significantly increased by earlier queries for nonsevere symptoms and conditions such as bloating, sweating, hives, heartburn, dizziness, bulging eyes, impotence, and back ache. However, several severe symptoms significantly increased the likelihood of subsequent AA or NA queries, although it was unclear if they were primarily medical, psychiatric, or related to substance withdrawal.

We also hypothesized that the symptom queries that were significantly linked to immediate/near-immediate seeking of 12-step information would be severe or medically dangerous; severe symptoms would not require persistence over time to produce treatment motivation. Surprisingly, this second hypothesis was not confirmed. No medical symptom queries were significantly linked to immediate seeking of NA information or concurrent NA and AA information (eg, co-occurring drug and alcohol addiction treatment). In addition, the symptom queries that were significantly linked to the immediate seeking of AA information primarily consisted of common and nonsevere medical conditions such as cramp, rash, toothache, and dry mouth (the exceptions were hallucination and phobia, which could also be psychiatric or related to withdrawal).

Last, we hypothesized that the persistent symptom queries that were significantly linked to subsequent seeking of information on the 12-step programs would be common or nonsevere in nature; a common or nonsevere symptom would need to be persistent to sufficiently compel a search for addiction treatment information. This third hypothesis was only partially confirmed, as several common and nonsevere symptoms prompted near-immediate search for AA (but not for NA or for NA and AA) information. However, as hypothesized, most persistent and near-persistent symptom queries were about common or nonsevere symptoms (ie, blister, itch, swelling, tinnitus, dizziness, heartburn, and diarrhea for AA and cough, cramp, constipation, fever, and back pain for NA).

Notably, some symptoms were vague regarding severity. Specifically, regarding blurred vision for AA, pain and weight loss for NA, and pain for NA and AA, the extent of pain and weight loss were not specified and blurred vision could potentially range from mild (yet persistent) to severely impairing vision loss. Queries about the vague symptom of “deformity” was also nearly persistent for AA information seekers, which may reflect concerns about appearance with increasing alcohol misuse but may also reflect concerns about pregnancy and alcohol-related birth defects.

In contrast, symptoms such as hallucination, akathisia, amnesia, blindness, and cataplexy (ie, the brief loss of voluntary muscle movement) can more clearly be categorized as severe, although in the absence of comprehensive psychiatric interviews, it was not possible to discern whether these symptoms reflected a primary medical condition/psychiatric disorder or were secondary to substance intoxication, substance withdrawal, or the adverse effects of psychiatric medications (see Limitations).

Textbox 1. Symptom queries preceding the first search for information on the 12-step programs. Medical symptoms and conditions ranked from the highest to lowest according to the ratio of the probability that the symptom will be queried by those who later search for information on the 12-step programs, compared to the rest of all United States Bing users. A Chi-square test found no significant differences between medical symptom queries that preceded searches for information on the 12-step programs versus meeting-specific searches for the 12-step programs. For symptoms appearing in both general and meeting-only categories, the differences in the number of searches is statistically significant.

General or meeting information

Queries for alcoholics anonymous information

- Exophthalmos (bulging eyes)
- Pyrosis (heartburn)
- Hallucination
- Impotence
- Back ache
- Urticaria (hives)
- Akathisia (inner restlessness, inability to stay still)
- Diaphoresis (heavy sweating)
- Agoraphobia
- Cataplexy (brief loss of voluntary muscle movement)

Queries for narcotics anonymous information:

- Pyrosis (heartburn)
- Hallucination
- Impotence
- Urticaria (hives)
- Agoraphobia
- Bloating
- Blindness
- Amnesia
- Dizziness

Meeting information only

Queries for alcoholics anonymous information:

- Exophthalmos (bulging eyes)
- Pyrosis (heartburn)
- Hallucination
- Impotence
- Back ache
- Akathisia (inner restlessness, inability to stay still)
- Urticaria (hives)
- Diaphoresis (heavy sweating)
- Agoraphobia
- Amnesia

Queries for narcotics anonymous information:

- Bloating

Table 2. Medical symptom search increasing in a statistically significant manner over the 30 days preceding seeking of information on the 12-step programs.

Symptom persistence	Category of the 12-step treatment information sought		
	Alcoholics anonymous	Narcotics anonymous	Both
Persistent (manifesting at all three time points)	- Anxiety[a,b,c] - Blister[a,b,c] - Blurred vision[a,b,c] - Itch[a,b,c] - Swelling[a,b,c] - Tinnitus (ears ringing)[a,b,c]	- Anorexia[a,b,c] - Cough[a,b,c] - Cramp[a,b,c] - Depression[a,b,c] - Pain[a,b,c] - Weight loss[a,b,c]	- Pain[a,b,c]
Near persistent (manifesting at two of the three time points)	- Dizzy[a,b] - Deformity[a,b] - Pyrosis (heartburn)[a,b] - Paresthesia (skin crawling)[a,b] - Diarrhea[b,c] - Malaise[b,c] - Phobia[b,c] - Tired[b,c] - Depression[a,c]	- Constipation[a,b] - Fever[a,b] - Back pain[b,c]	- Anxiety[b,c]
Immediately before 12-step information seeking (manifesting at the last time point only)	- Cramp[c] - Hallucination[c] - Phobia[c] - Rash[c] - Toothache[c] - Xerostomia (dry mouth)[c]	—[d]	—
Distal to 12-step information seeking (manifesting at the first time point only)	- Dysphagia (trouble swallowing)[a]	- Paranoia[a,b] - Perspiration[a] - Diarrhea[b]	- Paranoia[a]

[a] Symptom search occurred on days 0-7 in the 30-day lead up to a 12-step program information query on day 30.

[b] Symptom search occurred on days 7-14 in the 30-day lead up to a 12-step program information query on day 30.

[c] Symptom search occurred on days 14-30 in the 30-day lead up to a 12-step program information query on day 30.

[d] None.

Discussion

Principal Results

The present study used anonymous Web-search log data analysis to examine which medical symptom queries best motivated subsequent searches for the general and meeting information on the 12-step programs. Our results suggest that queries about more severe or dangerous medical consequences as well as more nonsevere or common symptoms were important in motivating individuals toward subsequently seeking information on the 12-step programs.

Our findings support past work demonstrating the high prevalence of co-occurring psychiatric disorders in addiction [60,61] and the current knowledge base supporting the efficacy of physician advice about severe long-term medical consequences in motivating individuals towards addiction treatment options [62].

These results underscore an important point: Our results do not support the notion that more severe, long-term medical consequences of drugs and alcohol should, in any way, be ignored during brief advice by a physician, but rather, that adding on to current BI practices with additional emphasis on common nonsevere symptoms may be additionally beneficial.

Potential Underlying Mechanisms

Our findings raise questions about the possible underlying mechanisms that might make common or nonsevere medical symptoms persuasive for people to seek information on 12-step programs. One possible underlying mechanism is that more common symptoms may better allow individuals to envision the negative medical consequences of AUDs and SUDs compared to more chronic or dangerous symptoms. In other words, since a symptom like heartburn is common, it may be easy to identify with the experience of the symptom, notice it worsening with increasing substance use, and envision it becoming even more persistent or chronic. In contrast, it may be difficult for an individual to envision obtaining a cancer diagnosis or its likelihood may be discounted as too distant for concern; moreover, for those who can envision a cancer diagnosis, such envisioning may be so frightening that it fosters denial rather than motivation. Emphasizing the potential for worsening frequency and discomfort of otherwise common symptoms (eg, heartburn, dizziness, hives, and back ache) in addition to severe medical consequences could help further motivate individuals to accept treatment after being newly identified in primary care settings as having an AUD or SUD.

A second possible mechanism underlying our findings may be related to the social embarrassment associated with certain medical symptoms. For example, symptoms such as bloating, bulging eyes, and impotence may impact physical attractiveness. Social conformity has been shown to be an important motive for alcohol and drug use [63], with alcohol, in particular, being considered a social lubricant [64]. Adding to BI's current practices, highlighting how the medical consequences of substance use can be socially embarrassing may also help motivate patients newly diagnosed with AUD or SUD to consider treatment options including 12-step programs. Although discussion of the social repercussions of drinking alcohol is already a component of the brief motivational interviewing that is sometimes incorporated into BI, our results suggest it may provide additional motivation to discuss how medical and nonmedical consequences can interact to intensify social embarrassment in patients with newly identified AUD or SUD.

In sum, our results suggest that many common or nonsevere medical symptoms and conditions motivate subsequent engagement with 12-step programs. As a result, BI could be restructured in multiple ways to maximize patients' motivation to engage in treatment, including accompanying current BI practices and highlighting the common medical symptoms that could immediately worsen with increasing substance misuse and how some of those worsening symptoms could exacerbate social embarrassment. Such findings may help inform SBIRT efforts to screen individuals who present with medical complaints for AUD, provide brief interventions that motivate individuals toward considering treatment options, and provide referrals to AUD treatments. Specifically, our findings suggest that it may be beneficial to modify SBIRT's BIs to add focus on the contribution of alcohol to immediately discomforting and socially embarrassing medical symptoms.

Limitations

A primary limitation of the present study was the anonymous nature of Web-search log data, which did not allow us to collect information on multiple demographic factors including age, sex, ethnicity, and socioeconomic status. Another primary limitation was our inability to follow-up on querying behavior to gauge whether search queries translated into real-world action. For example, although someone in the Prochaska and DiClemente [65] precontemplation stage of change would be very unlikely to run a search query for "AA information" (and even more unlikely to query for "AA meeting near me"), we cannot be certain whether search queries denote a contemplation, preparation, or action stage of change. In other words, we were unable to determine if searching for "AA meeting near me" was a query soon followed by AA meeting attendance or whether it was only gathering information for those contemplating or preparing to make a change but not yet ready to take action.

We also noted that in a minority of cases, people may be searching for others and multiple people may be using the same browser, thus conflating multiple people to the same user identifier. However, although this unlikely possibility cannot be ruled out entirely, it is estimated based on past work [66] that these searches are limited to only a small fraction of the data. For instance, although the keyword "alanon" was included, which typically reflects family members searching for loved ones, we found that only 1.4% of queries we studied contained this keyword.

Our study also lacked comprehensive psychiatric interviews, and thus, it was not possible to discern if certain symptoms were suggestive of co-occurring psychiatric conditions, the adverse side effects of psychiatric medications, substance-withdrawal symptoms, severe medical consequences of substance misuse, the use of multiple substances, physical comorbidities related to the searched-for symptoms, or the social consequences of excessive drinking. For instance, symptoms like hallucination and akathisia could reflect substance withdrawal (ie, withdrawal-related hallucinations and severe agitation) or a co-occurring psychotic disorder alongside the adverse effects of the antipsychotic medication used to treat it. Similarly, phobia queries may signify a co-occurring anxiety disorder or withdrawal-related anxiety increases. Further, queries about amnesia may reflect concerns about blackout drinking, withdrawal-related states of delirium, or the severe cognitive repercussions of long-term substance misuse (ie, memory loss).

Our anonymous online search data also grouped all people with NA queries together, and thus, our data did not allow us to differentiate between subgroups of narcotics being misused. For example, stimulants and sedatives have different and often opposing medical consequences. Thus, we did not have access to data that could have improved our precision in delineating the medical symptoms that preceded NA information queries.

Our data were exclusively derived from Bing searches and did not include data from other search engines such as Google or Yahoo. Future studies are needed to replicate our findings in other search engines as well as at a smaller scale where demographic characteristics can be more comprehensively collected (and statistically compared and contrasted) rather than estimated based on profoundly large usage data.

Our study scope was limited to an investigation of how delay discounting may impact online treatment information seeking, and indeed, there are many barriers to addiction treatment beyond delay discounting that our study did not investigate. This limited scope reflected our need to maintain study focus. Future studies are needed to expand on other potential barriers. Further, the conceptual scope of our study was also limited by the technical limitations of online search data analysis, which relies on an adequately large number of single keywords or short phrases that are well defined and unambiguous in their meaning and ability to be linked together. First, our treatments were limited to twelve step programs without other forms of evidence-based treatment such as counseling or medication. A prerequisite of studying online search data is that any chosen query sets must be reasonably likely to have sufficient public visibility to gather a large enough number of responses for analysis. As a result, we narrowed our scope to 12-step programs, the most highly utilized form of addiction treatment [67] and therefore the most likely to have the greatest number of search queries. Second, we limited our predictors to adverse medical symptoms without other kinds of addiction-related problems such as social or legal consequences. Online search data analysis is not yet able to adequately capture more nuanced metrics such as social and legal problems without a significant number of false positives. To address this limitation, we opted to solely focus on medical symptoms because of pre-existing mapping of Web-search queries to medical symptoms that in the past work, has successfully limited false positives and avoided off-base random exploration [68]. Although accommodating these technical limitations narrowed our study focus, elucidating how medical symptoms impact seeking information on the 12-step program can offer unique insights critical to SBIRT and other early intervention efforts in primary care settings. Nevertheless, future studies are needed to expand upon the scope of our line of research inquiry to include other forms of treatment and consequences of addiction.

Conclusions

In conclusion, our anonymous search log data analysis indicated that querying about the 12-step program was preceded by the occurrence of a number of nonsevere medical consequences of alcohol and drug misuse. The persistence of these nonsevere symptoms also appeared to play an important role in motivating individuals to query about the 12-step program. These findings can help inform modifications to current SBIRT protocols by concurrently emphasizing the long-term medical consequences of alcohol and drug misuse alongside shorter-term sets of symptoms that are more immediately discomforting and socially embarrassing. Such modified SBIRT protocols may apply BI to improve the motivation of the ~17% and 22.7% of individuals who screen positively during primary care encounters (using brief screening instrument scores) for risky/problematic alcohol and drug use, respectively [69,27] in order to accept referrals to 12-step programs or other outside addiction-treatment resources.

Conflicts of Interest

EYT is an employee of Microsoft Corp, the owner of Bing. No other conflicts of interest are declared.

Multimedia Appendix 1

Alcoholics Anonymous and Narcotics Anonymous webpages queried using the Bing search engine.

[PDF File (Adobe PDF File), 91KB - jmir_v21ie10946_app1.pdf]

References

1. Aalto M, Seppä K, Mattila P, Mustonen H, Ruuth K, Hyvärinen H, et al. Brief intervention for male heavy drinkers in routine general practice: a three-year randomized controlled study. Alcohol Alcohol 2001;36(3):224-230. [Medline: 11373259]
2. Nilsen P, Aalto M, Bendtsen P, Seppä K. Effectiveness of strategies to implement brief alcohol intervention in primary healthcare. A systematic review. Scand J Prim Health Care 2006 Mar;24(1):5-15. [doi: 10.1080/02813430500475282] [Medline: 16464809]
3. Reid MC, Fiellin DA, O'Connor PG. Hazardous and harmful alcohol consumption in primary care. Arch Intern Med 1999;159(15):1681-1689. [Medline: 10448769]
4. Chou SP, Grant BF, Dawson DA. Medical consequences of alcohol consumption--United States, 1992. Alcohol Clin Exp Res 1996 Nov;20(8):1423-1429. [Medline: 8947320]
5. Corrao G, Bagnardi V, Zambon A, La Vecchia C. A meta-analysis of alcohol consumption and the risk of 15 diseases. Prev Med 2004 May;38(5):613-619. [doi: 10.1016/j.ypmed.2003.11.027] [Medline: 15066364]
6. Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS Med 2009 Apr 28;6(4):e1000058 [FREE Full text] [doi: 10.1371/journal.pmed.1000058] [Medline: 19399161]
7. Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ, Comparative Risk Assessment Collaborating Group. Selected major risk factors and global and regional burden of disease. Lancet 2002 Nov 02;360(9343):1347-1360. [doi: 10.1016/S0140-6736(02)11403-6] [Medline: 12423980]
8. Rehm J, Baliunas D, Borges GLG, Graham K, Irving H, Kehoe T, et al. The relation between different dimensions of alcohol consumption and burden of disease: an overview. Addiction 2010 May;105(5):817-843 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.02899.x] [Medline: 20331573]
9. Shield KD, Parry C, Rehm J. Chronic diseases and conditions related to alcohol use. Alcohol Res 2013;35(2):155-173 [FREE Full text] [Medline: 24881324]
10. Patra J, Taylor B, Irving H, Roerecke M, Baliunas D, Mohapatra S, et al. Alcohol consumption and the risk of morbidity and mortality for different stroke types--a systematic review and meta-analysis. BMC Public Health 2010 May 18;10:258 [FREE Full text] [doi: 10.1186/1471-2458-10-258] [Medline: 20482788]

11. Puddey IB, Croft KD. Alcohol, stroke and coronary heart disease. Are there anti-oxidants and pro-oxidants in alcoholic beverages that might influence the development of atherosclerotic cardiovascular disease? Neuroepidemiology 1999;18(6):292-302. [doi: 10.1159/000026224] [Medline: 10545781]
12. Samokhvalov AV, Irving HM, Rehm J. Alcohol consumption as a risk factor for atrial fibrillation: a systematic review and meta-analysis. Eur J Cardiovasc Prev Rehabil 2010 Dec;17(6):706-712 [FREE Full text] [doi: 10.1097/HJR.0b013e32833a1947] [Medline: 21461366]
13. Taylor B, Irving HM, Baliunas D, Roerecke M, Patra J, Mohapatra S, et al. Alcohol and hypertension: gender differences in dose-response relationships determined through systematic review and meta-analysis. Addiction 2009 Dec;104(12):1981-1990. [doi: 10.1111/j.1360-0443.2009.02694.x] [Medline: 19804464]
14. Samokhvalov AV, Irving HM, Rehm J. Alcohol consumption as a risk factor for pneumonia: a systematic review and meta-analysis. Epidemiol Infect 2010 Dec;138(12):1789-1795. [doi: 10.1017/S0950268810000774] [Medline: 20380771]
15. Rehm J, Samokhvalov AV, Neuman MG, Room R, Parry C, Lönnroth K, et al. The association between alcohol use, alcohol use disorders and tuberculosis (TB). A systematic review. BMC Public Health 2009 Dec 05;9:450 [FREE Full text] [doi: 10.1186/1471-2458-9-450] [Medline: 19961618]
16. Samokhvalov AV, Irving H, Mohapatra S, Rehm J. Alcohol consumption, unprovoked seizures, and epilepsy: a systematic review and meta-analysis. Epilepsia 2010 Jul;51(7):1177-1184 [FREE Full text] [doi: 10.1111/j.1528-1167.2009.02426.x] [Medline: 20074233]
17. Baliunas DO, Taylor BJ, Irving H, Roerecke M, Patra J, Mohapatra S, et al. Alcohol as a risk factor for type 2 diabetes: A systematic review and meta-analysis. Diabetes Care 2009 Nov;32(11):2123-2132 [FREE Full text] [doi: 10.2337/dc09-0227] [Medline: 19875607]
18. Irving HM, Samokhvalov AV, Rehm J. Alcohol as a risk factor for pancreatitis. A systematic review and meta-analysis. JOP 2009 Jul 06;10(4):387-392 [FREE Full text] [Medline: 19581740]
19. Kristiansen L, Grønbaek M, Becker U, Tolstrup JS. Risk of pancreatitis according to alcohol drinking habits: a population-based cohort study. Am J Epidemiol 2008 Oct 15;168(8):932-937. [doi: 10.1093/aje/kwn222] [Medline: 18779386]
20. Samokhvalov AV, Rehm J, Roerecke M. Alcohol Consumption as a Risk Factor for Acute and Chronic Pancreatitis: A Systematic Review and a Series of Meta-analyses. EBioMedicine 2015 Dec;2(12):1996-2002 [FREE Full text] [doi: 10.1016/j.ebiom.2015.11.023] [Medline: 26844279]
21. Yadav D, Hawes RH, Brand RE, Anderson MA, Money ME, Banks PA, North American Pancreatic Study Group. Alcohol consumption, cigarette smoking, and the risk of recurrent acute and chronic pancreatitis. Arch Intern Med 2009 Jun 08;169(11):1035-1045. [doi: 10.1001/archinternmed.2009.125] [Medline: 19506173]
22. Rehm J, Taylor B, Mohapatra S, Irving H, Baliunas D, Patra J, et al. Alcohol as a risk factor for liver cirrhosis: a systematic review and meta-analysis. Drug Alcohol Rev 2010 Jul;29(4):437-445. [doi: 10.1111/j.1465-3362.2009.00153.x] [Medline: 20636661]
23. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 2006 May 27;367(9524):1747-1757. [doi: 10.1016/S0140-6736(06)68770-9] [Medline: 16731270]
24. Donoghue K, Patton R, Phillips T, Deluca P, Drummond C. The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: a systematic review and meta-analysis. J Med Internet Res 2014;16(6):e142 [FREE Full text] [doi: 10.2196/jmir.3193] [Medline: 24892426]
25. Bertholet N, Daeppen J, Wietlisbach V, Fleming M, Burnand B. Reduction of alcohol consumption by brief alcohol intervention in primary care: systematic review and meta-analysis. Arch Intern Med 2005 May 9;165(9):986-995. [doi: 10.1001/archinte.165.9.986] [Medline: 15883236]
26. Bray JW, Del Boca FK, McRee BG, Hayashi SW, Babor TF. Screening, Brief Intervention and Referral to Treatment (SBIRT): rationale, program overview and cross-site evaluation. Addiction 2017 Dec;112 Suppl 2:3-11. [doi: 10.1111/add.13676] [Medline: 28074566]
27. Madras BK, Compton WM, Avula D, Stegbauer T, Stein JB, Clark HW. Screening, brief interventions, referral to treatment (SBIRT) for illicit drug and alcohol use at multiple healthcare sites: comparison at intake and 6 months later. Drug Alcohol Depend 2009 Jan 01;99(1-3):280-295 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.08.003] [Medline: 18929451]
28. Vendetti J, Gmyrek A, Damon D, Singh M, McRee B, Del Boca F. Screening, brief intervention and referral to treatment (SBIRT): implementation barriers, facilitators and model migration. Addiction 2017 Dec;112 Suppl 2:23-33. [doi: 10.1111/add.13652] [Medline: 28074571]
29. McCabe SE, Cranford JA, West BT. Trends in prescription drug abuse and dependence, co-occurrence with other substance use disorders, and treatment utilization: results from two national surveys. Addict Behav 2008 Oct;33(10):1297-1305 [FREE Full text] [doi: 10.1016/j.addbeh.2008.06.005] [Medline: 18632211]
30. MacKillop J, Amlung MT, Few LR, Ray LA, Sweet LH, Munafò MR. Delayed reward discounting and addictive behavior: a meta-analysis. Psychopharmacology (Berl) 2011 Aug;216(3):305-321 [FREE Full text] [doi: 10.1007/s00213-011-2229-0] [Medline: 21373791]

31. Bickel WK, Jarmolowicz DP, Mueller ET, Koffarnus MN, Gatchalian KM. Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence. Pharmacol Ther 2012 Jun;134(3):287-297 [FREE Full text] [doi: 10.1016/j.pharmthera.2012.02.004] [Medline: 22387232]
32. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
33. Paparrizos J, White RW, Horvitz E. Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results. J Oncol Pract 2016 Jun 7:737-744. [doi: 10.1200/JOP.2015.010504] [Medline: 27271506]
34. Paul MJ, White RW, Horvitz E. Search and Breast Cancer. ACM Trans Web 2016 May 25;10(2):1-27. [doi: 10.1145/2893481]
35. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]
36. Andalibi N, Ozturk P, Forte A. Depression-related imagery on Instagram. 2015 Feb 28 Presented at: The 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (pp.). ACM; Feb 28, 2015; Vancouver, BC, Canada p. 231-234. [doi: 10.1145/2685553.2699014]
37. Birnbaum ML, Ernala SK, Rizvi AF, De Choudhury M, Kane JM. A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals. J Med Internet Res 2017 Dec 14;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]
38. Chancellor S, Lin Z, Goodman EL, Zerwas S, De Choudhury M. Quantifying and predicting mental illness severity in online pro-eating disorder communities. 2016 Presented at: Proceedings of the 19th ACM Conference on computer-supported Cooperative Work & Social Computing; Feb 27-Mar 02, 2016; San Francisco, CA p. 1171-1184 URL: https://www.researchgate.net/publication/311493030_Quantifying_and_Predicting_Mental_Illness_Severity_in_Online_Pro-Eating_Disorder_Communities [doi: 10.1145/2818048.2819973]
39. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, MA URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6124/6351
40. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:2098-2110 [FREE Full text] [doi: 10.1145/2858036.2858207] [Medline: 29082385]
41. Harman G, Dredze M. Measuring post traumatic stress disorder in Twitter. 2014 Presented at: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 2098-2110 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8079
42. Kumar M, Dredze M, Coppersmith G, De Choudhury M. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides. HT ACM Conf Hypertext Soc Media 2015 Sep;2015:85-94 [FREE Full text] [doi: 10.1145/2700171.2791026] [Medline: 28713876]
43. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. 2018 Jun 05 Presented at: The 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality (pp.); 2015; Denver, CO p. 11-20.
44. Murnane E, Counts S. Unraveling abstinencerelapse: smoking cessation reflected in social media. 2014 Apr 26 Presented at: The SIGCHI Conference on Human Factors in Computing Systems; Apr 26, 2014; Toronto, ON, Canada p. 1345-1354. [doi: 10.1145/2556288.2557145]
45. Saha K, Weber I, Birnbaum ML, De Choudhury M. Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates. J Med Internet Res 2017 Dec 08;19(5):e156 [FREE Full text] [doi: 10.2196/jmir.6815] [Medline: 28483739]
46. Tamersoy A, De Choudhury M, Chau DH. Characterizing Smoking and Drinking Abstinence from Social Media. HT ACM Conf Hypertext Soc Media 2015 Sep;2015:139-148 [FREE Full text] [doi: 10.1145/2700171.2791247] [Medline: 26640831]
47. Yom-Tov E, Fernandez-Luque L, Weber I, Crain SP. Pro-anorexia and pro-recovery photo sharing: a tale of two warring tribes. J Med Internet Res 2012 Nov 07;14(6):e151 [FREE Full text] [doi: 10.2196/jmir.2239] [Medline: 23134671]
48. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. J Med Internet Res 2017 Oct 31;19(10):e353 [FREE Full text] [doi: 10.2196/jmir.6426] [Medline: 29089287]
49. Yom-Tov E. Crowdsourced Health: How What You Do on the Internet Will Improve Medicine. Cambridge, MA: MIT Press; 2016.
50. Mejova Y, Weber I, Fernandez-Luque L. Online Health Monitoring using Facebook Advertisement Audience Estimates in the United States: Evaluation Study. JMIR Public Health Surveill 2018 Mar 28;4(1):e30 [FREE Full text] [doi: 10.2196/publichealth.7217] [Medline: 29592849]
51. Frijters P, Johnston DW, Lordan G, Shields MA. Exploring the relationship between macroeconomic conditions and problem drinking as captured by Google searches in the U.S. Soc Sci Med 2013 May;84:61-68. [doi: 10.1016/j.socscimed.2013.01.028] [Medline: 23517705]

52. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Tracking Dabbing Using Search Query Surveillance: A Case Study in the United States. J Med Internet Res 2016 Sep 16;18(9):e252 [FREE Full text] [doi: 10.2196/jmir.5802] [Medline: 27637361]
53. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web Use for Symptom Appraisal of Physical Health Conditions: A Systematic Review. J Med Internet Res 2017 Jun 13;19(6):e202 [FREE Full text] [doi: 10.2196/jmir.6755] [Medline: 28611017]
54. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
55. Kelly JF, Greene MC, Bergman BG. Do Drug-Dependent Patients Attending Alcoholics Anonymous Rather than Narcotics Anonymous Do As Well? A Prospective, Lagged, Matching Analysis. Alcohol Alcohol 2014 Nov;49(6):645-653 [FREE Full text] [doi: 10.1093/alcalc/agu066] [Medline: 25294352]
56. Comscore. 2018. Latest Rankings URL: https://www.comscore.com/Insights/Rankings [accessed 2019-02-27] [WebCite Cache ID 76VI6qByb]
57. Rosenblum S, Yom-Tov E. Seeking Web-Based Information About Attention Deficit Hyperactivity Disorder: Where, What, and When. J Med Internet Res 2017 Dec 21;19(4):e126 [FREE Full text] [doi: 10.2196/jmir.6579] [Medline: 28432038]
58. Dong J, Yom-Tov E, Yom-Tov GB. The Impact of Delay Announcements on Hospital Network Coordination and Waiting Times. Management Science 2018 Jul 30:1-26. [doi: 10.1287/mnsc.2018.3048]
59. Yom-Tov E, Marino B, Pai J, Harris D, Wolf M. The Effect of Limited Health Literacy on How Internet Users Learn About Diabetes. J Health Commun 2016 Oct;21(10):1107-1114. [doi: 10.1080/10810730.2016.1222033] [Medline: 27684112]
60. Compton WM, Thomas YF, Stinson FS, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. Arch Gen Psychiatry 2007 May;64(5):566-576. [doi: 10.1001/archpsyc.64.5.566] [Medline: 17485608]
61. Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry 2007 Jul;64(7):830-842. [doi: 10.1001/archpsyc.64.7.830] [Medline: 17606817]
62. Gaume J, McCambridge J, Bertholet N, Daeppen J. Mechanisms of action of brief alcohol interventions remain largely unknown - a narrative review. Front Psychiatry 2014;5:108 [FREE Full text] [doi: 10.3389/fpsyt.2014.00108] [Medline: 25206342]
63. Simons J, Correia CJ, Carey KB. A comparison of motives for marijuana and alcohol use among experienced users. Addict Behav 2000;25(1):153-160. [Medline: 10708331]
64. Monahan J, Lannutti P. Alcohol as social lubricant. Human Comm Res 2000 Apr;26(2):175-202. [doi: 10.1111/j.1468-2958.2000.tb00755.x]
65. Prochaska JO, DiClemente CC. Transtheoretical therapy: Toward a more integrative model of change. Psychotherapy: Theory, Research & Practice 1982;19(3):276-288. [doi: 10.1037/h0088437]
66. Yom-Tov E, Borsa D, Hayward AC, McKendry RA, Cox IJ. Automatic identification of Web-based risk markers for health events. J Med Internet Res 2015 Jan 27;17(1):e29 [FREE Full text] [doi: 10.2196/jmir.4082] [Medline: 25626480]
67. Cohen E, Feinn R, Arias A, Kranzler HR. Alcohol treatment utilization: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions. Drug and Alcohol Dependence 2007 Jan 12;86(2-3):214-221. [doi: 10.1016/j.drugalcdep.2006.06.008]
68. Yom-Tov E, Muennig P, El-Sayed AM. Web-Based Antismoking Advertising to Promote Smoking Cessation: A Randomized Controlled Trial. J Med Internet Res 2016 Nov 21;18(11):e306 [FREE Full text] [doi: 10.2196/jmir.6563] [Medline: 27872032]
69. Pilowsky D, Wu L. Screening for alcohol and drug use disorders among adults in primary care: a review. Subst Abuse Rehabil 2012 Apr;3(1):25-34 [FREE Full text] [doi: 10.2147/SAR.S30057] [Medline: 22553426]

Abbreviations

AA: alcoholics anonymous
AUD: alcohol use disorder
BI: brief intervention
NA: narcotics anonymous
SBIRT: Screening, Brief Intervention, and Referral to Treatment
SUD: substance use disorder

Edited by G Eysenbach; submitted 02.05.18; peer-reviewed by M Meacham, R Zerbetto; comments to author 06.10.18; revised version received 28.11.18; accepted 26.01.19; published 03.05.19.

Please cite as:
Nitzburg G, Weber I, Yom-Tov E
Internet Searches for Medical Symptoms Before Seeking Information on 12-Step Addiction Treatment Programs: A Web-Search Log Analysis
J Med Internet Res 2019;21(5):e10946
URL: https://www.jmir.org/2019/5/e10946/
doi:10.2196/10946
PMID:31066685

Original Paper

Modeling Spatiotemporal Factors Associated With Sentiment on Twitter: Synthesis and Suggestions for Improving the Identification of Localized Deviations

Zubair Shah[1], PhD; Paige Martin[1], BE; Enrico Coiera[1], PhD, MBBS; Kenneth D Mandl[2,3], MD, MPH; Adam G Dunn[1], PhD

[1]Centre for Health Informatics, Australian Institute for Health Innovation, Macquarie University, Sydney, Australia
[2]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States
[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Zubair Shah, PhD
Centre for Health Informatics
Australian Institute for Health Innovation
Macquarie University
75 Talavera Road, Macquarie Park
Sydney, 2113
Australia
Phone: 61 404941319
Email: zubair.shah@mq.edu.au

Abstract

Background: Studies examining how sentiment on social media varies depending on timing and location appear to produce inconsistent results, making it hard to design systems that use sentiment to detect localized events for public health applications.

Objective: The aim of this study was to measure how common timing and location confounders explain variation in sentiment on Twitter.

Methods: Using a dataset of 16.54 million English-language tweets from 100 cities posted between July 13 and November 30, 2017, we estimated the positive and negative sentiment for each of the cities using a dictionary-based sentiment analysis and constructed models to explain the differences in sentiment using time of day, day of week, weather, city, and interaction type (conversations or broadcasting) as factors and found that all factors were independently associated with sentiment.

Results: In the full multivariable model of positive (Pearson r in test data 0.236; 95% CI 0.231-0.241) and negative (Pearson r in test data 0.306; 95% CI 0.301-0.310) sentiment, the city and time of day explained more of the variance than weather and day of week. Models that account for these confounders produce a different distribution and ranking of important events compared with models that do not account for these confounders.

Conclusions: In public health applications that aim to detect localized events by aggregating sentiment across populations of Twitter users, it is worthwhile accounting for baseline differences before looking for unexpected changes.

(J Med Internet Res 2019;21(5):e12881) doi:10.2196/12881

KEYWORDS

text mining; social media; public health

Introduction

Background

Data from social media are increasingly being used in the digital phenotyping of individual users and the characterization of population-level behaviors to answer health-related questions [1-7]. Sentiment analysis is a broad class of methods used to detect opinions or mood from text. Although there are a range of approaches used in context-specific situations to detect positive and negative opinions about a topic [8-12], here we restrict the definition to include the general sentiment analysis methods used to detect mood. Sentiment analysis has also been used for applications in public health to evaluate reactions and attitudes to certain current events [13], health interventions such as vaccination [14], human mobility [15], and outcomes such as seasonal affective disorder and obesity [16-18].

When using sentiment analysis tools to observe or find signals of changes in the sentiment of a population, researchers must navigate the complicated interactions between the tools they use and the spatiotemporal and social factors that are known to modify mood and emotion. For example, the positive and negative affect measured by sentiment analysis has been shown to be associated with the time of day and day of week [19-21], weather [22-25], and the quality of social interactions [26].

Studies applying sentiment analysis to Twitter data have confirmed the periodicity of positive and negative affect by time of day and day of week [16,25,27-29]. However, the results and conclusions vary from study to study, and these differences may depend on the methods used to aggregate sentiment across sets of tweets or users, differences in the ways the investigators sampled the data, differences in the sentiment analysis algorithms or tools used, or because of challenges associated with validating results against external information. In comparison, studies examining variation in sentiment by geography or weather are relatively rare compared with those that measure temporal variation [30-34]. Studies that report analyses for social interactions on Twitter—tweets that mention, reply to, or quote other users—do not appear to have focused on measuring differences in the sentiment relative to tweets that broadcast a message [35].

Objectives

The aim of this study was to construct simple models of positive and negative sentiment using time of day, day of week, interaction type, weather, or city as factors to understand how each of the different modifying factors might distort the results of public health studies that use sentiment analysis to study Twitter data. We then used the model and degenerate versions of the model to measure the magnitude of the differences between expected and observed sentiment over time and show how accounting for spatiotemporal differences affects the ranking of the importance of individual events.

Methods

This study was an analysis of tweets posted by Twitter users in 100 cities. To address our aims, we aggregated sentiment scores for each hour in each of the 100 cities and constructed multivariable models to explain differences in the proportion of tweets, expressing positive or negative sentiment using city, interaction type, weather, time of day, and day of week as factors. We selected each of these factors because they have been shown to be associated with sentiment in past research and are relatively easily and accurately inferred from Twitter data.

Twitter Data

We used the Twitter streaming application programming interface (API) to collect tweets between July 13 and November 30, 2017, without using any keywords. The retrieved tweets represent an approximate 1% sample of all tweets produced globally. Each tweet contains information about the user including name, location, tweet counts, follower counts, and following counts and the information about the tweet itself such as timestamp and the users it mentions.

Information in the tweet also provides information about whether it was a reply to a previous tweet, a retweet, or includes a link (quotes) to another tweet. We used this information to label each tweet as either broadcast (quotes, retweets, and tweets that do not mention other users) or social (replies and direct mentions of other users in the tweet).

Location Data

Identifying the home locations of users on Twitter is a challenging task owing to the low number of posts with precise location information (geotags) and the need to parse user-defined location information using a gazetteer. Fewer than 0.5% tweets are geotagged, and fewer than 50% of Twitter users have provided useful home locations in their profiles [36]. To identify the location of the tweets from where it has been posted, we took the user-defined text from the location field in Twitter user profiles and used Nominatim, a gazetteer that returns a JavaScript Object Notation (JSON) object containing structured geographical information and a score associated with the confidence in the answer. Rather than filtering Nominatim results using a threshold on the confidence score, we found that Nominatim produces better results if we filter addresses based on type field of the return JSON object; therefore, we used type field in the returned JSON object to accept the top first address having type as city, county, village, suburb, hamlet, state, or country. This helped us to filter out other types of addresses without needing to use a specific threshold.

Not all Twitter accounts represent individuals; some are brands or organizations where tweets may be posted by humans or bots. Rahimi et al [37] used a simple but effective approach to removing *celebrities* in a study on location inference, in which they removed tweets from accounts that had more than 300,000 followers. After examining a set of Twitter users on either side of this threshold in our training data, we followed the same approach and removed all users with at least 300,000 followers.

Timing Data

Past studies examining temporal patterns in sentiment on social media have found clear patterns [16,20,21,27]. However, those patterns vary substantially from study to study: some observed the most negative sentiment on Mondays and the most positive sentiment on Fridays or Saturdays. Some observed the strongest negative sentiment between 2 am and 5 am, whereas others observed the same between 8 pm and 11 pm.

As Twitter no longer includes a localized timestamp for users in the metadata of tweets, we used the identified location of the users posting the tweets to convert the timestamps of tweets from Universal Time Coordinated to local time. In what follows, all tweets are considered relative to the local time of the city in which the user is believed to be located.

Weather Data

Past studies examining weather and sentiment on Twitter have produced variable results, but most observe one or more associations [31-33]. We collected hourly weather data for the top 100 cities using the API from the Open Weather website

[38]. The information provided by the Open Weather website includes detailed weather information, such as temperature and humidity, and weather descriptions. We then mapped weather for each hour in each city to one of 7 values: clear, clouds, fog, haze, rain, snow, or storm.

Sentiment Measures

Sentiment analysis of written text is a widely studied problem in natural language processing [39-41]. In this study, we have considered sentiment in a simple form—the presence of positive or negative affect—and applied SentiStrength [42], a widely used open-source Java library designed for sentiment analysis of tweets. It has been evaluated manually and compared with a range of advanced machine learning and statistical methods in several studies [42-44]. SentiStrength is a dictionary-based method, using a lexicon of words categorized as positive or negative with a score for its polarity and strength. For a given tweet, SentiStrength identifies the presence of sentiment terms from its lexicon and computes the sentiment of the text based upon the scores of the words found. SentiStrength produces 2 scores for each tweet, one indicating positive sentiment (from 1 to 5, least positive to most positive, respectively) and one indicating negative sentiment (from 1 to 5, least negative to most negative, respectively). As SentiStrength uses a score of +1 or −1 for neutral words, we considered scores from 2 to 5 for both positive and negative sentiments. In addition, as SentiStrength identifies positive and negative words independently, it is possible for a tweet to be labeled as having positive, negative, or both positive and negative sentiment.

We aggregated sentiment scores across a set of tweets using the proportion of tweets that have a positive sentiment score (a score from 2 to 5 in positive sentiment) or the proportion of tweets that have a negative sentiment score (a score from 2 to 5 in negative sentiment). Methods for aggregating scores across groups of tweets are important because they can influence the interpretation and lead to different conclusions. To aggregate sentiment scores, researchers have used counts, averages, proportions, ratios, and weighted averages [16,27,28,45-50]. Some have combined positive and negative scores to create a single measure [13,27,28,48,49], whereas others have kept positive and negative scores separate [46,47,51]. Following Scott et al [16], we used positive and negative sentiment scores separately because the positive and negative affect can coexist [52,53] and because when aggregated, a population can exhibit higher levels of both positive and negative sentiment at the same time. Thus, a low positive score indicates the absence of positive emotion across a set of tweets not the presence of negative emotion.

Analysis and Modeling

In the first part of the analysis, we examined how each of the factors—interaction type, time of day, day of week, weather, and city—were associated with differences in the proportions of tweets that expressed positive or negative sentiment in a city in an hour. To do this, we constructed multivariable regression models using each of the factors individually and then in combination. We chose to use multivariable regression models because they are a simple way of capturing the baseline patterns of sentiment, and models built using individual factors and their combinations can be directly compared. For each model, we reported the r-squared value as a percentage, representing the percentage of the variance in sentiment that can be explained by each model.

In our evaluation of the models on unseen data, we then reported the correlation (Pearson r) between the values predicted by the model and the observed data in a set of testing data, distinct from the period of observation used to construct the models. These comparisons tell us how important each of the factors are as independent predictors of the sentiment for a city-hour pair and can provide guidance on which of the factors may be useful to control for when analyzing sentiment to detect changes or anomalies.

In the second part of the analysis, we have used the models constructed in the first part of the analysis as a baseline for detecting deviations from the expected proportions of positive and negative sentiment tweets per city per hour. The objective was to determine whether baseline differences in spatiotemporal and social factors would introduce biases in the detection of extreme deviations in sentiment that occur during major localized news events and if accounting for them in a baseline model could address these biases. To do this, we compared the expected and observed proportions of positive and negative sentiment tweets per city per hour using a chi-square test and then used the resulting *P* value as an indicator of the magnitude of the deviation.

Rather than defining an explicit threshold to label hour-city pairs as events or nonevents, we used the magnitude of the deviation in sentiment to rank all hour-city pairs in descending order based on the chi-square test. To make it easier to understand the expected frequency of the events, we defined a recurrence interval: the number of days of observation divided by the frequency of an event of that magnitude across the set of all cities in the analyses. For example, given 60 days of observation in the test period, a recurrence interval of 30 days is an event with a test statistic that was exceeded only twice during the 60 days. A recurrence interval of 1 day is an event with a magnitude that was exceeded 60 times in a 60-day period.

To characterize an event by its magnitude, we also needed to account for extreme sentiment that persisted for multiple hours or was expressed across multiple cities within a country. To do this, we merged events that produced significant differences between the observed and predicted number of positive or negative sentiment tweets and labeled them using the highest test statistic in the period. Similarly, we merged cities within a country if significant events occurred at the same time. As a result, hour-city pairs could be merged to produce day-city, day-country, or multi–day-country events depending on how many of the ranked deviations were traversed.

We then compared the events identified from the full model with the events produced by degenerate forms of the full model (eg, excluding city or interaction type as a factor). We used these differences to evaluate how the use of baseline spatiotemporal modeling affected the identification and ranking of extreme sentiment events. The expectation was that the degenerate forms of the models would introduce a bias in the distribution of events toward certain cities or times of day.

Results

On average, we received 3.66 million tweets a day for 141 days, for a total of 507.60 million tweets from 27.61 million unique users. In the dataset, Twitter tagged 29.78% (151.21/507.60 million) as English language. Of these, 65.67% (99.30/151.21 million) had location information available in the users' profiles.

After removing celebrity/brand accounts, we ranked cities based on the total number of English language tweets posted by users with locations that the gazetteer was able to resolve. We identified the 100 cities with the highest numbers of English language tweets posted during the study period. These included 52 cities in North America (45 from the United States, 6 from Canada, and 1 from Mexico), 11 cities in the United Kingdom, 6 cities from Europe, 16 cities in Asia and Southeast Asia, 9 cities in Africa, 3 cities in Australasia, 2 cities from the Middle East, and 1 city in South America. We were able to resolve 16.61% (16.50/99.30 million) of the English language tweets to one of the 100 cities (Figure 1). We used these tweets as the basis for the study.

Analysis of Spatiotemporal and Social Factors

The training data used to construct the multivariable models comprised 8.39 million tweets from the first 81 days of data collection (July 13 to September 30, 2017). Of these, we found that 39.69% (3.33 million) expressed positive sentiment and 28.13% (2.36 million) expressed negative sentiment. Users across the 100 cities posted more tweets on Monday to Thursday and slightly fewer tweets from Friday to Sunday. The hour in which users were typically most active was between 12 noon and 1 pm (an average of 7652 tweets across the 100 cities), and users were least active between 4 am and 5 am (an average of 1745 tweets across the 100 cities). The number of tweets in each category of weather varied from snow (230 tweets) and storms (189,201 tweets) to cloudy weather (3,247,680 tweets). Relative to the average proportions of positive and negative sentiment, early morning hours exhibited lower proportions for both positive and negative sentiment, with the highest proportions of positive sentiment between 9 pm and 10 pm and highest rates of negative sentiment in the hours between 11 pm and 1 am, with an additional smaller peak between 7 am and 8 am (Figure 2). Fridays exhibited the highest proportion of positive sentiment and the lowest proportion of negative sentiment.

We constructed each model to estimate the proportion of tweets that expressed positive or negative sentiment in a city in an hour and have presented results based on the correlation between the estimated and observed proportions within the training data (Tables 1 and 2).

Figure 1. From 507.6 million tweets, 16.5 million were labelled as English language and attributed to users in 100 cities.

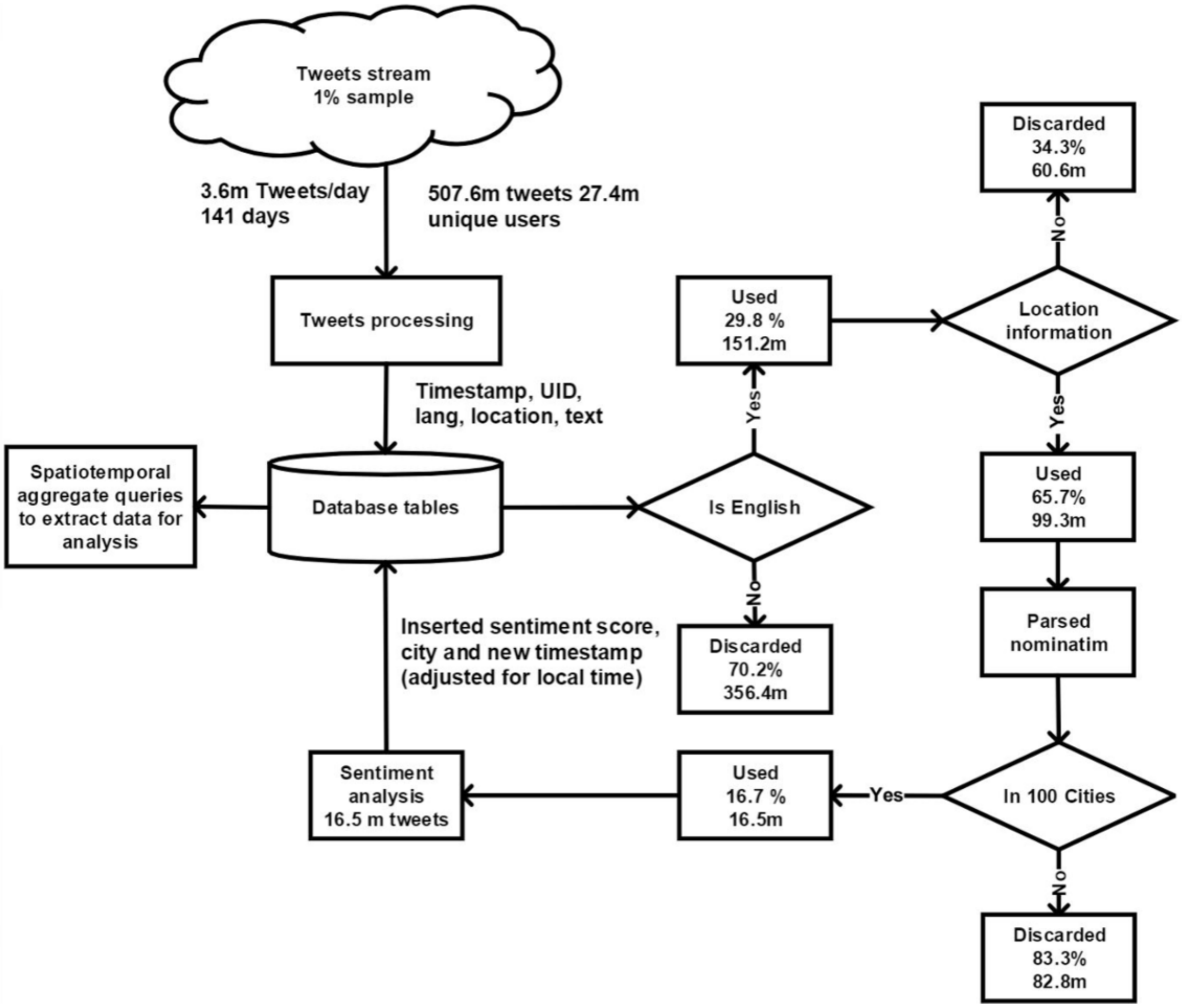


Figure 2. Observed proportions of positive and negative sentiment aggregated for all city-hour pairs by hour of the day (top), day of the week (center), and weather type (below). CIs are an indication of the number of city-hour pairs that contributed and the variability in proportion for that value. All values are categorical, so dotted lines are for visual interpretation only.

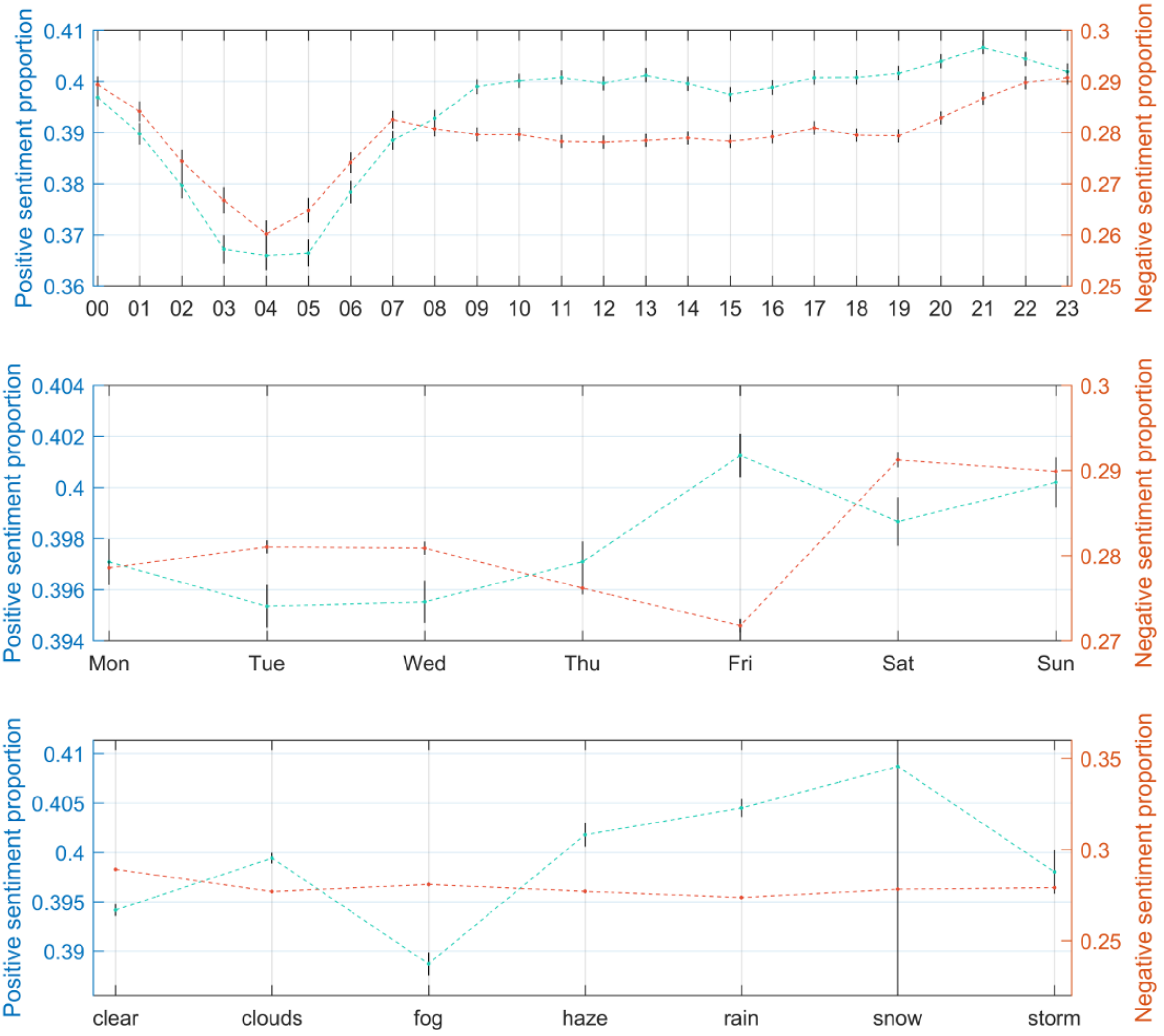


Table 1. Final model coefficient estimates for models of the proportion of tweets that exhibited negative sentiment in an hour.

Factor	Number of coefficients (number *P*<.05)	r-squared in the training period, %	Pearson r (95% CI) in the testing period
Multiple factor models			
All factors	136 (108)	9.345	0.306 (0.301-0.310)
Social, city, hour, day	130 (107)	9.338	0.306 (0.301-0.310)
Social, city	101 (80)	8.831	0.297 (0.292-0.302)
Hour, day	30 (26)	0.486	0.070 (0.065-0.075)
Single factor models			
City	100 (81)	8.736	0.296 (0.291-0.300)
Hour of day	24 (20)	0.298	0.055 (0.049-0.060)
Day of week	7 (7)	0.191	0.044 (0.039-0.049)
Weather	7 (5)	0.193	0.044 (0.039-0.049)
Social proportion	2 (2)	0.010	0.010 (0.005-0.015)

Table 2. Final model coefficient estimates for models of the proportion of tweets that exhibited positive sentiment in an hour.

Factor	Number of coefficients (number *P*<.05)	r-squared in the training period, %	Pearson r (95% CI) in the testing period
Multiple factor models			
All factors	136 (107)	5.584	0.236 (0.231-0.241)
Social, city, hour, day	130 (107)	5.580	0.236 (0.231-0.241)
Social, city	101 (85)	4.671	0.216 (0.211-0.221)
Hour, day	30 (26)	1.330	0.115 (0.110-0.133)
Single factor models			
City	100 (90)	3.732	0.193 (0.188-0.198)
Hour of day	24 (21)	1.271	0.113 (0.108-0.118)
Day of week	7 (6)	0.053	0.023 (0.018-0.028)
Weather	7 (5)	0.170	0.041 (0.036-0.046)
Social proportion	2 (2)	1.387	0.118 (0.113-0.123)

A model combining both temporal factors was significantly correlated with the proportion of tweets expressing negative sentiment (r=0.070; 95% CI 0.065-0.070). The association was stronger with the proportion of tweets expressing positive sentiment (r=0.115; 95% CI 0.110-0.133) and explained 5% of the variance. For both positive and negative sentiment outcomes, adding the day of the week to the hour of the day in the model produced a significant improvement in the model.

Positive and negative sentiment also varied by interaction type, where social tweets (tweets that mention or reply to another user) were much more likely to be expressions of positive sentiment relative to nonsocial tweets (tweets that do not mention or reply to another user). In hours where higher proportions of tweets were social interactions, the proportion of tweets that expressed positive sentiment were higher (r=0.118; 95% CI 0.113-0.123) and the proportion of tweets that expressed negative sentiment were lower (r=0.010; 95% CI 0.005-0.015) but this was a much weaker association. Adding the proportion of tweets that were social interactions as a factor in multivariable models made a significant improvement to the performance of the model in all cases.

The median number of tweets per city during the testing period was 48,974 and the number varied from 24,825 (Istanbul, Turkey) to 856,471 (New York City, United States). The numbers of tweets generally matched with the populations of the cities (Figure 3) and was lower for countries where languages other than English are used. Cities in the United States tended to have higher proportions of negative sentiment tweets and lower proportions of positive sentiment tweets (Figure 4). Models using only city information exhibited the strongest correlation with the proportion of positive and negative sentiment tweets in an hour compared with all other factors, explaining 8.73% of the variance in negative sentiment (r=0.296; 95% CI 0.291-0.300) and 3.70% of the variance in positive sentiment (r=0.193; 95% CI 0.188-0.198).

Weather exhibited weak associations with the proportions of tweets expressing positive (r=0.041; 95% CI 0.036-0.046) or negative sentiment (r=0.044; 95% CI 0.039-0.049). Its addition to the multivariable model including all other factors significantly improved the performance. However, as the coefficients for weather were orders of magnitude smaller than other factors such as city and social proportion, weather did not appear to be a useful addition to the baseline models used in the detection of variation in sentiment caused by exogenous factors.

Figure 3. The number of tweets identified per city relative to the population of the city. Population data were manually collected from Wikipedia in December 2017, using the most recent metropolitan values available. Cities in the United States are highlighted in red and cities are partially labelled.

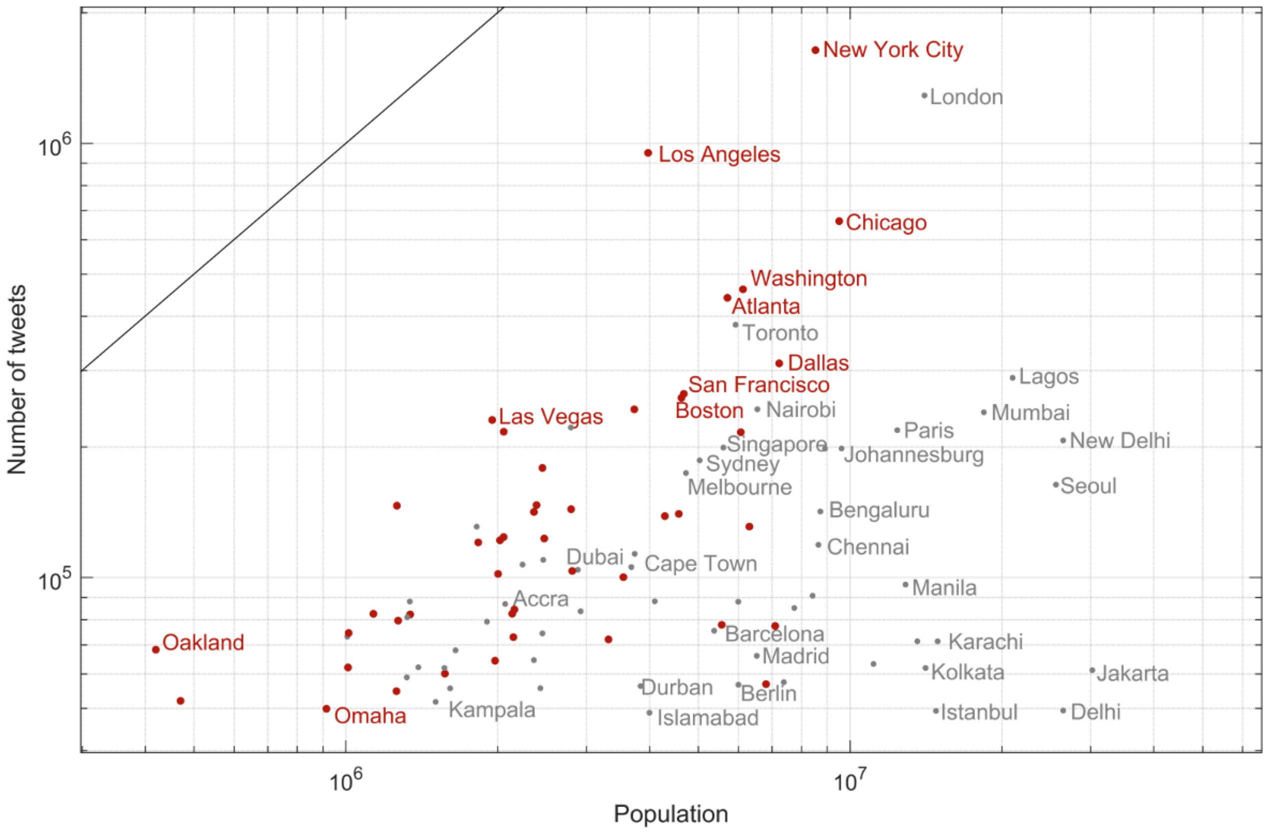
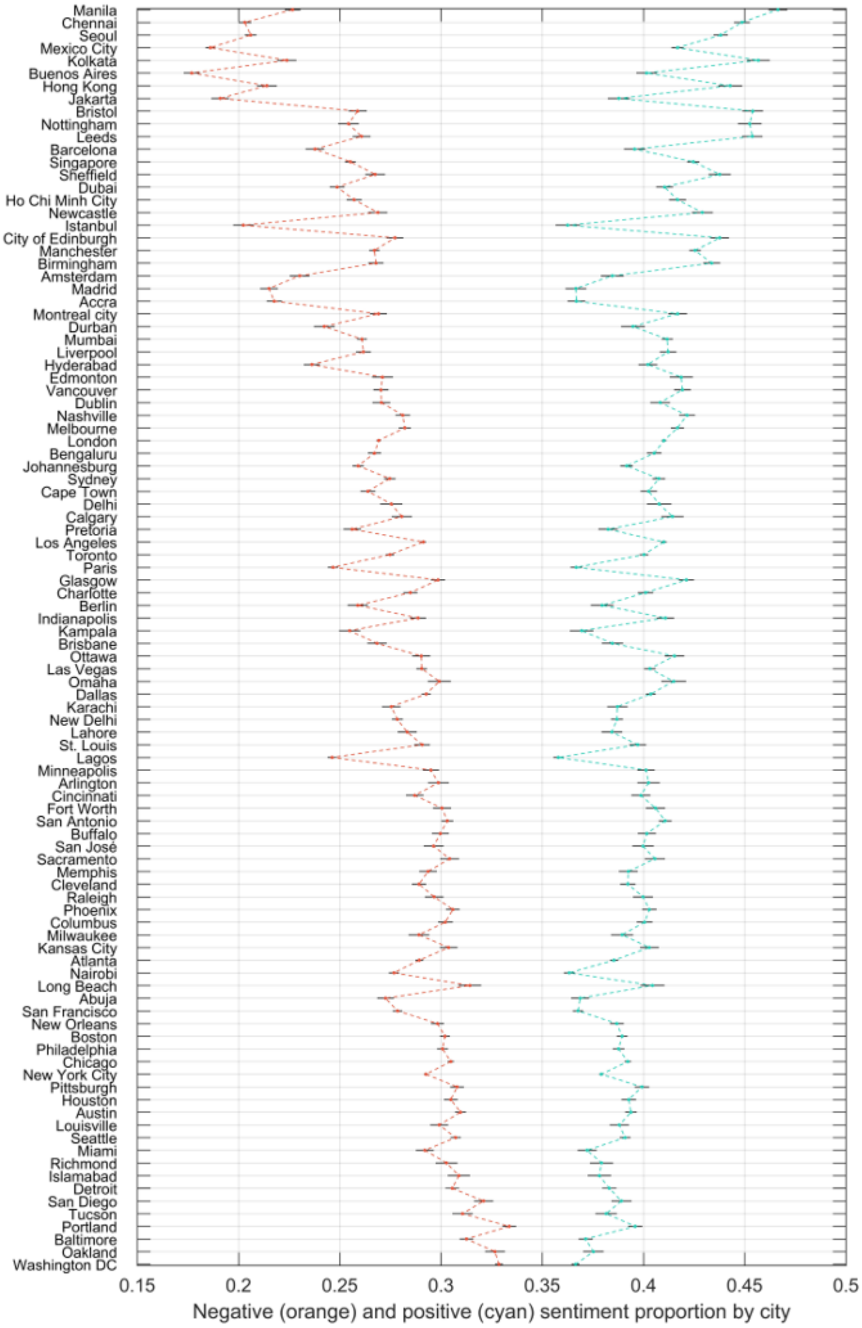


Figure 4. Sentiment by city in the training period, by proportion of positive (cyan) and negative (orange) sentiment tweets. Cities are ranked in decreasing order of the mean of the proportion of tweets with positive sentiment minus the proportion of tweets with negative sentiment.



Detecting Deviations in City-Level Expression of Positive or Negative Sentiment

We then used the models constructed above to predict the expected sentiment in city-hour pairs constructed from a separate set of 8.02 million tweets from the following 60 days (October 1 to November 30, 2017). We found similar proportions of tweets expressing positive sentiment (3.20/8.02 million, 39.90%) or negative sentiment (2.28/8.02, 28.43%) as we found in the training data. For every hour-city pair, we determined the magnitude of localized deviations by measuring the difference between the expected and observed proportions of positive and negative sentiment tweets.

Using the full model to identify unexpected deviations in the proportion of positive or negative sentiment tweets in the test period, we ranked events based on the magnitude of the deviation (Figure 5). As the number of events that might be considered important may vary depending on application, we have used the rank set of all city-hour pairs and traverse the list from the most extreme deviations to the least extreme deviations.

The top examples of localized deviations are listed in Table 3. We aggregated hour-city pairs across contiguous hours and cities wherever possible by reporting the most extreme deviation and merging any subsequent (less extreme) deviation that was on the same day (eg, extreme deviations in sentiment in the same direction on the same day in the same city are merged and reported as a day event) or cities in the same country (eg, 10 am in New York City and 10 am in Los Angeles is reported as 10 am in the United States). This was also extended to merge over both dimensions to report events by country and day. Where contiguous days reported events in the same direction, these events were merged as multi-day events.

After accounting for city-level differences in baseline proportions of positive and negative sentiment tweets, we found that the highest ranked events were distributed across 7 countries and could be retrospectively matched with major news stories that were specific to each of the cities. Using the degenerate models that do not account for city-level baseline differences, the United States accounted for a lower proportion of extreme positive events (Figure 6). This occurs because cities in the United States tend to exhibit higher rates of negative sentiment and lower rates of positive sentiment than cities in other countries. Models that do not take this baseline difference into account may overestimate the number of important negative events in the United States (which also has the effect of making violence in Barcelona or Nairobi seem less important) or underestimate the number of positive events in the United States (shifting down positive sentiment events such as Thanksgiving Day parade in New York City, New York or the World Series win in Houston, Texas).

Figure 5. The set of all city-hour pairs for negative sentiment (left) and positive sentiment (right), ordered by decreasing the chi-square test statistic value. Note that there are thousands of city-hour pairs for which the test produces a P value under .05 (red). The recurrence interval for each city-hour pair is given by the value on the horizontal axis divided by the observation period in days (60 days).

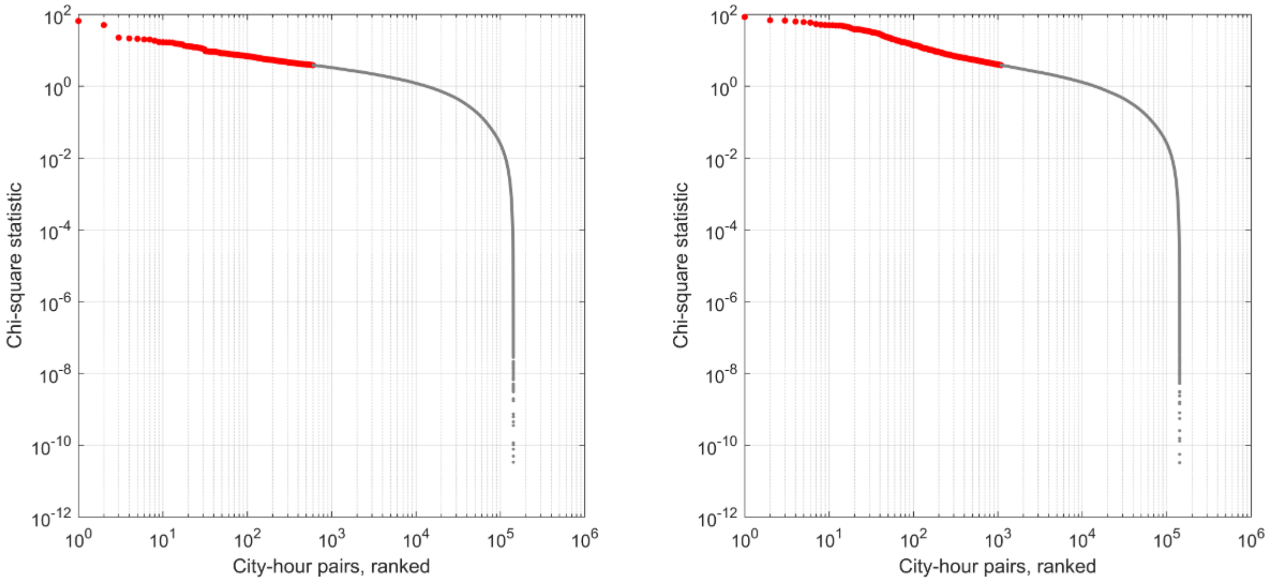


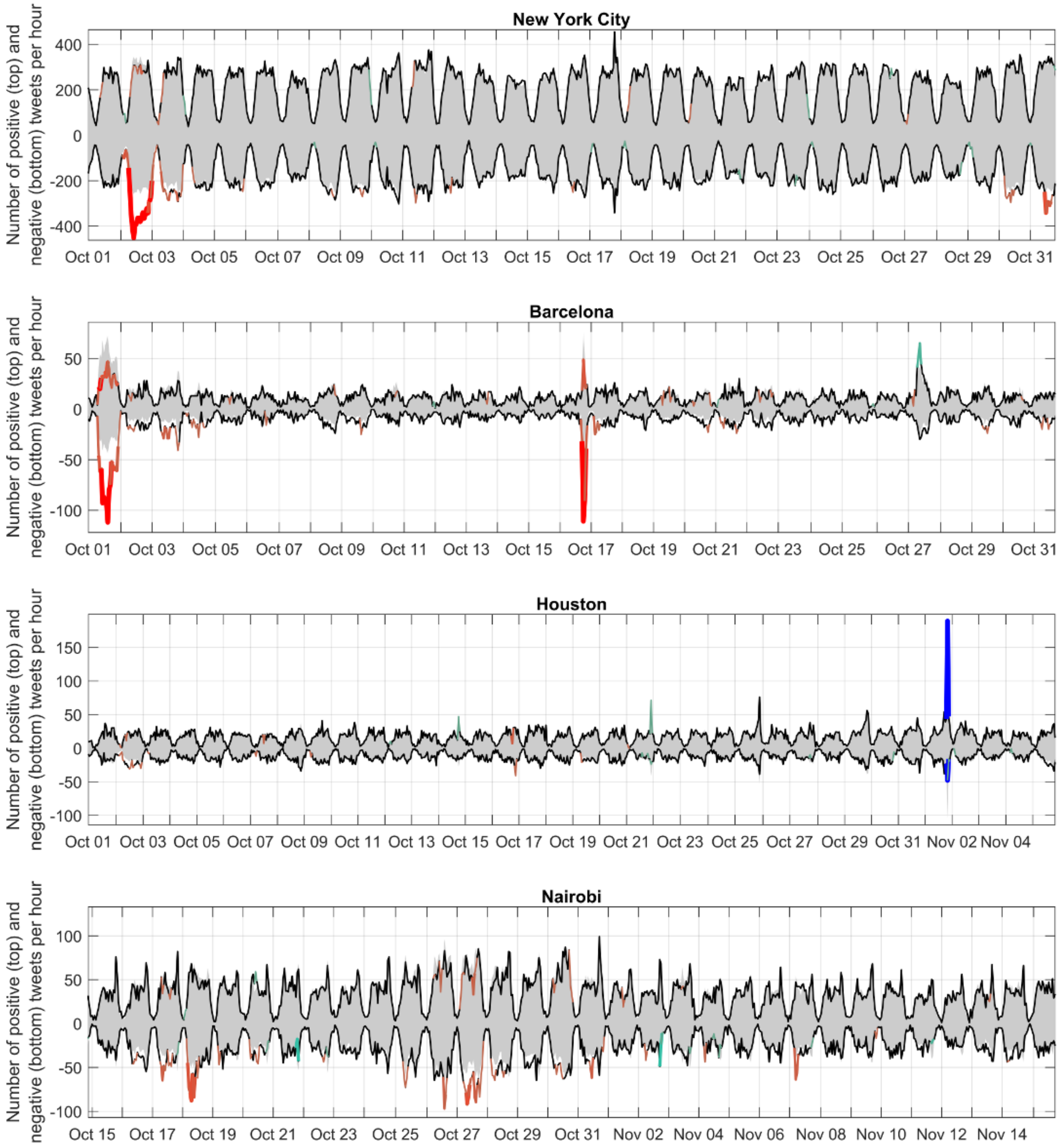
Table 3. Examples of extreme city-level events with large deviations detected in sentiment.

Time and location	Percentage of negative sentiment tweets (% expected)	Percentage of positive sentiment tweets (% expected)	Recurrence interval (global; days)	Corresponding news event in the period
October 2, 2017 in the United States	49.56 (28.70)	31.30 (38.14)	>60	Coverage following Las Vegas shooting
November 25-27 in Manila	12.13 (22.91)	73.20 (45.67)	30	Miss Universe pageant
October 1-2, 2017 in Las Vegas	61.51 (30.72)	48.32 (40.52)	20	Shooting terror event at a music festival
October 1, 2017 in Barcelona	60.89 (23.78)	14.67 (39.56)	12	Voting for Catalonian independence
October 16, 2017 in Barcelona	67.41 (23.78)	17.8 (39.67)	10	Catalonian independence events
November 2, 2017 in Houston	14.41 (31.61)	56.60 (38.20)	8.6	Houston Astros win world series
November 23, 2017 in New York City	20.40 (29.01)	50.51 (37.40)	7.5	Thanksgiving Day parade
October 19, 2017 in Dubai	8.11 (25.01)	92.13 (39.02)	6	Diwali festival
October 27, 2017 in Nairobi	48.50 (26.52)	22.13 (37.32)	5.5	Riots following election
November 27, 2017 in Seoul	8.02 (21.01)	71.67 (43.30)	5	2 North Korean embarrassments
November 24, 2017 in London	35.50 (26.51)	47.12 (37.89)	4.6	False terror scare in Oxford Circus

Figure 6. Most positive events for the 3 models aggregated where possible over hours, days, and cities. Note that compared with the full model (left), events from the United States tend to be moderated by the baseline tendency away from positive sentiment in the model without cities as factors (centre), and the null model (right).

From among the examples listed in Table 3, the visualization of the extreme events illustrates different types of deviations from the baseline (Figure 7). In each example, the expected baseline is the expected proportion of positive sentiment and negative sentiment tweets in an hour multiplied by the number of tweets from that city. Unexpected deviations occur when the observed number of positive or negative sentiment tweets is much higher or much lower than the baseline (in Figure 7, colored in red or blue). There were visible differences in the patterns indicating events that occur over a period of time (eg, riots after an election in Nairobi and a day of attempted voting in Barcelona) and events that occur within 1 or several hours (Houston Astros winning a baseball final). Other events not pictured include the outpouring of grief across multiple cities in the United States after a mass shooting, which decay more slowly over a period of days.

Figure 7. Examples of individual cities (New York City, Barcelona, Houston, and Nairobi) by the expected (gray areas) and observed (black and colored lines) sentiment. The color of the line indicates the magnitude of the deviation (darker red: more negative or fewer positive tweets; darker blue: more positive or fewer negative tweets).



Discussion

When applying sentiment analysis tools to Twitter data to characterize a population over time, it is useful to account for baseline spatiotemporal differences before attempting to detect deviations in mood. The first contribution of this study was to show that hour of day, day of week, the proportion of social tweets, the locations of the users posting the tweets, and the weather are each independently correlated with both positive and negative sentiment. Second, although these factors together account for less than 10% of the variance in positive and negative sentiment, ignoring them can affect the detection of unexpected deviations. Finally, we confirmed that in studies aggregating across populations (ecological designs), positive and negative sentiment can rise and fall independently and aggregating them into a single measure may mean losing important information that helps characterize the mood of a population.

Comparisons With Past Literature and Implications

A range of studies have applied sentiment analysis tools to social media data to examine changes in mood or emotion in relation to current events, weather and season, or circadian and daily rhythms. Our results extend these analyses to demonstrate the relative importance of each of these factors.

We found that the time of day and day of week were more closely correlated with positive sentiment than with negative sentiment. For positive sentiment, models built using these temporal factors typically explain less of the variance than models that used social interactions and cities as factors. Previous studies investigating hourly and daily patterns of sentiment on Twitter vary in structure from cohort designs, where individual users are followed [16,54], to ecological designs where signals from a population are aggregated [27,29,55]. The results of these studies and the conclusions they draw appear to be related to design choices including the tools used to measure sentiment and the methods used to aggregate measures of sentiment across populations.

The results of the study are consistent with previous studies that have found associations between weather and sentiment on Twitter [31-33]. Despite the observed independent correlations between weather and sentiment, weather explained little of the variance in positive or negative sentiment. These results should not be confused with seasonal variation in weather or sunlight; our results did not extend across a full range of seasons, and other studies have examined the use of Twitter data for its potential to observe seasonal affective disorder [16,17]. Mitchell et al [30] examined the geography of happiness in 373 cities in the United States using Twitter data and found that happiness was correlated with socioeconomic status and health-related census data, among other factors. We found that negative sentiment was more common and positive sentiment less common in tweets from many cities in the United States and suggest that future research in the area would benefit from studying international differences in sentiment associated with culture and patterns of living and working that might influence the expression of sentiment on social media.

Tweets that involve social interactions on Twitter (typically replies and mentions) are common in applications of network science. Our results show a strong positive correlation between the proportion of social interactions in a city in an hour and positive sentiment and a weak correlation with negative sentiment. Future applications that couple network analysis with sentiment tools may benefit from recognizing and potentially accounting for the differences between tweets that are social in nature, relative to those that are broadcasting information.

Twitter and other social media platforms offer the opportunity to undertake naturalistic studies of human behaviors at unprecedented scales [56-59]. However, studies in the area are at risk of producing incomparable results and inconsistent conclusions if sampling methods vary in ways that skew toward certain locations or certain times of the day or week. Practitioners in the area are already aware of the risks of selecting only geotagged tweets [60], but the spatiotemporal differences we highlight here are typically not discussed or accounted for in applications that use Twitter data to answer public health questions.

Limitations and Future Work

The study has several limitations. First, Twitter users represent a biased sample of countries and a biased sample of the population within countries [60-64], and we did not infer the demographics nor apply any reweighting methods to adjust for differences between the users posting English language tweets and the demographics of the cities we examined. Furthermore, users who include enough biographical information to be located within a city may represent a biased subset of the overall Twitter population, and we did not use location inference methods that take advantage of location-indicative words or social network structure [65-68] because these could introduce further sampling biases (eg, the overlapping of words in the dictionary and those that are useful in predicting a location). For these reasons, the study only captures deviations that might be expected to be important to population-level (epidemiological) studies.

Second, we used SentiStrength as a measure of sentiment and did not consider alternatives, sentiment in languages other than English, or ensembles combining multiple tools [69-72]. We think our use of SentiStrength is justified because it is a commonly used tool in studies in public health and has been examined for sentence-level sentiment and on individual tweets previously [42,44,45]. Although we did not test multiple sentiment detection methods to confirm, we expect that the need to account for baseline spatiotemporal differences is likely to be useful across all other sentiment detection approaches.

Third, certain events are less localized and affect multiple cities or even multiple countries and others may extend across many hours, days, or weeks. Methods for dealing with the spatiotemporal granularity of these events would be a useful addition to the sets of methods used in analyses of sentiment (or other measures that can be observed in social media datasets). Real-time event detection on Twitter is an active area of research [73,74], and our aim was not to add to this literature. Rather, we sought to develop a way to improve the robustness of observational studies that use sentiment analysis of Twitter to make sense of how populations react to real-world events.

Finally, we selected a set of factors that were known to be associated with sentiment on Twitter and used a relatively simple approach to modeling their associations. Other user-level factors and more sophisticated models may improve our ability to account for baseline differences in sentiment, including heterogeneity of individual-level differences that are apparent at population-level scales. For example, other factors that could have been included are gender, age, and number of followers; and other modeling pipelines might consider feature selection or dimensionality reduction and cross-validation techniques to avoid overfitting and improve generalization.

Conclusions

In this study we showed that in applications that use population-level measures of sentiment on Twitter, it is useful to account for baseline differences in sentiment by time of day, day of week, location, weather, and interaction type. Doing so could improve the accuracy of methods that use sentiment to detect localized events or changes in mood. The first contribution of this research is the consistent evaluation of a broad set of factors—making it easier to compare the importance of location, time, and social interactions on positive and negative sentiment. The second contribution is the use of these factors to construct a simple and interpretable model of the expected variation in positive and negative sentiment on Twitter.

Acknowledgments

This research was supported by funding from the National Health and Medical Research Council (Project APP1128968).

Conflicts of Interest

None declared.

References

1. Centola D. Social media and the science of health behavior. Circulation 2013 May 28;127(21):2135-2144. [doi: 10.1161/CIRCULATIONAHA.112.101816] [Medline: 23716382]
2. Salathé M, Bengtsson L, Bodnar T, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]
3. Dredze M. How social media will change public health. IEEE Intell Syst 2012;27(4):81-84. [doi: 10.1109/MIS.2012.76]
4. Paul M, Dredze M. Association for the Advancement of Artificial Intelligence. 2011. You are what you Tweet: Analyzing Twitter for public health URL: https://www.cs.jhu.edu/~mdredze/publications/twitter_health_icwsm_11.pdf [accessed 2019-04-17] [WebCite Cache ID 77iSu3Co2]
5. Coppersmith G, Dredze M, Harman C. Association for Computational Linguistics. 2014. Quantifying mental health signals in Twitter URL: https://www.cs.jhu.edu/~mdredze/publications/2014_acl_mental_health.pdf [accessed 2019-04-17] [WebCite Cache ID 77iT5hB1a]
6. Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci '13; May 02-04, 2013; Paris, France. [doi: 10.1145/2464464.2464480]
7. Althouse B, Scarpino S, Meyers L, Ayers JW, Bargsten M, Baumbach J, et al. Enhancing disease surveillance with novel data streams: challenges and opportunities. EPJ Data Sci 2015;4:- [FREE Full text] [doi: 10.1140/epjds/s13688-015-0054-0] [Medline: 27990325]
8. Tumasjan A, Sprenger T, Sandner P, Welpe I. Predicting elections with twitter: What 140 characters reveal about political sentiment. 2010 Presented at: Fourth international AAAI conference on weblogs and social media; 2010; Washington, USA p. 178-185.
9. Bermingham A, Smeaton A. On using Twitter to monitor political sentiment and predict election results. 2011 Presented at: Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP); 2011; Chiang Mai, Thailand p. 2-10.
10. Asur S, Huberman B. Predicting the future with social media. 2010 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology; 31 August-3 September, 2010; Toronto, ON, Canada p. 492-499. [doi: 10.1109/WI-IAT.2010.63]
11. Jain V. Prediction of movie success using sentiment analysis of tweets. IJSCE 2013;3(3):308-313. [doi: 10.7321/jscse.v3.n3.46]
12. Pasek J, Yan H, Conrad F, Newport F, Marken S. The stability of economic correlations over time: identifying conditions under which survey tracking polls and Twitter sentiment yield similar conclusions. Public Opin Q 2018;82(3):470-492. [doi: 10.1093/poq/nfy030]
13. Bollen J, Mao H, Pepe A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: ICWSM 2011; 17-21 July, 2011; Barcelona,Spain p. 450-453.
14. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
15. Frank M, Mitchell L, Dodds P, Danforth C. Happiness and the patterns of life: a study of geolocated tweets. Sci Rep 2013;3:2625 [FREE Full text] [doi: 10.1038/srep02625] [Medline: 24026340]
16. Golder S, Macy M. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: 10.1126/science.1202775] [Medline: 21960633]
17. Coppersmith G, Dredze M, Harman C, Hollingshead K. Association for Computational Linguistics. 2015. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses URL: https://www.aclweb.org/anthology/W15-1201 [accessed 2019-04-17] [WebCite Cache ID 77iU1VNnQ]
18. Gore R, Diallo S, Padilla J. You are what you Tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]

19. Super DE. A life-span, life-space approach to career development. J Vocat Behav 1980;16(3):282-298. [doi: 10.1016/0001-8791(80)90056-1]
20. Stone AA, Schneider S, Harter JK. Day-of-week mood patterns in the United States: on the existence of 'Blue Monday', 'Thank God it's Friday' and weekend effects. J Posit Psychol 2012;7(4):306-314. [doi: 10.1080/17439760.2012.691980]
21. Egloff B, Tausch A, Kohlmann CW, Krohne HW. Relationships between time of day, day of the week, and positive mood: exploring the role of the mood measure. Motiv Emot 1995;19(2):99-110. [doi: 10.1007/BF02250565]
22. Howarth E, Hoffman M. A multidimensional approach to the relationship between mood and weather. Br J Psychol 1984 Feb;75(Pt 1):15-23. [Medline: 6704634]
23. Denissen J, Butalid L, Penke L, van Aken MA. The effects of weather on daily mood: a multilevel approach. Emotion 2008 Oct;8(5):662-667. [doi: 10.1037/a0013497] [Medline: 18837616]
24. Klimstra T, Frijns T, Keijsers L, Denissen JJ, Raaijmakers QA, van Aken MA, et al. Come rain or come shine: individual differences in how weather affects mood. Emotion 2011 Dec;11(6):1495-1499. [doi: 10.1037/a0024649] [Medline: 21842988]
25. Baylis P. Energy Institute at HAAS. 2015. Temperature and Temperament: Evidence from a Billion Tweets URL: https://ei.haas.berkeley.edu/research/papers/WP265.pdf [accessed 2019-04-18] [WebCite Cache ID 77iVouI2S]
26. Berry D, Hansen J. Positive affect, negative affect, and social interaction. J Pers Soc Psychol 1996;71(4):796-809. [doi: 10.1037/0022-3514.71.4.796]
27. Dodds P, Harris K, Kloumann I, Bliss C, Danforth C. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 2011;6(12):e26752 [FREE Full text] [doi: 10.1371/journal.pone.0026752] [Medline: 22163266]
28. O'Connor B, Balasubramanyan R, Routledge B, Smith N. Association for the Advancement of Artificial Intelligence. 2010. From tweets to polls: Linking text sentiment to public opinion time series URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1536/1842 [accessed 2019-04-18] [WebCite Cache ID 77iVzqWLO]
29. Larsen M, Boonstra T, Batterham P, O'Dea B, Paris C, Christensen H. We feel: mapping emotion on Twitter. IEEE J Biomed Health Inform 2015;19(4):1246-1252. [doi: 10.1109/JBHI.2015.2403839]
30. Mitchell L, Frank M, Harris K, Dodds P, Danforth C. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PLoS One 2013;8(5):e64417 [FREE Full text] [doi: 10.1371/journal.pone.0064417] [Medline: 23734200]
31. Park K, Lee S, Kim E, Park M, Park J, Cha M. Association for the Advancement of Artificial Intelligence. 2013. Mood and weather: Feeling the heat? URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6068/6330 [accessed 2019-04-18] [WebCite Cache ID 77iWC82Tl]
32. Hannak A, Anderson E, Barrett L, Lehmann S, Mislove A, Riedewald M. Association for the Advancement of Artificial Intelligence. 2012. Tweetin' in the Rain: Exploring Societal-scale Effects of Weather on Mood URL: http://www.ccs.neu.edu/home/amislove/publications/Weather-ICWSM.pdf [accessed 2019-04-18] [WebCite Cache ID 77iWHGNe6]
33. Li J, Wang X, Hovy E. What a Nasty Day: Exploring Mood-Weather Relationship from Twitter. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. 2014 Presented at: CIKM '14; November 03-07, 2014; Shanghai, China p. 1309-1318. [doi: 10.1145/2661829.2662090]
34. Padilla J, Kavak H, Lynch C, Gore R, Diallo S. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857. [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
35. Giachanou A, Crestani F. Like it or not: a survey of Twitter sentiment analysis methods. ACM Comput Surv 2016;49(2):-. [doi: 10.1145/2938640]
36. Mahmud J, Nichols J, Drews C. Home location identification of twitter users. ACM Trans Intell Syst Technol 2014;5(3):-. [doi: 10.1145/2528548]
37. Rahimi A, Cohn T, Baldwin T. Association for Computational Linguistics. 2015. Twitter User Geolocation Using a Unified Text and Network Prediction Model URL: https://www.aclweb.org/anthology/P15-2104 [accessed 2019-04-18] [WebCite Cache ID 77iWeiIQq]
38. Open Weather Map. Weather API URL: https://openweathermap.org/api [accessed 2018-11-22] [WebCite Cache ID 746qWakXy]
39. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl Based Syst 2015;89:14-46. [doi: 10.1016/j.knosys.2015.06.015]
40. Ribeiro F, Araújo M, Gonçalves P, Gonçalves M, Benevenuto F. Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci 2016;5:23. [doi: 10.1140/epjds/s13688-016-0085-1]
41. Reagan A, Danforth C, Tivnan B, Williams J, Dodds P. Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. EPJ Data Sci 2017;6:28. [doi: 10.1140/epjds/s13688-017-0121-9]
42. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. J Assoc Inf Sci Technol 2010;61(12):2544-2558.
43. Gonçalves P, Araújo M, Benevenuto F, Cha M. Comparing and combining sentiment analysis methods. In: Proceedings of the first ACM conference on Online social networks. 2013 Presented at: COSN '13; October 07-08, 2013; Boston, Massachusetts, USA p. 27-38.

44. Thelwall M, Buckley K, Paltoglou G. Sentiment in Twitter events. J Assoc Inf Sci Technol 2011;62(2):406-418.
45. Thelwall M. University of Wolverhampton. 2014. Sentiment analysis and time series with Twitter URL: http://mozdeh.wlv.ac.uk/resources/TwitterTimeSeriesAndSentimentAnalysis.pdf [accessed 2019-04-18] [WebCite Cache ID 77iXpW733]
46. Alves A, de Souza BC, Firmino A, de Oliveira MG, de Paiva AC. A spatial and temporal sentiment analysis approach applied to Twitter microtexts. JIDM 2016;6(2):118-129.
47. Balog K, Mishne G, De Rijke M. Why are they excited?: identifying and explaining spikes in blog mood levels. In: Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations. 2006 Presented at: EACL '06; April 05-06, 2006; Trento, Italy p. 207-210.
48. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J Comput Sci 2011;2(1):1-8. [doi: 10.1016/j.jocs.2010.12.007]
49. Antweiler W, Frank M. Is all that talk just noise? The information content of internet stock message boards. J Finance 2004;59(3):1259-1294.
50. Adam DI. An unobtrusive behavioral model of "gross national happiness". In: Proceedings of the SIGCHI conference on human factors in computing systems. 2010 Presented at: CHI '10; April 10-15, 2010; Atlanta, Georgia, USA.
51. Eric G, Karrie K. Widespread Worry and the Stock Market. 2010 Presented at: Proceedings of the International Conference on Weblogs and Social; 2010; Washington, USA.
52. Diener E, Emmons R. The independence of positive and negative affect. J Pers Soc Psychol 1984 Nov;47(5):1105-1117. [Medline: 6520704]
53. Clark L, Watson D. Mood and the mundane: relations between daily life events and self-reported mood. J Pers Soc Psychol 1988 Feb;54(2):296-308. [Medline: 3346815]
54. Bollen J, Gonçalves B, van de Leemput I, Ruan G. The happiness paradox: your friends are happier than you. EPJ Data Sci 2017;6(1):4. [doi: 10.1140/epjds/s13688-017-0100-1]
55. Burnap P, Williams M. Us and them: identifying cyber hate on Twitter across multiple protected characteristics. EPJ Data Sci 2016;5:11.
56. An J, Quercia D, Cha M, Gummadi K, Crowcroft J. Sharing political news: the balancing act of intimacy and socialization in selective exposure. EPJ Data Sci 2014;3:12. [doi: 10.1140/epjds/s13688-014-0012-2]
57. Salathé M, Vu D, Khandelwal S, Hunter D. The dynamics of health behavior sentiments on a large online social network. EPJ Data Sci 2013;2:4. [doi: 10.1140/epjds16]
58. Volkova S, Charles L, Harrison J, Corley C. Uncovering the relationships between military community health and affects expressed in social media. EPJ Data Sci 2017;6:9. [doi: 10.1140/epjds/s13688-017-0102-z]
59. Dunn AM, Coiera E. Social media interventions for precision public health: promises and risks. NPJ Digit Med 2018;1:-. [doi: 10.1038/s41746-018-0054-0]
60. Sloan L, Morgan J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of Geoservices and Geotagging on Twitter. PLoS One 2015;10(11):e0142209 [FREE Full text] [doi: 10.1371/journal.pone.0142209] [Medline: 26544601]
61. Sloan L, Morgan J, Burnap P, Williams M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. PloS one 2015;10(3):e0115545. [doi: 10.1371/journal.pone.0115545]
62. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the Demographics of Twitter Users. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: AAAI Press; 2011; Barcelona, Spain p. 554-557.
63. Sadah A, Shahbazi M, Wiley T, Hristidis V. A study of the demographics of web-based health-related social media users. J Med Internet Res 2015 Aug 06;17(8):e194 [FREE Full text] [doi: 10.2196/jmir.4308] [Medline: 26250986]
64. Malik M, Lamba H, Nakos C, Pfeffer J. Population bias in Geotagged tweets. 2015 Presented at: Ninth International AAAI Conference on Web and Social Media; May 26–29, 2015; Oxford, England p. 531.
65. Zhang Y, Szabo C, Sheng Q. Sense and focus: towards effective location inference and event detection on Twitter. 2015 Presented at: International Conference on Web Information Systems Engineering; 2015; Miami, FL, USA p. 463-477. [doi: 10.1007/978-3-319-26190-4_31]
66. Compton R, Jurgens D, Allen D. Geotagging one hundred million twitter accounts with total variation minimization. 2014 Presented at: 2014 IEEE International Conference on Big Data (Big Data); October 27-30, 2014; Washington, USA p. 393-401. [doi: 10.1109/BigData.2014.7004256]
67. Jurgens D. HRL Laboratories. 2013. That's What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships URL: http://jurgens.people.si.umich.edu/docs/icwsm-2013-slides.pdf [accessed 2019-04-18] [WebCite Cache ID 77iZOwXyC]
68. Ajao O, Hong J, Liu W. A survey of location inference techniques on Twitter. J Inf Sci 2015;41(6):855-864. [doi: 10.1177/0165551515602847]
69. Koto F, Adriani M. A comparative study on twitter sentiment analysis: Which features are good? 2015 Presented at: International Conference on Applications of Natural Language to Information System; 2015; Salford, United Kingdom p. 453-457. [doi: 10.1007/978-3-319-19581-0_46]

70. Saif H, Fernandez M, He Y, Alani H. Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold. 2013 Presented at: 1st Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013); December 3, 2013; Turin, Italy.
71. Gonçalves P, Dalip D, Costa H, Gonçalves M, Benevenuto F. On the combination of "off-the-shelf" sentiment analysis methods. In: Proceedings of the 31st Annual ACM Symposium on Applied Computing. 2016 Presented at: SAC '16; April 04-08, 2016; Pisa, Italy p. 1158-1165.
72. Yan Y, Yang H, Wang HM. Two simple and effective ensemble classifiers for Twitter sentiment analysis. 2017 Presented at: 2017 Computing Conference; July 18-20, 2017; London, UK p. 1386-1393. [doi: 10.1109/SAI.2017.8252275]
73. Atefeh F, Khreich W. A survey of techniques for event detection in twitter. Comput Intell 2015;31(1):132-164. [doi: 10.1111/coin.12017]
74. Weng J, Lee BS. Event detection in Twitter. In: Proceedings of the 2nd International Workshop on Social Computing. 2011 Presented at: IWSC '18; 2011; Barcelona, Spain p. 401-408.

Abbreviations

API: application programming interface
JSON: JavaScript Object Notation

Edited by G Eysenbach; submitted 29.11.18; peer-reviewed by R Gore, J Du, T Boonstra; comments to author 24.01.19; revised version received 19.03.19; accepted 29.03.19; published 08.05.19.

Please cite as:
Shah Z, Martin P, Coiera E, Mandl KD, Dunn AG
Modeling Spatiotemporal Factors Associated With Sentiment on Twitter: Synthesis and Suggestions for Improving the Identification of Localized Deviations
J Med Internet Res 2019;21(5):e12881
URL: https://www.jmir.org/2019/5/e12881/
doi:10.2196/12881
PMID:31344669

Original Paper

Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus

Ashlynn R Daughton[1,2], MPH; Michael J Paul[2], PhD

[1]Analytics, Intelligence, and Technology, Los Alamos National Laboratory, Los Alamos, NM, United States
[2]Information Science, University of Colorado, Boulder, Boulder, CO, United States

Corresponding Author:
Ashlynn R Daughton, MPH
Information Science
University of Colorado, Boulder
Technology Learning Center
1045 18th Street, UCB 315
Boulder, CO, 80309
United States
Phone: 1 303 735 7581
Email: adaughton@lanl.gov

Abstract

Background: An estimated 3.9 billion individuals live in a location endemic for common mosquito-borne diseases. The emergence of Zika virus in South America in 2015 marked the largest known Zika outbreak and caused hundreds of thousands of infections. Internet data have shown promise in identifying human behaviors relevant for tracking and understanding other diseases.

Objective: Using Twitter posts regarding the 2015-16 Zika virus outbreak, we sought to identify and describe considerations and self-disclosures of a specific behavior change relevant to the spread of disease—travel cancellation. If this type of behavior is identifiable in Twitter, this approach may provide an additional source of data for disease modeling.

Methods: We combined keyword filtering and machine learning classification to identify first-person reactions to Zika in 29,386 English-language tweets in the context of travel, including considerations and reports of travel cancellation. We further explored demographic, network, and linguistic characteristics of users who change their behavior compared with control groups.

Results: We found differences in the demographics, social networks, and linguistic patterns of 1567 individuals identified as changing or considering changing travel behavior in response to Zika as compared with a control sample of Twitter users. We found significant differences between geographic areas in the United States, significantly more discussion by women than men, and some evidence of differences in levels of exposure to Zika-related information.

Conclusions: Our findings have implications for informing the ways in which public health organizations communicate with the public on social media, and the findings contribute to our understanding of the ways in which the public perceives and acts on risks of emerging infectious diseases.

(J Med Internet Res 2019;21(5):e13090) doi:10.2196/13090

KEYWORDS

social media; travel; behavior; communicable diseases; zika virus; public health; epidemiology; information science; travel-related illness

Introduction

Social Media for Public Health

Internet data, including data from social media platforms such as Twitter, have been used extensively in recent years to study health patterns and better understand infectious disease outbreaks [1]. Although it is known that social media usage is demographically biased [2], these data are thought to be fundamentally changing health care [3]. Social media data have been studied to provide insights into public health discourse [4,5] and concerns [6,7].

A particularly successful area of research has used internet data to improve the forecasting of disease outbreaks. Several studies have found that these data, when combined with traditional sources of epidemiological data, can improve the surveillance and forecasting of seasonal diseases such as flu [8-12] and mosquito-borne diseases such as dengue [13,14] and West Nile [15].

In this study, we have considered disease epidemics from the perspective of human *behaviors* that can affect a disease outbreak. We studied the recent outbreak of Zika virus, a mosquito-borne virus that has recently been linked to birth defects and other disorders, as a domain for studying disease-relevant behavior. We have focused on a specific behavior, decisions to change travel to avoid areas affected by Zika, because of the extensive literature that travel contributes significantly to infectious disease emergence [16,17]. We have used a combination of content analysis and supervised machine learning techniques to understand first-person accounts of travel-related decisions during the Zika outbreak. This study aimed to answer the following research questions (RQ):

1. *RQ1:* Can we identify individuals who report they changed their travel behavior in response to concerns about Zika?
2. *RQ2:* What are the characteristics of Twitter users who change or consider changing their travel behavior? In particular, we wished to know:

- *2(a):* Are there temporal, geospatial, or gender-based patterns in users who change their behavior?
- *2(b):* Are there linguistic differences in messages posted by these individuals compared with users selected at random from Twitter?
- *2(c):* Are these individuals exposed to more information about Zika on Twitter?

We have answered these questions by analyzing a collection of 29,386 English-language tweets filtered for keywords describing Zika and travel. We used a cascade of 3 machine learning classifiers to identify behavior mentions in tweets, and we have proposed a method of incorporating classifier error into our statistical analyses to test our hypotheses.

Zika Emergence

Mosquito-borne infections have long been known to cause large outbreaks that result in substantial morbidity and mortality. An estimated 3.9 billion individuals live in a location endemic for common mosquito-borne diseases, for example, dengue, chikungunya, and now, Zika [18]. Although Zika emerged only recently in Central, South, and North America, the virus was originally discovered in 1947 in Uganda [19]. Through the 20th century, documented outbreaks were rare. The first outbreaks occurred in 2007 in Gabon and the Federated States of Micronesia [19]. Furthermore, 6 years later, French Polynesia experienced the first large outbreak, and there was a documented association between neurological symptoms and Zika [19]. The subsequent emergence of Zika in South America in 2015 marked the largest known Zika outbreak and caused hundreds of thousands of infections [19-21]. Between 2015 and 2017, the Pan American Health Organization (PAHO) reported over half a million suspected Zika cases in South and Central America [22].

For the overwhelming majority, Zika is a mild infection; the majority of cases are asymptomatic [23]. However, for some, Zika infection can lead to more serious complications, including the neurological syndrome, Guillain-Barré [24], and birth defects in fetuses infected in-utero [25].

Importantly, these causal relationships have only been recently established. In October 2015, Brazil reported an association between Zika cases and microcephaly, a condition where an infant’s head circumference is extremely small and is accompanied by severe developmental and health complications [26], and others noted a possible association with Guillain-Barré syndrome in adults [24]. As evidence mounted that there was a causal relationship, the World Health Organization (WHO) and PAHO issued alerts in December 2015 about the association between Zika, neurological syndromes, and birth defects. The United States responded to these alerts in mid-January 2016 by issuing a travel advisory for pregnant women, which cautioned against traveling to locations with local Zika transmission [21]. Zika was then declared a public health emergency by the WHO in February 2016 [21].

Travel and Infectious Disease

Travel advisories are an important public health intervention because of the documented impact of travel on the emergence of infectious diseases [16,17]. Historical case studies describe imported cases of diseases that led to large outbreaks as early as the 1500s [16]. In the present day, there are many outbreaks that have been attributed to travel from other regions. For example, genetic data from the 2009 H1N1 outbreak show that the movement of swine around Mexico was responsible for outbreaks in various provinces [27]. Genetic evidence further indicates that H1N1 was probably introduced to the United States from both Mexico and Asia [27].

Simulations find that the impact of travel on disease spread varies based on a number of factors. For example, Bajardi et al found that travel restrictions could reduce cases but probably only minimally [28]. However, research done by Huizer et al finds that air travel could have dramatically changed the 1968 pandemic influenza in Hong Kong [29]. In general, travel is thought to play an important but variable role in disease transmission. Current recommendations are to implement travel-related control measures as necessary [30,31].

Social Media and Zika

Internet data have been used to better understand individual health behaviors and health discourse on the Web. Studies have found evidence that users publicly discuss a variety of ailments [4], as well as particular behaviors used to prevent ailments. For example, Signorini et al observed discussions of behaviors such as hand washing and wearing masks to prevent the flu [10]. Paul and Dredze similarly note that individuals often report medications used for symptom relief [4].

As the largest known Zika outbreak occurred recently, researchers are only now beginning to investigate the use of internet data to understand this particular disease. McGough et al used an autoregressive modeling approach to combine epidemiological data from PAHO, Twitter, Google search queries, and reports from HealthMap to build short-term forecasts for several Central and South American countries.

They found that the lowest error models were produced when using Google search query volumes [32].

Others have found important information in Twitter data. Stefanidis et al used tweets from the first 3 months of the outbreak to characterize discourse around Zika [33]. These data were used to look at the emergence of spatial clusters in online discussions of Zika on Twitter and to identify distinct geospatial communities that participated in the conversation early in the outbreak. In particular, they found that Twitter users tended to use public health organizations to find information and did not generally use Twitter as a way to interact with organizations directly [33]. Using data encompassing more of the outbreak, Miller et al used Twitter to identify tweets about treatment, transmission, and prevention of Zika and noted the use of Twitter as a way to monitor concerns in the general population [34].

Sharma et al investigated information dispersion on Facebook and specifically noted that inaccurate or misleading posts were more popular than those with scientifically sound information [35]. This observation is consistent with previous work which identified rumors and health misinformation on Twitter [36]. Similarly, Gui et al noted that even official sources of information were unreliable during the outbreak because of incomplete information and observed that the internet provided spaces that allowed individuals to frame risk and decisions [37].

Seltzer et al used Instagram to look at image-sharing practices around Zika [38]. They found that health-related images related to Zika were predominately about transmission and prevention and suggested that Instagram could be used to track sentiment with regard to Zika [38].

Motivation and Contributions

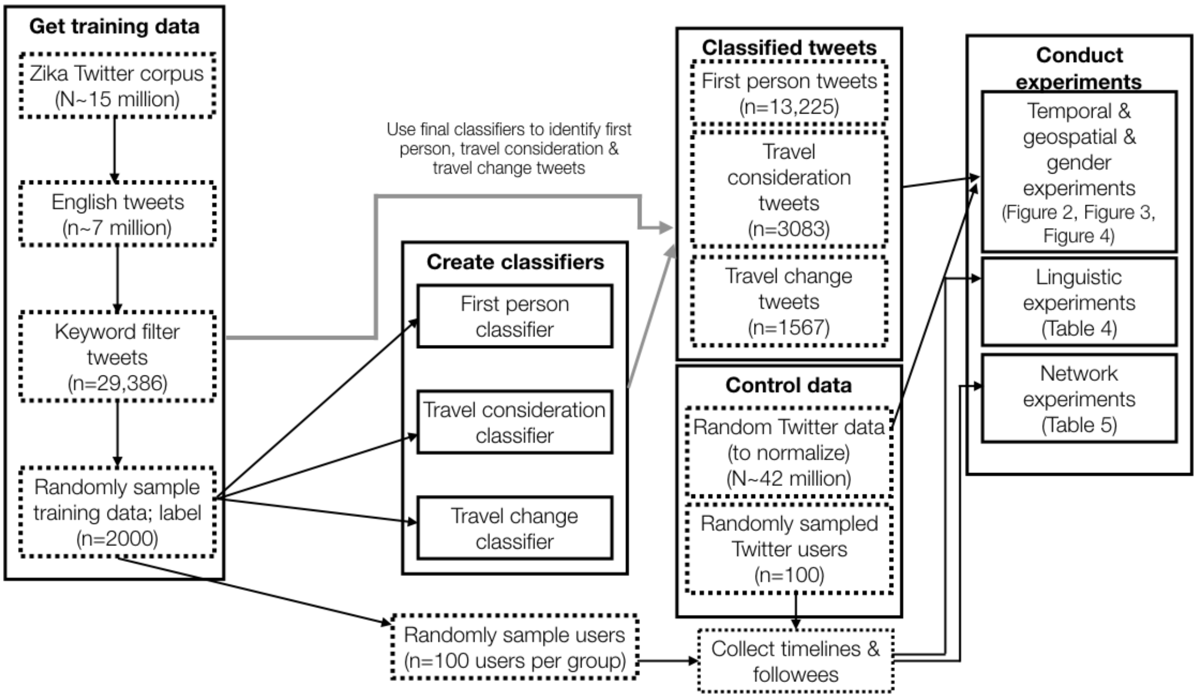
Zika is likely to continue to be an emerging illness of concern with considerable impacts in South, Central, and North America [39]. In contrast to previous work that has focused on the discourse on different platforms or the possible utility of various internet data sources for modeling forecasting, this study focused on identifying a particular behavior of impact on the spread of the disease—travel change. As a first step, this study aimed to identify individuals on Twitter who chose to change their behaviors (RQ1), understand the characteristics of those individuals (RQ2a), and test explanations for any patterns observed (RQ2b and 2c).

Human behaviors directly impact disease transmission [40,41]. Previous work has found that travel patterns are important for transmission but difficult to quantify because of a general lack of available data [41]. A long-term goal of this study is to incorporate behavior change data into disease-forecasting models. This initial study focused on the important first step of identifying travel behaviors and characterizing the factors that affect them.

Methods

This section describes the process used to identify relevant tweets and the techniques used to train and tune the classifiers. We then provide details on the collection of the Twitter timeline and followee data used in later analyses. Figure 1 summarizes the various datasets and methods.

Figure 1. Data processing and experimental overview. Dotted boxes show datasets and corresponding sizes where applicable. Solid boxes show methods used and reference relevant text figures or tables. Black arrows show the flow of data through the pipeline. The gray arrows denote that the final classifiers were used to identify first person, travel consideration, and travel change tweets from the keyword filtered tweets.



Identifying Relevant Tweets

Our data come from a set of 15 million Zika-related tweets from March 1, 2015, to October 31, 2016, with about 7 million in English, described in Daughton et al [42]. This collection includes all tweets mentioning *Zika* and related terms during this time period.

Qualitatively, we observed that the bulk of these Zika tweets were sharing news or other information, usually with links to external articles. However, we also observed a number of English-language tweets describing personal or shared experiences with Zika, including behavior changes in response to concerns about Zika (eg, changing travel plans or buying a mosquito repellent). This section describes our approach to identifying such personal mentions of travel-related behavior through a pipeline of keyword filtering and supervised machine learning.

Keyword Filtering

As personal mentions of travel behavior are a very small proportion of the dataset, we first filtered the dataset to provide a subset with a higher fraction of relevant tweets. This is a standard approach in many social media applications to obtain a large enough fraction of relevant instances to build a reasonably balanced training set [43-47]. In our study, tweets were filtered for those that contained (1) at least one of a set of first-person pronouns to target personal behaviors and (2) at least one of a large set of travel-related terms (see Multimedia Appendix 1).

To be as comprehensive as possible when constructing the list of travel-related terms, we included all major airlines in the United States and all airlines with flights to South America [48,49]. Twitter handles of the airline, cruise, and travel agency companies, including official Twitter handles as well as handles used for negative feedback, were included. These were manually curated by searching for the company on Twitter and identifying associated handles.

After filtering and excluding retweets, 29,386 English-language tweets matched these criteria.

Classification

After keyword filtering, we still observed a variety of tweet topics in the data. This included mentions of changes in travel, opinions about the Olympics (which were hosted in Brazil during the outbreak), opinions about quarantining travelers, and general worry about Zika. The filters also captured tweets that were neither first person nor about travel, such as the headline, *Spraying Mosquitoes by Plane Ain't Perfect, But It's the Best We've Got for Zika - WIRED*.

To further filter the dataset to tweets of relevance to this study—tweets in which people express that they are personally changing or thinking about changing their travel behavior—we constructed 3 binary classifiers:

1. *First person:* Tweets where someone makes their own comment related to Zika in contrast to sharing external content. This can include jokes, opinions, observations, and questions. This category does not include headlines, promotion or solicitation for articles or events, or generic requests for congresspersons to fund Zika.
2. *Travel consideration:* First-person tweets that are about the tweeter's travel plans. This can include tweets that explicitly express the desire to change or not change travel, as well as tweets that are concerned but undecided about travel change.
3. *Travel change:* Travel consideration tweets that explicitly indicate that the tweeter has changed travel plans or is actively trying to change their travel plans. We also attempted to categorize tweets that explicitly said the user would not change travel, but we were unable to build a reliable classifier (F1=.35) and, therefore, excluded it from this study. Messages such as *I want a refund for my trip* would be labeled as travel change whereas messages such as *I'm interested in your refund policy* would not.

Each category only applies to tweets positively labeled with the previous category—travel consideration tweets must also be first-person tweets, and travel change tweets must also be travel consideration tweets.

Annotation

To create a training set for learning supervised classifiers, we randomly sampled 2000 English-language tweets from the keyword-filtered dataset and annotated them with the 3 categories above. Furthermore, 2 researchers independently annotated all tweets to measure agreement. As tweets were only labeled for travel consideration and travel change when they were labeled with the previous category, we only calculated agreement for these categories when annotators also agreed on the previous category. This can be interpreted as measuring: in the cases where annotators agreed on first person, what was their agreement on travel consideration?

Examples of each category, frequency, and agreement are shown in Table 1. To create the final set of labeled data, the 2 annotators discussed the disagreements and updated category criteria to resolve disagreements. For example, annotators disagreed on whether promotion or solicitation of articles or events, as well as requests for congresspersons to fund Zika should be in the first-person category. After discussion, we clarified the criteria to exclude those types of tweets. Using these updated criteria, disagreements were resolved, and the final labels were selected.

Table 1. Label frequency (%), annotator agreement (Cohen's κ), and example tweets for each classification category.

Category	Example (paraphrased)	% (n/N)	κ
First person	When Zika explodes after the Olympics, I'm going to say I told you so!	41.15% (823/2000)	.52
Travel consideration	Thinking about going to Rio for honeymoon. Will I be safe with Zika?	17.5% (350/2000)	.76
Travel change	So mad I had to cancel my island babymoon because of Zika	10.8% (216/2000)	.66

Training and Evaluation

All classifiers were binary logistic regression classifiers built using the Python package scikit-learn (version 0.19.1) [50], where the 3 classifiers were used in a pipeline. Binary logistic regression is an attractive method because outputs are easily interpretable and can be easily tuned for optimal precision and recall. Furthermore, this is a common method used in other health surveillance work [9,51]. Twenty percent (400/2000) of the initial dataset was reserved for testing. This is a standard method used in machine learning to avoid overfitting models [52]. On the training data, we used a grid search to learn the best regularization parameter and feature set, using 5-fold cross-validation to measure the validation performance. For all classifiers, we tested features that included 1-, 2-, and 3-grams. Unigrams (1-grams) consistently outperformed longer n-grams or combinations of n-grams. We also experimented with feature selection using a chi-square test in an attempt to improve classifier metrics [53]. The best results were obtained when all features were used (first person and travel consideration) and when the top 70% of features were used (travel change; see Multimedia Appendix 2). Tweets were preprocessed to remove emojis, punctuation, and consecutive identical characters (eg, vowel elongation) and to replace URLs and usernames with generic tokens.

Performance results on the held-out test data are shown in Table 2. Note that the F1 values shown here differ from those shown in Multimedia Appendix 2 because Multimedia Appendix 2 was generated using cross-validation on the training data, whereas the final metrics were generated using the testing data. We observed that the travel consideration classifier performs the weakest. We also compared the pipeline approach with stand-alone travel consideration and travel change classifiers. However, this method resulted in significantly worse F1 scores (.63 and .65, respectively), and thus, we proceeded with a pipelined approach. The next subsection describes how we account for the cascade of classifier errors in our statistical analyses.

Precision is a measurement of type I error and describes the number of selected items that are actually relevant (percent of those classified positive that are actually positive). Recall, related to type II error, instead describes how many relevant items are selected (percent of positive instances in the full dataset that are classified positive). F1 then combines these 2 metrics, using a harmonic mean, to describe the system overall. We show both F1 using the pipelined approach (the final classifier) as well as the F1 score if each classifier is built independently (see Table 2).

Table 2. Final precision, recall, and F1 of the 3 classifiers.

Classifier	Precision	Recall	F1	F1 (no pipeline)
First person	0.89	0.94	0.92	0.92
Travel consideration	0.61	0.74	0.67	0.63
Travel change	0.66	0.81	0.73	0.65

Statistical Analysis

Our analyses involve measuring the proportion of tweets classified as the various categories along different dimensions. When appropriate, we have provided CIs of these estimates. Our CIs are based on *bootstrap resampling* [54], a nonparametric technique that works as follows. A single bootstrapped estimate of the desired statistic (eg, proportion of tweets) is estimated by resampling the dataset with replacement (bootstrapping) and calculating the statistic from the randomly sampled version of the dataset. This is repeated many times (1000 times in our experiments) to construct a distribution of bootstrapped estimates, and the middle 95% of the estimates are taken as a 95% CI for that statistic [55].

We further modify this approach to account for the uncertainty present in the classifier, using the negative predictive value (NPV) and the positive predictive value (PPV). The NPV is the ratio of true negatives to the sum of true negatives and false negatives whereas the PPV (equivalent to precision in classification) is the ratio of true positives to the sum of true positives and false positives (see [56] for an extensive description of the method). By using this method, we are able to account for the inaccuracies of the individual classifiers and avoid propagating error through the pipeline. We refer to this method as a *weighted bootstrapped CI* in all relevant figures.

Timeline and Followee Collection

Owing to the widespread attention the Zika outbreak received in the media, we wanted to identify if there are other characteristics that differentiate users who changed or considered changing travel as compared with users who tweeted about Zika but did not discuss travel plans.

Using our labeled training data, we collected a set of 100 users sampled at random for each of the 3 classification categories. To construct comparison groups, we also sampled 100 users from the entire set of English-language Zika tweets, as well as 100 English-language users selected at random from all of Twitter. When sampling, we excluded verified users, as the inclusion of celebrities and other prolific accounts could bias the results. We then identified 3 sets of 100 users at random for each classifier. For each group, we collected the Twitter timelines of the users and the list of individuals they follow (their *followees*) using Tweepy [57]. These data were downloaded in January 2018.

Owing to Twitter's application programming interface (API) restrictions on user timelines, we were only able to collect the most recent 3200 tweets for each user. This means that we were not able to collect tweets during the time period of the Zika outbreak especially frequently. This could affect the analyses but will be a close approximation as long as these users have not substantially changed their tweeting behavior since 2016.

Tweets were preprocessed in the same manner as described in the Classification section.

Results

Applying the classifiers to the keyword-filtered tweets resulted in a final dataset of 13,225 first-person tweets, 3083 travel consideration tweets, and 1567 travel change tweets. This section describes the results of our analyses of these tweets and the users who posted these tweets.

Temporal Patterns

Temporal trends in the 3 datasets are shown in Figure 2. Two major spikes corresponding to important events during the outbreak are evident. The first occurred in February 2016 during the time of initial travel advisories by the WHO and the Centers for Disease Control and Prevention [21]. The second, more gradual peak occurs in the summer of 2016 and appears to correspond to the summer Olympics in Rio de Janeiro. We noticed an increase in travel change tweets primarily during the initial set of travel advisories, rather than sustained interest in travel throughout the course of the outbreak.

We also explored temporal differences in the destinations of the users' cancelled travel. To do this, we manually labeled the destinations in all 1567 tweets that were classified in the *travel change* category as international or domestic with regard to the United States. Many tweets were not specific about the location of travel plans; we were able to identify 34% of *travel change* destinations. We found 2 distinct peaks in decisions to change travel (Figure 3). International change spikes sharply in conjunction with the initial travel advisories of February 2016, whereas domestic change spikes sharply in August 2016. The latter spike aligns in time with evidence of local Zika transmission in Florida that was first identified in July 2016 [58] and may also correspond to the increase in cases in US territories such as Puerto Rico [59]. There is an additional peak in the international change tweets in September 2016. These tweets primarily discuss canceling travel to Singapore, which had started to identify local cases in late August 2016 [60]. Although the volume of tweets is small, they show a timely response to the news that Zika had emerged and was circulating locally, within a week of the initial official Ministry of Health report [60].

Figure 2. Temporal trends in classifications by week.

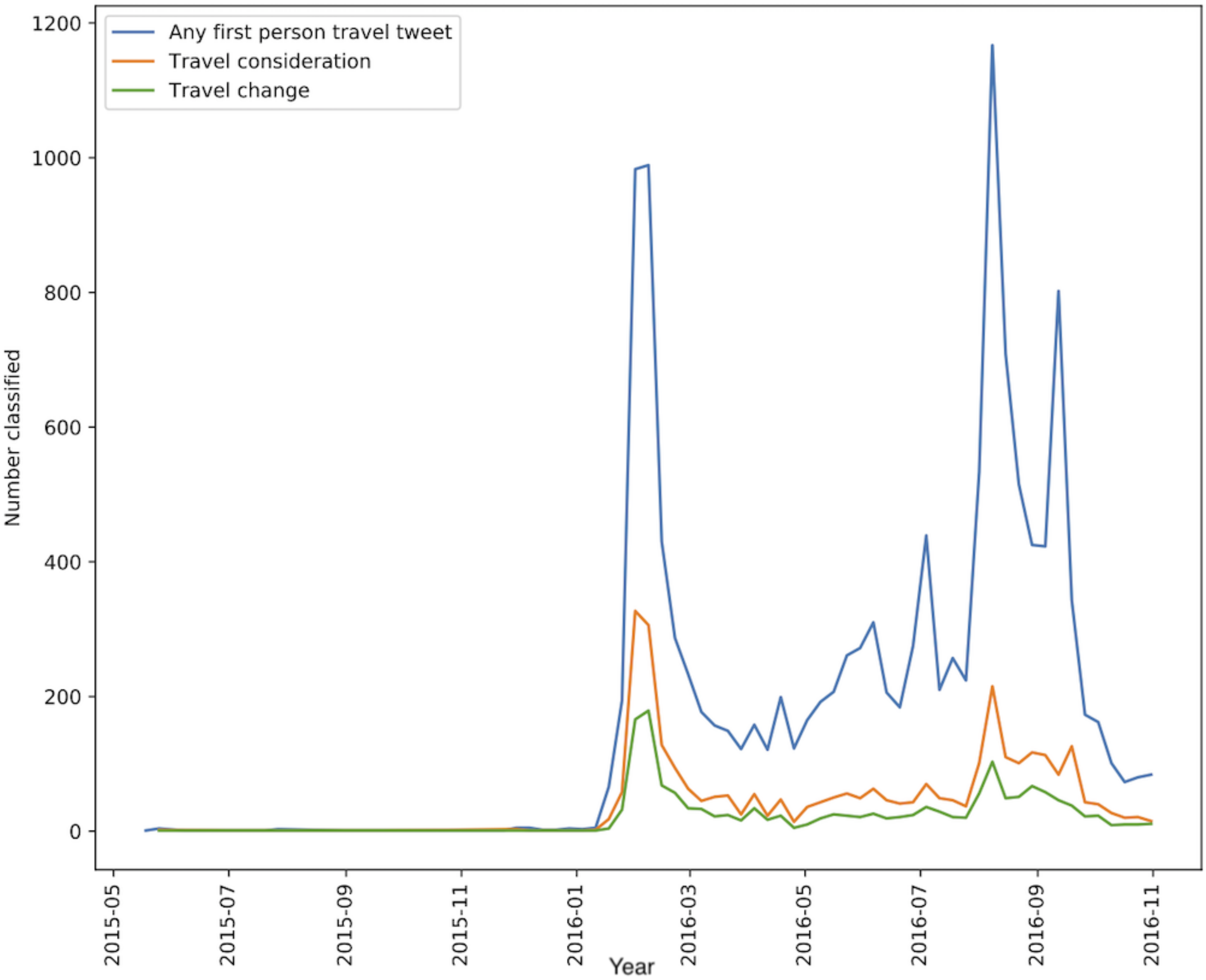
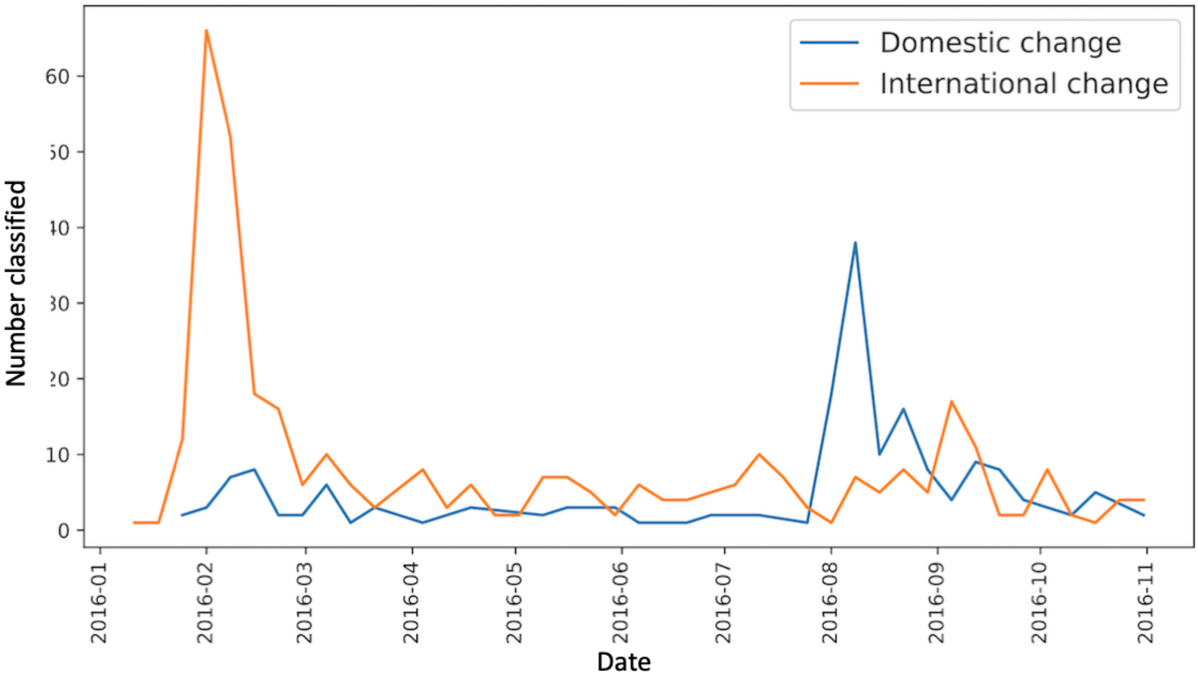


Figure 3. Temporal trends in decisions to change international (outside of the United States) and domestic (within the United States) travel.



Geospatial and Gender Patterns

Geospatial Variability

To evaluate spatial trends, we geolocated tweets using Carmen [61], which resolves tweets to structured locations using geographic coordinates when available and user profile information if not.

We grouped tweets into geographic regions defined by the US Department of Health and Human Services (HHS). HHS Regions are regional groupings of states in the United States that are commonly used to aggregate states for health studies. As the traditional HHS Regions group geographically disparate states together (eg, Hawaii and island territories are grouped with mainland regions), we modified the HHS Regions as follows:

1. R1: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont.
2. R2: New Jersey, New York.
3. R3: Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, West Virginia.
4. R4: Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee.
5. R5: Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin.
6. R6: Arkansas, Louisiana, New Mexico, Oklahoma, Texas.
7. R7: Iowa, Kansas, Missouri, Nebraska.
8. R8: Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming.
9. R9: Arizona, California, Nevada.
10. R10: Alaska, Idaho, Oregon, Washington.
11. Caribbean Islands: Puerto Rico, US Virgin Islands.
12. Pacific Islands: Hawaii, American Samoa, Northern Mariana Islands, Federated States of Micronesia, Guam, Marshall Islands, Republic of Palau.

We ultimately excluded both the Pacific Islands and Caribbean Islands from this analysis because there were not enough tweets classified in these regions (fewer than 50 tweets each).

As tweet volume varies by location, we created a type of per-capita estimate to adjust for the overall popularity of Twitter in each region. We collected a 1% sample of tweets from the Twitter streaming API over approximately 10 nonconsecutive days throughout December 2017 and January 2018 to normalize the estimates (42.1 million tweets). The number of tweets classified from each region was then divided by the total number of tweets from that region in the random sample.

Figure 4 shows a wide variation in the weighted volume of tweets across different spatial regions of the United States. Regions 1, 7, 8, and 10 have the highest relative volume of tweets considering and changing travel plans. These regions predominantly consist of landlocked states in the center of the country and include individuals who would have only been at risk of Zika infection if they traveled to an area with local transmission. Interestingly, regions that included states where Zika transmission occurred (Florida—Region 4 and Texas—Region 6) were among the lowest in weighted volume of tweets. It could be that individuals in these locations were not tweeting about travel change because they were at a more acute risk of infection. It is also possible that more granular (eg, state-level) observations are obscured by aggregation to the HHS level.

Gender Variability

As Zika is primarily a concern for women who are pregnant or trying to become pregnant, we investigated the relative percentage of women tweeting versus men (Figure 5). Gender was inferred using the *Demographer* tool [62], which infers gender of Twitter users with an estimated 94% accuracy based on character n-grams of the persons' names.

Figure 4. Weighted volume of classified tweets by modified US Department of Health and Human Services Region. Bars show median weighted volume. Error bars represent 95% confidence intervals obtained using weighted bootstrapped sampling.

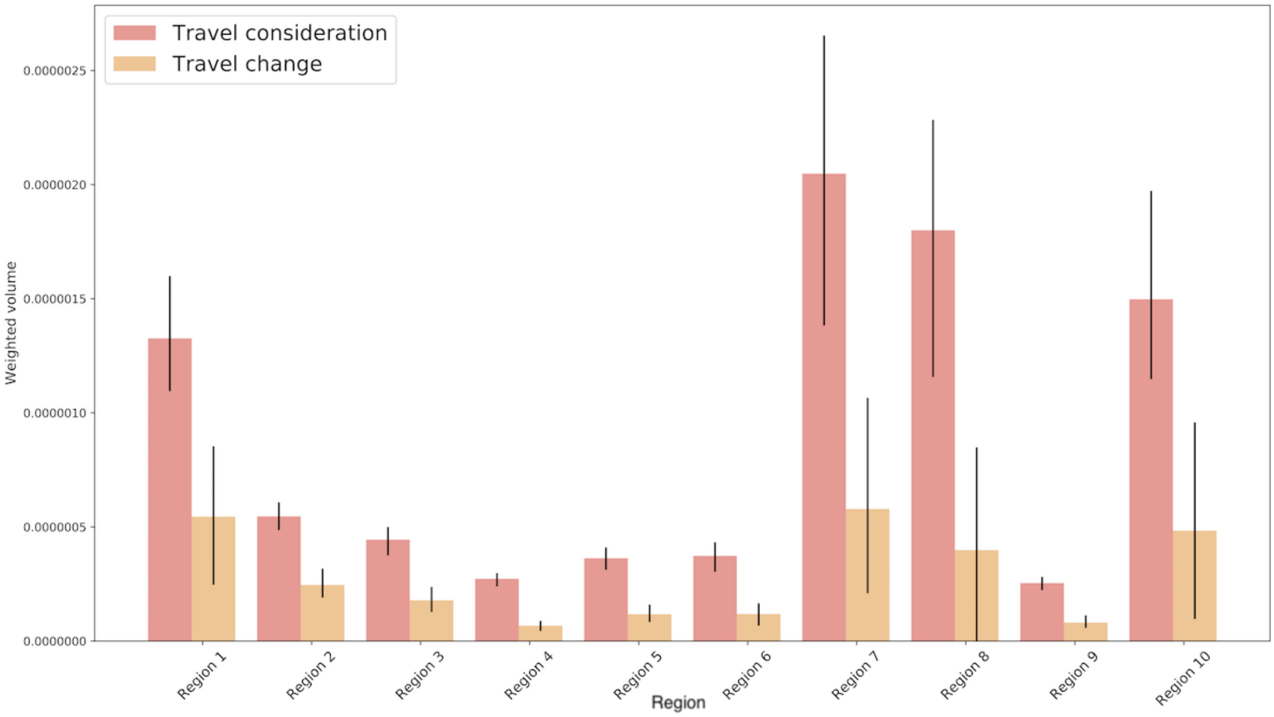
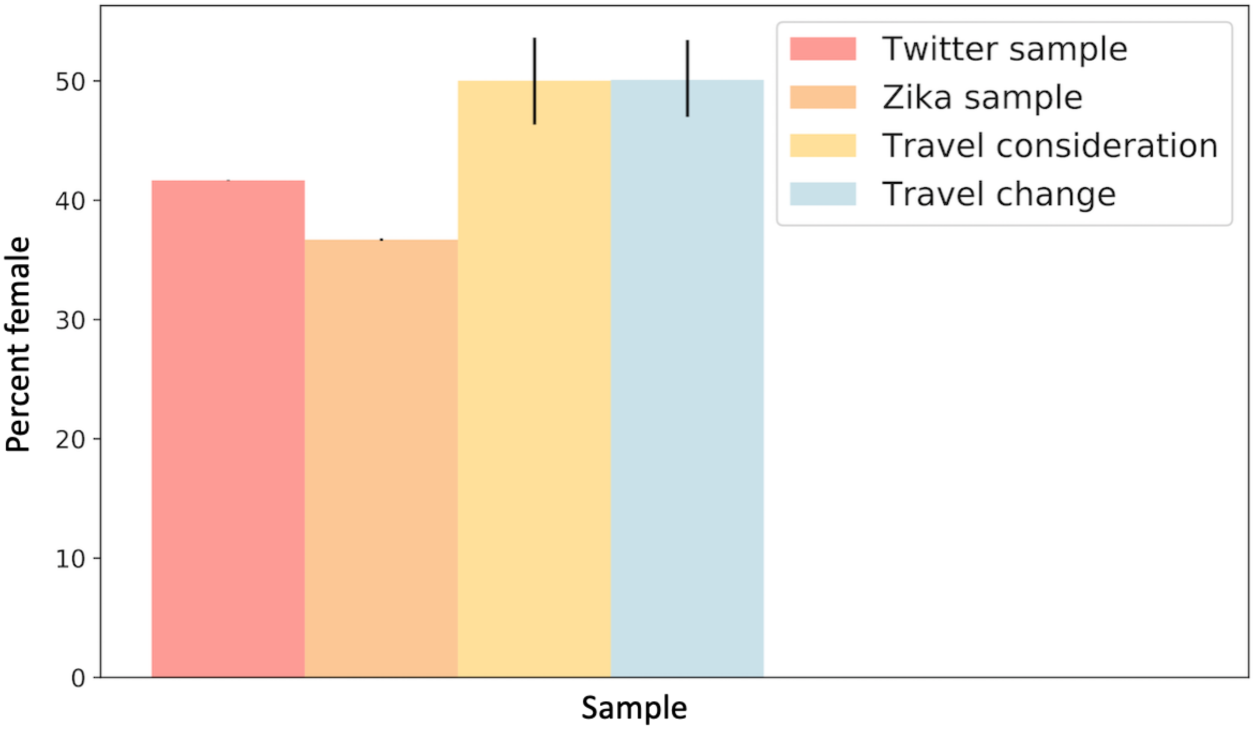


Figure 5. Relative percent of women in a sample of Twitter (red), English Zika dataset (orange), travel consideration dataset (yellow), and the travel change dataset (blue). Bars show 95% weighted bootstrapped confidence intervals.



Linguistic Comparison

To better understand the factors that contribute to a decision to change travel, we compared the style and content of messages between users in the travel consideration and travel change groups with the random sample of Twitter users. We hypothesized that those who discuss Zika travel are more likely to talk about health in general than typical Twitter users and that those who consider changing travel may have higher levels of fear or anxiety.

We used Linguistic Inquiry Word Count (LIWC) [63], which maps various English-language terms to linguistic and psychological constructs. We selected a subset of LIWC categories related to our hypotheses (health and anxiety), as well as categories related to various personal concerns as a way of categorizing other general content of discussion. We also created a category specifically for pregnancy- related terms, using the regular expression *pregnan**, because of the relevance of Zika to a developing pregnancy. Although pregnancy is included in the LIWC biological processes category, that category is much broader than pregnancy specifically.

For each user timeline and each LIWC category, we calculated the percentage of tweets that contain a term from the category. In this calculation, we excluded the tweets mentioning Zika so that the analysis does not reflect the same data used to select users. In addition, we restricted the analysis to timelines with a minimum of 10 tweets across the timeline. Finally, for each category, we calculated the average percentage across all timelines in each user group. The results are shown in Table 3.

Compared with a random sample of Twitter users, users who tweeted about changing or considering changing travel in reaction to Zika are significantly more likely to use past and present tense, as well as terms indicating social processes, perhaps indicating increased planning. Travel consideration users are significantly more likely to use personal pronouns and singular first-person pronouns and were significantly higher in the anxiety category. Travel change users were significantly more likely to use plural first-person pronouns, had higher inhibition, and tweeted more about pregnancy. There are no significant differences between the travel consideration and travel change groups.

Contrary to our expectations, the travel groups do not tweet significantly differently from the overall Twitter population about health or bodily functions. This indicates that the users we identified as part of this behavior change pipeline were uniquely concerned about Zika and did not appear to be generally more aware or interested in discussing health-related topics on social media (with the important exception of pregnancy). It would be useful to explore more on this line of inquiry in future work, as understanding who talks about infectious diseases (and how) is of immediate interest to the disease surveillance community [64].

Table 3. Average percent of Linguistic Inquiry Word Count category prevalence per group.

Type	Category	All Twitter	Consideration	Change
Linguistic processes	Personal pronouns	0.6080	*0.7495* [a]	0.7501
Linguistic processes	1st singular	0.2788	*0.3673* [a]	0.3214
Linguistic processes	1st plural	0.0458	0.0699	*0.0895* [a]
Linguistic processes	3rd singular	0.0692	0.0699	*0.0895* [a]
Linguistic processes	3rd plural	0.0474	0.0561	0.0571
Linguistic processes	Past tense	0.1794	*0.2538* [a]	*0.2665* [a]
Linguistic processes	Future tense	0.0648	0.0842	0.0871
Linguistic processes	Present tense	0.6053	*0.7711* [a]	*0.7947* [a]
Psychological processes	Social processes	0.7181	*0.8867*	*0.9760*
Psychological processes	Affective processes	0.6648	0.7362	0.7587
Psychological processes	Positive emotion	0.4323	0.5106	0.5105
Psychological processes	Negative emotion	0.2290	0.2225	0.2440
Psychological processes	Anxiety	0.0246	*0.0364* [a]	0.0331
Psychological processes	Tentativeness	0.1556	0.2019	0.2075
Psychological processes	Certainty	0.1203	0.1437	0.1375
Psychological processes	Inhibition	0.0470	0.0633	*0.0680* [a]
Psychological processes	Biological processes	0.2230	0.2712	0.2401
Psychological processes	Body	0.0705	0.0787	0.0674
Psychological processes	Health	0.0495	0.0744	0.0734
Psychological processes	Sexual	0.0857	0.0648	0.0526
Other (non- Linguistic Inquiry Word Count)	Pregnancy	0.0004	0.0106	*0.0016* [a]

[a] Instances where there are significant differences from the random sample. Significance is estimated using an unpaired 2-sided *t* test with a significance level of *P*<.05 after Bonferroni correction.

Network Comparison

As a final experiment, we look at the number of followees each of the randomly selected users had that were also present elsewhere in the Zika dataset—that is, the accounts a user follows that had at least one Zika-related tweet. Table 4 shows the number of Zika followees in each group as well as the number of tweets those followees tweeted that were about Zika. We calculate both the raw counts as well as normalized counts in which we divide the number of Zika followees and number of Zika tweets by each user's total number of followees. This allows us to measure both the raw number of Zika tweets an individual could have been exposed to and the relative likelihood of exposure based on the proportion of their feeds that contained Zika content.

Although it is impossible to replicate Twitter's algorithm for showing information on the timeline, we have the unique capability to look at network effects because we have 100% of the tweets during the time period that explicitly mentions either *zika* or *zikv*. We reasoned that individuals who have many followees who appear in the Zika corpus (ie, they follow accounts that are also tweeting about Zika) were more likely to have tweets about Zika appear in their feed. If we were to find that individuals who follow many accounts that appear in the dataset are more likely to appear in the travel change group, we would then further question the role that Twitter plays in catalyzing and informing decisions about behavior change.

Indeed, we did find that those individuals who considered or changed their travel plans had a higher number of followees and tweets that they could have been exposed to in the sample. Although the travel groups had higher counts under every metric when compared with the control group, the difference is only significant under the normalized metrics.

Table 4. The number of followees an individual user has who are also in the dataset, and the number of tweets that followees tweeted that are also in the dataset. We normalized to the number of total followees for each individual. Values in italics are significant (*P*≤.05).

Metric	All Twitter, median (95% CI)	Consideration, median (95% CI)	Change, median (95% CI)
Number of followees (raw)	92.8 (58.3-135.4)	111.6 (71.1-170.9)	122.2 (82.3-177.4)
Number of followees (normalized)	0.08 (0.06-0.11)	*0.15 (0.12-0.17)*	*0.17 (0.14-0.20)*
Number of tweets (raw)	93.6 (56.2-141.2)	111.3 (67.7-169.8)	122.7 (79.6-179.0)
Number of tweets (normalized)	1.71 (1.02-2.62)	*5.7 (3.41-8.74)*	*7.99 (3.47-14.98)*

Discussion

Principal Findings

In an age where infectious diseases are emerging and re-emerging rapidly [65], the ability to identify groups of individuals who might be at increased risk of contracting or contributing to the spread of infection can inform methods of risk communication, infectious disease interventions, and policies at a broader level.

We present supervised classifiers that identify evidence of behavior changes with regard to concerns and changes in travel plans owing to Zika on Twitter. Although previous work has observed that individuals mention protective health behaviors on social media [10], to the best of our knowledge, this is the first work to study a specific behavior change in depth. We examined temporal and demographic patterns in travel behaviors, as well as psycholinguistic markers and information exposure (as approximated through lexical and network analyses, respectively) of individuals changing behavior compared with a randomly sampled control group. More concretely, we considered 4 research questions. Their respective conclusions are discussed below.

RQ1: Can we identify individuals who report they changed travel behavior in response to Zika? We conclude that tweets about changing travel and considering changing travel can be identified with high recall. Furthermore, we are able to account for the comparatively lower precision achieved here using our weighted bootstrap resampling method.

RQ2(a): Are there temporal, geospatial, or gender-based patterns in users who change their behavior? We observed temporal patterns in travel consideration and travel change tweets, including the destination of travel, which correspond to important events in the Zika outbreak. We are encouraged that temporal trends correspond with events that we would expect to be reflected in this data stream.

We additionally find significant differences in the gender distribution of users tweeting about travel consideration and change compared with the general population of Twitter. In particular, we find that the relative proportion of women engaging in conversation indicating travel change behaviors on Twitter is higher than men. This, in combination with the results of RQ2(b) discussed below, is evidence that pregnancy was playing a role in these considerations.

For comparison with existing knowledge on this subject, we discuss 2 small surveys (85 and 121 participants) conducted in New York (NY) [66] and Miami [67] about the knowledge around Zika and travel and included related questions about behaviors. In NY, researchers found that roughly a third of women surveyed were not aware of the travel advisories in place during their travel, almost half were not aware that Zika was being transmitted in the location that they traveled to, and a relatively large number (about one-third) did not know they were pregnant at the time of travel [66]. In Miami, the vast majority of respondents were aware of Zika and reported that they changed their behaviors to avoid the disease; however, only 27% were aware that they were at risk of infection where they lived [67]. Although these survey data exclude men, they do find evidence that women were aware of the disease and that some women (though not all) took measures like changing travel plans to avoid exposure to the disease. Indeed, these surveys highlight the importance of more work in this area to further understand behavior changes over larger spatiotemporal regions.

RQ2(b): Are there linguistic differences in messages posted by these individuals compared with users selected at random from Twitter? We found that users in the travel categories do not appear to tweet more often about health than Twitter users in general. However, travel change users do tweet more often about pregnancy, which suggests this may be a factor in considering travel changes. In addition, travel consideration users tweet words associated with anxiety more than the general population.

RQ2(c): Are individuals who change their behavior exposed to more information about Zika? Travel consideration and change users have a statistically higher fraction of Zika-related followees and tweets in the sample, indicating that these users had greater exposure to Zika-related information. This is evidence that exposure to information about Zika may play a role in this decision-making process.

Limitations

There are several limitations of the data and our methodology that must also be considered. First, it is known that Twitter is a demographically biased data source [68] and as such may not be representative of the broader population. However, this research contributes to the vast literature that uses the Twitter platform to understand health behaviors [4,10,69]. The data are additionally biased in that data only includes tweets in English, which are predominately from the United States. However, we note that data from the United States are appropriate for studying travel behaviors because there was minimal Zika transmission in the United States, as the mosquito vector is absent in the majority of the country. As such, the main method of exposure was through travel to locations with local transmission. We believe our framework could be applied to other behaviors that are only applicable in places with local transmission, such as the use of mosquito repellent, but the classifiers would need to be trained in other languages such as Latin American Spanish.

Second, we recognize the lack of external validity owing to the absence of comparable ground truth data. We view this as a motivation for this research, where findings from this study can be viewed as hypotheses to test with future experiments. It is well known that human behaviors directly impact disease transmission [40,41,70]. For example, Lau et al find that the Severe Acute Respiratory Syndrome epidemic changed individuals' travel patterns [71], and substantial research has shown that beliefs and behaviors about vaccinations dramatically impact disease occurrence [72]. However, travel-related research data are currently sparse [41]. Although we cannot say that the findings from this research are generalizable, the fact that they exist on Twitter is evidence they do exist. As such, these data can be viewed as motivations for larger survey experiments to confirm the findings and to evaluate if Twitter is a viable alternative data stream. Future work could also aim to validate behavior estimates indirectly by verifying their utility in an external task such as disease forecasting.

Third, machine learning classifiers introduce error [73], which could be further amplified by using a pipeline approach. However, we use weighted bootstrap sampling to appropriately account for these errors in downstream analyses. As our results showed significant differences even after accounting for errors, we did not make it a priority to build the best possible classifier in this work, but instead relied on standard tools.

Finally, there are some limitations of our labeled dataset. It is relatively small compared with some previous work. We specifically chose not to scale up the annotation process with crowdsourcing [74] so that the annotations were done by researchers with domain expertise. However, it is possible that a larger training set could lead to better classifier performance. Similarly, our ability to identify statistically significant differences between user groups is limited by having only 100 timelines per group. However, the rate limits of the Twitter API make it difficult to collect large numbers of user timelines. Furthermore, although small sample sizes may affect the power of the analyses, this does not affect the correctness of the approach, which correctly constructs CIs.

In addition, the labeling criteria we used could introduce bias. In particular, we can only capture people who explicitly state that they are canceling travel and that they are doing so because of Zika. Research in this field is limited, but initial work on self-reports of cold and flu illness indicates that it is rare for individuals to tweet about their health concerns [77], and it is currently unknown how this could bias the distribution of labels. However, the experiments presented here do not try to measure overall levels of travel cancellation because of these issues. Instead, we focus on comparisons across groups, which are valid if these data biases are consistent across groups (eg, gender and geography).

Implications

The results of this study show that people do describe first-person behavior changes on Twitter and that such tweets can be classified and analyzed at scale. In particular, we find that our behavior change classifier produces a dataset that corresponds to events during the outbreak and shows evidence of geographic and gender-based differences in the behavior change.

These data support hypotheses that social media can play a role in an individual’s health choices. Other research has shown that an important predictor of population health is knowledge and that this knowledge can be disproportionate across different geographical areas based on access to health care expertise [75]. Research on ways in which social media can facilitate promotion of accurate and important health messages, thus, has clear applications.

Eventually, we envision these types of algorithms being used within the disease surveillance community. There is substantial previous work using internet data to gather traces of information about individuals’ health to monitor and forecast infectious disease outbreaks (eg, search query volumes used for Google Flu Trends). In principle, social media–derived data about behaviors that affect the spread of disease could be incorporated into forecasting models to better describe disease transmission dynamics. As part of this study, we plan to eventually incorporate this type of data into such models.

In addition to monitoring and forecasting, data and conclusions from studies such as this work can inform preventative health messaging. Previous research has found that the ways infectious diseases are framed contribute in important ways to the public perception of the event's severity [76]. Gui et al describe the way in which individuals frame their personal risk from Zika amid uncertain or unclear public health recommendations [37]. They noted that even official sources of information were unreliable during the outbreak because of incomplete information and observed that the internet provided spaces that allowed individuals to frame risk and decisions [37]. We qualitatively observed in our data that there were many instances of individuals who were at low risk of complications resulting from Zika but were highly concerned about their personal risk from Zika. Future work in understanding how individuals frame personal risk from infectious diseases could contribute to our understanding of ways to improve public health risk communication.

Acknowledgments

ARD and MJP conceptualized the work, developed methodology, wrote associated software, performed the analyses and wrote and edited the original manuscript. MJP provided supervision and project administration.

The LANL publication number is LA-UR-18-24423.

Conflicts of Interest

MJP serves on the advisory board to Sickweather, a company that uses social media to forecast illness.

Multimedia Appendix 1

Travel-related keywords used to filter tweets.

[DOCX File, 14KB - jmir_v21i5e13090_app1.docx]

Multimedia Appendix 2

Cross-validated F1 scores for classifiers stratified by n-gram range and percentage of features used (based on chi-square).

[DOCX File, 20KB - jmir_v21i5e13090_app2.docx]

References

1. Paul MJ, Dredze M. Social monitoring for public health. In: Marchionini G, editor. Synthesis Lectures on Information Concepts, Retrieval, and Services. USA: Morgan & Claypool; Aug 31, 2017:1-183.
2. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
3. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368. [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
4. Paul M, Dredze M. You are what you tweet: analyzing Twitter for public health. ICWSM 2011:1-8 [FREE Full text]
5. Bakal G, Kavuluru R. On Quantifying Diffusion of Health Information on Twitter. 2017 Feb Presented at: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); February 16-19, 2017; Orlando, Florida, USA p. 485-488 URL: http://europepmc.org/abstract/MED/28736772 [doi: 10.1109/BHI.2017.7897311]
6. Ji X, Chun S, Geller J. Monitoring Public Health Concerns Using Twitter Sentiment Classifications. In: Proceedings of the 2013 IEEE International Conference on Healthcare Informatics. 2013 Presented at: ICHI'13; September 9-11, 2013; Philadelphia, PA, USA p. 335-344. [doi: 10.1109/ICHI.2013.47]
7. Kanhabua N, Nejdl W. Understanding the Diversity of Tweets in the Time of Outbreaks. In: Proceedings of the 22nd International Conference on World Wide Web. 2013 Presented at: WWW'13 Companion; May 13-17, 2013; Rio de Janeiro, Brazil. [doi: 10.1145/2487788.2488172]
8. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. In: Proceedings of the First Workshop on Social Media Analytics. 2010 Presented at: SOMA'10; July 25-28, 2010; Washington DC, District of Columbia. [doi: 10.1145/1964858.1964874]
9. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014;6:1-12 [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
10. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
11. Lampos V, Cristianini N. Tracking the flu pandemic by monitoring the social web. 2010 Presented at: 2nd International Workshop on Cognitive Information Processing; June 14-16, 2010; Elba Island, Italy. [doi: 10.1109/CIP.2010.5604088]

12. Brownstein JS, Freifeld CC, Chan EH, Keller M, Sonricker AL, Mekaru SR, et al. Information technology and global surveillance of cases of 2009 H1N1 influenza. N Engl J Med 2010 May 6;362(18):1731-1735 [FREE Full text] [doi: 10.1056/NEJMsr1002707] [Medline: 20445186]
13. Generous N, Fairchild G, Deshpande A, del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. PLoS Comput Biol 2014 Nov;10(11):e1003892 [FREE Full text] [doi: 10.1371/journal.pcbi.1003892] [Medline: 25392913]
14. Gomide J, Veloso A, Meira W. Dengue Surveillance Based on a Computational Model of Spatio-temporal Locality of Twitter. In: Proceedings of the 3rd International Web Science Conference. 2011 Presented at: WebSci'11; June 15-17, 2011; Koblenz, Germany.
15. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
16. Wilson M. Travel and the emergence of infectious diseases. Emerg Infect Dis 1995;1(2):39-46 [FREE Full text] [doi: 10.3201/eid0102.952001] [Medline: 8903157]
17. Ni S, Weng W. Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. Phys Rev E Stat Nonlin Soft Matter Phys 2009 Jan;79(1 Pt 2):016111. [doi: 10.1103/PhysRevE.79.016111] [Medline: 19257111]
18. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis 2012;6(8):e1760 [FREE Full text] [doi: 10.1371/journal.pntd.0001760] [Medline: 22880140]
19. Weaver S, Costa F, Garcia-Blanco M, Ko AI, Ribeiro GS, Saade G, et al. Zika virus: history, emergence, biology, and prospects for control. Antiviral Res 2016 Dec;130:69-80 [FREE Full text] [doi: 10.1016/j.antiviral.2016.03.010] [Medline: 26996139]
20. Patterson J, Sammon M, Garg M. Dengue, Zika and Chikungunya: emerging arboviruses in the New World. West J Emerg Med 2016 Nov;17(6):671-679 [FREE Full text] [doi: 10.5811/westjem.2016.9.30904] [Medline: 27833670]
21. Kindhauser M, Allen T, Frank V, Santhana RS, Dye C. Zika: the origin and spread of a mosquito-borne virus. Bull World Health Organ 2016 Sep 1;94(9):675-86C [FREE Full text] [doi: 10.2471/BLT.16.171082] [Medline: 27708473]
22. PAHO. Pan American Health Organization. 2017. Zika Cumulative Cases URL: http://www.webcitation.org/74avyZugu [accessed 2017-11-02] [WebCite Cache ID 74avyZugu]
23. Moghadas SM, Shoukat A, Espindola AL, Pereira RS, Abdirizak F, Laskowski M, et al. Asymptomatic transmission and the dynamics of Zika infection. Sci Rep 2017 Jul 19;7(1):5829 [FREE Full text] [doi: 10.1038/s41598-017-05013-9] [Medline: 28724972]
24. Parra B, Lizarazo J, Jiménez-Arango JA, Zea-Vera AF, González-Manrique G, Vargas J, et al. Guillain-Barré syndromea associated with Zika virus infection in Colombia. N Engl J Med 2016 Dec 20;375(16):1513-1523. [doi: 10.1056/NEJMoa1605564] [Medline: 27705091]
25. Melo AS, Malinger G, Ximenes R, Szejnfeld PO, Sampaio S, de Filippis AM. Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: tip of the iceberg? Ultrasound Obstet Gynecol 2016 Jan;47(1):6-7 [FREE Full text] [doi: 10.1002/uog.15831] [Medline: 26731034]
26. Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, et al. Zika virus associated with microcephaly. N Engl J Med 2016 Mar 10;374(10):951-958. [doi: 10.1056/NEJMoa1600651] [Medline: 26862926]
27. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernández R, Lara-Puente JH, et al. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife 2016 Dec 28;5:e16777 [FREE Full text] [doi: 10.7554/eLife.16777] [Medline: 27350259]
28. Bajardi P, Poletto C, Ramasco JJ, Tizzoni M, Colizza V, Vespignani A. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. PLoS One 2011 Jan 31;6(1):e16591 [FREE Full text] [doi: 10.1371/journal.pone.0016591] [Medline: 21304943]
29. Grais RF, Ellis JH, Glass GE. Assessing the impact of airline travel on the geographic spread of pandemic influenza. Eur J Epidemiol 2003;18(11):1065-1072. [Medline: 14620941]
30. Huizer YL, Swaan CM, Leitmeyer KC, Timen A. Usefulness and applicability of infectious disease control measures in air travel: a review. Travel Med Infect Dis 2015;13(1):19-30. [doi: 10.1016/j.tmaid.2014.11.008] [Medline: 25498904]
31. LaRocque RC, Jentes ES. Health recommendations for international travel: a review of the evidence base of travel medicine. Curr Opin Infect Dis 2011 Oct;24(5):403-409. [doi: 10.1097/QCO.0b013e32834a1aef] [Medline: 21788892]
32. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. PLoS Negl Trop Dis 2017 Jan;11(1):e0005295 [FREE Full text] [doi: 10.1371/journal.pntd.0005295] [Medline: 28085877]
33. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: temporal variations of locations, actors, and concepts. JMIR Public Health Surveill 2017 Apr 20;3(2):e22 [FREE Full text] [doi: 10.2196/publichealth.6925] [Medline: 28428164]
34. Miller M, Banerjee T, Muppalla R, Romine W, Sheth A. What are people tweeting about Zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. JMIR Public Health Surveill 2017 Jun 19;3(2):e38 [FREE Full text] [doi: 10.2196/publichealth.7157] [Medline: 28630032]

35. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 1;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
36. Ghenai A, Mejova Y. Catching Zika Fever: Application of Crowdsourcing Machine Learning for Tracking Health Misinformation on Twitter. 2017 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); August 23-26, 2017; Park City, UT p. 518. [doi: 10.1109/ICHI.2017.58]
37. Gui X, Kou Y, Pine K, Chen Y. Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI'17; May 2017; Denver, CO p. 4520-4533 URL: https://dl.acm.org/citation.cfm?id=3025891 [doi: 10.1145/3025453.3025891]
38. Seltzer EK, Horst-Martz E, Lu M, Merchant RM. Public sentiment and discourse about Zika virus on Instagram. Public Health 2017 Sep;150:170-175. [doi: 10.1016/j.puhe.2017.07.015] [Medline: 28806618]
39. Fauci A, Morens D. Zika virus in the Americas--yet another Arbovirus threat. N Engl J Med 2016 Feb 18;374(7):601-604. [doi: 10.1056/NEJMp1600297] [Medline: 26761185]
40. Moran K, Fairchild G, Generous N, Hickmann K, Osthus D, Priedhorsky R, et al. Epidemic forecasting is messier than weather forecasting: the role of human behavior and internet data streams in epidemic forecast. J Infect Dis 2016 Dec 1;214(suppl_4):S404-S408 [FREE Full text] [doi: 10.1093/infdis/jiw375] [Medline: 28830111]
41. Meloni S, Perra N, Arenas A, Gómez S, Moreno Y, Vespignani A. Modeling human mobility responses to the large-scale spreading of infectious diseases. Sci Rep 2011 Aug 12;1(1):1-7. [doi: 10.1038/srep00062]
42. Daughton A, Pruss D, Arnot B. Characteristics of Zika behavior discourse on Twitter. 2017 Presented at: AMIA Workshop on Social Media Mining for Health Applications; November 2017; Washington, DC URL: https://pdfs.semanticscholar.org/6a65/3d92e43258092a51839d5106e6da904564bd.pdf
43. Rizoiu M, Graham T, Zhang R. DEBATENIGHT: The Role and Influence of Socialbots on Twitter During the First 2016 U.S. Presidential Debate. 2018 Presented at: International AAAI Conference on Web and Social Media; June 2018; Stanford, CA URL: https://arxiv.org/pdf/1802.09808.pdf
44. Saha K, Weber I, De Choudhury M. A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses. 2018 Presented at: International AAAI Conference on Web and Social Media; June 25-28, 2018; Stanford, CA.
45. Olteanu A, Castillo C, Boy J, Varshney K. The Effect of Extremist Violence on Hateful Speech Online. 2018 Presented at: International AAAI Conference on Web and Social Media; June 25-28, 2018; Stanford, CA URL: https://arxiv.org/pdf/1804.05704.pdf
46. Tay Y, Tuan L, Hui S. COUPLENET: Paying Attention to Couples with Coupled Attention for Relationship Recommendation. 2018 Presented at: International AAAI Conference on Web and Social Media; June 25-28, 2018; Stanford, CA.
47. El Sherief M, Kulkarni V, Nguyen D. Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media. 2018 Presented at: International AAAI Conference on Web and Social Media; June 25-28, 2018; Stanford, CA URL: https://arxiv.org/pdf/1804.04257.pdf
48. Wikipedia. 2018. Major Airlines of the United States URL: https://en.wikipedia.org/wiki/Major_airlines_of_the_United_States [accessed 2018-12-11] [WebCite Cache ID 74avc9J8f]
49. Skyscanner. Airlines to South America URL: https://www.skyscanner.com/flights-to/s/airlines-to-south-america.html [accessed 2018-12-11] [WebCite Cache ID 74avlcPPz]
50. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
51. Lamb A, Paul M, Dredze M. Separating fact from fear: tracking flu infections on Twitter. 2013 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2013; Atlanta, GA p. 789-795.
52. Kotsiantis S. Supervised Machine Learning: A Review of Classification Techniques. 2007 Presented at: Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies; 2007; Amsterdam, The Netherlands p. 3-24.
53. Manning C, Raghavan P, Schütze H. Introduction to Information Retrieval. New York: Cambridge University Press; 2008.
54. Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7(1):26 [FREE Full text]
55. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall; 1993.
56. Daughton A, Paul M. Constructing Accurate Confidence Intervals when Aggregating Social Media Data for Public Health Monitoring. 2019 Presented at: AAAI International Workshop on Health Intelligence (W3PHIAI); January 2019; Honolulu, HI URL: https://cmci.colorado.edu/~mpaul/files/w3phiai19_confidence
57. Roesslein J. Tweepy. 2009. URL: http://docs.tweepy.org/en/v3.5.0/ [accessed 2019-04-25] [WebCite Cache ID 77u3J5TU6]
58. Likos A, Griffin I, Bingham A, Stanek D, Fischer M, White S, et al. Local mosquito-borne transmission of Zika virus—Miami-Dade and Broward Counties, Florida, June–August 2016. MMWR Morb Mortal Wkly Rep 2016 Sep 30;65(38):1032-1038 [FREE Full text] [doi: 10.15585/mmwr.mm6538e1] [Medline: 27684886]
59. Adams L, Bello-Pagan M, Lozier M, Ryff KR, Espinet C, Torres J, et al. Update: ongoing Zika virus transmission - Puerto Rico, November 1, 2015-July 7, 2016. MMWR Morb Mortal Wkly Rep 2016 Aug 5;65(30):774-779 [FREE Full text] [doi: 10.15585/mmwr.mm6530e1] [Medline: 27490087]

60. Maurer-Stroh S, Mak T, Ng Y, Phuah SP, Huber RG, Marzinek JK, et al. South-east Asian Zika virus strain linked to cluster of cases in Singapore, August 2016. Euro Surveill 2016 Sep 22;21(38):- [FREE Full text] [doi: 10.2807/1560-7917.ES.2016.21.38.30347] [Medline: 27684526]
61. Dredze M, Paul M, Bergsma S, Tran H. Carmen: A Twitter Geolocation System with Applications to Public Health. 2013 Presented at: AAAI 2013 Workshop; July 14–15, 2013; Bellevue, WA p. 20-24 URL: https://pdfs.semanticscholar.org/9bc4/6fb12f2c7fae0e9e56e734e6efb9ca07fd98.pdf
62. Knowles R, Carroll J, Dredze M. Demographer: Extremely Simple Name Demographics. 2016 Presented at: 2016 EMNLP Workshop on Natural Language Processing and Computational Social Science; November 5, 2016; Austin, TX p. 108-113 URL: https://www.aclweb.org/anthology/W16-5614
63. Pennebaker J, Booth R, Francis M. Linguistic Inquiry and Word Count: LIWC2007. Austin, TX: LIWC.net; 2007. URL: http://liwc.wpengine.com/ [accessed 2019-04-28] [WebCite Cache ID 77zDCBZpw]
64. Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: a systematic review. Milbank Q 2014 Mar;92(1):7-33 [FREE Full text] [doi: 10.1111/1468-0009.12038] [Medline: 24597553]
65. O'Dowd A. Infectious diseases are spreading more rapidly than ever before, WHO warns. Br Med J 2007 Sep 1;335(7617):418 [FREE Full text] [doi: 10.1136/bmj.39318.516968.DB] [Medline: 17762021]
66. Whittemore K, Tate A, Illescas A, Saffa A, Collins A, Varma JK, et al. Zika virus knowledge among pregnant women who were in areas with active transmission. Emerg Infect Dis 2017 Dec;23(1):164-166 [FREE Full text] [doi: 10.3201/eid2301.161614] [Medline: 27855041]
67. Chandrasekaran N, Marotta M, Taldone S, Curry C. Perceptions of community risk and travel during pregnancy in an area of Zika transmission. Cureus 2017 Jul 26;9(7):e1516 [FREE Full text] [doi: 10.7759/cureus.1516] [Medline: 28959511]
68. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the demographics of Twitter users. 2011 Presented at: International Conference on Weblogs and Social Media; July 2011; Barcelona, Spain URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2816
69. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. 2014 Presented at: International Conference on Weblogs and Social Media; June 2014; Ann Arbor, MI.
70. Funk S, Salathé M, Jansen V. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 2010 Sep 6;7(50):1247-1256 [FREE Full text] [doi: 10.1098/rsif.2010.0142] [Medline: 20504800]
71. Lau J, Yang X, Pang E, Tsui HY, Wong E, Wing YK. SARS-related perceptions in Hong Kong. Emerg Infect Dis 2005 Mar;11(3):417-424 [FREE Full text] [doi: 10.3201/eid1103.040675] [Medline: 15757557]
72. Salathé M, Bonhoeffer S. The effect of opinion clustering on disease outbreaks. J R Soc Interface 2008 Dec 6;5(29):1505-1508 [FREE Full text] [doi: 10.1098/rsif.2008.0271] [Medline: 18713723]
73. Forman G. Quantifying counts and costs via classification. Data Min Knowl Disc 2008 Jun 10;17(2):164-206. [doi: 10.1007/s10618-008-0097-y]
74. Snow R, O'Connor B, Jurafsky D, Ng A. Cheap and Fast -- but is It Good?valuating Non-expert Annotations for Natural Language Tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2008 Presented at: EMNLP'08; October 25-27, 2008; Honolulu, HI p. 254-263.
75. Shipman S, Lan J, Chang C, Goodman D. Geographic maldistribution of primary care for children. Pediatrics 2011 Jan;127(1):19-27. [doi: 10.1542/peds.2010-0150] [Medline: 21172992]
76. Gilman SL. Moral panic and pandemics. Lancet 2010 May;375(9729):1866-1867. [doi: 10.1016/S0140-6736(10)60862-8]
77. Daughton AR, Paul MJ, Chunara R. What do people tweet when they’re sick? A preliminary comparison of symptom reports and Twitter timelines. In: ICWSM Workshop on Social Media and Health. 2018.

Abbreviations

API: application programming interface
HHS: US Department of Health and Human Services
LIWC: Linguistic Inquiry Word Count
NPV: negative predictive value
NY: New York
PAHO: Pan American Health Organization
PPV: positive predictive value
RQ: research question
WHO: World Health Organization

Edited by L Tudor Car; submitted 18.12.18; peer-reviewed by A Louren, P Sambaturu, D Broniatowski; comments to author 20.02.19; revised version received 18.03.19; accepted 02.04.19; published 13.05.19.

Please cite as:
Daughton AR, Paul MJ
Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus
J Med Internet Res 2019;21(5):e13090
URL: https://www.jmir.org/2019/5/e13090/
doi:10.2196/13090
PMID:31094347

Original Paper

Areas of Interest and Stigmatic Attitudes of the General Public in Five Relevant Medical Conditions: Thematic and Quantitative Analysis Using Twitter

Miguel Angel Alvarez-Mon[1*], MD; María Llavero-Valero[2*], MD; Rodrigo Sánchez-Bayona[3], MD; Victor Pereira-Sanchez[1], MD; Maria Vallejo-Valdivielso[1], MD; Jorge Monserrat[4], PhD; Guillermo Lahera[4,5,6], MD, PhD; Angel Asunsolo del Barco[5,7,8*], MD, PhD; Melchor Alvarez-Mon[5,9,10*], MD, PhD

[1]Department of Psychiatry, Clinica Universidad de Navarra, Pamplona, Spain
[2]Department of Endocrinology and Nutrition, Clinica Universidad de Navarra, Pamplona, Spain
[3]Department of Oncology, Clinica Universidad de Navarra, Pamplona, Spain
[4]Department of Medicine and Medical specialities, University of Alcala, Madrid, Spain
[5]Instituto Ramon y Cajal de Investigaciones Sanitarias, Madrid, Spain
[6]Center for Biomedical Research in the Mental Health Network, Madrid, Spain
[7]Department of Surgery, Medical and Social Sciences, University of Alcala, Madrid, Spain
[8]Department of Epidemiology & Biostatistics. Graduate School of Public Health and Health Policy, University of New York, New York, NY, United States
[9]Department of Medicine and Medical Specialities, University of Alcala, Madrid, Spain
[10]Service of Internal Medicine, Autoimmune Diseases and Rheumatology, Hospital Universitario Principe de Asturias, Alcala de Henares, Spain
[*]these authors contributed equally

Corresponding Author:
Miguel Angel Alvarez-Mon, MD
Department of Psychiatry
Clinica Universidad de Navarra
Avda Pio XII, 36
Pamplona, 31008
Spain
Phone: 34 948255400
Email: maalvarezdemon@icloud.com

Abstract

Background: Twitter is an indicator of real-world performance, thus, is an appropriate arena to assess the social consideration and attitudes toward psychosis.

Objective: The aim of this study was to perform a mixed-methods study of the content and key metrics of tweets referring to psychosis in comparison with tweets referring to control diseases (breast cancer, diabetes, Alzheimer, and human immunodeficiency virus).

Methods: Each tweet's content was rated as nonmedical (NM: testimonies, health care products, solidarity or awareness and misuse) or medical (M: included a reference to the illness's diagnosis, treatment, prognosis, or prevention). NM tweets were classified as positive or pejorative. We assessed the appropriateness of the medical content. The number of retweets generated and the potential reach and impact of the hashtags analyzed was also investigated.

Results: We analyzed a total of 15,443 tweets: 8055 classified as NM and 7287 as M. Psychosis-related tweets (PRT) had a significantly higher frequency of misuse 33.3% (212/636) vs 1.15% (853/7419; $P<.001$) and pejorative content 36.2% (231/636) vs 11.33% (840/7419; $P<.001$). The medical content of the PRT showed the highest scientific appropriateness 100% (391/391) vs 93.66% (6030/6439; $P<.001$) and had a higher frequency of content about disease prevention. The potential reach and impact of the tweets related to psychosis were low, but they had a high retweet-to-tweet ratio.

Conclusions: We show a reduced number and a different pattern of contents in tweets about psychosis compared with control diseases. PRT showed a predominance of nonmedical content with increased frequencies of misuse and pejorative tone. However, the medical content of PRT showed high scientific appropriateness aimed toward prevention.

(J Med Internet Res 2019;21(5):e14110) doi:10.2196/14110

KEYWORDS

social stigma; social media; psychosis; breast cancer; HIV; dementia; public opinion; diabetes

Introduction

Psychotic disorders are among the world's leading causes of disability [1,2]. The estimated lifetime rate of suffering any psychotic disorder is 2 to 3% [3]. The societal and economic burden of schizophrenia is very high [4]. Schizophrenia is associated with at least 10 to 15 years of potential life lost, with no indication of a decline in this trend [5].

Despite many decades of research, the treatment of psychotic disorders remains only partially effective, and their etiology is not fully understood [6,7]. Currently, patients are encouraged to take an active role in the development of an active and meaningful life while growing beyond the misfortune of mental illness [8,9]. The traditional clinical and societal view of schizophrenia is of a debilitating and deteriorating disorder, with a poor outcome [10]. There is evidence of persisting stigma about mental illness that leads to negative stereotyping and to discriminatory behavior toward people with schizophrenia [11]. Stigma may cause affected patients to experience rejection and to feel shame about their condition, reducing their self-esteem and limiting their opportunities [12-14].

Background

In recent years, the internet and social media have become pivotal instruments for sharing knowledge [15]. Accordingly, the internet has radically modified how most people communicate, share, and seek out information regarding health and medical conditions [16,17]. Twitter, one of the most popular and widely used platforms of social media, is currently considered an effective channel of communication [18]. Different players in health and medicine have realized Twitter's potential for acquiring and distributing medical information [19]. Furthermore, American mainstream media outlets and the general public demonstrate a preferential interest for psychiatric disorders on Twitter [20]. A third of patients with schizophrenia use social networking sites, including Twitter, at least daily [21]. The analysis of distributed tweets is increasingly appreciated in health research [22]. The utilization of Twitter data has enabled researchers to study health-related attitudes toward behaviors and diseases, predict the incidence of both communicable and noncommunicable diseases, or have an insight of patients' medical experience [23-25]. The utilization of online data for health care purposes has led to the development of an emerging field defined as infodemiology [26].

Moreover, the analysis of tweets about psychiatric disorders is a recent relevant area of study for understanding the sentiments of society, patients, and health players [27-33]. Concerning results have been reported about the trivialization, stigmatization, and mockery of schizophrenia and other psychiatric disorders by Twitter users [34-39]. The areas of medical and nonmedical interest of Twitter users about psychosis spectrum disorders have not been established. The reach and impact of psychosis-related tweets (PRT) remain unknown.

Objectives

The aims of this multidisciplinary study were to investigate the medical knowledge and social consideration of Twitter users toward psychosis in comparison with 2 prevalent causes of death worldwide (breast cancer and diabetes mellitus), a relevant cause of severe neurocognitive impairment (Alzheimer disease) and a socially relevant disease (human immunodeficiency virus, HIV infection), as well as their areas of medical and nonmedical interest. In addition, we investigated the potential impact and reach derived from tweets and retweets of each condition.

Methods

Research Strategy

In this observational quantitative and qualitative study, we focused on searching for tweets that referred to psychosis over a period of 8 consecutive days in 2018. As controls, we studied in parallel the tweets related to breast cancer, diabetes mellitus, Alzheimer disease, and HIV infection. In this study, we focused on tweets with the following hashtags: #psychosis, #psychotic, #schizophrenic, #schizophrenia, #diabetes, #diabetic, #breastcancer, #hiv, and #alzheimer [40]. Content was limited to English-language tweets. Data collection spanned from Monday, February 26 to Monday, March 5, 2018. This period of time has at least 2 months of separation from any major international awareness month for any of the diseases we studied and was selected to avoid potential bias in the type of disease-related tweets.

Search Tool and Data Collection

In this study, we used the Twitter Firehose data stream, which is managed by Gnip and allows access to 100% of all public tweets that match a set of "search" criteria (query) [41]. In our study, the search criteria were the previously mentioned hashtags. Tweet Binder, the search engine we employed, uses automatic machine-learning text analysis algorithms, as well as node.js and the PHP language, which enables an analysis of tweets in the json format (used by Gnip).

Content Analysis Process

All of the collected tweets were classified using qualitative content analysis methods as a systematic method for making inferences from the text to summarize the content of communication [42]. In this study, we used a codebook specifically created by the members of the research team. All of the team's members who qualitatively analyzed the content were medical doctors specialized in psychiatry, medical oncology, internal medicine, immunology, or endocrinology, with clinical practice in university hospitals. The analysis strategy included a series of steps. First, to achieve reliability, raters reviewed an initial subset of 100 tweets to apply initial classifications of each category. Differences in categorization and other discrepancies between the evaluators were discussed until a consensus was reached, and classification criteria were adapted to reflect the initial rating experience. Second, researchers grouped by pairs, independent and blinded, rated a second training set of 300 tweets using the improved codebook. The obtained reliability was higher than 90% for tweet content analysis, and a final version of the codebook was established. Third, all tweets were analyzed separately by 2 blinded researchers. If discrepancies in the classification of a tweet occurred between both raters (less than 10% of the cases), the whole group of researchers analyzed the tweet's content and reached a final decision by a consensus of at least two-thirds of the research team. Tweets that included unclassifiable content were excluded.

Each tweet, depending on the content, was rated as medical or nonmedical. Medical tweets included a reference to the illness and its diagnosis, treatment, prognosis, or prevention. We also assessed if the content was medically appropriate or inappropriate according to the current medical knowledge. Nonmedical tweets were classified into 4 categories: (1) patient, family, or caregiver testimony; (2) information about medical health providers and scientific meetings; (3) solidarity, support campaigns, and advocacy; and (4) misuse. Nonmedical tweets were also classified as positive or pejorative, depending on the tone of the tweet. Classification criteria and examples of tweets by category are shown in Multimedia Appendix 1.

Measuring Influence on Twitter: Retweets and Hashtags' Reach and Impact

We analyzed the number of retweets generated by each tweet as an indicator of the user interest in a given topic [43-45]. We also measured the potential reach and impact of the hashtags analyzed. Impact is a numerical value representing the potential views a tweet may receive. To calculate impact, we multiplied, for each user who contributed to the hashtag, the number of followers by the number of tweets posted, and finally, we added this number for all such users. Reach is a numerical value measuring the potential audience of the hashtag (how many people could have seen it). To calculate reach, we measured the number of followers of each user who contributed to the hashtags and added them all together. We collected the 10 hashtags most frequently associated with the hashtags we studied.

Ethical Considerations

This study received the approval of the University of Navarra Research Ethics Committee and was compliant with the research ethics principles of the Declaration of Helsinki (7th revision, 2013). However, this study did not directly involve human subjects nor include any intervention but instead used publicly available tweets. Nevertheless, we have taken care to not reveal any username and to avoid citing the tweets that could reveal it.

Statistical Analysis

A descriptive study of the sample was performed, describing the variables by their absolute and relative frequencies. The percentages found were compared using the chi-square test. The mean numbers of retweets per original tweet about the different diseases were compared by analysis of variance. The Tamhane test was performed for a posteriori comparison between diseases.

Results

Increased Pejorative Sentiment and Misuse Content in Psychosis-Related Tweets

The number of tweets generated about Alzheimer disease and psychosis were lower than that of diabetes, HIV infection, or breast cancer (Table 1). Of the total of 15,443 tweets analyzed, 101 were excluded according to the criteria of the study. Thus, 15,342 tweets were classified into 2 categories according to their medical or nonmedical content, and the frequencies of both categories between the different diseases were significantly different (*P*<.001; Table 1). The percentage of PRT with nonmedical content was higher than those in the groups related to diabetes, HIV infection, or Alzheimer disease and lower than that in the breast cancer group. The percentage of tweets with medical content was higher among those related to HIV infection and diabetes than those related to the other diseases analyzed.

Interestingly, different patterns of distribution of the nonmedical tweets among the 4 categorizes of contents were found between the diseases (*P*<.001; Figure 1). Remarkably, in PRT, the category with highest frequency of tweets was misuse, which was significantly higher than that found in the control diseases 33.3% (212/636) vs 1.15% (853/7419); (*P*<.001). In contrast, misuse was absent or minimal in breast cancer, diabetes, and HIV infection. The frequency of PRT with misuse content was 12 times higher than in those related to HIV infection. The frequency of tweets with solidarity and advocacy content related to HIV infection was the highest. The frequency of tweets with content about medical health providers and scientific meetings was lower in those related to psychosis, breast cancer, and HIV infection compared with Alzheimer disease and diabetes.

We analyzed the tone of the 8055 nonmedical tweets (Table 2). The frequencies of positive and nonpositive tweet contents were significantly different between the different diseases (*P*<.001). The frequency of PRT with positive content was significantly lower than what was found in the control diseases 63.7% (405/636) vs 88.67% (6522/7354; *P*<.001). The percentage of pejorative tweets related to psychosis 36.3% (231/636) doubled that of breast cancer 15.0% (365/2424) and diabetes 12.75% (677/3115) and was 5 times higher than that of Alzheimer 7.6% (38/506) and HIV 2.72% (37/1364). In psychosis and in the control diseases, the frequency of tweets with positive content was significantly lower in the misuse category than those found in the other 3 categories (*P*<.001).

Table 1. Number and content of tweets about psychosis and control diseases. Percentages (%) were calculated with respect to the total number of tweets generated about the 5 diseases. Number of tweets with nonmedical and medical contents generated about the diseases. Percentages (%) were calculated with respect to the total number of tweets generated about each disease.

Medical condition	Tweets generated, n (%)	Content[a]	
		Nonmedical, n (%)	Medical, n (%)
Psychosis	1029 (6.66)	636 (61.81)	393 (38.19)
Breast cancer	3703 (23.98)	2434 (65.98)	1255 (34.02)
Diabetes	6467 (41.88)	3115 (48.65)	3288 (51.35)
Alzheimer	930 (6.02)	506 (54.64)	420 (45.36)
HIV	3314 (21.46)	1364 (41.40)	1931 (58.60)
Total	15,443 (100)	8055 (52.50)	7287 (47.50)

[a] Test chi-square; *P*<.001. 101 tweets nonclassifiable (99.35% analyzed).

Figure 1. Different percentages (%) of tweets with nonmedical and medical content generated about psychosis and control diseases. Percentages (%) were calculated with respect to the total number of tweets generated about each disease.

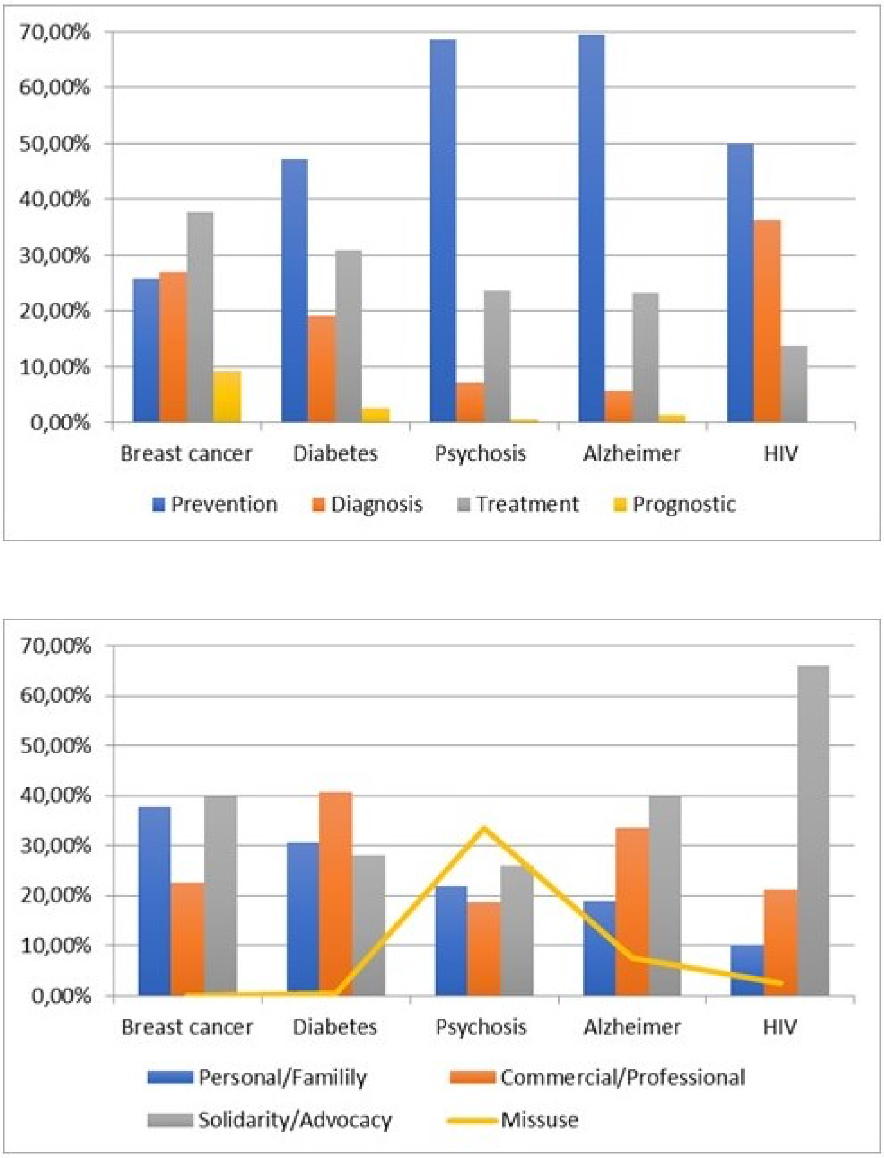


Table 2. Number of tweets with nonmedical, positive-tone content about psychosis and control diseases. Percentages (%) were calculated with respect to the total number of nonmedical contents tweets generated about each disease. Number of tweets with nonmedical, positive-tone content in the testimonies, medical health providers, solidarity/advocacy, or misuse categories generated about the diseases. Percentages (%) were calculated with respect to the total number of tweets generated about each category and disease.

Medical condition	Nonmedical content positive[a], n (%)	Positive sentiment[b]				
		Personal/family, n (%)	Commercial/professional, n (%)	Solidarity/advocacy, n (%)	Misuse, n (%)	*P* value
Psychosis	405 (63.7)	112 (80.58)	105 (88.24)	144 (87.27)	44 (20.75)	<.001
Breast cancer	2070(85.05)	741 (80.81)	501 (91.26)	828 (85.54)	0 (0)	<.001
Diabetes	2703 (87.25)	713 (74.89)	1236 (97.86)	746 (86.04)	8 (50)	<.001
Alzheimer	457 (92.40)	94 (100)	164 (98.20)	194 (97.98)	5 (13.89)	<.001
HIV	1287 (97.28)	129 (100)	262 (100)	818 (99.63)	2 (6.45)	<.001
Total	6927 (86.70)	1789 (80.19)	2268 (96.10)	2730 (90.43)	59 (20)	<.001

[a] Test chi-square; *P*<.001. 65 tweets not classifiable.

[b] 85 tweets not classifiable.

The Medical Content of Psychosis-Related Tweets Showed High Scientific Appropriateness Geared Toward Prevention

We investigated the scientific appropriateness and areas of interest of the 7287 tweets with medical content, and we excluded 8 tweets according to the analysis criteria. The frequency of appropriateness between the 5 diseases was significantly different (*P*<.001; Table 3). According to the scientific evaluation, the content of the 391 PRT analyzed was correct. This frequency of scientific appropriateness was higher than that found in the 4 control diseases 100% (391/391) vs 93.66% (6030/6439; *P*<.001). The scientific appropriateness found in the diabetes- and HIV infection–related tweets was higher than that found in breast cancer and Alzheimer disease.

Furthermore, the tweets were classified according to the area of interest of the medical content: diagnosis, prognosis, treatment, and prevention (Figure 1). We found a different pattern of distribution of the 4 categories of medical content between tweets related to psychosis and tweets related to the control diseases (*P*<.001). Interestingly, the frequency of tweets with content about disease prevention were higher in those related to psychosis and Alzheimer disease than in those related to diabetes, HIV infection, and breast cancer. Opposite results were observed in diagnosis-related tweets. Tweets with treatment content were higher in breast cancer 37.84% (475/1255). As shown in Table 3, the lowest frequencies of scientifically appropriate medical content were found in tweets related to treatment and prognosis of Alzheimer disease.

Psychosis-Related Tweets Showed High Frequency of Retweets

We measured the number of retweets generated about each disease (Table 4). We found that the retweet-to-tweet ratio, and thus, the probability of being retweeted for the PRT, was significantly higher than that found for the control diseases. We did not find significant differences (*P*=.49) in the frequency of retweets between those with pejorative or positive tone related to the different health conditions analyzed. We did not find significant differences in the frequency of retweets between those with misuse content and the rest of the nonmedical tweets (*P*=.08).

Psychosis-Related Tweets Showed Limited Reach and Impact

As shown in Table 4, we found that the potential impact and reach (7,738,305 and 5,360,995, respectively) of PRT were less than those of breast cancer (62,348,473 and 20,930,244, respectively), diabetes (92,770,714 and 46,143,068, respectively), Alzheimer disease (10,019,729 and 7,118,104, respectively), and HIV infection (101,643,088 and 52,072,034, respectively). Finally, we analyzed the 10 hashtags most frequently associated with the hashtags of the different diseases analyzed. We found that the hashtags most frequently associated with #psychosis were psychosis, mentalhealth, schizophrenia, cannabis, bipolar, depression, mental illness, schoolshooting, ptsd, and wpatc18. In the case of the control diseases, the hashtags most frequently associated with #breastcancer were cancer, bcsm, mastectomy, breastreconstruction, BreastCancerAwareness, bccww, health, blog, and chemo; the hashtags most frequently associated with #diabetes were health, obesity, t1d, cancer, diet, diabetic, insulin, t2d, and type1diabetes; the hashtags most frequently associated with #Alzheimer were dementia, health, brain, care, memory, caregiver, aging, science, and alzheimers; and the hashtags most frequently associated with #HIV were aids, PrEP, stigma, USA, health, Philippines, std, vaccine, and tuberculosis.

Table 3. Number of tweets with medically appropriate content about psychosis and control diseases. Percentages (%) were calculated with respect to the total number of tweets generated with medical content about each disease. Number of tweets with medically appropriate content about diagnosis, treatment, prognosis, and prevention generated in the different diseases. Percentages (%) were calculated with respect to the total number of tweets generated about each different medical content and disease.

Medical condition	Medical content accuracy[a], N (%)	Scientific accuracy[b]				
		Diagnosis, N (%)	Treatment, N (%)	Prognostic, N (%)	Prevention, N (%)	*P* value
Psychosis	391 (100)	28 (100)	91 (100)	2 (100)	264 (100)	—[c]
Breast cancer	1034 (82.52)	285 (84.07)	400 (84.39)	99 (84.62)	250 (77.40)	.05
Diabetes	3126 (95.57)	627 (100.00)	895 (88.61)	85 (97.70)	1519 (98.19)	<.001
Alzheimer	374 (89.05)	23 (95.83)	60 (61.22)	4 (66.67)	287 (98.29)	<.001
HIV	1905 (98.76)	685 (98)	258 (96.99)	2 (100)	958 (99.58)	.004
Total	6830 (94.03)	1648 (95.98)	1704 (87.88)	192 (89.72)	3278 (96.75)	<.001

[a] Test chi-square; *P*<.001. 23 Tweets not classifiable.

[b] 8 tweets not classifiable.

[c] It is not possible to calculate the p value because in Psychosis the four categories (Diagnosis, Treatment, Prognosis and Prevention) had the same value (100).

Table 4. Potential impact, potential reach, and number of retweets generated by psychosis- and control disease–related tweets.

Medical condition	Potential impact	Potential reach	Contributors, n	Followers per contributor, n	Retweets per original tweet, mean (SE)	*P* value[a]
Psychosis	7,738,305	5,360,995	1155	19,409	0.23 (1.22)	Ref[b]
Breast cancer	62,348,473	20,930,244	3161	6621	0.03 (0.29)	<.001
Diabetes	92,770,714	46,143,068	5087	9071	0.11 (0.01)	.002
Alzheimer	10,019,729	7,118,104	1105	6442	0.04 (0.32)	<.001
HIV	101,643,088	52,072,034	7308	11,029	0.08 (0.59)	.02

[a] Analysis of variance; *P*<.001. Numbers are Tamhane test between psychosis and each disease.

[b] Ref: reference category.

Discussion

Principal Findings

In this work, we investigated all the tweets generated about psychosis during 8 consecutive days in the winter of 2018. As controls, we studied 2 prevalent causes of death worldwide (breast cancer and diabetes mellitus), a relevant cause of severe neurocognitive impairment (Alzheimer disease), and a socially relevant disease (HIV infection) [46]. We found a different pattern of content in tweets about psychosis with respect to those related to control diseases. PRT showed a predominance of nonmedical content with increased frequency of misuse and pejorative tone with respect to the control diseases. However, the medical content of PRT showed high scientific appropriateness geared toward prevention. The potential reach and impact of the tweets related to psychosis were low but showed a high retweet-to-tweet ratio.

The search tool utilized for data collection allows access to 100% of all public tweets. Thus, the conclusions were obtained from the results measured the *total* population of tweets, and they are not deduced from the analysis of a reduced sample (previous health-related studies utilizing Twitter have generally focused on the analysis of a 1% sample of the total number of tweets available). To our knowledge, this is the first study that analyzed *all tweets* about psychosis in particular in a defined period of time.

Our data show a differential pattern of information and opinions expressed in the contents of the PRT in comparison with those relating to the different control diseases. The majority of PRT with nonmedical content were focused on misuse, with a small proportion expressing solidarity. This bias observed in the content of PRT was further supported by the finding that more than a third of these nonmedical tweets had pejorative content about the disease and/or patients. Unfortunately, psychosis is still employed as an insult in a relevant proportion of tweets. Our findings about the elevated misuse and pejorative tone toward psychosis on Twitter are consistent with previous studies that analyzed schizophrenia in selected samples of tweets [35-38]. A recent study found that the terms psychosis/psychotic are associated with a significantly higher number of tweets with negative contents than schizophrenia/schizophrenic [34].

The relevance of this evident and extended misuse and pejorative content found in PRT is supported by the comparison of these results with those found in the investigated control diseases. Misuse in psychosis was 4 times greater than in Alzheimer disease and was marginal in breast cancer, diabetes, and HIV infection. The frequency of pejorative psychosis tweets was 5 times greater than the frequencies found in tweets related to

Alzheimer disease and HIV infection and doubled those of breast cancer and diabetes. The fact that #schoolshooting was among the most frequently associated hashtags with #psychosis also reflects the negative and incorrect stereotyping of psychosis by a relevant number of Twitter users. All together, these twitter data show that psychosis patients are targets of negative ideas, feelings or judgments by twitter users, demonstrating the persistence of social stigma for psychiatric diseases, in general, and psychosis, in particular [47-50]. The bad social habit of using "schizophrenia" or "psychosis" to refer to a "madness" of some kind might have an impact on these negative results found in PRT [51]. Social stigma has major adverse effects on the lives of people with mental health conditions [52]. Stigma has also been common in portrayals of physical conditions. HIV infection has been a paradigmatic example of an organic condition suffering not only social stigma but also stigma from health care providers and professionals [53,54]. Interestingly, our results show a marginal frequency of HIV-related tweets with misuse and negativity content. We found that fewer than 3% of the HIV-related tweets had stigmatizing content, and this low number was 6 times lower than the frequency recently described [35]. This reduction may be explained by different factors, including the size and the selection criteria of the sample and the temporal gap between the 2 studies. These results support the notion of evolution in the social attitudes about diseases.

The cause of stereotypes about psychosis in Twitter is multifactorial. It may reflect the persistence of social negative stereotyping and stigmatizing attitudes toward people with psychosis [11]. Furthermore, the use of Twitter for the distribution of health care information carries some risks that are even more pronounced in the field of mental health: high rate of misinformation, sources of questionable reliability, overwhelmingly high volumes of information available, and concerns about professionalism [55]. Due to the small number of characters required, tweets are often brief and must omit key information and may lead to fruitless discussions [56]. In this context, terms such as "psychotic" or "mentally ill" can be used to disparage or ridicule someone, thus spread social stigma to the social network. Furthermore, the massive and immediate response to nonexpert opinions or news related with mental disorders can convert Twitter on an "echo chamber of ideas," representing shared opinions rather than balanced facts because of the ease of quoting or retweeting. Furthermore, other players may lead to the production of stigma in Twitter. The public rely on the media as a key source of information about mental illness turns the news media in a strong influence on public discourse and attitudes about mental health issues as well as to medical decisions, health service utilization, or consumption of antidepressants [57,58]. Selecting the topics they cover and highlighting certain aspects, they contribute to the creation of stereotypes. It has been shown that news stories referring to mental illness frequently emphasize on violence, although rates of violence among those with mental illnesses are very low, and people with schizophrenia are more likely to be victims of violence rather than perpetrators [59,60].

Twitter appears to be a relevant communication tool for distributing and acquiring medical information. Thus, it was relevant to investigate the scientific accuracy and the areas of interest of the medical content on Twitter pertaining to psychosis and the control disorders. Interestingly, the scientific content of all the PRT was surprisingly correct, and the rate was higher than those found in the control diseases. This accuracy of tweets containing medical content starkly contrasts the high rates of misuse and negative tone in the nonmedical tweets about psychosis. These data suggest Twitter is also used to communicate medical content about psychosis by users with correct medical information. The areas of interest of the medical information varied between psychosis and control diseases. Interestingly, the frequency of tweets with content about disease prevention was markedly higher in those related to psychosis or Alzheimer disease than in those of diabetes, HIV infection, and breast cancer. Different reasons may support this differential pattern of medical interest. The social perception of limited effectiveness for the medical treatment of psychosis and Alzheimer disease might support the special interest in preventive strategies. The chronic and severe impact of patients with both diseases on their families and caregivers might also explain the high interest in the prevention of these disorders [10,61,62]. The absence of established analytical or image criteria for the diagnosis of psychosis might also contribute to the limited interest in the diagnosis of the disease.

Finally, we investigated the interest and diffusion of the tweets generated about psychosis. In the period of time analyzed, the frequency of retweets generated by PRT was higher than that found in the control diseases. This parameter is considered an indicator of the user interest in the topic of each tweet [43-45]. However, the metrics of the tweets related to psychosis were small compared with those of the control diseases. The impact of the PRT measured as the potential views that the tweet may receive was clearly lower than that of the control diseases. Similarly, poor diffusion of the PRT was found when we calculated the potential audience or reach of the PRT. These metrics are used for the quantification of the diffusion of tweets and potential influence in society. Thus, our data imply that the stigmatization of and limited social support for psychosis are reflected in the low impact and reach of the PRT.

Strengths and Limitations

Although this study improves and expands previous research on the communication of psychosis in a popular and widely used form of social media, there are still some limitations. The rating process had an inherent degree of subjectivity because of differences in the perceived context and emotional tone of some tweets. This was made particularly evident by words that had dual meanings. There was also a degree of selection bias, as stigmatizing and trivializing tweets were more likely to be lacking in context and/or grammatical correctness, rendering them less likely to be considered for analysis. We minimized the effects of these issues through our robust rating criteria and binary rating system, which were chosen for the analysis performed by expert clinicians in the medical fields analyzed. To achieve maximum reliability, the qualitative analysis of the disease-related tweets required a manual input that was time-consuming and required expert involvement. We followed what we considered a gold-standard qualitative analysis strategy.

Conclusions

Twitter is a tool for developing interventions and strategies of information aimed at modifying health-related social and individual behaviors [63]. Our results support the dynamic and potentially positive evolution of the social stigmatization of health disorders, as can be observed in HIV infection and cancer. Thus, it is possible to expect a reduction of psychosis stigmatization. Moreover, mixed-methods research on Twitter and other social media may be a relevant strategy for measuring the effectiveness of the strategies and actions established for overcoming the social psychosis stigma [64]. A proactive sensitization by professionals, scientific societies, patient associations, and other social agents to use the promising platform of social media communication is needed. Furthermore, the respect and supportive in social media communication content may also impact a patient's life and treatment. Psychosis patients often use social media [21], and despite the anonymity of Twitter, many users identify themselves as patients [27,32].

Although stigmatization is significantly decreasing and societal consideration is improving in other disorders, such as breast cancer and HIV, the stigma regarding psychosis is not decreasing. On one hand, psychosis is used as hate-speech on Twitter, but on the other hand, Twitter is used as a beacon of medically accurate information for the disorder. Therefore, Twitter may be a great tool for antistigma campaigns and promotion of healthy habits.

Acknowledgments

This work was partially supported by grants from the Fondo de Investigación de la Seguridad Social, Instituto de Salud Carlos III (PI18/01726), Spain, and the Programa de Actividades de I+D de la Comunidad de Madrid en Biomedicina (B2017/BMD-3804), Madrid, Spain.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Classification criteria and examples of tweets by category.

[DOCX File, 27KB - jmir_v21i5e14110_app1.docx]

References

1. Hjorthøj C, Stürup AE, McGrath J, Nordentoft M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. Lancet Psychiatry 2017 Dec;4(4):295-301. [doi: 10.1016/S2215-0366(17)30078-0] [Medline: 28237639]
2. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry 2015 Apr;72(4):334-341 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.2502] [Medline: 25671328]
3. Owen MJ, Sawa A, Mortensen PB. Schizophrenia. Lancet 2016 Jul 2;388(10039):86-97 [FREE Full text] [doi: 10.1016/S0140-6736(15)01121-6] [Medline: 26777917]
4. Cloutier M, Sanon AM, Guerin A, Nitulescu R, Ramanakumar AV, Kamat S. The economic burden of Schizophrenia in the United States in 2013. J Clin Psychiatry 2016:764 [FREE Full text] [Medline: 27135986]
5. Westman J, Eriksson SV, Gissler M, Hällgren J, Prieto ML, Bobo WV, et al. Increased cardiovascular mortality in people with schizophrenia: a 24-year national register study. Epidemiol Psychiatr Sci 2018 Dec;27(5):519-527 [FREE Full text] [doi: 10.1017/S2045796017000166] [Medline: 28580898]
6. Lieberman J, First M. Psychotic disorders. Ropper AH 2018 Jul 19:270 [FREE Full text] [Medline: 30021088]
7. Radua J, Ramella-Cravaro V, Ioannidis JPA, Reichenberg A, Phiphopthatsanee N, Amir T, et al. What causes psychosis? An umbrella review of risk and protective factors. World Psychiatry 2018 Feb;17(1):49-66 [FREE Full text] [doi: 10.1002/wps.20490] [Medline: 29352556]
8. Gask L, Coventry P. Person-centred mental health care: the challenge of implementation. Epidemiol Psychiatr Sci 2012 Jun;21(2):139-144. [doi: 10.1017/S2045796012000078] [Medline: 22789160]
9. Fleischhacker WW, Arango C, Arteel P, Barnes TRE, Carpenter W, Duckworth K, et al. Schizophrenia--time to commit to policy change. Schizophr Bull 2014 Apr;40(Suppl 3):S165-S194 [FREE Full text] [doi: 10.1093/schbul/sbu006] [Medline: 24778411]
10. Mittendorfer-Rutz E, Rahman S, Tanskanen A, Majak M, Mehtälä J, Hoti F, et al. Burden for parents of patients with schizophrenia-a nationwide comparative study of parents of offspring with rheumatoid arthritis, multiple sclerosis, epilepsy, and healthy controls. Schizophr Bull 2018 Sep 3:1-10. [doi: 10.1093/schbul/sby130] [Medline: 30184197]
11. Imhoff R. Zeroing in on the effect of the schizophrenia label on stigmatizing attitudes: a large-scale study. Schizophr Bull 2016 Mar;42(2):456-463 [FREE Full text] [doi: 10.1093/schbul/sbv137] [Medline: 26409222]

12. Henderson C, Corker E, Lewis-Holmes E, Hamilton S, Flach C, Rose D, et al. England's time to change antistigma campaign: one-year outcomes of service user-rated experiences of discrimination. Psychiatr Serv 2012;63(5):451-457. [doi: 10.1176/appi.ps.201100422] [Medline: 22549532]
13. Carpiniello B, Pinna F. The reciprocal relationship between suicidality and stigma. Front Psychiatry 2017;8:35 [FREE Full text] [doi: 10.3389/fpsyt.2017.00035] [Medline: 28337154]
14. Picco L, Abdin E, Pang S, Vaingankar JA, Jeyagurunathan A, Chong SA, et al. Association between recognition and help-seeking preferences and stigma towards people with mental illness. Epidemiol Psychiatr Sci 2018 Dec;27(1):84-93. [doi: 10.1017/S2045796016000998] [Medline: 27927259]
15. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
16. Lee K, Hoti K, Hughes J, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec 2;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
17. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 4;362(9):859-860. [doi: 10.1056/NEJMc0909595] [Medline: 20200398]
18. Topf JM, Hiremath S. Social media, medicine and the modern journal club. Int Rev Psychiatry 2015 Apr;27(2):147-154. [doi: 10.3109/09540261.2014.998991] [Medline: 25906989]
19. Huang S, Martin LJ, Yeh CH, Chin A, Murray H, Sanderson WB, et al. The effect of an infographic promotion on research dissemination and readership: a randomized controlled trial. CJEM 2018 Dec;20(6):826-833. [doi: 10.1017/cem.2018.436] [Medline: 30289098]
20. Alvarez-Mon MA, Asunsolo Del Barco A, Lahera G, Quintero J, Ferre F, Pereira-Sanchez V, et al. Increasing interest of mass communication media and the general public in the distribution of tweets about mental disorders: observational study. J Med Internet Res 2018 Dec 28;20(5):e205 [FREE Full text] [doi: 10.2196/jmir.9582] [Medline: 29807880]
21. Miller BJ, Stewart A, Schrimsher J, Peeples D, Buckley PF. How connected are people with schizophrenia? Cell phone, computer, email, and social media use. Psychiatry Res 2015 Feb 28;225(3):458-463. [doi: 10.1016/j.psychres.2014.11.067] [Medline: 25563669]
22. Wakamiya S, Morita M, Kano Y, Ohkuma T, Aramaki E. Tweet classification toward Twitter-based disease surveillance: new data, methods, and evaluations. J Med Internet Res 2019 Feb 20;21(2):e12783 [FREE Full text] [doi: 10.2196/12783] [Medline: 30785407]
23. Wakamiya S, Matsune S, Okubo K, Aramaki E. Causal relationships among pollen counts, tweet numbers, and patient numbers for seasonal allergic rhinitis surveillance: retrospective analysis. J Med Internet Res 2019 Feb 20;21(2):e10450 [FREE Full text] [doi: 10.2196/10450] [Medline: 30785411]
24. Hendriks H, van den Putte B, Gebhardt W, Moreno M. Social drinking on social media: content analysis of the social aspects of alcohol-related posts on facebook and instagram. J Med Internet Res 2018 Dec 22;20(6):e226 [FREE Full text] [doi: 10.2196/jmir.9355] [Medline: 29934290]
25. Kolliakou A, Ball M, Derczynski L, Chandran D, Gkotsis G, Deluca P, et al. Novel psychoactive substances: an investigation of temporal trends in social media and electronic health records. Eur Psychiatry 2016 Dec;38:15-21. [doi: 10.1016/j.eurpsy.2016.05.006] [Medline: 27611330]
26. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
27. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Online communication about depression and anxiety among twitter users with schizophrenia: preliminary findings to inform a digital phenotype using social media. Psychiatr Q 2018 Dec;89(3):569-580. [doi: 10.1007/s11126-017-9559-y] [Medline: 29327218]
28. Thackeray R, Burton S, Giraud-Carrier C, Rollins S, Draper C. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]
29. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Dec 5;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
30. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. #Proana: pro-eating disorder socialization on Twitter. J Adolesc Health 2016 Dec;58(6):659-664. [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]
31. 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 Dec;15(1 Pt B):210-217. [doi: 10.1016/j.jacr.2017.09.043] [Medline: 29154103]
32. Lachmar E, Wittenborn A, Bogen K, McCauley H. #MyDepressionLooksLike: examining public discourse about depression on twitter. JMIR Ment Health 2017 Oct 18;4(4):e43 [FREE Full text] [doi: 10.2196/mental.8141] [Medline: 29046270]

33. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on Twitter. J Med Internet Res 2014 Oct 15;16(10):e237 [FREE Full text] [doi: 10.2196/jmir.3765] [Medline: 25591063]
34. Passerello G, Hazelwood J, Lawrie S. Using Twitter to assess attitudes to schizophrenia and psychosis. BJPsych Bull 2019 Feb 20:1-9 [FREE Full text] [doi: 10.1192/bjb.2018.115] [Medline: 30784393]
35. Robinson P, Turk D, Jilka S, Cella M. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. Soc Psychiatry Psychiatr Epidemiol 2019 Jan;54(1):51-58 [FREE Full text] [doi: 10.1007/s00127-018-1571-5] [Medline: 30069754]
36. Reavley N, Pilkington P. Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. PeerJ 2014;2:e647 [FREE Full text] [doi: 10.7717/peerj.647] [Medline: 25374786]
37. Athanasopoulou C, Sakellari E. 'Schizophrenia' on Twitter: content analysis of Greek language tweets. Stud Health Technol Inform 2016;226:271-274. [Medline: 27350523]
38. Joseph AJ, Tandon N, Yang LH, Duckworth K, Torous J, Seidman LJ. #Schizophrenia: use and misuse on Twitter. Schizophr Res 2015;165:5. [Medline: 25937459]
39. Lydecker J, Cotter E, Palmberg A, Simpson C, Kwitowski M, White K, et al. Does this Tweet make me look fat? A content analysis of weight stigma on Twitter. Eat Weight Disord 2016 Jun;21(2):229-235. [doi: 10.1007/s40519-016-0272-x] [Medline: 27068174]
40. Chiang AL, Vartabedian B, Spiegel B. Harnessing the hashtag: a standard approach to GI dialogue on social media. Am J Gastroenterol 2016 Aug;111(8):1082-1084. [doi: 10.1038/ajg.2016.259] [Medline: 27349338]
41. Joseph K, Landwehr P, Carley K. Two 1%s Don't Make a Whole: Comparing Simultaneous Samples from Twitter's Streaming API. Switzerland: Springer, Cham; 2014 Presented at: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation; March 31-April 3, 2014; Washington, DC, USA p. 75-83 URL: http://link.springer.com/10.1007/978-3-319-05579-4_10
42. Hsieh H, Shannon S. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
43. Tsugawa S, Kito K. Retweets as a predictor of relationships among users on social media. PLoS One 2017;12(1):e0170279 [FREE Full text] [doi: 10.1371/journal.pone.0170279] [Medline: 28107489]
44. Fung IC, Jackson AM, Mullican LA, 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 2;4(2):e33 [FREE Full text] [doi: 10.2196/publichealth.8737] [Medline: 29610112]
45. Suh B, Hong L, Pirolli P, Chi E. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. 2010 Presented at: 2010 IEEE Second International Conference on Social Computing; August 20-22, 2010; Minneapolis, MN, USA p. 177-184 URL: http://ieeexplore.ieee.org/document/5590452/
46. Ayers JW, Althouse BM, Dredze M, Leas EC, Noar SM. News and internet searches about human immunodeficiency virus after Charlie Sheen's disclosure. JAMA Intern Med 2016 Apr;176(4):552-554. [doi: 10.1001/jamainternmed.2016.0003] [Medline: 26902971]
47. Hayhow B, Gaillard F, Velakoulis D, Walterfang M. Delayed diagnosis of multiple sclerosis in a patient with schizoaffective disorder: a case of 'diagnostic overshadowing'. Aust N Z J Psychiatry 2015 Feb;49(2):186-187. [doi: 10.1177/0004867414551067] [Medline: 25237140]
48. Corrigan P, Druss B, Perlick D. The impact of mental illness stigma on seeking and participating in mental health care. Psychol Sci Public Interest 2014 Oct;15(2):37-70. [doi: 10.1177/1529100614531398] [Medline: 26171956]
49. Webber M, Corker E, Hamilton S, Weeks C, Pinfold V, Rose D, et al. Discrimination against people with severe mental illness and their access to social capital: findings from the Viewpoint survey. Epidemiol Psychiatr Sci 2014 Jun;23(2):155-165. [doi: 10.1017/S2045796013000243] [Medline: 23683403]
50. Lasalvia A, Zoppei S, Van Bortel T, Bonetto C, Cristofalo D, Wahlbeck K, ASPEN/INDIGO Study Group. Global pattern of experienced and anticipated discrimination reported by people with major depressive disorder: a cross-sectional survey. Lancet 2013 Jan 5;381(9860):55-62. [doi: 10.1016/S0140-6736(12)61379-8] [Medline: 23083627]
51. Schomerus G, Matschinger H, Angermeyer MC. Familiarity with mental illness and approval of structural discrimination against psychiatric patients in Germany. J Nerv Ment Dis 2007 Jan;195(1):89-92. [doi: 10.1097/01.nmd.0000252314.45371.92] [Medline: 17220746]
52. Krajewski C, Burazeri G, Brand H. Self-stigma, perceived discrimination and empowerment among people with a mental illness in six countries: pan European stigma study. Psychiatry Res 2013 Dec 30;210(3):1136-1146. [doi: 10.1016/j.psychres.2013.08.013] [Medline: 23998361]
53. Roeloffs C, Sherbourne C, Unützer J, Fink A, Tang L, Wells KB. Stigma and depression among primary care patients. Gen Hosp Psychiatry 2003;25(5):311-315. [Medline: 12972221]
54. Greene K, Banerjee S. Disease-related stigma: comparing predictors of AIDS and cancer stigma. J Homosex 2006;50(4):185-209. [doi: 10.1300/J082v50n04_08] [Medline: 16723345]

55. Pershad Y, Hangge PT, Albadawi H, Oklu R. Social medicine: Twitter in healthcare. J Clin Med 2018 May 28;7(6):1-9 [FREE Full text] [doi: 10.3390/jcm7060121] [Medline: 29843360]
56. Widmer RJ, Engler NB, Geske JB, Klarich KW, Timimi FK. An academic healthcare twitter account: the Mayo Clinic experience. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):360-366. [doi: 10.1089/cyber.2015.0607] [Medline: 27327062]
57. Ross A, Morgan A, Jorm A, Reavley N. A systematic review of the impact of media reports of severe mental illness on stigma and discrimination, and interventions that aim to mitigate any adverse impact. Soc Psychiatry Psychiatr Epidemiol 2019 Jan;54(1):11-31. [doi: 10.1007/s00127-018-1608-9] [Medline: 30349962]
58. Green Lauridsen M, Kälvemark Sporrong S. How does media coverage effect the consumption of antidepressants? A study of the media coverage of antidepressants in Danish online newspapers 2010-2011. Res Social Adm Pharm 2018 Dec;14(7):638-644. [doi: 10.1016/j.sapharm.2017.07.011] [Medline: 28811152]
59. McGinty E, Kennedy-Hendricks A, Choksy S, Barry C. Trends in news media coverage of mental illness in the United States: 1995–2014. Health Aff (Millwood) 2016 Dec 1;35(6):1121-1129 [FREE Full text] [doi: 10.1377/hlthaff.2016.0011] [Medline: 27269031]
60. Caqueo-Urízar A, Boyer L, Urzúa A, Williams D. Self-stigma in patients with schizophrenia: a multicentric study from three Latin-America countries. Soc Psychiatry Psychiatr Epidemiol 2019 Feb 26:1-5. [doi: 10.1007/s00127-019-01671-4] [Medline: 30806726]
61. Gupta S, Isherwood G, Jones K, Van Impe K. Assessing health status in informal schizophrenia caregivers compared with health status in non-caregivers and caregivers of other conditions. BMC Psychiatry 2015 Jul 21;15:162 [FREE Full text] [doi: 10.1186/s12888-015-0547-1] [Medline: 26194890]
62. Daley S, Murray J, Farina N, Page TE, Brown A, Basset T, et al. Understanding the quality of life of family carers of people with dementia: development of a new conceptual framework. Int J Geriatr Psychiatry 2019 Dec;34(1):79-86. [doi: 10.1002/gps.4990] [Medline: 30251443]
63. Booth RG, Allen BN, Bray Jenkyn KM, Li L, Shariff SZ. Youth mental health services utilization rates after a large-scale social media campaign: population-based interrupted time-series analysis. JMIR Ment Health 2018 Apr 6;5(2):e27 [FREE Full text] [doi: 10.2196/mental.8808] [Medline: 29625954]
64. Saha K, Weber I, Birnbaum M, de Choudhury M. Characterizing awareness of schizophrenia among Facebook users by leveraging Facebook advertisement estimates. J Med Internet Res 2017 Dec 8;19(5):e156 [FREE Full text] [doi: 10.2196/jmir.6815] [Medline: 28483739]

Abbreviations

PRT: psychosis-related tweets

Edited by G Eysenbach; submitted 22.03.19; peer-reviewed by D Lampropoulos; comments to author 23.04.19; revised version received 29.04.19; accepted 30.04.19; published 28.05.19.

Please cite as:
Alvarez-Mon MA, Llavero-Valero M, Sánchez-Bayona R, Pereira-Sanchez V, Vallejo-Valdivielso M, Monserrat J, Lahera G, Asunsolo del Barco A, Alvarez-Mon M
Areas of Interest and Stigmatic Attitudes of the General Public in Five Relevant Medical Conditions: Thematic and Quantitative Analysis Using Twitter
J Med Internet Res 2019;21(5):e14110
URL: http://www.jmir.org/2019/5/e14110/
doi:10.2196/14110
PMID:31140438

Original Paper

Exploring the Extent of the Hikikomori Phenomenon on Twitter: Mixed Methods Study of Western Language Tweets

Victor Pereira-Sanchez[1*], MD; Miguel Angel Alvarez-Mon[1*], MD; Angel Asunsolo del Barco[2,3,4], MD, PhD; Melchor Alvarez-Mon[3,5,6], MD, PhD; Alan Teo[7,8,9], MD, MS

[1]Department of Psychiatry, Clinica Universidad de Navarra, Pamplona, Spain

[2]Department of Surgery, Medical and Social Sciences, University of Alcala, Madrid, Spain

[3]Instituto Ramon y Cajal de Investigaciones Sanitarias, Madrid, Spain

[4]Department of Epidemiology & Biostatistics, Graduate School of Public Health and Health Policy, University of New York, New York, NY, United States

[5]Department of Medicine and Medical Specialities, University of Alcala, Madrid, Spain

[6]Service of Internal Medicine, Autoimmune Diseases and Rheumatology, Hospital Universitario Principe de Asturias, Madrid, Spain

[7]Department of Psychiatry, Oregon Health & Science University, Portland, OR, United States

[8]School of Public Health, Oregon Health & Science University and Portland State University, Portland, OR, United States

[9]Center to Improve Veteran Involvement in Care, VA Portland Health Care System, Department of Veterans Affairs (VA), Portland, OR, United States

[*]these authors contributed equally

Corresponding Author:
Miguel Angel Alvarez-Mon, MD
Department of Psychiatry
Clinica Universidad de Navarra
Avda Pio XII, 36
Pamplona, 31008
Spain
Phone: 34 948255400
Email: maalvarezdemon@icloud.com

Abstract

Background: Hikikomori is a severe form of social withdrawal, originally described in Japan but recently reported in other countries. Debate exists as to what extent hikikomori is viewed as a problem outside of the Japanese context.

Objective: We aimed to explore perceptions about hikikomori outside Japan by analyzing Western language content from the popular social media platform, Twitter.

Methods: We conducted a mixed methods analysis of all publicly available tweets using the hashtag #hikikomori between February 1 and August 16, 2018, in 5 Western languages (Catalan, English, French, Italian, and Spanish). Tweets were first classified as to whether they described hikikomori as a problem or a nonproblematic phenomenon. Tweets regarding hikikomori as a problem were then subclassified in terms of the type of problem (medical, social, or anecdotal) they referred to, and we marked if they referenced scientific publications or the presence of hikikomori in countries other than Japan. We also examined measures of interest in content related to hikikomori, including retweets, likes, and associated hashtags.

Results: A total of 1042 tweets used #hikikomori, and 656 (62.3%) were included in the content analysis. Most of the included tweets were written in English (44.20%) and Italian (34.16%), and a majority (56.70%) discussed hikikomori as a problem. Tweets referencing scientific publications (3.96%) and hikikomori as present in countries other than Japan (13.57%) were less common. Tweets mentioning hikikomori outside Japan were statistically more likely to be retweeted (P=.01) and liked (P=.01) than those not mentioning it, whereas tweets with explicit scientific references were statistically more retweeted (P=.01) but not liked (P=.10) than those without that reference. Retweet and like figures were not statistically significantly different among other categories and subcategories. The most associated hashtags included references to Japan, mental health, and the youth.

Conclusions: Hikikomori is a repeated word in non-Japanese Western languages on Twitter, suggesting the presence of hikikomori in countries outside Japan. Most tweets treat hikikomori as a problem, but the ways they post about it are highly heterogeneous.

(J Med Internet Res 2019;21(5):e14167) doi:10.2196/14167

KEYWORDS

social isolation; loneliness; hikikomori; hidden youth; social media; Twitter; social withdrawal

Introduction

Hikikomori is the romanization of a Japanese concept referring to a complex phenomenon characterized by severe social withdrawal [1-3]. Definitions of hikikomori have evolved slightly over time [4], but in general, individuals with hikikomori are defined by their pattern of social isolation, remaining usually at their homes, which includes significant distress or functional impairment (eg, inability to maintain academic studies or a job) and a duration of at least 6 months [5-7].

This concept was developed in the last decades of the 20th century in Japan [2], where it is widely recognized as a heterogeneous psychiatric condition, with an estimated prevalence greater than 1% among adults in that country [8]. The extent of this condition has become a major source of concern among many health professionals and policymakers in Japan, as it usually affects young individuals vulnerable to psychological distress, social stigma or cultural marginalization [9], worse physical health [10], and loss of opportunities for education and work. Consequently, hikikomori implicates suffering for not only the affected individual but also his or her family [11], and, at a large scale, it jeopardizes the labor market and public health [12].

In recent years, clinicians and researchers across the world have reported the existence of patients with similar patterns of severe social withdrawal [13], including in Hong Kong [14], mainland China [15], South Korea, India and the United States [6], Oman [16], Spain [17-20], Italy [21], France [22,23], and Brazil [24]. Recently, researchers in Spain characterized 190 patients with social withdrawal meeting the definition of hikikomori, constituting one of the largest described such cohorts outside of Japan [17]. Together, these observations have challenged the view of hikikomori as a syndrome restricted to Japan, a revelation that was hinted at years ago [4]. These studies also raise questions as to whether the extent of hikikomori varies across countries and points in time.

Despite the growing observation of hikikomori globally, there has been an alternative viewpoint that hikikomori does not represent a form of psychopathology; rather, it should be considered a *nonproblematic* self-imposed *lifestyle* of isolation [25]. This consideration of hikikomori frequently falls in the discussion of the so-called *not in employment, education nor training* (NEET) individuals [9]. Accordingly, some authors have suggested that hikikomori might be a reaction to the pressures of a globalized and postindustrialized Japanese society from individuals who consciously refuse to adopt the mainstream cultural values, perhaps a *social pathology* rather than a psychiatric one [25,26].

Individuals with *hikikomori* are usually a hard-to-reach population owing to the fact that hikikomori's defining feature (social isolation) prevents or delays the presentation of these individuals to clinical care and their involvement in research. Some have hypothesized that hikikomori and internet addiction are closely correlated, under the presumption that individuals with hikikomori spend much of their time retreating to anonymous and impersonal Web spaces [27]. Accordingly, the internet, and particularly social media, may be a unique space where these individuals might seek peer help and where they could be reached, identified, and supported by mental health professionals.

Twitter is one of the most popular social media platforms in Western countries and allows users to publicly share and interact with short posts (tweets) [28]. This environment has been the focus of an increasing amount of quantitative and qualitative medical research, using a wide range of different approaches, from descriptive metrics of tweets to manual or machine-learning content analysis regarding disorders such as Alzheimer’s disease [29], epilepsy [30], breast cancer [31], anorexia nervosa [32], schizophrenia [33], and depression [34]. Twitter is a unique venue to study the feelings, beliefs, knowledge, and behaviors of large numbers of people, particularly young ones, and has been proposed as a great source of infodemiologic data to survey, track, and predict medical problems [35-38]. In addition, previous research from our team has highlighted the increasing interest among Twitter users in psychiatric disorders [28]. Thus, Twitter constitutes a rich social context and online community to explore mental disorders and to identify and reach otherwise hard-to-reach individuals [39,40]. Using Twitter in the study of hikikomori offers several potential strengths. Aspects such as language of tweets (as a proxy of users’ geographical and cultural background) and time trends can be examined. In addition, content analysis could be used to classify tweets into different topics or categories of interest. Finally, analyzing Twitter content related to hikikomori could inform the development of future social media–based interventions for individuals with hikikomori to provide users with accurate information, fight stigma, and reach a target population, which might be unlikely to leave their rooms to ask for help by themselves. These kinds of interventions, using other social media platforms, have started to prove effective in people with hikikomori [27] and other potentially margined populations such as the military veterans [41].

The primary aims of this study are to (1) describe and categorize the content of tweets regarding hikikomori in several Western languages; (2) identify what content related to hikikomori generates the most interest (retweets, likes, and associated hashtags); and (3) explore temporal trends in hikikomori on Twitter.

Methods

Study Design and Data Source

This study was designed as a mixed methods analysis of quantitative Twitter metrics and qualitative content from recent publicly available tweets about hikikomori in languages used in Western countries where hikikomori has been described. The inclusion criteria for tweets were (1) being public (nonprivate); (2) use of the hashtag #hikikomori; (3) posted between February 1 and August 16, 2018; and (4) text in English, Italian, Spanish, Catalan, or French. The exclusion criteria were (1) no identifiable language or (2) only contained a link (ie, spam tweets).

Twitter provides 3 primary sources of data: Twitter's Search application programming interface (API), Twitter's Streaming API, and Twitter's Firehose. Twitter's Firehose is the only one that has access to 100% of Twitter content. Twitter's Firehose formerly was handled by multiple data providers (eg, Gnip, DataSift, and Topsy), although, since August 2015, Twitter only allows access to Twitter's Firehose through Gnip [28,42]. Tweet Binder, the search engine we employed, uses Twitter Firehose [43] and allows access to 100% of all public tweets that match a set of search criteria (query) [44]. In addition, Tweet Binder provided a full report with general metrics from all the tweets with #hikikomori in the defined period of time. This includes the total number of tweets and retweets, date and time tweeted, as well as (for this study) other secondary metrics: the temporal trend of the tweets, potential reach (the number of unique users that could have read the hashtag), and potential impact (the number of times that somebody could have read the hashtags).

Figure 1 shows a flowchart illustrating the process we followed for the analysis of the tweets, along with the number of tweets included and excluded.

Content Analysis Process and Creation of the Codebook

All the retrieved tweets were directly inspected by 2 raters fluent in the included languages (VPS and MAM): both of them were psychiatric trainees and had previous experience in Twitter-related research, including the analysis of tweets. First, we scanned all of the tweets to classify them by language and excluded 386 of the total of 1042 tweets, according to our exclusion criteria. We created a codebook based on our research questions, previous experience analyzing tweets, and also determined by the most common themes we had observed reading the tweets. VPS and MAM analyzed 252 tweets separately to test the codebook. After an agreement on the codebook, the 2 raters classified, independently, a random set of 80 tweets (40 in English and 40 in Spanish), and the interrater reliability between both raters was assessed obtaining Kappa values ranging from 0.28 to 0.83 for the different categories and subcategories. Discrepancies were discussed between the raters and with the senior author (AT), and after revising the codebook, the interrater reliability was reassessed with a different set of 80 randomly selected tweets (40 in English and 40 in Spanish). As this resulted in adequate Kappa values ranging from 0.71 to 1.00 [45], the raters then proceeded to code the remaining tweets. For the remaining tweets, one of the raters (VPS) coded the tweets in Spanish, Italian, Catalan, and French, whereas the other rater (MAM) coded the tweets in English. One rater manually compiled the count of hashtags other than #hikikomori associated with the tweets.

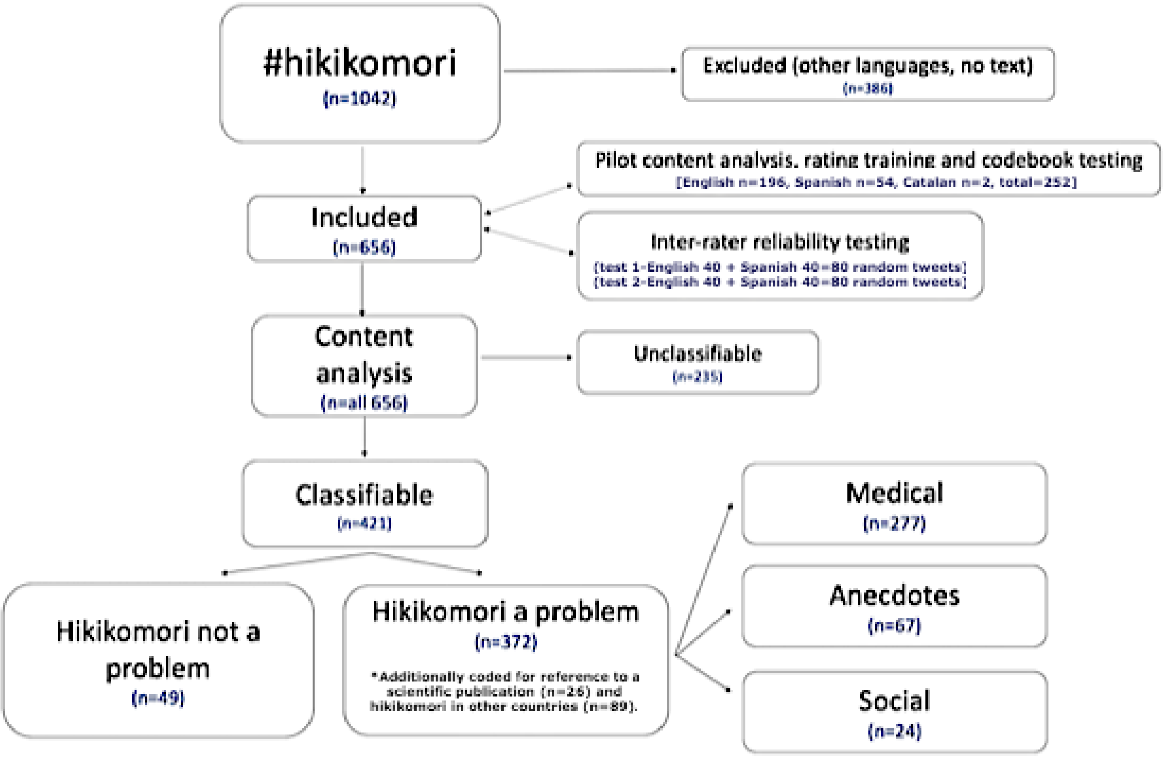


Figure 1. Flowchart of data management and content analysis.

Table 1. Category and subcategory definitions and examples of classification. Links and usernames within the tweets have been removed (usernames and personal names were replaced by XXXX). To seek for clarity, all the tweets reported here are in English (when tweets in other languages are reported, an English translation prepared by us is included in parentheses).

Category (or subcategory)	Definition	Examples of tweets
Unclassifiable	Not enough information, only links, spam, or random content.	"Furries is Love Furries is Life #Furries #Hikikomori"; "damn... that's why I play this game again... kawaii character :D; #WhoCares #Lunatic #Hikikomori"
Hikikomori not as a problem	Positive or indifferent thoughts, attitudes, or behaviors related to hikikomori.	"I'm a hikikomori but I think outside the room. But many extroverts think like birds in a cage. #hikikomori"; "About to go full #Hikikomori. No regrets"
Hikikomori as a problem		
Medical	Medical publications or reports, or events, campaigns, or interventions that present data or information related to hikikomori.	"Can we use #SocialMedia to identify socially withdrawn youth in China? Our latest paper on #hikikomori now out in @XXXX"; "#japan, Doctors began to observe it in the mid-1980s, with young men suffering from lethargy and refusing to communicate -- #hikikomori an insight #psychological ailment".
Anecdotes	Personal stories, testimonials, or third-person reports of people with hikikomori or related behaviors.	"#Hikikomori literally means 'withdrawal from others' in Japanese—follow the stories of a family affected by this modern-age social phenomenon as one day Nils decides to hide in his room and never leave. #Week53".
Social	Socially oriented issues related to hikikomori, including antistigma events, provision of social support, or related activities.	"If you are experiencing #hikikomori chat to others on #joinin247 #london #tokyo #osaka #kyoto #isolation #youarenotalone #endthesilence"
Scientific reference	Explicit reference to a scientific publication (paper, presentation) in tweets with hikikomori as a problem.	"Can we use #SocialMedia to identify socially withdrawn youth in China? Our latest paper on #hikikomori now out in @XXXX"; "Secure Base Script and Psychological Dysfunction in Japanese Young Adults (Umemura et al 2018) #hikikomori via @XXXX"
Other country	Explicit reference to hikikomori as a problem in a country other than Japan.	"#Hikikomori, è boom anche in Italia: migliaia di giovani si auto-recludono in casa". ("#Hikikomori is also a boom in Italy: thousands of youth are self-reclusive at home"); "Can we use #SocialMedia to identify socially withdrawn youth in China? Our latest paper on #hikikomori now out in @XXXX".

Table 1 illustrates the final content classification (codebook), providing the definitions and some examples of tweets coded in each category and subcategory. Regarding the tweet texts in the included languages, the first distinction was between *unclassifiable* and *classifiable* tweets. *Unclassifiable* tweets were no further analyzed, whereas the *classifiable* were then split between those referencing hikikomori *not as a problem* and those referring to hikikomori *as a problem*. The latter were subsequently coded in the subcategories of *medical*, *anecdotes*, and *social*. Finally, all the classifiable tweets were assessed to identify explicit references to scientific contents and reports to the existence of hikikomori or related behaviors in countries other than Japan (in that case, we recorded the countries explicitly referenced). In case of finding contents repeated exactly or almost identically in different tweets, they were classified in the same way as the first tweet encountered.

All the tweets were statistically analyzed to describe the number of tweets, retweets, and likes per language and category (and subcategory), considering retweets and likes as indices for reflecting the users' interests in particular topics. We had previously reported the value of retweets in this regard [28], so we further calculated the Spearman correlation between retweets and likes for all the tweets to assess whether the likes could provide similar information. These analyses were conducted with the software packages STATA v14 (StataCorp) and SPSS Statistics v23.0.0.0 (IBM Corp).

This study received the approval of the University of Navarra Research Ethics Committee (October 11, 2018, modified on December 13, 2018) and is compliant with the research ethics principles of the Declaration of Helsinki (seventh revision, 2013). This study did not directly involve human subjects, nor did it include any intervention; instead uses only publicly available tweets (subject to universal access through the internet according to the Terms of Service that all users in Twitter accept). Nevertheless, we have taken care to not directly reveal in this report any username, and we have avoided citing tweets that could be offensive or compromised to someone.

Results

Our search tool provided 1042 original tweets using #hikikomori in the established period, with 1433 retweets, a potential reach of 7,974,329, 10,613,856 potential impacts, and 908 contributors (ie, total number of different users posting with a given hashtag). As shown in Figure 1, our content analysis included 656 tweets (62.3% of the initial dataset). Of the total of 656 included tweets, 421 (64.8%) were considered classifiable. From these 421 classifiable tweets, 372 (88.36%) considered hikikomori as a problem.

Table 2 shows the numbers and percentages of tweets per language and (sub)category. The distribution of tweets in each category was significantly different ($P<.001$) among languages.

English and Italian were the most used languages (76.36%). The proportion of tweets referring to hikikomori as a nonproblematic behavior or lifestyle was highest in Spanish/Catalan tweets and lowest in Italian tweets. Italian was the language with the highest percentage of tweets considering hikikomori as a problem. Among tweets considering hikikomori as a problem, the medical contents were higher in all the languages in comparison with contents related to anecdotes or social. The language with the highest proportion of medical content was Italian, anecdotes was English, and social content was Italian. The proportion of tweets with explicit scientific references was very low in all languages; the highest proportion was found in English tweets. Only a minority of the tweets mentioned hikikomori as a problem outside Japan in all the languages except for Italian, where almost a third of the tweets had this reference. Figure 2 shows a world map including the countries other than Japan where hikikomori was explicitly reported as a problem.

The probability of retweet and like per category and subcategory is presented in Table 3. Tweets with an explicit reference to hikikomori outside Japan were statistically more retweeted and liked than those without that reference, whereas tweets with explicit scientific references were statistically more retweeted (but not more liked) than those tweets without that reference.

Table 2. Descriptive characteristics of the original tweets included in the analysis, categorized by total amount per language and category. For each language and category or subcategory, total number of tweets (n) and relative proportions (%) are provided. In the first row, percentages of the total tweets in each language is calculated over the total of included tweets (ie, percentage of tweets in a given language among the total number of included tweets, 656); in the following rows, percentages are calculated for each category over the figures in the first row and in the same column (ie, percentage of tweets from a category among the total tweets in the given language or percentage of total tweets in each category among the total number of included tweets). Percentages are rounded to two decimals.

Category	English, n (%)	Italian, n (%)	Spanish and Catalan, n (%)	French, n (%)	Total, n (%)
Total tweets	290 (44.2)	211 (32.16)	61 (9.29)	94 (14.32)	656 (100)
Unclassifiable	125 (43.1)	54 (25.56)	19 (31.14)	37 (39.36)	235 (35.82)
Classifiable	165 (56.89)	157 (74.44)	42 (68.86)	57 (60.64)	421 (64.18)
Not a problem	30 (10.34)	0 (0)	13 (21.31)	6 (6.38)	49 (7.46)
Problem					
Any	132 (45.51)	156 (73.93)	29 (47.54)	51 (54.25)	372 (56.7)
Medical	79 (27.24)	132 (62.56)	19 (31.14)	47 (50)	277 (42.22)
Anecdotes	43 (14.83)	13 (6.16)	8 (13.11)	3 (3.19)	67 (10.21)
Social	10 (3.45)	11 (5.21)	2 (3.28)	1 (1.06)	24 (3.66)
Scientific reference	18 (6.21)	5 (2.37)	1 (1.64)	2 (2.13)	26 (3.96)
Other country	16 (5.18)	64 (30.33)	2 (3.28)	7 (7.45)	89 (13.57)

Figure 2. World map with countries where the presence of hikikomori was explicitly referenced by the tweets (Italy=63 tweets, United States=7 tweets, France=6 tweets, China=4 tweets, United Kingdom=3 tweets, and South Korea, India, Tunisia, and Spain=1 tweet each).

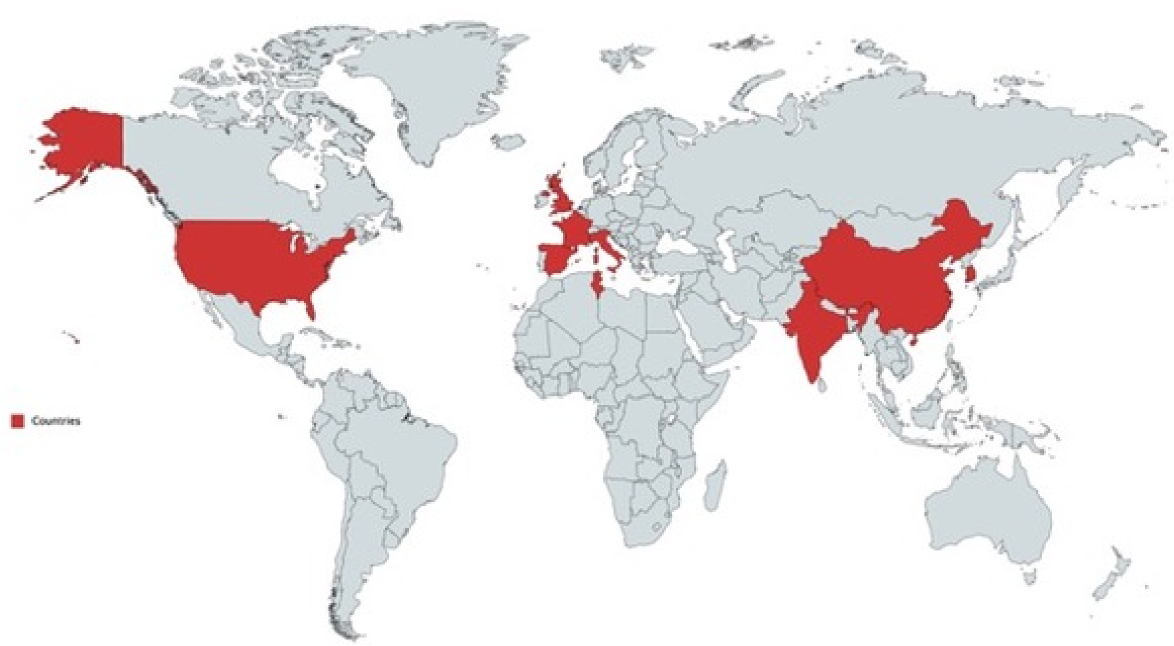


Table 3. Retweet to tweet ratio per category and subcategory.

Category and subcategory	Tweets, N	Retweets				Likes			
		N	Mean (SD)	Median	*P* value	N	Mean (SD)	Median	*P* value
Problem					.23				.82
No	49	140	2.86 (11.00)	1		171	3.49 (16.54)	0	
Yes	372	663	1.78 (2.66)	1		507	1.36 (3.84)	0	
As a problem					.51				.65
Medical	281	518	1.84 (2.82)	1		379	1.35 (4.11)	0	
Anecdotes	67	109	1.63 (2.38)	1		110	1.64 (3.25)	0	
Social	24	36	1.5 (1.10)	1		18	0.75 (0.90)	0.5	
Scientific reference					*.01*				.10
No	346	595	1.72 (2.67)	1		413	1.19 (3.21)	0	
Yes	26	67	2.58 (2.56)	1.5		90	3.46 (8.46)	1	
Reference to another country					*.02*				*.02*
No	283	476	1.68 (2.81)	1		330	1.17 (3.41)	0	
Yes	91	187	2.05 (2.11)	1		174	1.91 (4.89)	1	

[a] Number of retweets and likes per category and subcategory of all tweets in the included languages (here not separated by language), along with its mean (SD) and median values. Mann-Whitney *U* and Kruskall-Wallis tests were conducted to assess for statistical differences for the codification of each category. Statistical significance was considered when *P*<.05, and significant values are italicized.

There were no differences in retweets and likes ratios among the different subcategories that considered hikikomori as a problem. The distribution of retweets per language followed an asymmetric distribution, with some individual tweets receiving a large number of retweets and likes; therefore, a statistical comparison of tweets and likes per language would be unreliable. Interestingly, accounting for all the tweets in the sample, and a secondary research result, the indices of retweet and like showed a moderate (Spearman’s rho: 0.54) and significant (*P*<.001) correlation (Figure 3).

Table 4 includes the *top 5* most frequently associated hashtags per language group. Looking at all languages, the most repeated words include references to Japan, the youth, and mental health or psychology.

The number of times each hashtag appears is shown in parentheses. Less than 5 are reported if the total number of associated hashtags per language was lesser than that number. More than 5 are reported when a tie occurred. For hashtags in languages other than English, an English translation is provided in parentheses. An interpretation of the hashtags is provided in the Discussion section.

Finally, Figure 4 depicts the graphical timeline of tweets during the time frame selected, also as a secondary analysis. No temporal time trend pattern can be clearly identified.

Figure 3. Correlation between retweets and likes among all the tweets in the sample (1042), showing a moderate (Spearman's rho: 0.54) and significant (*P*<.001) correlation between the 2 indices.

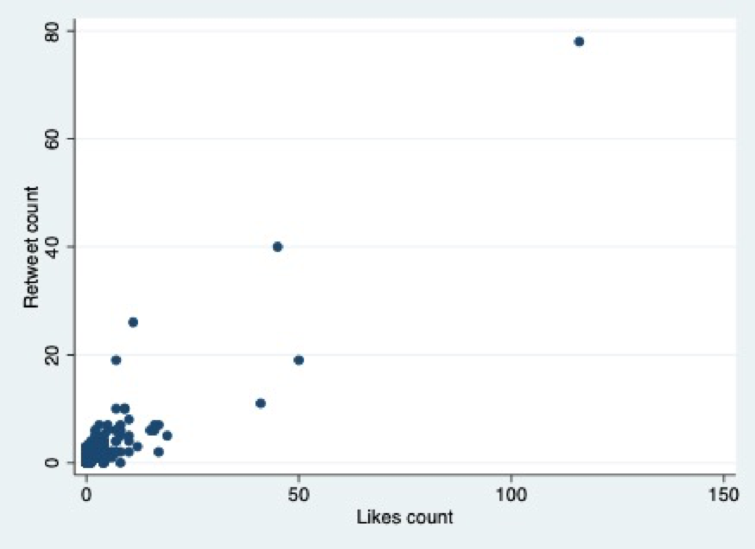
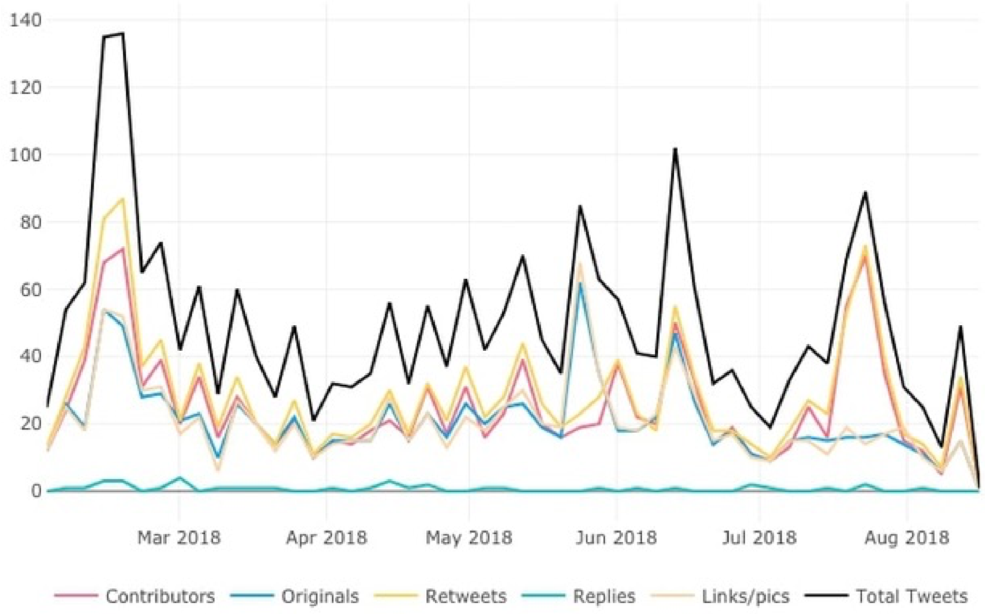


Table 4. Top 5 hashtags associated with classifiable #hikikomori tweets, according to the language of the tweet. The values in brackets represent the number of tweets in the sample with the corresponding hashtags.

Language	*Top*[a] related hashtags
English	#japan (32), #neet (10), #culture (9), #otaku (7), #mentalhealth (7)
Italian	#giovani (youth) (24), #italia (Italy) (16), giappone (Japan) (9), #isolamento (isolation) (8), #adolescenti (adolescents), #asocialita (asociality), #stareindisparte (being apart) and #autorrecludono (self-reclusion) (5)
Spanish	#japon (Japan) (5), #psicologia (psychology), #neetlife (4), #depresion (depression), #videojuegos (videogames), #anime, #adiccionalosvideojuegos (addiction to the videogames) (2)
French	#societe (society) (4), #sante (health) (2)

[a] *Top* hashtags refer to the most prevalent hashtags (different from #hikikomori) included in the tweets of the study.

Figure 4. Time trend of all the tweets with #hikikomori in the period of study. The graph includes the number of contributors (different Twitter users publishing with this hashtag), original tweets, retweets, replies (tweets published as replies to the tweets with the hashtag), links and pictures included, and total number of tweets (original+retweets).



Discussion

Comments on the Results

In this study, we investigated the tweets generated about hikikomori in several Indo-European languages (English, Italian, Spanish, Catalan, and French) which are mainly spoken in Western countries and altogether account for more than a billion native speakers. Our analysis strategy with Twitter Firehose data stream allows access to 100% of all public tweets within the search limits of words and time [44]. Thus, the conclusions were obtained from the results measured in the total population of tweets in those languages between Februray and mid-August 2018, and were not deduced from the analysis of a reduced sample.

To our best knowledge, this is the first study related to hikikomori in Twitter and the first to apply a mixed quantitative-qualitative approach to analyze tweets in different languages. English and Italian were the most used languages among those analyzed. A fair amount of the tweets could not be analyzed, which might be due to lack of information or context, or a random or nonsense use of the word hikikomori (in several cases, probably, as spam). Among the classified tweets, a minority described a perception of hikikomori or related behaviors as nonproblematic, perhaps describing a self-imposed lifestyle. Conversely, the majority of the tweets reported hikikomori as a problem, mainly in general terms (as an alarming social phenomenon or as a psychopathology), and at a lesser extent, some included first- or third-person testimonials, whereas the fewer were related to solidarity/activism. Explicit scientific references were barely found, but more than one-tenth of the tweets reported hikikomori in countries other than Japan, with a striking majority referring to Italy. Among the classifiable tweets, top associated hashtags were mainly related to Japan, the youth, the mental health psychology, and the society and culture. The time trend did not show markedly differentiated peaks of activity related to the hashtag.

Recent publications have proven the role of social media as a target for medical research and interventions [46,47]. Mental health disorders and conditions are topics of increased interest in Twitter [28], and this platform could be an enormous niche where many young people and socially withdrawn individuals could be reached. Our study tried to identify users reporting hikikomori-like symptoms or behaviors in themselves or others, in the search of potential targets for social media–based interventions. We did find tweets reporting these symptoms and behaviors in the categories of *not a problem* and *as a problem, anecdotes*. The similar number of tweets found in both categories (combined, up to 18% of the included tweets) shows that many users disclosing these symptoms or behaviors do not perceive them as a problem: this fact might be related with the common perspective of hikikomori as a self-imposed nonpathological lifestyle [25]. It furthermore provides an insight of the potential value of Twitter as an arena where affected individuals seek to revendicate their way of living or seeking for help.

In addition, Twitter might be considered as an indicator of real-time public opinion [46,48], as a source of general and medical information, and a framework for peer support through an online social community [49]. In this context, we thought that a mixed quantitative and qualitative analysis of tweets in Western languages related to hikikomori would be able to explore the public perceptions regarding this condition and the extent of this phenomenon within a global perspective. Our results suggest that hikikomori is not univocally perceived by users, probably existing a wide confusion in how it is widely understood. They also reflect that, whereas it is often perceived as a problem related to Japan, it might exist in other countries in the world, with particularly alarming references to Italy.

It should be noted that, currently, the literature related to medical research in hikikomori is also heterogenous and the concept has not yet reached an international consensus regarding its nature as a social phenomenon, a cultural-bond syndrome, or a psychopathological symptom or disorder per se [4]. Despite the increased evidence of its existence in countries other than Japan, especially in East Asia and in the West [1], and its frequent identification in mental health settings, the conceptual diversity, ignorance, and confusion in relation to how this problem is perceived in the psychiatric community seem high and should be addressed with cross-cultural international research.

Although few tweets included personal testimonies, their existence proves the value of Twitter as a means of communicating this kind of contents to a large public in spite of the still present public and self-stigma toward psychiatric conditions [50]. However, as Twitter allows for anonymity, it might be preferred by people with real or perceived personal or social restrictions, what makes this social network an ideal arena to discuss this topic. The lack of tweets with explicit scientific contents, which is probably, in part, related to our unavailability to explore the links attached, contrasted with its statistically significative higher chance to be retweeted (but not liked). These results are a stimulus to encourage researchers, editorials, and mass media outlets to explicitly highlight the results of scientific works in the social media, thus both offering evidence-based information to the general public and increasing the academic impact (ie, increasing the chance of citation) of the tweeted papers [51].

The global extension of the hikikomori phenomenon was confirmed in our analysis, as this term was used to name a problem existing in some countries across the world, and, despite the relatively low amount of tweets with these contents, they elicited a significant interest among users, as reflected by their statistically higher chance to be retweeted and liked. The increased interest in the emergence of hikikomori in countries other than Japan (especially in Italy, with some tweets reporting very high figures of prevalence in that country) reflected in retweets and likes of these contents contrasts with the still scarce published scientific literature of hikikomori and the lack of epidemiological figures and intervention guidelines in non-Japanese countries. Twitter is a fast social media reflecting real-time events and opinions; so, despite the heterogeneity and unnacuracy of many contents, clinicians, researchers, and policymakers should take them into account to address the problems and worries reflected in this social media and to rapidly detect and intervene on various health conditions [39].

We could not futher characterize the contents of the tweets addressing hikikomori as a medical problem, but they included news, epidemiological figures, opinions, research reports, and interventions. It would be valuable to explore the medical contents more in detail and compare them with the Twitter research in other health conditions. In addition, hikikomori's core symptoms make social media a valuable tool to reach these patients [27]. As it has been shown that interventions in Twitter can modify health-related behavior [52], this platform could be a good scenario to promote a healthy lifestyle among at-risk individuals and encourage hikikomori patients to get in touch with offline health providers.

In accordance with previous research in Twitter, we considered retweets as a measure of users' particular interest in a topic, which might be associated to the emotions elicited by the tweets in them [53]. In addition, we have explored the value of likes for the same purpose and found a moderate positive correlation between both indices. Further research could assess if retweets and likes are interchangeable metrics of interest or present different particularities between them.

Limitations and Strengths

Some limitations should be noted in this study and need to be discussed in the light of its strengths. Most importantly, the codebook design and text analysis by 2 raters (with a third author as a supervisor), altogether with the mixed nature of this study (with a qualitative plus a quantitative approach), imply a degree of subjectivity, disagreement, and human error and constitute a challenge for its potential reproducibility by other authors. To address this issue, the study comprised a series of steps of initial review, design, and test of pilot codebooks and measurements of the interrater reliability. Thus, the process has been consistent, and the final ratings were expected to be reliable between the 2 raters. Consequently, in our opinion, this methodology is consistent with previous medical research studies in Twitter [54,55] and could be applied to different topics and by different authors. Although computerized machine-learning methods have been tested to automatically identify and classify topics in medical research in social media [56], we counted on the clinical expertise of the raters in Mental Health, which constitutes a qualitative advantage in relation to these automatized strategies.

To note, #hikikomori is not probably the unique word used in Twitter to reference this phenomenon (particularly outside Japan, where this concept may be unknown and where other words such as social isolation and withdrawal might be preferred), and tweets in Japanese (where one would expect to find the majority of the contents related to hikikomori) were not considered in this study and thus remain as a source of future cross-cultural research. However, although hikikomori is used as a limited term outside Japan (people might use other words to name it), this concept could also be a potential self-identification term for people who suffer from this problem and have no other words to name it [4].

The time frame was somewhat arbitrary but was selected to expect a reasonable number of tweets to classify. Apparently, the time trends of the retrieved tweets did not suggest the existence of a clear temporal pattern. A time trend was described in this study to rule out the presence of evident peaks in the activity of Twitter users regarding this topic, which has been observed in other medical conditions such as breast cancer during the so-called *awareness months* [57]. Many tweets were initially discarded owing to its language, according to the possibilities of the raters, so we could not analyze the contents of many tweets. Finally, the content analysis of tweets found the expected challenge of retrieving enough information to classify them. As tweets are limited in words and require no standard of linguistic or content correction to be published by anyone, and as we decided to avoid using information not included in the text itself (such as links or pictures, to limit the subjectivity in the rating, and for the sake of analytic efficiency), many tweets were rated as unclassifiable because of lacking enough information. Finally, the relatively small sample of tweets considered for content analysis precluded the possibility to separate independent samples for pilot testing of codebooks and interrater reliability, so the final classification might be somewhat biased. Nevertheless, the consistency of our method and the wide variety of tweets make us confident of having conducted a valuable mixed quantitative-qualitative exploration of this topic in a social media global context.

Conclusions

In conclusion, hikikomori is a repeated word in different Western languages in Twitter, and despite its frequent use for uncertain or nonsense purposes, it is markedly perceived as a problem with strong associations with Japan, the society, the youth, and isolation. In addition, Twitter is a means to report personal stories, scientific publications, and the presence of this problem in non-Japanese societies. Our results provide a framework to take advantage of Twitter to provide users with accurate information, fight stigma, and reach a target population which might be unlikely to leave their rooms to ask for help by themselves. These kinds of interventions, using other social media platforms, have started to prove effective in people with hikikomori [27] and other potentially margined populations such as the military veterans [41]. Further research should take a cross-cultural perspective looking at tweets and other kinds of social media contents in Japanese and other Eastern and Western languages and test the potential of Web-based interventions to reach individuals with hikikomori or related behaviors and offer them support.

Acknowledgments

AT’s work was supported in part by a Career Development Award from the Veterans Health Administration Health Service Research and Development (HSR&D; CDA 14-428). The US Department of Veterans Affairs had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. The findings and conclusions in this document are those of the authors who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the US Department of Veterans Affairs or the United States government. We want to acknowledge the Japanese Society of Psychiatry and Neurology for the 2018 Fellowship Award granted to VPS and for its encouragement to work on international research studies on hikikomori. Ms Teresa Abrego and Ms Maite Muruzabal, from Tweetbinder, SA, collaborated significantly in the retrieval of tweets through their search engine.

This work was partially supported by grants from the *Fondo de Investigación de la Seguridad Social*, Instituto de Salud Carlos III (PI18/01726), Spain, and the *Programa de Actividades de I+D de la Comunidad de Madrid en Biomedicina* (B2017/BMD-3804), Madrid, Spain.

Authors' Contributions

VPS and MAM participated equally as principal contributors in the research design, content analysis, statistical supervision and manuscript writing, review, and submission. AA conducted and reported the statistical analysis. MA contributed as the reviewer of the manuscript. AT contributed as the main supervisor of all the cited stages, with a special involvement in the study design and manuscript writing.

Conflicts of Interest

None declared.

References

1. Kato TA, Kanba S, Teo AR. Hikikomori: experience in Japan and international relevance. World Psychiatry 2018 Feb;17(1):105-106 [FREE Full text] [doi: 10.1002/wps.20497] [Medline: 29352535]
2. Teo AR. A new form of social withdrawal in Japan: a review of hikikomori. Int J Soc Psychiatry 2010 Mar;56(2):178-185 [FREE Full text] [doi: 10.1177/0020764008100629] [Medline: 19567455]
3. Kato TA, Kanba S, Teo AR. A 39-Year-Old "Adultolescent": Understanding Social Withdrawal in Japan. Am J Psychiatry 2016 Feb 01;173(2):112-114 [FREE Full text] [doi: 10.1176/appi.ajp.2015.15081034] [Medline: 26844793]
4. Teo AR, Gaw AC. Hikikomori, a Japanese culture-bound syndrome of social withdrawal?: A proposal for DSM-5. J Nerv Ment Dis 2010 Jun;198(6):444-449 [FREE Full text] [doi: 10.1097/NMD.0b013e3181e086b1] [Medline: 20531124]
5. Li T, Wong P. Youth social withdrawal behavior (hikikomori): a systematic review of qualitative and quantitative studies. Aust N Z J Psychiatry 2015 Jul;49(7):595-609. [doi: 10.1177/0004867415581179] [Medline: 25861794]
6. Teo A, Fetters M, Stufflebam K, Tateno M, Balhara Y, Choi T, et al. Identification of the hikikomori syndrome of social withdrawal: psychosocial features and treatment preferences in four countries. Int J Soc Psychiatry 2015 Feb;61(1):64-72 [FREE Full text] [doi: 10.1177/0020764014535758] [Medline: 24869848]
7. Hayakawa K, Kato T, Watabe M, Teo A, Horikawa H, Kuwano N, et al. Blood biomarkers of Hikikomori, a severe social withdrawal syndrome. Sci Rep 2018 Feb 13;8(1):2884 [FREE Full text] [doi: 10.1038/s41598-018-21260-w] [Medline: 29440704]
8. Koyama A, Miyake Y, Kawakami N, Tsuchiya M, Tachimori H, Takeshima T, World Mental Health Japan Survey Group, 2002-2006. Lifetime prevalence, psychiatric comorbidity and demographic correlates of "hikikomori" in a community population in Japan. Psychiatry Res 2010 Mar 30;176(1):69-74. [doi: 10.1016/j.psychres.2008.10.019] [Medline: 20074814]
9. Uchida Y, Norasakkunkit V. The NEET and Hikikomori spectrum: assessing the risks and consequences of becoming culturally marginalized. Front Psychol 2015;6:1117 [FREE Full text] [doi: 10.3389/fpsyg.2015.01117] [Medline: 26347667]
10. Yuen J, Yan Y, Wong V, Tam W, So K, Chien W. A Physical Health Profile of Youths Living with a. Int J Environ Res Public Health 2018 Dec 11;15(2):1-13 [FREE Full text] [doi: 10.3390/ijerph15020315] [Medline: 29439488]
11. Funakoshi A, Miyamoto Y. Significant factors in family difficulties for fathers and mothers who use support services for children with hikikomori. Psychiatry Clin Neurosci 2015 Apr;69(4):210-219 [FREE Full text] [doi: 10.1111/pcn.12230] [Medline: 25131493]
12. Kato TA, Shinfuku N, Sartorius N, Kanba S. Are Japan's hikikomori and depression in young people spreading abroad? Lancet 2011 Sep 17;378(9796):1070. [doi: 10.1016/S0140-6736(11)61475-X] [Medline: 21924990]
13. Kato T, Tateno M, Shinfuku N, Fujisawa D, Teo A, Sartorius N, et al. Does the 'hikikomori' syndrome of social withdrawal exist outside Japan? A preliminary international investigation. Soc Psychiatry Psychiatr Epidemiol 2012 Jul;47(7):1061-1075 [FREE Full text] [doi: 10.1007/s00127-011-0411-7] [Medline: 21706238]
14. Wong P, Li T, Chan M, Law Y, Chau M, Cheng C, et al. The prevalence and correlates of severe social withdrawal (hikikomori) in Hong Kong: a cross-sectional telephone-based survey study. Int J Soc Psychiatry 2015 Jun;61(4):330-342. [doi: 10.1177/0020764014543711] [Medline: 25063752]
15. Wong PW, Liu LL, Li TM, Kato TA, Teo AR. Does hikikomori (severe social withdrawal) exist among young people in urban areas of China? Asian J Psychiatr 2017 Dec;30:175-176. [doi: 10.1016/j.ajp.2017.10.026] [Medline: 29080513]
16. Sakamoto N, Martin R, Kumano H, Kuboki T, Al-Adawi S. Hikikomori, is it a culture-reactive or culture-bound syndrome? Nidotherapy and a clinical vignette from Oman. Int J Psychiatry Med 2005;35(2):191-198. [doi: 10.2190/7WEQ-216D-TVNH-PQJ1] [Medline: 16240975]
17. Malagón-Amor Á, Martín-López LM, Córcoles D, González A, Bellsolà M, Teo A, et al. A 12-month study of the hikikomori syndrome of social withdrawal: clinical characterization and different subtypes proposal. Psychiatry Res 2018 Dec;270:1039-1046. [doi: 10.1016/j.psychres.2018.03.060] [Medline: 29615267]
18. Malagón-Amor Á, Córcoles-Martínez D, Martín-López LM, Pérez-Solà V. Hikikomori in Spain: a descriptive study. Int J Soc Psychiatry 2015 Aug;61(5):475-483. [doi: 10.1177/0020764014553003] [Medline: 25303955]
19. García-Campayo J, Alda M, Sobradiel N, Sanz Abós B. [A case report of hikikomori in Spain]. Med Clin (Barc) 2007 Sep 8;129(8):318-319. [Medline: 17878030]
20. Ovejero S, Caro-Cañizares I, de León-Martínez V, Baca-Garcia E. Prolonged social withdrawal disorder: a hikikomori case in Spain. Int J Soc Psychiatry 2014 Sep;60(6):562-565. [doi: 10.1177/0020764013504560] [Medline: 24101742]
21. De Michele F, Caredda M, Delle Chiaie R, Salviati M, Biondi M. [Hikikomori (ひきこもり): a culture-bound syndrome in the web 2.0 era]. Riv Psichiatr 2013;48(4):354-358. [doi: 10.1708/1319.14633] [Medline: 24056835]
22. Chauliac N, Couillet A, Faivre S, Brochard N, Terra JL. Characteristics of socially withdrawn youth in France: a retrospective study. Int J Soc Psychiatry 2017 Jun;63(4):339-344. [doi: 10.1177/0020764017704474] [Medline: 28446040]
23. Furuhashi T, Figueiredo C, Pionnié-Dax N, Fansten M, Vellut N, Castel PH. [Pathology seen in French "Hikikomori"]. Seishin Shinkeigaku Zasshi 2012;114(10):1173-1179. [Medline: 23234197]
24. Gondim FA, Aragão AP, Holanda Filha JG, Messias EL. Hikikomori in Brazil: 29 years of voluntary social withdrawal. Asian J Psychiatr 2017 Dec;30:163-164. [doi: 10.1016/j.ajp.2017.10.009] [Medline: 29065363]
25. Norasakkunkit V, Uchida Y, Takemura K. Evaluating Distal and Proximal Explanations for Withdrawal: A Rejoinder to Varnum and Kwon's "The Ecology of Withdrawal". Front Psychol 2017;8:2085 [FREE Full text] [doi: 10.3389/fpsyg.2017.02085] [Medline: 29238319]

26. Toivonen T, Norasakkunkit V, Uchida Y. Unable to conform, unwilling to rebel? Youth, culture, and motivation in globalizing Japan. Front Psychol 2011;2:207 [FREE Full text] [doi: 10.3389/fpsyg.2011.00207] [Medline: 21949510]
27. Liu L, Li T, Teo A, Kato T, Wong P. Harnessing social media to explore youth social withdrawal in three major cities in China: cross-sectional web survey. JMIR Ment Health 2018 May 10;5(2):e34 [FREE Full text] [doi: 10.2196/mental.8509] [Medline: 29748164]
28. Alvarez-Mon MA, Asunsolo Del Barco A, Lahera G, Quintero J, Ferre F, Pereira-Sanchez V, et al. Increasing interest of mass communication media and the general public in the distribution of tweets about mental disorders: observational study. J Med Internet Res 2018 Dec 28;20(5):e205 [FREE Full text] [doi: 10.2196/jmir.9582] [Medline: 29807880]
29. Robillard JM, Johnson TW, Hennessey C, Beattie BL, Illes J. Aging 2.0: health information about dementia on Twitter. PLoS One 2013;8(7):e69861 [FREE Full text] [doi: 10.1371/journal.pone.0069861] [Medline: 23922827]
30. Meng Y, Elkaim L, Wang J, Liu J, Alotaibi NM, Ibrahim GM, et al. Social media in epilepsy: a quantitative and qualitative analysis. Epilepsy Behav 2017 Dec;71(Pt A):79-84. [doi: 10.1016/j.yebeh.2017.04.033] [Medline: 28554148]
31. Nastasi A, Bryant T, Canner JK, Dredze M, Camp MS, Nagarajan N. Breast cancer screening and social media: a content analysis of evidence use and guideline opinions on Twitter. J Cancer Educ 2018 Jun;33(3):695-702. [doi: 10.1007/s13187-017-1168-9] [Medline: 28097527]
32. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. #Proana: pro-eating disorder socialization on Twitter. J Adolesc Health 2016 Dec;58(6):659-664. [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]
33. Joseph AJ, Tandon N, Yang LH, Duckworth K, Torous J, Seidman LJ, et al. #Schizophrenia: use and misuse on Twitter. Schizophr Res 2015 Jul;165(2-3):111-115. [doi: 10.1016/j.schres.2015.04.009] [Medline: 25937459]
34. Cavazos-Rehg P, Krauss M, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related Tweets. Comput Human Behav 2016 Jan 1;54:351-357 [FREE Full text] [doi: 10.1016/j.chb.2015.08.023] [Medline: 26392678]
35. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
36. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
37. Wakamiya S, Matsune S, Okubo K, Aramaki E. Causal relationships among pollen counts, Tweet numbers, and patient numbers for seasonal allergic rhinitis surveillance: retrospective analysis. J Med Internet Res 2019 Feb 20;21(2):e10450 [FREE Full text] [doi: 10.2196/10450] [Medline: 30785411]
38. Wakamiya S, Morita M, Kano Y, Ohkuma T, Aramaki E. Tweet classification toward Twitter-based disease surveillance: new data, methods, and evaluations. J Med Internet Res 2019 Feb 20;21(2):e12783 [FREE Full text] [doi: 10.2196/12783] [Medline: 30785407]
39. Prieto V, Matos S, Álvarez M, Cacheda F, Oliveira J. Twitter: a good place to detect health conditions. PLoS One 2014;9(1):e86191 [FREE Full text] [doi: 10.1371/journal.pone.0086191] [Medline: 24489699]
40. Young S, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
41. Teo A, Liebow S, Chan B, Dobscha S, Graham A. Reaching those at risk for psychiatric disorders and suicidal ideation: Facebook advertisements to recruit military veterans. JMIR Ment Health 2018 Jul 5;5(3):e10078 [FREE Full text] [doi: 10.2196/10078] [Medline: 29980498]
42. Alcaraz Mateos E, Guerra Pastrián L, Pijuan Andújar L, López Solache L, Zucchiatti A, García Ángel R, et al. [Pathology in social media networks. Recruitment campaign]. Rev Esp Patol 2018;51(1):6-13. [doi: 10.1016/j.patol.2017.09.002] [Medline: 29290326]
43. Lienemann B, Unger J, Cruz T, Chu KH. Methods for coding tobacco-related Twitter data: a systematic review. J Med Internet Res 2017 Dec 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]
44. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In: Association for the Advancement of Artificial Intelligence.: AAAI; 2013 Jun 21 Presented at: 7th International AAAI Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, Massachusetts, USA p. 400-408 URL: http://arxiv.org/abs/1306.5204
45. McHugh M. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
46. Reavley N, Pilkington P. Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. PeerJ 2014;2:e647 [FREE Full text] [doi: 10.7717/peerj.647] [Medline: 25374786]
47. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Dec;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
48. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J Comput Sci 2011 Mar;2(1):1-8 [FREE Full text] [doi: 10.1016/j.jocs.2010.12.007]

49. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Dec 5;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
50. Corrigan P, Rao D. On the self-stigma of mental illness: stages, disclosure, and strategies for change. Can J Psychiatry 2012 Aug;57(8):464-469 [FREE Full text] [doi: 10.1177/070674371205700804] [Medline: 22854028]
51. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
52. Booth R, Allen B, Jenkyn KM, Li L, Shariff S. Youth mental health services utilization rates after a large-scale social media campaign: population-based interrupted time-series analysis. JMIR Ment Health 2018 Apr 6;5(2):e27 [FREE Full text] [doi: 10.2196/mental.8808] [Medline: 29625954]
53. Chen J, Liu Y, Zou M. User emotion for modeling retweeting behaviors. Neural Netw 2017 Dec;96:11-21. [doi: 10.1016/j.neunet.2017.08.006] [Medline: 28946003]
54. Xu WW, Chiu I, Chen Y, Mukherjee T. Twitter hashtags for health: applying network and content analyses to understand the health knowledge sharing in a Twitter-based community of practice. Qual Quant 2014 Jul 8;49(4):1361-1380. [doi: 10.1007/s11135-014-0051-6]
55. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, Maher C. Creating engaging health promotion campaigns on social media: observations and lessons from Fitbit and Garmin. J Med Internet Res 2018 Dec 10;20(12):e10911 [FREE Full text] [doi: 10.2196/10911] [Medline: 30530449]
56. Birnbaum ML, Ernala SK, Rizvi AF, de Choudhury M, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. J Med Internet Res 2017 Dec 14;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]
57. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]

Abbreviations

API: application programming interface
NEET: not in employment, education nor training

Edited by G Eysenbach; submitted 27.03.19; peer-reviewed by Á Malagón-Amor, T Kato; comments to author 19.04.19; revised version received 28.04.19; accepted 29.04.19; published 29.05.19.

Please cite as:
Pereira-Sanchez V, Alvarez-Mon MA, Asunsolo del Barco A, Alvarez-Mon M, Teo A
Exploring the Extent of the Hikikomori Phenomenon on Twitter: Mixed Methods Study of Western Language Tweets
J Med Internet Res 2019;21(5):e14167
URL: http://www.jmir.org/2019/5/e14167/
doi:10.2196/14167
PMID:31144665

Original Paper

A Telemedicine-Based Registration System for the Management of Renal Anemia in Patients on Maintenance Hemodialysis: Multicenter Study

Zhaohui Ni[1,2*], MD; Haijiao Jin[1,2*], MD; Gengru Jiang[2,3], MD; Niansong Wang[2,4], MD; Ai Peng[5], MD; Zhiyong Guo[6], MD; Shoujun Bai[7], MD; Rong Zhou[8], MD; Jianrao Lu[9], MD; Yi Wang[10], MD; Ying Li[11], MD; Shougang Zhuang[12], MD; Chen Yu[13], MD; Yueyi Deng[14], MD; Huimin Jin[15], MD; Xudong Xu[16], MD; Junli Zhang[17], MD; Junli Zhao[18], MD; Xiuzhi Yu[19], MD; Xiaoxia Wang[2,20], MD; Liming Zhang[21], MD; Jianying Niu[22], MD; Kun Liu[23], MD; Xiaorong Bao[24], MD; Qin Wang[25], MD; Jun Ma[26], MD; Chun Hu[2,27], MD; Xiujuan Zang[28], MD; Qing Yu[2,29], MD

[1]Department of Nephrology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[2]Clinical Research Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Department of Nephrology, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[4]Department of Nephrology, The Sixth People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai, China

[5]Department of Nephrology, Tenth People's Hospital of Tongji University, Shanghai, China

[6]Department of Nephrology, Changhai Hospital of Shanghai, Shanghai, China

[7]Department of Nephrology, Qingpu Branch of Zhongshan Hospital Affiliated to Fudan University, Shanghai, China

[8]Department of Nephrology, Yangpu Hospital Affiliated to Tongji University, Shanghai, China

[9]Department of Nephrology, Seventh People's Hospital of Shanghai University of Traditional Chinese Medicine, Shanghai, China

[10]Department of Nephrology, Yueyang Hospital of Integrated Traditional Chinese Medicine and Western Medicine Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai, China

[11]Department of Nephrology, Central Hospital of Shanghai Jiading District, Shanghai, China

[12]Department of Nephrology, Shanghai East Hospital, Shanghai, China

[13]Department of Nephrology, Tongji Hospital, Tongji University School of Medicine, Shanghai, China

[14]Department of Nephrology, Longhua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China

[15]Department of Nephrology, Shanghai Pudong Hospital, Shanghai, China

[16]Department of Nephrology, Minhang Hospital, Fudan University, Shanghai, China

[17]Department of Nephrology, Jiulingwu Hospital, Shanghai, China

[18]Department of Nephrology, Shanghai Pudong New District Zhoupu Hospital, Shanghai, China

[19]Department of Nephrology, Navy Characteristic Medical Center, Shanghai, China

[20]Department of Nephrology, Tongren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[21]Department of Nephrology, Zhabei Central Hospital, Jingan District, Shanghai, China

[22]Department of Nephrology, The Fifth People's Hospital of Shanghai, Fudan University, Shanghai, China

[23]Department of Nephrology, Jinshan Branch of Shanghai No. 6 People's Hospital, Shanghai, China

[24]Department of Nephrology, Jinshan Hospital, Fudan University, Shanghai, China

[25]Department of Nephrology and Rheumatology, Shanghai Fengxian Central Hospital, Shanghai, China

[26]Department of Nephrology, Jingan District Central Hospital of Shanghai, Shanghai, China

[27]Department of Nephrology, No. 9 People Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[28]Department of Nephrology, Shanghai Songjiang District Central Hospital, Shanghai, China

[29]Department of Nephrology, Shanghai General Hospital, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Zhaohui Ni, MD
Department of Nephrology, Renji Hospital
School of Medicine, Shanghai Jiao Tong University
Dongfang Road 1630
Shanghai,
China
Phone: 86 02168383121

Fax: 86 02168383834
Email: profnizh@126.com

Abstract

Background: Renal anemia is one of the most important complications in patients on maintenance hemodialysis (MHD). Telehealth-based dialysis registration systems have the advantage of real-time monitoring and have gradually been applied to the management of chronic diseases.

Objective: The objective of our study was to evaluate the impact of a telehealth-based dialysis registration system on patients on MHD in terms of renal anemia control.

Methods: The Red China project aimed to develop a dialysis registration system based on the WeChat mobile platform. Demographic and baseline laboratory parameters such as age, gender, primary disease, dialysis age, and baseline creatinine levels were recorded using this system. In addition, the hemoglobin and hematocrit levels were recorded monthly. The platform then generated a hemoglobin and hematocrit statistics report for each hemodialysis center monthly, including the detection rate, target rate, and distribution of hemoglobin and released it to physicians via the WeChat mobile phone app. The physicians were then able to treat the individual's anemia appropriately by changing the doses of erythropoiesis-stimulating agents or iron use on the basis of this report. We analyzed the demographic and baseline laboratory parameters, detection rate, target rate, and average level and distribution of hemoglobin 28 months after the launch of the project.

Results: A total of 8392 patients on MHD from 28 hemodialysis centers in Shanghai were enrolled from June 2015 to October 2017. The detection rate of hemoglobin increased from 54.18% to 73.61% ($P<.001$), the target rate of hemoglobin increased from 47.55% to 56.07% ($P<.001$), and the mean level of hemoglobin increased from 10.83 (SD 1. 60) g/dL to 11.07 (SD 1.60) g/dL ($P<.001$). In addition, the proportion of patients with hemoglobin levels ≥11 g/dL but <13 g/dL increased from 40.40% to 47.48%.

Conclusions: This telehealth-based dialysis registration system can provide timely reporting of the anemia status in patients on MHD, which may improve the awareness of anemia and the attention to and compliance with anemia monitoring.

(J Med Internet Res 2019;21(5):e13168) doi:10.2196/13168

KEYWORDS

telemedicine; dialysis registration system; hemodialysis; renal anemia; end-stage renal disease

Introduction

Telehealth technologies offer advantages of accessibility, convenience, and time-effectiveness and are thus being increasingly adopted to aid the management of long-term conditions worldwide. Increasing evidence has demonstrated promising results for telecare or telehealth medicine in the management of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer [1]; however, data on its use in patients on maintenance hemodialysis (MHD) are still sparse.

Recent Quality and Outcomes Framework Disease Register data comparing data from 2006-2007 to 2010-2011 showed a 45% increase in the prevalence of chronic kidney disease, second only to the increase in cancer (79%) [2]. The management of this long-term condition is increasingly challenging when it develops into end-stage renal disease (ESRD) or requires MHD. Renal anemia is extremely common among patients on MHD and often underlies symptoms including fatigue, depression, reduced exercise tolerance, and dyspnea; increased morbidity and mortality related to cardiovascular disease; an increased risk of hospitalization; and an increased length of hospital stay. Patient mortality and hospitalization risks were shown to decrease by 10%-12% for every 1 g/dL increase in mean facility-level hemoglobin [3]. MHD patients should thus be monitored for anemia in a timely manner and managed carefully; according to the Kidney Disease Improving Global Outcomes (KDIGO) guidelines, hemoglobin levels should be monitored at least monthly [4]. However, the actual management is still not satisfactory. Data from the Dialysis Outcomes and Practice Patterns Study (DOPPS) showed that more frequent hemoglobin monitoring was associated with lower facility-level variations in the hemoglobin levels [5]. Data from Chinese DOPPS facilities also showed that a large proportion of patients on MHD did not meet the expressed hemoglobin target and that less frequent and substantial increases in the doses of erythropoiesis-stimulating agents (ESAs) were associated with hemoglobin levels <9 g/dL [6]. Another cohort study enrolled a total of 2388 patients with ESRD (1775 patients on MHD) from nine centers in the largest dialysis facilities in six cities around China and found that about 60% of the patients did not reach the hemoglobin target of 11 g/dL, even though 85.0% of them were treated with erythropoietin [7].

Several renal dialysis registration data systems exist worldwide, including the United States Renal Data System (USRDS), the European Renal-European Dialysis and Transplantation Association Registry, the Australia and New Zealand Dialysis and Transplant Registry, the State of Chronic Dialysis Therapy in Japan, the Hong Kong Renal registration, and the China National Renal Data System. However, these registration systems collect data from dialysis units and issue a dialysis report every year only for quality measures. In China, a registration system that can reflect the dynamic, real-time anemia status of patients on hemodialysis is still lacking, thus preventing their timely treatment.

We therefore established the Red China project using telehealth technology in June 2015 with the aim of improving renal anemia in patients on MHD, allowing timely reporting to nephrologists, facilitating the early recognition and resolution of anemia, and benefiting the long-term prognosis of patients on MHD.

Methods

Participants

The Red China project developed a dialysis registration system based on the WeChat mobile platform. All patients with ESRD undergoing MHD at the 28 centers between June 1, 2015, and October 31, 2017, were enrolled in the study. There were no exclusion criteria, apart from patients not willing to participate. Demographic and baseline laboratory parameters such as age, gender, primary disease, dialysis age, and baseline creatinine levels were recorded in this system. Hemoglobin and hematocrit levels were recorded monthly. We analyzed the demographic and baseline laboratory parameters and the detection rate, target rate, average level, and distribution of hemoglobin from June 2015 to October 2017 after the launch of the project. Detection rate refers to the proportion of patients for whom the hemoglobin level was recorded for one month, and target rate refers to the proportion of patients who achieved a hemoglobin level of 11g/dL in that one month.

Telehealth System

The project was developed using an online platform based on WeChat, currently regarded as the most popular instant-messaging platform in China. This app is the largest social app in China and has met the requirements of international authoritative certification standards in terms of information security. The platform aimed to help medical staff record and monitor hemoglobin levels, hematocrit levels, and other physiological indicators in real time.

Clinical User Interface

Data entry was conducted by full-time personnel with research qualifications. The hemoglobin data were collected from laboratories in each hemodialysis center. The platform generated monthly hemoglobin and hematocrit statistics reports for each hemodialysis center, including the detection rate, target rate, and distribution of hemoglobin. The system highlighted patients who were outside the target and released this information to physicians via the WeChat mobile phone app. The physicians were then able to adjust the patient’s treatment to resolve their anemia individually, on the basis of this report.

Statistical Analysis

Baseline characteristics were expressed as the mean (SD) for normally distributed data and as frequencies and percentages for categorical data. Comparisons between baseline and 28 months after the project were performed using paired *t* tests or Chi-square tests, as applicable. All analyses were performed using SPSS for Windows (version 19.0; SPSS, Inc, Chicago, IL). A *P* value <.05 was considered to be statistically significant.

Results

Baseline Characteristics

This study included 8392 patients on MHD from 28 hemodialysis centers in Shanghai. The baseline characteristics are shown in Table 1.

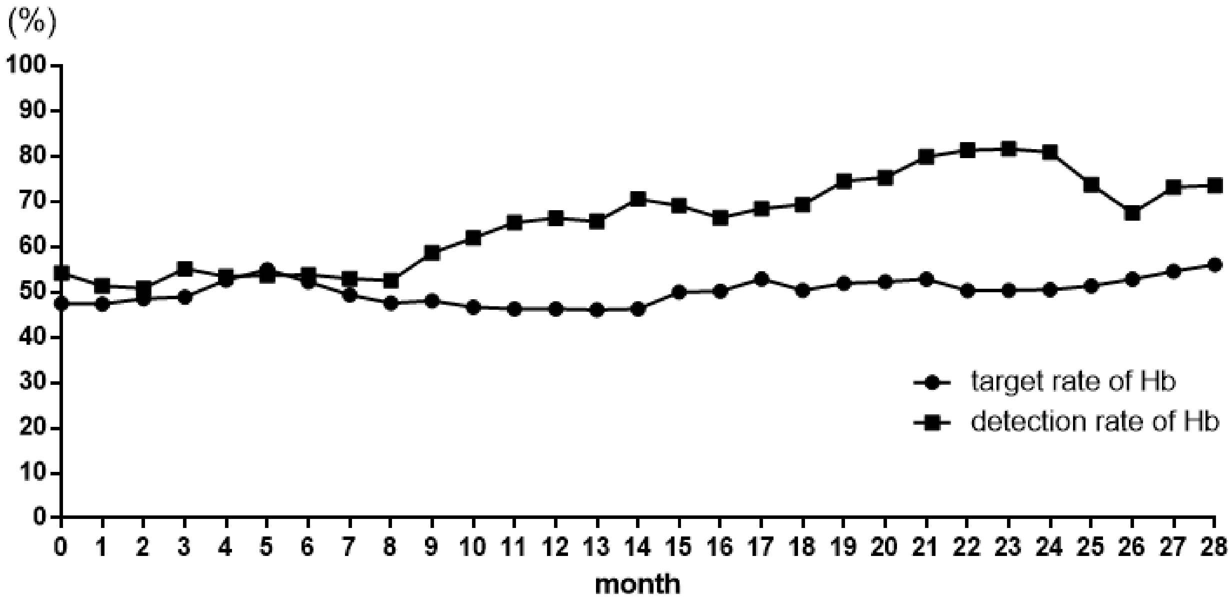
Table 1. Baseline characteristics of patients on maintenance hemodialysis.

Characteristic	Value
Sex - men, n (%)	5059 (60.28)
Age (years), mean (SD)	60.5 (13.7)
Dialysis duration, n (%)	
<3 months	1220 (14.54)
3 months to 1 year	3359 (40.02)
1 to 5 years	3029 (36.09)
5 to 10 years	744 (8.87)
>10 years	40 (0.48)
Primary disease, n (%)	
Glomerulonephritis	3880 (46.23)
Diabetic nephropathy	781 (9.31)
Hypertensive nephrosclerosis	843 (10.05)
Polycystic kidney disease	251 (2.99)
Others	2002 (23.86)
Unknown	635 (7.57)
Serum creatinine level (μmol/L), mean (SD)	853.43 (341.59)
Hemoglobin level (g/L), mean (SD)	10.83 (1.60)

Detection Rate and Target Rate of Hemoglobin

The detection and target rates of hemoglobin were both significantly higher in the 28 months after the project was set up compared with the period before the beginning of the project (detection rates: 73.61% vs 54.18%; target rates: 56.07% vs 47.55%, both *P*<.001; Figure 1).

Figure 1. Detection and target rates of hemoglobin in patients on maintenance hemodialysis. Hb: hemoglobin.



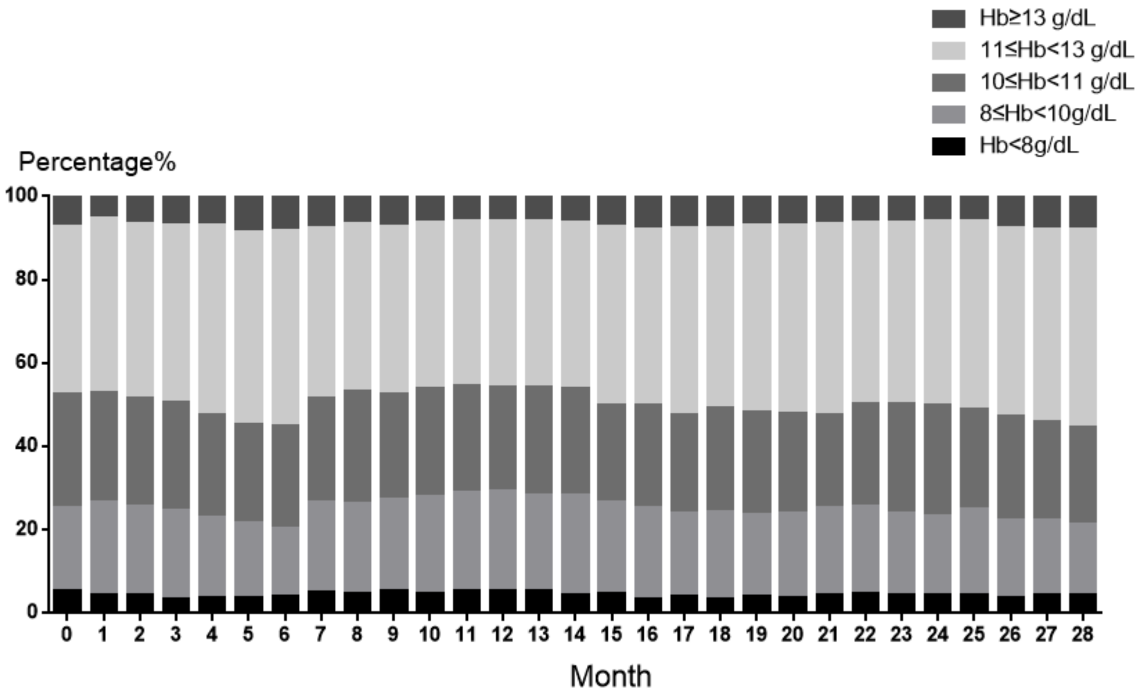
Improvements in Hemoglobin and Hematocrit Levels

Hemoglobin and hematocrit levels were both significantly higher in the 28 months after the project was set up compared to the period before the start of the project (hemoglobin: mean 11.07, SD 1.60 g/dL vs mean 10.83, SD 1.60 g/dL; hematocrit: mean 34.08%, SD 4.89% vs mean 33.51%, SD 5.12%; both *P*<.001).

Distribution of Hemoglobin

The monthly distribution of hemoglobin in patients on MHD is shown in Figure 2. During the 28-month follow-up, the proportion of patients with hemoglobin levels ≥8 g/dL but <10 g/dL decreased from 20.14% to 17.07%, the proportion of patients with hemoglobin levels <8 g/dL decreased from 4.96% to 4.08%, and the proportion of patients with hemoglobin levels ≥11 but <13 g/dL increased from 40.40% to 47.48%.

Figure 2. Distribution of hemoglobin in patients on maintenance hemodialysis. Hb: hemoglobin.



Discussion

Recently, there has been mounting evidence of the feasibility of smartphone apps in remote monitoring and flexible follow-up of patients in western countries [8-10]. However, the experience of using a smartphone app in the management of chronic diseases in China is limited. This study is the first in China to use telehealth technology to promote the management of renal anemia, assess user acceptability, and collect data on patients with ESRD on MHD.

The prevalence of ESRD is increasing worldwide. According to Nanjing Urban Employee Basic Medical Insurance data, the prevalence of ESRD is expected to increase by approximately 1.95% annually from 2015 to 2025, with a predictive value of 1505 per million population in 2025 [2], representing high financial and public health burdens. Furthermore, the increasing number of patients with ESRD accessing hemodialysis is associated with increasing challenges in terms of managing the accompanying renal anemia. Anemia is extremely common among dialysis patients and underlies a range of symptoms including fatigue, depression, reduced exercise tolerance, and dyspnea; increased morbidity and mortality related to cardiovascular disease; an increased risk of hospitalization; and a longer hospital stay. According to a 2017 USRDS report, the mean hemoglobin level in ESRD patients was 9.5 g/dL [11], while the KDIGO guidelines, European Best Practices Guideline, and The National Kidney Foundation Kidney Disease Outcome Quality Initiative guidelines recommend a target hemoglobin level of 11-12 g/dL in patients on MHD [4,12,13].

Telehealth technologies offer advantages of accessibility, convenience, and time effectiveness and are thus being increasingly adopted to aid the management of long-term conditions worldwide. Increasing evidence has indicated promising results for telecare and telehealth medicine in the management of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer [1] and even chronic kidney disease and peritoneal dialysis [14-16]. Sobrinho et al reported that a mobile health app that aimed at assisting in the early diagnosis and self-monitoring of disease progression in patients with chronic kidney disease was associated with quality attributes such as safety, effectiveness, and usability [14]. Telemedicine is also a promising new tool for the remote management of automated peritoneal dialysis, allowing timely intervention prior to the development of more significant problems, reducing the frequency of in-person visits for emergency problems, and reducing health care resource utilization and associated costs [17-19]. However, data on its use in patients on MHD, especially for management of renal anemia, are still sparse.

We therefore developed the Red China project using telehealth technology in June 2015 with the aim of allowing timely reporting to nephrologists, to facilitate the early recognition and resolution of anemia and thus improve the long-term prognosis in patients on dialysis. Similar to the American practice, the Red China project collects and reports the patient hemoglobin test data on a regular basis, prompting the clinician to personally adjust the anemia management medication for patients on MHD, which can reduce the workload of medical staff. In addition to individualized management of patients, the project monitors the compliance status of the patient population in the dialysis center, assists clinicians in overall analysis, and improves MHD patients' hemoglobin level.

Our results showed that achieved rate of target hemoglobin levels in patients on MHD increased significantly from 47.55% to 56.07% during the 28 months following the introduction of the project, compared with another cohort study in China that showed a target rate of 40% [7]. Our results also showed that the proportion of patients with a hemoglobin level <8 g/dL decreased from 4.96% to 4.08% following the start of the project, compared with 12% in China's DOPPS research [6], 31.7% in Japan's DOPPS study [20], and 8.7% in North America's DOPPS study [20,21]. The mean hemoglobin level in this study increased from 10.83 (SD 1.60) g/dL to 11.07 (SD 1.60) g/dL, compared with the mean hemoglobin levels of 10.5 (SD 2.0) g/dL, 10.4 (SD 1.2) g/dL, and 11.5 (SD 1.2) g/dL in China, Japan, and North America, respectively. Although the Red China project has the potential to help clinicians improve anemia in patients on hemodialysis, there is a gap in performance between Shanghai and developed countries such as Japan and North America. Notably, we did not observe a consistent improvement in the hemoglobin levels over baseline for several months. This could be because many other factors played an important role in modulation of the hemoglobin levels, including dialysis adequacy, nutrient status, infection, and chronic gastrointestinal bleeding, which were not recorded in this project. In addition, management of renal anemia via the app took time to yield benefits.

In terms of privacy and security, the Red China project follows the "National Standard of the People's Republic of China: GB/T 35273-2017 Personal Information Security Specification of Information Security Technology." When collecting information, full informed consent was obtained from all patients on MHD and data entry was conducted by full-time personnel with research qualifications. Only researchers have access to the patient data in the center. The data are deidentified, and security measures such as encryption are used for data transmission and storage. The Red China project uses the largest social app in China—WeChat—for patient management. The WeChat platform uses information security technologies such as encryption and anonymization to ensure information security. WeChat has passed the assessment and filing of national network security–level protection and has met the requirements of international authoritative certification standards. The Red China project uses the Alibaba Cloud server to transmit normal and secure traffic back to the server by conducting recognition of malicious features and protecting the service traffic of the app; through the website HTTPS encryption, it can prevent hijacking and tampering, avoid malicious invasion of the website server, and ensure the security of core data.

This study had several limitations. We did not include information on serum iron levels, total iron-binding capacity, transferrin saturation, ferritin levels, or use of drugs such as iron and ESA. However, we aim to improve this platform and include other laboratory parameters such as iron metabolism indicators as well as iron and ESA use in the future. In addition, this telemedicine system has been limited to the physicians, and we believe it would be useful to send the information to patients as well and provide patient education to improve clinical outcomes in the future.

Telehealth technology offers a tool for improved monitoring and calibration of anemia to meet the recommended targets. This telemedicine system could be also used to analyze the reason why some patients could not meet the target and to help develop strategies to improve patient outcomes for anemia management as well as other clinical parameters such as dietary intervention for control of phosphate levels and blood pressure. Further studies in emerging dialysis practices serving large numbers of patients are needed to determine the effects of this technology on improving the achievement of anemia targets as well as the associations with patient outcomes.

In conclusion, telehealth technology offers a promising, feasible, and accessible tool for improving the management of renal anemia in patients on MHD.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81770666 and 81570604), the Research Project of Shanghai Municipal Commission of Health and Family Planning (201740037), and the Multicenter Clinical Research Project of Shanghai Jiao Tong University School of Medicine (DLY201805). We would like to thank Shenyang Sansheng Pharmaceutical Co, Ltd, for launching and supporting the Red China project. We would also like to thank Dr Mingli Zhu, Dr Wei Lu, Dr Xiaohua Sheng, Dr Xinhua Li, Dr Haiyang Wu, Dr Mingzheng Yang, Dr Xi Chen, Dr Jing Hu, Dr Min Chen, Dr Huiliang Jin, Dr Na Liu, Dr Jiangtao Li, Dr Rong Zhu, Dr Yanhong Gu, Dr Haidong He, Dr Chengxiao Yang, Dr Xia Shao, Dr Wei Gou, Dr Qiao Fu, Dr Bibo Wu, Dr Weifeng Fan, Dr Bing Shen, Dr Chengjun Wang, Dr Yi Zhou, Dr Bo Gu, Dr Jianxin Lu, Dr Mei Liu and Dr Jinfang Bao for their contribution to the project.

Conflicts of Interest

None declared.

References

1. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. J Med Internet Res 2017 Dec 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
2. Sun L, Zou LX, Han YC, Huang H, Tan Z, Gao M, et al. Forecast of the incidence, prevalence and burden of end-stage renal disease in Nanjing, China to the Year 2025. BMC Nephrol 2016 Dec 13;17(1):60 [FREE Full text] [doi: 10.1186/s12882-016-0269-8] [Medline: 27295981]
3. Pisoni R, Bragg-Gresham J, Young E, Akizawa T, Asano Y, Locatelli F, et al. Anemia management and outcomes from 12 countries in the Dialysis Outcomes and Practice Patterns Study (DOPPS). Am J Kidney Dis 2004 Jul;44(1):94-111. [Medline: 15211443]
4. KDIGO work Group. Chapter 1: Diagnosis and evaluation of anemia in CKD. Kidney Int Suppl (2011) 2012 Aug;2(4):288-291 [FREE Full text] [doi: 10.1038/kisup.2012.33] [Medline: 25018948]
5. Pisoni R, Bragg-Gresham JL, Fuller DS, Morgenstern H, Canaud B, Locatelli F, et al. Facility-level interpatient hemoglobin variability in hemodialysis centers participating in the Dialysis Outcomes and Practice Patterns Study (DOPPS): Associations with mortality, patient characteristics, and facility practices. Am J Kidney Dis 2011 Feb;57(2):266-275. [doi: 10.1053/j.ajkd.2010.11.003] [Medline: 21251541]
6. Zuo L, Wang M, Hou F, Yan Y, Chen N, Qian J, et al. Anemia Management in the China Dialysis Outcomes and Practice Patterns Study. Blood Purif 2016;42(1):33-43 [FREE Full text] [doi: 10.1159/000442741] [Medline: 27045519]
7. Zhou Q, Jiang J, Wu S, Tian J, Chen J, Yu X, et al. Current pattern of Chinese dialysis units: a cohort study in a representative sample of units. Chin Med J (Engl) 2012 Oct;125(19):3434-3439. [Medline: 23044302]
8. Stockwell K, Alabdulqader E, Jackson D, Basu A, Olivier P, Pennington L. Feasibility of parent communication training with remote coaching using smartphone apps. Int J Lang Commun Disord 2019 Mar;54(2):265-280. [doi: 10.1111/1460-6984.12468] [Medline: 30851010]
9. Henriksson P, Sandborg J, Blomberg M, Alexandrou C, Maddison R, Silfvernagel K, et al. A Smartphone App to Promote Healthy Weight Gain, Diet, and Physical Activity During Pregnancy (HealthyMoms): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Mar 01;8(3):e13011 [FREE Full text] [doi: 10.2196/13011] [Medline: 30821695]
10. Olalla J, García de Lomas JM, Márquez E, González FJ, Del Arco A, De La Torre J, et al. Experience of Using an App in HIV Patients Older Than 60 Years: Pilot Program. JMIR Mhealth Uhealth 2019 Mar 06;7(3):e9904 [FREE Full text] [doi: 10.2196/mhealth.9904] [Medline: 30839281]
11. USRDS: United States Renal Data System. 2017. Chapter 1: Incidence, Prevalence, Patient Characteristics, and Treatment Modalities URL: https://www.usrds.org/2017/download/2017_Volume_2_ESRD_in_the_US.pdf [accessed 2019-04-22]

12. KDOQI. KDOQI Clinical Practice Guideline and Clinical Practice Recommendations for anemia in chronic kidney disease: 2007 update of hemoglobin target. Am J Kidney Dis 2007 Sep;50(3):471-530. [doi: 10.1053/j.ajkd.2007.06.008] [Medline: 17720528]
13. Locatelli F, Aljama P, Bárány P, Canaud B, Carrera F, Eckardt KU, European Best Practice Guidelines Working Group. Revised European best practice guidelines for the management of anaemia in patients with chronic renal failure. Nephrol Dial Transplant 2004 May;19 Suppl 2:ii1-i47. [Medline: 15206425]
14. Sobrinho A, da Silva LD, Perkusich A, Pinheiro M, Cunha P. Design and evaluation of a mobile application to assist the self-monitoring of the chronic kidney disease in developing countries. BMC Med Inform Decis Mak 2018 Dec 12;18(1):7 [FREE Full text] [doi: 10.1186/s12911-018-0587-9] [Medline: 29329530]
15. Lee YL, Cui YY, Chang P. A Content Analysis of Mobile Apps for Chronic Kidney Disease Patient Care: Searching in English and Chinese. Stud Health Technol Inform 2017;245:1232. [Medline: 29295319]
16. Welch J, Astroth K, Perkins S, Johnson C, Connelly K, Siek K, et al. Using a mobile application to self-monitor diet and fluid intake among adults receiving hemodialysis. Res Nurs Health 2013 Jun;36(3):284-298 [FREE Full text] [doi: 10.1002/nur.21539] [Medline: 23512869]
17. Drepper V, Martin P, Chopard C, Sloand J. Remote Patient Management in Automated Peritoneal Dialysis: A Promising New Tool. Perit Dial Int 2018;38(1):76-78. [doi: 10.3747/pdi.2017.00054] [Medline: 29311200]
18. Milan Manani S, Crepaldi C, Giuliani A, Virzì GM, Garzotto F, Riello C, et al. Remote Monitoring of Automated Peritoneal Dialysis Improves Personalization of Dialytic Prescription and Patient's Independence. Blood Purif 2018;46(2):111-117. [doi: 10.1159/000487703] [Medline: 29694954]
19. Makhija D, Alscher MD, Becker S, D'Alonzo S, Mehrotra R, Wong L, et al. Remote Monitoring of Automated Peritoneal Dialysis Patients: Assessing Clinical and Economic Value. Telemed J E Health 2018 Dec;24(4):315-323. [doi: 10.1089/tmj.2017.0046] [Medline: 29024613]
20. Akizawa T, Pisoni R, Akiba T, Saito A, Fukuhara S, Asano Y, et al. Japanese haemodialysis anaemia management practices and outcomes (1999-2006): results from the DOPPS. Nephrol Dial Transplant 2008 Nov;23(11):3643-3653 [FREE Full text] [doi: 10.1093/ndt/gfn346] [Medline: 18577535]
21. Fishbane S, Berns JS. Hemoglobin cycling in hemodialysis patients treated with recombinant human erythropoietin. Kidney Int 2005 Sep;68(3):1337-1343 [FREE Full text] [doi: 10.1111/j.1523-1755.2005.00532.x] [Medline: 16105069]

Abbreviations

DOPPS: Dialysis Outcomes and Practice Patterns Study
ESA: erythropoiesis-stimulating agents
ESRD: end-stage renal disease
KDIGO: Kidney Disease Improving Global Outcomes
MHD: maintenance hemodialysis
USRDS: United States Renal Data System

Edited by G Eysenbach; submitted 17.12.18; peer-reviewed by M Singh, C He, Z Su, T Jamieson; comments to author 24.01.19; revised version received 19.03.19; accepted 05.04.19; published 08.05.19.

Please cite as:
Ni Z, Jin H, Jiang G, Wang N, Peng A, Guo Z, Bai S, Zhou R, Lu J, Wang Y, Li Y, Zhuang S, Yu C, Deng Y, Jin H, Xu X, Zhang J, Zhao J, Yu X, Wang X, Zhang L, Niu J, Liu K, Bao X, Wang Q, Ma J, Hu C, Zang X, Yu Q
A Telemedicine-Based Registration System for the Management of Renal Anemia in Patients on Maintenance Hemodialysis: Multicenter Study
J Med Internet Res 2019;21(5):e13168
URL: https://www.jmir.org/2019/5/e13168/
doi:10.2196/13168
PMID:31344676

Original Paper

Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention

Allison Lindauer[1*], PhD, NP; Glenise McKenzie[2*], PhD, RN; David LaFazia[3,4*], PhD, MSW; Loriann McNeill[5*], BS; Kate Mincks[1*], BA; Natasha Spoden[1*], MS; Marcella Myers[1*], LPN; Nora Mattek[1*], MPH; Linda L Teri[4,6*], PhD

[1]Layton Aging and Alzheimer's Disease Center, Oregon Health & Science University, Portland, OR, United States
[2]School of Nursing, Oregon Health & Science University, Portland, OR, United States
[3]School of Social Work, University of Washington, Seattle, WA, United States
[4]Northwest Research Group on Aging, School of Nursing, University of Washington, Seattle, WA, United States
[5]Family Caregiver Support Program, Multnomah County, Oregon, Portland, OR, United States
[6]Northwest Roybal Center, School of Nursing, University of Washington, Seattle, WA, United States
[*]all authors contributed equally

Corresponding Author:
Allison Lindauer, PhD, NP
Layton Aging and Alzheimer's Disease Center
Oregon Health & Science University
3181 SW Sam Jackson Park Road, CR-131
Portland, OR, 97239
United States
Phone: 1 503 494 4347
Fax: 1 503 494 7499
Email: lindauer@ohsu.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/4/e18032/

Abstract

Background: Families living with Alzheimer disease and related dementias have more access to support thanks to the development of effective telehealth-based programs. However, as technological science grows, so does the risk that these technology-based interventions will diverge from foundational protocols, diluting their efficacy. Strategies that ensure programs are delivered as intended, with fidelity to guiding protocols, are needed across the intervention spectrum—from development to wide-scale implementation. Few papers address fidelity in their technology-based work. Here, we present our translated telehealth intervention, Tele-STAR, with our fidelity findings.

Objective: This study aimed to assess the preliminary efficacy of Tele-STAR on reducing family caregiver burden and depression. Across the implementation phases, we assessed the fidelity of a caregiver education intervention, STAR-C, as it was translated into a telehealth option (Tele-STAR).

Methods: A total of 13 family caregivers consented to participate in an 8-week, videoconference-based intervention (Tele-STAR). Tele-STAR efficacy in reducing the affective burden of caregiving was assessed using pre- and postintervention paired t tests. Content experts assessed program fidelity by reviewing and rating Tele-STAR materials for adherence to the original STAR-C protocol. These experts assessed treatment fidelity by viewing videos of the intervention and rating adherence on a checklist.

Results: Tele-STAR reduced caregiver burden and retained good program and treatment fidelity to STAR-C.

Conclusions: We found Tele-STAR reduced caregiver burden and had good fidelity to the original protocol. Assessing fidelity is a complex process that requires incorporation of these procedures early in the research process. The technology used in this study facilitated the accrual of informative data about the fidelity of our translated intervention, Tele-STAR.

(J Med Internet Res 2019;21(5):e13599) doi:10.2196/13599

KEYWORDS

dementia; caregiving; fidelity

Introduction

Background

Caring for one of the world's 47 million adults with Alzheimer disease or a related dementia can be both rewarding and taxing for family members [1,2]. For some, caregiving can have detrimental effects on their mental health, leading to depression and a sense of burden [3]. The strain of caregiving can also have negative effects on quality of life for the person with dementia. Stressed families are more likely to consider long-term care placement for those with dementia, which may alleviate some of the caregiving tasks but often not the affective symptoms of depression and pre-death grief [4-6].

Interventions such as educational programs and support groups are available to family caregivers for those with dementia. These interventions are often effective [7] but caregivers report access challenges. Distance, cost, neuropsychiatric symptoms of dementia and time constraints all hinder caregiver engagement in these opportunities [8,9]. Real-time, internet-based videoconferencing technology (also known as telehealth) makes education and support interventions accessible for families living with dementia. Here, we report on our work developing and assessing an internet-based intervention to provide accessible support for family members.

Telehealth-based interventions are appealing because caregivers can receive help and support in their own homes. These interventions have small to moderate effects on reducing caregiver burden and depression and good consumer acceptance [10,11]. However, despite evidence that family caregivers prefer individualized interventions with real-time counselors, most telehealth interventions are group-based, automated, and not tailored to stages of disease [11-14].

To address the need for personalized, real-time educational interventions for families caring for those with dementia, we designed Tele-STAR. Tele-STAR uses videoconferencing to connect nurse consultants with family caregivers to guide caregivers in strategies to reduce the emotional, cognitive, and physical effects of distressing behavioral symptoms of dementia. Tele-STAR was developed from the in-person intervention, STAR-C [15]. STAR-C employs cognitive behavioral techniques to increase caregiver awareness of care partner behavior and their reactions to them. The efficacy of the 8-week STAR-C intervention has been established in earlier work [15,16]. STAR-C is implemented face-to-face in families' homes. However, in rural states such as Oregon, implementation for families outside metropolitan areas requires a different approach [16]. Consequently, we revised the STAR-C intervention into a telehealth option to increase access to anyone with a computer and internet connection.

To preserve the efficacy of STAR-C, we sought to retain its essential components as it was translated into Tele-STAR. Our aims were to (1) explore the preliminary efficacy of Tele-STAR and (2) assess the fidelity of Tele-STAR to the original STAR-C [15], after we converted it from an in-person intervention to a telehealth-based version.

Fidelity Assessment

Fidelity assessment ascertains if an intervention adheres to prescribed protocols and treatments across the implementation spectrum, from early-stage pilots to full-fledged wide-scale interventions [17,18]. High fidelity to an intervention increases confidence in internal validity so that scientists (and other consumers) can trust that the effects of the intervention resulted from the intervention and not extraneous factors [17]. A fidelity assessment can be used to examine adherence to a novel intervention in the early stages of implementation, to evaluate if a translated program is faithful to the original protocol across the stages of an intervention, or to prevent divergence from the protocol and ensure consistent delivery in large-scale programs [18].

Recognizing the value of fidelity, the US National Institute on Aging encourages scientists to consider fidelity assessments in their behavioral research [19]. Onken et al (2014) [19] argue that behavioral interventions cannot be adequately implemented if a fidelity assessment plan is not in place. Without a fidelity plan, community providers lack guidelines to ensure an intervention is delivered as intended.

We sought to determine if Tele-STAR retained the important components of the original intervention (STAR-C) [15], if the nurse consultants adhered to the intervention protocol, and if caregivers received and enacted the information provided. Thus, we assessed 2 types of fidelity: program and treatment, to evaluate how closely Tele-STAR aligned with STAR-C [15].

Our fidelity assessment, modeled on Teri's (2010) and Griffiths' (2016) work, examined both program fidelity and treatment fidelity. A program fidelity assessment examines if a program retains the theoretical foundation and essential components of an intervention when it is translated into a telehealth intervention. Treatment fidelity assessment explores if a treatment is implemented and received as planned, based on the program design [17,20,21].

Methods

Participants

We recruited family caregivers from the local community and our dementia clinic. Participants had to provide care at least 4 hours/day for a family member with Alzheimer disease or a related dementia, speak English, and have access to a functional computer. They were not required to live with their family member but all did. We accepted participants from Oregon and Southwest Washington, without regard to their distance from the study center. Caregivers were not paid for their work, nor did they receive any financial compensation for study participation. Each caregiver had to report 3 or more care recipient behaviors that were distressing for them. These standardized questions were embedded in the initial telephone screening interview [15].

Most caregivers owned a computer with internet connectivity (broadband, Wi-Fi, or cellular); we lent a laptop with cellular service to those who did not have a functional connection. All caregivers consented to having the visits recorded. We did not query caregivers about the type of computer they used (eg, computer, laptop, and smartphone), but we did assess the age of their device. We assessed caregivers' level of comfort and knowledge with computer use.

To further characterize the sample of caregivers, we asked them how often they contacted their health care provider (from once a year or less to more than once a week) and if they attended a support group. We also queried caregivers on how often they contacted their care recipient's health care provider and how often they gave their care recipient "as needed" medications for behaviors. The computer, health care–usage and satisfaction surveys were all emailed to caregivers, allowing them to send anonymous replies using Qualtrics [22].

All participants provided consent over the telephone. Although the care recipients did not participate in the intervention, we required their consent for their caregiver to participate. If they were unable to consent, we followed Oregon Health and Science University's (OHSU's) Decisionally Impaired protocol [23] and sought assent from the family member. All components of this study were approved by OHSU's Institutional Review Board (IRB# 17526).

Tele-STAR Procedures

Measures used in previous STAR-C studies [15,16,24] were used in Tele-STAR (Table 1). Caregivers were assessed before, during (after the 4th visit), and after the full 8-week intervention and then 2 months after the intervention. Demographic information was collected before the initiation of the intervention. We asked caregivers to estimate the number of daily hours of caregiving they provided per week. The research assistant administered all the measures in Table 1 via the direct-to-home videoconferencing interface.

A total of 2 consultants (AL and MM) provided the Tele-STAR intervention to the family caregivers over 8 weekly sessions. One consultant is a licensed practical nurse (LPN); the other, the principal investigator (PI), is a PhD-prepared gerontological nurse practitioner (GNP). The GNP, an experienced interventionist, trained the LPN in the Tele-STAR protocol using a training outline developed by the PI. The LPN received instruction in the booklets, forms, and telehealth approach. To ensure the LPN provided the intervention as intended, the GNP observed the LPN with her first 2 caregivers and reviewed the videos from these visits.

After the training was completed, each consultant met one-to-one with a caregiver for 8 weekly, hour-long sessions via Health Insurance Probability and Accountability Act (HIPAA)–secure, internet-based videoconferencing. The consultants followed the Tele-STAR manual, which guided them through the weekly sessions [23]. Caregivers used a workbook, which was a revision of the original STAR-C workbook [32]. The bound workbook, which contained all information, handouts, and space for writing notes, was mailed to the caregivers before session 1. Caregivers were encouraged to write in the workbook and keep it for later reference. At each session, caregivers were asked to read their notations to the consultant and/or show them their written work. Recognizing that shared authority is essential to adult learning, we asked caregivers to show us their written work on the video screen, but this was not a requirement for study participation [33].

At the first session, caregivers identified 2 or 3 care recipient target behaviors that they found distressing. Over the next 7 weeks, consultants guided the caregivers through a process in which they identified a plan to address the *a* ctivators of the behaviors, the *b* ehaviors, and the *c* onsequences (ABC). The consultants, following the Tele-STAR manual, provided information on communication strategies, pleasant events, and caregiver health. After the second assessment (2 months after completing the intervention), caregivers joined 1 final videoconferencing meeting in teams of 3 dubbed "Tele-STAR Trios," to meet each other, and if they so desired, exchange contact information. All Tele-STAR sessions were recorded then stored in the HIPAA-secure cloud-based site.

Table 1. Tele-STAR measures.

Measure[a,b]	Description
Revised Memory and Behavior Problems Checklist [25]	• 24 items • Documents the frequency of distressing care recipient behaviors and rates caregivers' reactions to these behaviors • Among the most commonly used measures of burden in caregiver research [26] • Excellent reliability when used with telehealth technology (ICC[b]=.80) [27]
Zarit Burden Interview [28]	• 4 items • Reliable (ICC=.79) when used with telehealth [27]
Desire to Institutionalize, Revised [24,29]	• 5 dichotomous items and 1 modified item that rates the likelihood of placement on a 5-point Likert scale (1: "not at all likely" to 5: "very likely")
Quality of Life in Alzheimer's Disease (QOLAD) [30]	• 13 items, option for comments
Montreal Cognitive Assessment [31]	• 30-point assessment that measures cognitive function that is reliable when used with telehealth (ICC=.93) [27] • Used with care recipient only, if no formal assessment within 1 year of study start

[a] All data were collected on all participants because of the research assistant's ability to connect with the participants via videoconferencing.

[b] ICC: intraclass correlation coefficient.

To foster treatment fidelity, the 2 consultants met every week to review the process and discuss challenges. The consultants met with the content experts 3 times over the course of the study to discuss the protocol and the intervention approach.

Fidelity Procedures

We structured our fidelity assessment framework based on Griffiths et al's (2016) [12] and Teri et al's (2010) [20] approaches (Table 2). Like these authors, we engaged experts who were very familiar with the original protocol. The senior expert (LT), who designed the original STAR-C [15], provided assistance with developing the assessment criteria and mentoring the PI (AL). A total of 3 content experts (DL, LAM, GM) volunteered their time to assist with the development of the assessment criteria, assess the written materials, and view a sample of Tele-STAR videos. All 3 experts have extensive experience with the STAR-C program and with fidelity assessment of the program.

Before implementing Tele-STAR, criteria for program and treatment fidelity were identified and reviewed by the senior expert and content experts [15,17,20]. We then gave the content experts a packet with the materials to assess and work sheets to document their reviews. They had the option to complete the assessments either online or traditionally with paper and pencil.

To assess program fidelity, the content experts assessed 30 components of 4 Tele-STAR domains: General principles (12 components), homework and handouts (9 components), consultant documents (6 components), and Tele-STAR-specific documents (3 components). They were asked to use the worksheet to rate the materials in each domain with the following scale (1=inconsistent with STAR-C, 2=Same as STAR-C, and 3=adds to STAR-C) [12].

Content experts assessed treatment fidelity by viewing 12 videos of Tele-STAR sessions, which recorded the participants and consultants as they progressed through the protocol. The content experts were given access to the videos via to the university's secure, HIPAA-compliant, cloud-based system for storing and sharing documents and videos. We used a store-and-forward process to provide the content experts with video recordings of a subsample of all Tele-STAR sessions [34]. The content experts accessed and viewed videos at a later date at their convenience. Using the Consultant Adherence Checklist worksheet [20], they rated how well the treatment information was provided by the consultants and enacted by the caregivers (0=not applicable, 1=not at all/some, 2=moderately, and 3=extensively). The 3 content experts each reviewed and rated the same sessions (2 and 5) for the same 2 caregivers, assessing a total of 12 videos. They then rated 7 more videos of their choice for 4 more caregiver participants. The content experts could choose from several participants but could only view sessions 2 and 5 to maintain consistency across reviews.

For each video, the content experts assessed 6 components: General homework review, behavior change planning, maintaining and focusing on current and observable behavior, assisting caregiver in developing own solutions, responsiveness to caregiver's current needs, and overall quality of sessions. We calculated the percentage of caregivers who completed all sessions and prescribed homework.

Data Analysis

This was a pilot study, and thus, data analyses were for exploratory purposes. Efficacy was assessed by comparing the measure scores before, during, and after the intervention using paired *t* tests [35]. As the consultants (the LPN and GNP) had different training backgrounds, we examined if there were differences in Revised Memory and Behavior Problems Checklist (RMBPC) change [25] by consultant using simple *t* tests.

For program fidelity, we calculated the percentage that each component was rated 1=inconsistent with STAR-C, 2=same as STAR-C, or 3=adds to STAR-C. We compared scores across the 3 expert consultants to identify agreement trends. We planned to use the Fleiss kappa to assess inter-rater agreement [36,37].

For treatment fidelity, we assessed consultant adherence to the weekly session content and the percent of caregivers attending 8 or more treatment sessions. We compared the proportion of content coverage for each consultant in Tele-STAR to content coverage in the original STAR-C study [15]. We calculated the percentage that each component on the Consultant Adherence Checklist was rated: 0=not applicable, 1=not at all/some, 2=moderately, 3=extensively. As assessing the videos could be subjectively interpreted, we used Cohen kappa coefficient to calculate inter-rater agreement, with the following parameters: almost perfect 0.81-1.00, substantial 0.61-0.80, moderate 0.41-0.60, fair 0.21-0.40, and slight/poor <0.00 [38].

Table 2. Tele-STAR fidelity assessment components.

Component	Element
Program fidelity [12]	• Guiding principles: Do the materials reflect the guiding theory and principles of the original program (STAR-C)? • Homework and handouts: Consistent with those used in STAR-C? • Consultant documents: Align with the principles and goals of STAR-C? • Tele-STAR-specific documents: Does the Tele-STAR training outline and logo follow STAR-C principles?
Treatment fidelity [20]	• Was the Tele-STAR intervention delivered following the STAR-C approach?: Consultant Adherence Checklist (used with videos) • Did forms reflect treatment receipt?: (1) Attendance records (% complete), (2) Content checklist (% complete), (3) Participant compliance measure (% complete)

Results

Participants

The majority (77%) of the caregivers and 31% of the care recipients were women, and all were white (Table 3). All of the caregivers were spouses or siblings. Care recipients were, for the most part, in the moderate stages of dementia. A third of the families lived in rural areas. Of the 13 caregivers who consented to Tele-STAR, 1 dropped out and was lost to follow-up after the second visit. This caregiver had a higher depression score, but not burden score, than the other caregivers.

The caregivers had newer computer models (1-3 years) and were moderately comfortable using telehealth videoconferencing (average 3.6; range 1: not at all comfortable to 5: extremely comfortable). The majority found Tele-STAR very convenient and, given the choice, would prefer a telehealth option over an in-home one (Table 4).

Table 3. Demographics (n=13; one lost to follow-up).

Participants	Mean (SD); range
Caregivers	
Age	67.1 (7.4); 56-78
Number of years caregiving	3.4 (2.5); 0.5-9
Care recipients	
Age	71.5 (9); 56-83
Number of years with Alzheimer Disease and Related Dementias	3.6 (1.9); 1-7
Montreal Cognitive Assessment	15.2 (3); 10-20
Both[a]	
Miles from study center	39.1 (53.6); 1-169

[a] Miles each dyad lived from the study.

Table 4. Participant experience rating (n=11; surveys emailed to caregivers; 1 did not respond).

Question and response	n (%)
Overall, how convenient was your Tele-STAR experience?	
Very convenient	11 (90)
Somewhat convenient	1 (10)
Not convenient	0 (0)
I had good technical support from study team	
Agree	11 (90)
Neutral	0 (0)
Disagree	0 (0)
No answer	1 (10)
I could easily see and hear the consultant	
Agree	11 (100)
Neutral	0 (0)
Disagree	0 (0)
It was easy to connect with my consultant via videoconferencing	
Agree	11 (100)
Neutral	0 (0)
Disagree	0 (0)
If you had the option of in-home or telehealth training, which would you prefer?	
In-home	2 (18)
Telehealth	3 (27)
Combination of both	6 (55)

Caregivers contacted their own health care providers once a year or less (average 5.3; range 1-6, 1=more than once a week to 6=once a year or less). Of the 13 participants, 7 attended support groups. There were no significant differences in the RMBPC change between those who attended support groups and those who did not.

Caregivers contacted their care recipient’s health care provider about once per month (average 4.7; range 1-6). A total of 4 caregivers gave “as needed” medications daily to their family members; 10 did not give any as needed medications.

Results: Tele-STAR

Caregivers’ depression and burden decreased (improved) slightly, but not significantly, by session 4 (Table 5). Burden improved significantly after session 8 as indicated by improvements in caregiver reactivity to upsetting behaviors, measured on the RMBPC [25]. This effect was sustained at 2 months postintervention. The frequency and reactivity to the target behaviors reduced significantly and was appreciated clinically by the caregivers. Ratings on the RMBPC did not differ significantly by consultant type (LPN vs GNP; *P*=.50 and .47, respectively).

Results: Program Fidelity

The 3 content experts rated 30 Tele-STAR components for a total of 90 ratings (3 per component) (Figure 1, top). Of these, 72 out of the 90 ratings (80%) were labeled as “same as” STAR-C, and 16 of the 90 ratings (17%) were labeled “adds to” STAR-C. One rating was labeled “inconsistent with STAR-C,” and 1 data point was missing. There was total agreement between 3 content experts for 18 of the 30 (60%) of the components. There was some disagreement in 11 (36%) of the rated components, but the disagreement was between whether the components were “same as STAR-C” or “adds to STAR-C,” indicating 96% agreement that Tele-STAR was the same as or adds to STAR-C. Content experts noted the Tele-STAR trios “added to” the intervention. The lack of variability in the ratings made a Fleiss kappa statistic unstable. We chose not to report it because the value is not meaningful [37]. Content experts spent about 2 to 4 hours each on this assessment. These findings indicate the Tele-STAR adhered to the theoretical foundation and components of STAR-C.

Table 5. Tele-STAR results (n=12).

Variable[a]	Baseline, mean (SD)	Post session 4, mean (SD)[b]	Post session 8, mean (SD)	*P* value	2 months postintervention, mean (SD)[c]	*P* value[c]
CESD-10[d]	10.9 (3.4)	11.1 (4.7)	10.7 (5.5)	.88	11.8 (4.4)	.46
ZBI[e]	9.1 (2.2)	8.1 (2.5)	8.3 (2.4)	.22	8.9 (1.6)	.38
QOL-AD[f]	32.5 (7.2)	N/A[g]	33.1 (7.7)	.55	31.4 (5.5)	.37
Overall RMBPC[h]-frequency	41.8 (10.4)	N/A	41.1 (11.1)	.73	41.7 (12.6)	.81
Overall RMBPC-reactivity[i]	32.3 (12.5)	N/A	28.2 (10.4)	.04[k]	30.4 (8.9)	.41
Target symptom 1 frequency[j]	3.3 (1.1)	N/A	2.3 (1.7)	.15	—[m]	—
Target symptom 1 reactivity[k]	2.7 (0.9)	N/A	1.3 (1.1)	.0001[l]	—	—
Target symptom 2 frequency[j]	3.7 (0.5)	N/A	2.4 (1.3)	.01[l]	—	—
Target symptom 2 reactivity[k]	3.0 (0.7)	N/A	1.5 (1.2)	.001[l]	—	—

[a] Excludes baseline data for one caregiver who dropped out.

[b] Only Center for Epidemiological Studies Depression Scale-10-item (CES-D 10) and Zarit Burden Interview (ZBI) were assessed after session 4.

[c] Change from post session 8.

[d] CESD-10: Center for Epidemiological Studies Depression Scale-10-item.

[e] ZBI: Zarit Burden Interview.

[f] QOL-AD: Quality of Life-AD.

[g] N/A: not applicable.

[h] RMBPC-frequency: Revised Memory and Behavior Problems Checklist, frequency of behavioral symptoms.

[i] RMBPC-reactivity: Revised Memory and Behavior Problems Checklist, Caregiver’s reactions to behavioral symptoms.

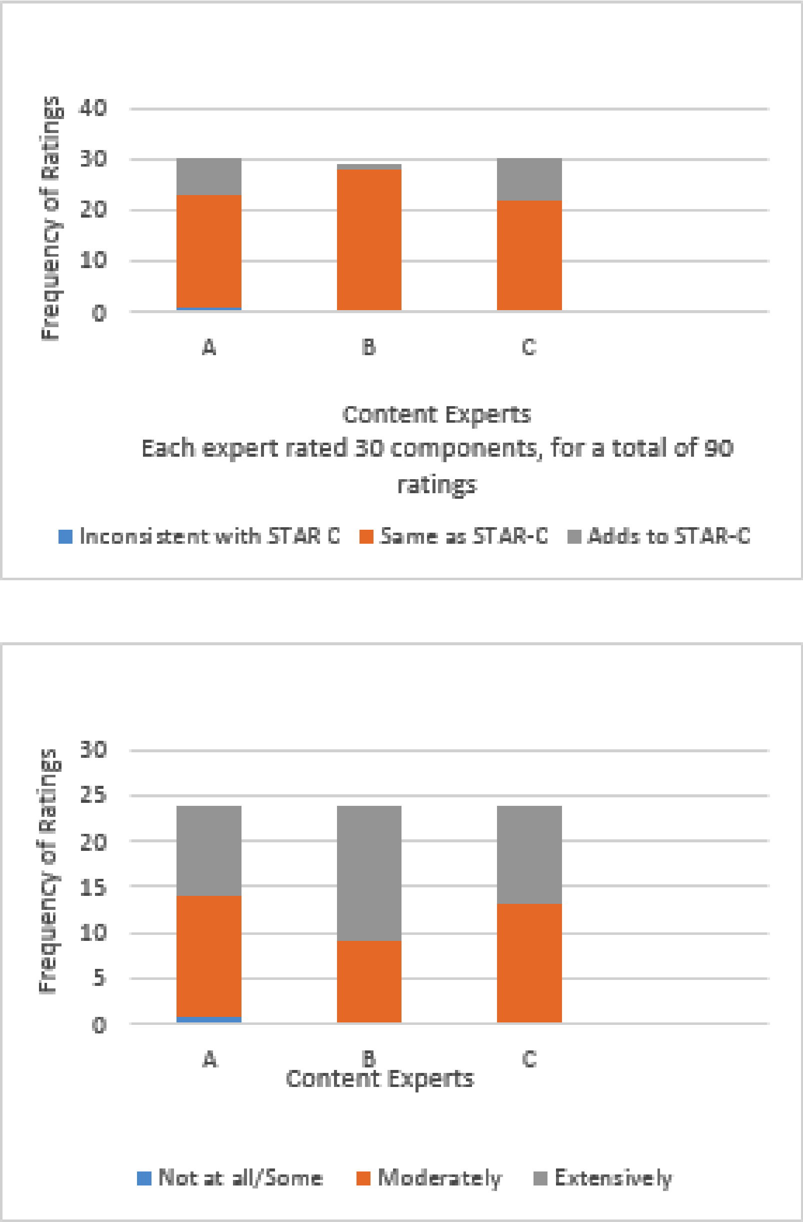
[j] Behavioral symptoms identified by caregivers, frequency of behavioral symptoms.

[k] Caregiver’s reactions to behavioral symptoms.

[l] *P* value <.05.

[m] Target symptoms frequency and reactivity were measured only post session 8.

Figure 1. Program fidelity ratings (top) and treatment fidelity ratings (bottom).



Results: Treatment Fidelity

Similar to Teri et al's findings (2005) [15] for STAR-C, 92% (12 out of 13) of the caregivers completed all Tele-STAR sessions. Similarly, 92% of the caregivers (the 12 that completed all the sessions) completed the homework assignments and developed at least one ABC plan. These findings indicate that the vast majority of the caregivers were able to understand and engage in the treatment protocol. The content experts rated consultant adherence to the protocol as "moderately" for 56 out of the 114 (49%) of the components and "extensively" for 57 of the 114 components (50%), with 1 expert rating 1 component ("Assisting Caregiver in Developing Own Solutions") as "not all or some" for 1 video only (Figure 1, bottom). Inter-rater agreement was moderate (kappa=.43) [38]. Content experts spent about 1.5 to 2.5 hours assessing each video. These findings show that the consultants were able to follow the protocol and implement it as designed.

Discussion

This pilot study assessed the preliminary efficacy of a revised telehealth-based intervention (Tele-STAR) and the fidelity of Tele-STAR to the original caregiver intervention (STAR-C) [15]. We found that the Tele-STAR intervention in this small sample reduced burden on the RMBPC [25] but did not improve on depression scores (Table 5). Tele-STAR had good program and treatment fidelity to STAR-C. Our fidelity assessment suggests that the Tele-STAR intervention adhered to the original STAR-C protocol and that it was implemented as designed. The implication being that the caregiver burden was reduced by the intervention and not by extraneous factors caused by divergence from the STAR-C program.

Assessing fidelity is a complex process involving time, expertise, and resources. Perepletchikova and others (2009) identified multiple barriers to fidelity assessment and found, in their review of 147 randomized trials, that only 3.5% of the papers adequately described comprehensive fidelity assessments embedded in the trials. The authors argued that to improve consumer trust, scientists need to implement, and then report, fidelity assessment findings. However, these authors also reported that fidelity assessment demands resources that many scientists lack, namely, money, time, and senior expertise. In contrast to these findings, we found that the use of videoconferencing allowed for a feasible fidelity assessment. The technological strategies modulated financial demands, maintained quality, minimized demands on content experts' time and allowed access to nationally recognized experts.

As we had access to the university-based telehealth system that allowed video-recording of the intervention sessions, we were able to complete this assessment with minimal cost. Unlike Teri's fidelity assessment (2010), we did not need to use direct observation strategies. If we had performed direct observation, we would have had to pay staff for the time and travel, costing a minimum of US $1500.00 [24]. None of our content experts or staff had to travel to any site for this fidelity project, making the assessment financially feasible.

Using video recordings allowed for high-quality, multidimensional assessment of the intervention sessions. The video recordings displayed participant and consultant nonverbal cues and body language [34] along with the audio. Participants consented to the recording, but the unobtrusive nature of the recording was, for the most part, ignored by the participants. Consultants were able to view the videos on their own schedules. This store-and-forward approach allowed the content experts to perform quality treatment fidelity assessments via a secure link, at their own pace. The literature supports the fidelity of assessing interactive sessions using store-and-forward strategies. An Agency for Healthcare Research and Quality study found that telehealth fidelity assessments work best when there is a high degree of human interaction (as in the Tele-STAR intervention) and less so when the activity involves complex diagnostic testing [34]. This ability to record the sessions speaks to a maturation of technology that can facilitate fidelity assessment at a lower cost, while at the same time retain quality.

Our study provided a unique opportunity to evaluate fidelity with a program (STAR-C) [15] that had undergone previous fidelity assessments [20]. Thus, we had published guidelines and skilled experts to facilitate the process. Further, even though some of the experts lived in other states, the technology allowed for content expert assessment and team meetings, minimizing time demands for all.

The videoconferencing technology not only allowed for implementation of the intervention, it provided a medium for expert access, indicating that content experts do not need to be on-site to provide valuable feedback and guidance. In our case, we were able to engage remote content experts who were intimately familiar with the earth-bound STAR-C program [15]. Their deep knowledge of the theoretical framework, the applied intervention, and the appropriate expectations for the consultants allowed for sophisticated mentoring for the PI (AL), skilled assessment of fidelity, and ready feedback for consultants in a timely manner.

Taken together, the usable technology, expert guidance, and published protocols facilitated a fidelity assessment that was informative and laid the groundwork for future assessments. Through this process, the limitations to our pilot study were highlighted. Specifically, the ordinal assessment scales prevented sophisticated statistical analyses for the fidelity assessment. Also of concern is the racial homogeneity of the sample. Although we were able to recruit underrepresented rural participants, we fell short of our goals in recruiting African American and Asian participants. Despite these limitations, we learned that our intervention adhered to the STAR-C [15] intervention and was effective in reducing caregiver burden.

Embarking on a fidelity assessment can be nerve-wracking for scientists. Researchers may be reluctant to perform a fidelity assessment because the assessment may identify poor fidelity to the protocols, which can undermine the scientific credibility [39]. This concern speaks to the need to include fidelity assessment in the pilot phases of early-stage research [19] and to engage experts in process. Our pilot study allowed for testing of our fidelity assessment, identified areas for improvement, and laid the foundation for more sophisticated assessments of future iterations of Tele-STAR.

As the use of technology for caregiver support expands, so does the risk for interventions to become detached from the original protocols. Yet, technology also allows for creative solutions that can both change an intervention or improve upon it. In this rapidly evolving world, tactics to maintain fidelity are essential to long-term programmatic success.

Acknowledgments

This study was funded by the Collins Medical Trust and the National Institute on Aging: P30AG008017; P30AG024978.

The authors would like to extend their gratitude to the families that participated in this study.

Conflicts of Interest

None declared.

References

1. Prince M. Alzheimer's Disease International. 2016. World Alzheimer Report 2016: Improving healthcare for people living with dementia URL: https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf [accessed 2019-05-15] [WebCite Cache ID 78OdfuL90]

2. Lindauer A, Harvath T. The meanings caregivers ascribe to dementia-related changes in care recipients: a meta-ethnography. Res Gerontol Nurs 2015;8(1):39-48. [doi: 10.3928/19404921-20141121-01] [Medline: 25490777]
3. van der Lee J, Bakker TJ, Duivenvoorden HJ, Dröes RM. Multivariate models of subjective caregiver burden in dementia: a systematic review. Ageing Res Rev 2014 May;15:76-93. [doi: 10.1016/j.arr.2014.03.003] [Medline: 24675045]
4. Gaugler JE, Yu F, Krichbaum K, Wyman JF. Predictors of nursing home admission for persons with dementia. Med Care 2009 Feb;47(2):191-198. [doi: 10.1097/MLR.0b013e31818457ce] [Medline: 19169120]
5. Kiely DK, Prigerson H, Mitchell SL. Health care proxy grief symptoms before the death of nursing home residents with advanced dementia. Am J Geriatr Psychiatry 2008 Aug;16(8):664-673 [FREE Full text] [doi: 10.1097/JGP.0b013e3181784143] [Medline: 18669945]
6. Lindauer A, Harvath TA. Pre-death grief in the context of dementia caregiving: a concept analysis. J Adv Nurs 2014 Oct;70(10):2196-2207. [doi: 10.1111/jan.12411] [Medline: 24702153]
7. Selwood A, Johnston K, Katona C, Lyketsos C, Livingston G. Systematic review of the effect of psychological interventions on family caregivers of people with dementia. J Affect Disord 2007 Aug;101(1-3):75-89. [doi: 10.1016/j.jad.2006.10.025] [Medline: 17173977]
8. Karlin NJ, Bell PA, Noah JL, Martichuski DK, Knight BL. Assessing Alzheimer's support group participation: a retrospective follow-up. Am J Alzheimers Dis (Columbia) 1999;14(6):326-333. [doi: 10.1177/153331759901400607]
9. Brodaty H, Thomson C, Thompson C, Fine M. Why caregivers of people with dementia and memory loss don't use services. Int J Geriatr Psychiatry 2005 Jun;20(6):537-546. [doi: 10.1002/gps.1322] [Medline: 15920707]
10. Boots LM, de Vugt ME, van Knippenberg RJ, Kempen GI, Verhey FR. A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. Int J Geriatr Psychiatry 2014 Apr;29(4):331-344. [doi: 10.1002/gps.4016] [Medline: 23963684]
11. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. J Med Internet Res 2018 Dec 12;20(6):e216 [FREE Full text] [doi: 10.2196/jmir.9548] [Medline: 29895512]
12. Griffiths PC, Whitney MK, Kovaleva M, Hepburn K. Development and implementation of Tele-Savvy for dementia caregivers: a Department of Veterans Affairs clinical demonstration project. Gerontologist 2016 Feb;56(1):145-154. [doi: 10.1093/geront/gnv123] [Medline: 26566806]
13. Austrom MG, Geros KN, Hemmerlein K, McGuire SM, Gao S, Brown SA, et al. Use of a multiparty web based videoconference support group for family caregivers: innovative practice. Dementia (London) 2015 Sep;14(5):682-690 [FREE Full text] [doi: 10.1177/1471301214544338] [Medline: 25062788]
14. Marziali E, Garcia LJ. Dementia caregivers' responses to 2 internet-based intervention programs. Am J Alzheimers Dis Other Demen 2011 Feb;26(1):36-43. [doi: 10.1177/1533317510387586] [Medline: 21282276]
15. Teri L, McCurry SM, Logsdon R, Gibbons LE. Training community consultants to help family members improve dementia care: a randomized controlled trial. Gerontologist 2005 Dec;45(6):802-811. [doi: 10.1093/geront/45.6.802] [Medline: 16326662]
16. McCurry SM, Logsdon RG, Mead J, Pike KC, La Fazia DM, Stevens L, et al. Adopting evidence-based caregiver training programs in the real world: outcomes and lessons learned from the STAR-C Oregon translation study. J Appl Gerontol 2017 Dec;36(5):519-536. [doi: 10.1177/0733464815581483] [Medline: 25873454]
17. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychol 2004 Sep;23(5):443-451. [doi: 10.1037/0278-6133.23.5.443] [Medline: 15367063]
18. Rixon L, Baron J, McGale N, Lorencatto F, Francis J, Davies A. Methods used to address fidelity of receipt in health intervention research: a citation analysis and systematic review. BMC Health Serv Res 2016 Dec 18;16(1):663 [FREE Full text] [doi: 10.1186/s12913-016-1904-6] [Medline: 27863484]
19. Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M. Reenvisioning clinical science: unifying the discipline to improve the public health. Clin Psychol Sci 2014 Jan 01;2(1):22-34 [FREE Full text] [doi: 10.1177/2167702613497932] [Medline: 25821658]
20. Teri L, McKenzie GL, Pike KC, Farran CJ, Beck C, Paun O, et al. Staff training in assisted living: evaluating treatment fidelity. Am J Geriatr Psychiatry 2010 Jun;18(6):502-509 [FREE Full text] [doi: 10.1097/JGP.0b013e3181c37b0e] [Medline: 19910884]
21. Gearing RE, El-Bassel N, Ghesquiere A, Baldwin S, Gillies J, Ngeow E. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. Clin Psychol Rev 2011 Feb;31(1):79-88. [doi: 10.1016/j.cpr.2010.09.007] [Medline: 21130938]
22. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2019-05-13] [WebCite Cache ID 78LXO9iiL]
23. Oregon Health & Science University. Decisionally Impaired (DI) Adults Supplement URL: https://ohsu.ellucid.com/documents/view/7690/?security=60109bca402a46fe68909634c9d8f8010f6f6529 [accessed 2019-05-13] [WebCite Cache ID 78LXfZroB]

24. Lindauer A, Croff R, Mincks K, Mattek N, Shofner SJ, Bouranis N, et al. "It Took the Stress out of Getting Help": the STAR-C-Telemedicine mixed methods pilot. Care Wkly 2018;2:7-14 [FREE Full text] [doi: 10.14283/cw.2018.4] [Medline: 30393783]
25. Teri L, Truax P, Logsdon R, Uomoto J, Zarit S, Vitaliano PP. Assessment of behavioral problems in dementia: the revised memory and behavior problems checklist. Psychol Aging 1992 Dec;7(4):622-631. [Medline: 1466831]
26. Jones C, Edwards RT, Hounsome B. Health economics research into supporting carers of people with dementia: a systematic review of outcome measures. Health Qual Life Outcomes 2012 Nov 26;10:142 [FREE Full text] [doi: 10.1186/1477-7525-10-142] [Medline: 23181515]
27. Lindauer A, Seelye A, Lyons B, Dodge HH, Mattek N, Mincks K, et al. Dementia care comes home: patient and caregiver assessment via telemedicine. Gerontologist 2017 Dec 01;57(5):e85-e93 [FREE Full text] [doi: 10.1093/geront/gnw206] [Medline: 28158415]
28. Bédard M, Molloy DW, Squire L, Dubois S, Lever JA, O'Donnell M. The Zarit Burden Interview: a new short version and screening version. Gerontologist 2001 Oct;41(5):652-657. [doi: 10.1093/geront/41.5.652] [Medline: 11574710]
29. Morycz RK. Caregiving strain and the desire to institutionalize family members with Alzheimer's disease. Possible predictors and model development. Res Aging 1985 Sep;7(3):329-361. [doi: 10.1177/0164027585007003002] [Medline: 4059641]
30. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. Psychosom Med 2002;64(3):510-519. [Medline: 12021425]
31. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
32. Teri L, Schmidt A. Understanding Alzheimer's: A guide for families, friends, and healthcare providers. Seattle, WA: University of Washington; 2010.
33. Knowles MS, Holton EF, Swanson RA. The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development. Burlington, MA: Elsevier; 2005.
34. Locatis C, Ackerman M. Three principles for determining the relevancy of store-and-forward and live interactive telemedicine: reinterpreting two telemedicine research reviews and other research. Telemed J E Health 2013 Jan;19(1):19-23 [FREE Full text] [doi: 10.1089/tmj.2012.0063] [Medline: 23186085]
35. Munro BH. Statistical Methods For Health Care Research, 5th Edition. Philadelphia, PA: Lippincott Williams & Wilkins; 2004.
36. Fleiss J. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76(5):378-382. [doi: 10.1037/h0031619]
37. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [Medline: 2348207]
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
39. Perepletchikova F, Hilt LM, Chereji E, Kazdin AE. Barriers to implementing treatment integrity procedures: survey of treatment outcome researchers. J Consult Clin Psychol 2009 Apr;77(2):212-218. [doi: 10.1037/a0015232] [Medline: 19309181]

Abbreviations

ABC: activators of the behaviors, the behaviors, and the consequences
GNP: gerontological nurse practitioner
HIPAA: Health Insurance Probability and Accountability Act
LPN: licensed practical nurse
OHSU: Oregon Health and Science University
PI: principal investigator
RMBPC: Revised Memory and Behavior Problems Checklist

Edited by G Eysenbach; submitted 05.02.19; peer-reviewed by M Gately, H Hah, C Kerssens; comments to author 28.03.19; revised version received 15.04.19; accepted 16.04.19; published 24.05.19.

Please cite as:
Lindauer A, McKenzie G, LaFazia D, McNeill L, Mincks K, Spoden N, Myers M, Mattek N, Teri LL
Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention
J Med Internet Res 2019;21(5):e13599
URL: http://www.jmir.org/2019/5/e13599/
doi:10.2196/13599
PMID:31127721

Original Paper

Predictors of Postal or Online Response Mode and Associations With Patient Experience and Satisfaction in the English Cancer Patient Experience Survey

Tra My Pham[1*], PhD; Gary A Abel[2*], PhD; Mayam Gomez-Cano[2*], PhD; Georgios Lyratzopoulos[1*], MD

[1]Epidemiology of Cancer Healthcare and Outcomes Group, Department of Behavioural Science and Health, University College London, London, United Kingdom

[2]University of Exeter Medical School (Primary Care), University of Exeter, Exeter, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Georgios Lyratzopoulos, MD
Epidemiology of Cancer Healthcare and Outcomes Group
Department of Behavioural Science and Health
University College London
London,
United Kingdom
Phone: 44 2076791719
Email: y.lyratzopoulos@ucl.ac.uk

Abstract

Background: Patient experience surveys are important tools for improving the quality of cancer services, but the representativeness of responders is a concern. Increasingly, patient surveys that traditionally used postal questionnaires are incorporating an online response option. However, the characteristics and experience ratings of online responders are poorly understood.

Objective: We sought to examine predictors of postal or online response mode, and associations with patient experience in the (English) Cancer Patient Experience Survey.

Methods: We analyzed data from 71,186 patients with cancer recently treated in National Health Service hospitals who responded to the Cancer Patient Experience Survey 2015. Using logistic regression, we explored patient characteristics associated with greater probability of online response and whether, after adjustment for patient characteristics, the online response was associated with a more or less critical evaluation of cancer care compared to the postal response.

Results: Of the 63,134 patients included in the analysis, 4635 (7.34%) responded online. In an adjusted analysis, male (women vs men: odds ratio [OR] 0.50, 95% confidence interval [CI] 0.46-0.54), younger (<55 vs 65-74 years: OR 3.49, 95% CI 3.21-3.80), least deprived (most vs least deprived quintile: OR 0.57, 95% CI 0.51-0.64), and nonwhite (nonwhite vs white ethnic group: OR 1.37, 95% CI 1.24-1.51) patients were more likely to respond online. Compared to postal responders, after adjustment for patient characteristics, online responders had a higher likelihood of reporting an overall satisfied experience of care (OR 1.24, 95% CI 1.16-1.32). For 34 of 49 other items, online responders more frequently reported a less than positive experience of care (8 reached statistical significance), and the associations were positive for the remaining 15 of 49 items (2 reached statistical significance).

Conclusions: In the context of a national survey of patients with cancer, online and postal responders tend to differ in their characteristics and rating of satisfaction. Associations between online response and reported experience were generally small and mostly nonsignificant, but with a tendency toward less than positive ratings, although not consistently. Whether the observed associations between response mode and reported experience were causal needs to be examined using experimental survey designs.

(J Med Internet Res 2019;21(5):e11855) doi:10.2196/11855

KEYWORDS

cancer; patient survey; satisfaction; experience; online response; Web; internet

Introduction

Patient experience is an important aspect of quality of care [1]. In England, the National Health Service (NHS) has embarked on major policy initiatives regarding the measurement of patient experience through nationwide surveys since 2009 [2,3]. Relatedly, the National Cancer Strategy indicates that patient experience should be considered as "being on a par with clinical effectiveness and safety" [4].

The English Cancer Patient Experience Survey (CPES), a major nationwide survey of patients recently treated for cancer, was first undertaken in 2010. Internationally, it is the largest annual nationwide survey of patients with cancer [5]. Survey items assess key domains of patient experience [6-8], including the interpersonal skills of care providers, the provision of information about cancer diagnosis and its treatment, and the experience of access and timeliness of care, care coordination, and satisfaction with care. Its findings are reported publicly, both at the national level and for each hospital and health care commissioning organization. The survey has consistently had a high response rate across different waves (approximately 65%-67%) [9].

The CPES was chiefly a postal survey until 2015 when an online response option was introduced. In this new context, it is important to understand which groups of patients tend to use the online response mode, and whether patients using different response modes rate their care experience differently. Such understanding can help to establish whether comparisons between patient groups or between hospitals and over time might be impacted by the introduction of the online response option and variation in its use [10].

With these prior considerations, we aimed to examine the sociodemographic and cancer site predictors of the postal or online response mode in CPES 2015 and, subsequently, to examine the associations between response mode and key aspects of the cancer patient experience.

Methods

Data Source

We analyzed anonymous data from responders to CPES 2015 [11]. The survey was commissioned by NHS England and carried out by Quality Health, a specialist survey provider. The sampling frame consisted of patients aged 16 years and older who were treated for cancer in English NHS hospitals from April to June 2015. After relevant vital status checks, patients were mailed the questionnaire (with two reminders for nonresponders). Patients could complete and return the survey questionnaire by post or complete it online. The questionnaire could also be completed by phone via a freephone helpline, which also provided a translation and interpreting facility for patients whose first language was not English. Of the 108,269 initially sampled eligible patients, 71,186 completed the questionnaire (response rate=65.75%).

Variables

Information was available on responders' age group (<55 years, 55-64 years, 65-74 years, ≥75 years), sex, self-reported ethnic group (white, nonwhite), and deprivation status (based on quintiles of the Index of Multiple Deprivation [IMD] scores of the lower super output area of patients' residence) [12]. Patients' cancer diagnoses were categorized into the 11 major cancer sites (endometrial, melanoma, leukemia, rectal, lung, non-Hodgkin lymphoma, bladder, colon, multiple myeloma, prostate, breast) and an "other" group using the *International Classification of Diseases, Tenth Revision*, diagnosis code (based on hospital records) [13]. To ensure conformity with a strict anonymization standard (ie, regarding the minimum count of patients in a given stratum), ethnicity information in our analysis sample was suppressed by the data provider for 3064 (4.30%) patients; these were patients with melanoma, renal, and thyroid cancers (Multimedia Appendix 1). Information on the survey response mode was also available and categorized into three groups (postal, online, and other).

The survey consisted of 49 evaluative questions reflecting different aspects of the cancer care experience, with a question on overall satisfaction with care (#59: "Overall, how would you rate your care?") (M Gomez-Cano et al, unpublished data, 2019). Of the 49 evaluative questions, 7 had binary response options and 42 used a Likert-response format. We used binary forms of the latter items ("positive" and "less than positive" experience categories), consistent with public reporting conventions of the survey [14]. Overall satisfaction with care was rated using scores 0 (very poor) to 10 (very good); answers to this question were dichotomized into two categories of "satisfied" (scores 9 or 10) or "less than satisfied" (scores 0 to 8). We were a priori interested in the question regarding overall satisfaction with cancer care (#59) separately to the other (experience) items, considering satisfaction as an outcome of care experience (M Gomez-Cano et al, unpublished data, 2019).

We excluded individuals who responded to the questionnaire with modes other than postal or online, had missing or suppressed ethnic group, or had missing deprivation information (ie, information on sex and age group was complete). Data from 63,134 responders were used for all analyses, representing 88.69% of the total responders' sample (Multimedia Appendix 2). For associations between response mode and reported experience, responders with missing or noninformative answers (eg, "don't know/can't say") to the survey questions were further excluded, resulting in variation in sample size across the different questions.

Statistical Analyses

To examine predictors of online response, crude proportions of responders using the postal and online response modes were calculated by sociodemographic characteristic and cancer site variables. Univariable and multivariable logistic regression models were used to obtain (1) the unadjusted odds ratios (ORs) of online response, and (2) the ORs of online response adjusted for all patient characteristic (age group, sex, ethnic group, deprivation status) and cancer site variables considered. From the latter model, covariate-adjusted proportions of online responders were predicted and compared with the corresponding crude proportions.

Similarly, to examine the associations between postal or online response mode and reported experience, for each of the 50 survey questions, univariable and multivariable logistic regression models were used to obtain (1) the unadjusted ORs of reporting a satisfied/positive experience for response mode, and (2) the OR of reporting a satisfied/positive experience for response mode, adjusted for all patient characteristic and cancer site variables considered.

Supplementary Analyses

In supplementary analyses examining the association between response mode and overall satisfaction with care (#59), the sensitivity of the main analysis findings to another cut-off choice was explored, using scores 8 to 10 for a satisfied experience instead of scores 9 or 10 as in the main analysis. For this item, we also examined pairwise interactions between response mode and each sociodemographic variable in the multivariable logistic regression model. All analyses were conducted using Stata version 15.1 (StataCorp LP, College Station, TX, USA).

Results

Predictors of Online Response

Of the 63,134 patients included in this analysis, 58,499 (92.66%) completed the survey by post and 4635 (7.34%) online. In univariable analyses, online response mode was less likely among women (OR 0.68, 95% confidence interval [CI] 0.64-0.72 for women versus men) (Table 1). Increasing age was associated with lower likelihood of online response (OR 0.40, 95% CI 0.36-0.44 for ≥75 versus 65-74 years). Increasing level of deprivation was similarly associated with lower likelihood of online response (OR 0.70, 95% CI 0.63-0.78 for IMD quintile 5 versus quintile 1). Nonwhite patients were more likely to respond online compared to white patients (OR 1.62, 95% CI 1.47-1.78). There was also evidence for variation in the odds of online response between patients across different cancer sites (joint *P* value <.001), with leukemia associated with the greatest odds and lung cancer with the lowest odds of online response compared with rectal cancer.

Similar patterns of variation and related estimates to those obtained in the univariable analyses were also observed in the adjusted analysis, suggestive of an overall small degree of confounding between cancer site and sociodemographic variables (Table 1 and Figure 1).

Associations Between Response Mode and Reported Experience

There were 60,921 patients in the analysis sample who answered the question about overall satisfaction with cancer care (#59). Of these, 22,030 (36.16%) patients gave a response in the “less than satisfied” category (defined as scores 0 to 8) and 38,891 (63.84%) patients responded in the “satisfied” category (defined as scores 9 or 10). Online responders were more likely to report a satisfied experience compared to those who responded by post (OR 1.16, 95% CI 1.09-1.24). Female, younger, more deprived, and nonwhite responders were more likely to report a less than satisfied experience (Table 2). There were also differences in overall satisfaction across cancer sites (joint *P* value <.001), with non-Hodgkin lymphoma being associated with the greatest likelihood and bladder cancer with the lowest likelihood of reporting a satisfied experience of care (Table 2 and Figure 2).

Adjusting for demographic characteristic, cancer site, and response mode variables led to some changes in the estimated associations for response mode, sex, and cancer site. In particular, adjustment accentuated the difference between online and postal responders (OR 1.24, 95% CI 1.16-1.32) (Table 2 and Figure 2).

For each of the remaining 49 questions, the overall percentage of responders reporting a positive experience ranged from 28.89% (17,167/59,430) for question #58 (“patient asked to take part in cancer research”) to 95.63% (52,663/55,067) for question #42 (“cancer doctor had the right documents at outpatient appointment”) (Multimedia Appendix 3). Figure 3 presents the ORs of reporting a positive experience for responders who completed the questionnaire online compared with those who responded by post, both unadjusted and adjusted for patient demographic characteristic and cancer site variables. In general, the unadjusted associations tended to be more negative than the adjusted ones, possibly indicating a degree of confounding by the differences in the sociodemographic characteristics of online and postal responders, which became attenuated when age and other factors were taken into account.

Considering the adjusted analyses, there was evidence for an association between response mode and reported experience for 10 of 49 questions examined, although in opposite directions: for 8 questions, the online response mode was associated with a less than positive experience. These consisted of questions about “patient given all information needed about chemotherapy treatment” (#47), “overall rating of administration of care” (#56), “different people treating and caring work well together” (#54), “doctors and nurses asked what name patient preferred to be called by” (#33), “patient had confidence and trust in ward nurses” (#31), “patient told they could bring family/friend when first told they had cancer” (#8), “easy to contact clinical nurse specialist” (#18), and “patient found hospital staff to talk to about worries and fears during hospital visit” (#35). Conversely, for two questions online response mode was associated with a positive experience. Specifically, the questions about “doctors and nurses gave family/someone close to patient all information to help care at home” (#49) and “doctors and nurses talked in front of patient as if they were not there” (#28) (Figure 3).

Table 1. Predictors of online response: frequency and percentage of online response by patient characteristic and cancer site variables and related crude and adjusted odds ratios of online response (N=63,134).

Variable	Frequency and percentage of online response		Univariable logistic regression models		Multivariable logistic regression model	
	n (%)	Total, N	OR[a] (95% CI)	*P* value[b]	aOR[c] (95% CI)	*P* value[b]
Sex				<.001		<.001
Male	2543 (8.78)	28,973	1		1	
Female	2092 (6.12)	34,161	0.68 (0.64-0.72)		0.50 (0.46-0.54)	
Age group (years)				<.001		<.001
<55	1457 (15.28)	9535	2.82 (2.61-3.05)		3.49 (3.21-3.80)	
55-64	1377 (10.66)	12,913	1.87 (1.72-2.02)		2.05 (1.89-2.22)	
65-74	1347 (6.01)	22,395	1		1	
≥75	454 (2.48)	18,291	0.40 (0.36-0.44)		0.39 (0.35-0.43)	
IMD score				<.001		<.001
Quintile 1 (least deprived)	1231 (8.06)	15,264	1		1	
Quintile 2	1163 (7.84)	14,832	0.97 (0.89-1.05)		0.96 (0.88-1.04)	
Quintile 3	1016 (7.33)	13,853	0.90 (0.83-0.98)		0.85 (0.78-0.93)	
Quintile 4	749 (6.83)	10,963	0.84 (0.76-0.92)		0.75 (0.68-0.83)	
Quintile 5 (most deprived)	476 (5.79)	8222	0.70 (0.63-0.78)		0.57 (0.51-0.64)	
Ethnic group				<.001		<.001
White	4083 (7.03)	58,067	1		1	
Nonwhite	552 (10.89)	5067	1.62 (1.47-1.78)		1.37 (1.24-1.51)	
Cancer site[d]				<.001		.01
Prostate	527 (9.21)	5725	1.23 (1.05-1.44)		1.19 (1.01-1.41)	
Leukemia	220 (9.27)	2373	1.24 (1.02-1.50)		1.15 (0.95-1.41)	
Endometrial	78 (5.55)	1406	0.71 (0.55-0.93)		1.11 (0.84-1.46)	
Non-Hodgkin lymphoma	350 (7.88)	4444	1.04 (0.87-1.23)		1.10 (0.92-1.31)	
Colon	308 (6.80)	4530	0.89 (0.74-1.06)		1.03 (0.86-1.24)	
Other cancers	1265 (7.76)	16,292	1.02 (0.88-1.18)		1.01 (0.87-1.18)	
Rectal	235 (7.61)	3087	1		1	
Breast	949 (7.35)	12,904	0.96 (0.83-1.12)		1.00 (0.85-1.18)	
Multiple myeloma	298 (6.54)	4555	0.85 (0.71-1.02)		0.99 (0.83-1.19)	
Bladder	239 (5.39)	4437	0.69 (0.57-0.83)		0.96 (0.80-1.17)	
Lung	166 (4.91)	3381	0.63 (0.51-0.77)		0.83 (0.67-1.02)	
Total	4635 (7.34)	63,134				

[a] Unadjusted odds ratios (ORs) of online response from a series of univariable logistic regression models, conditional on each patient characteristic and cancer site variable considered.

[b] *P* values from joint Wald tests.

[c] Adjusted odds ratios (aORs) of online response from a multivariable logistic regression model, conditional on all patient characteristic and cancer site variables considered.

[d] Responders with renal and thyroid cancers (grouped into the “other” category) and melanoma skin cancer excluded due to their ethnic group being suppressed or missing.

Figure 1. Predictors of online response: crude and covariate-adjusted proportions of online response.

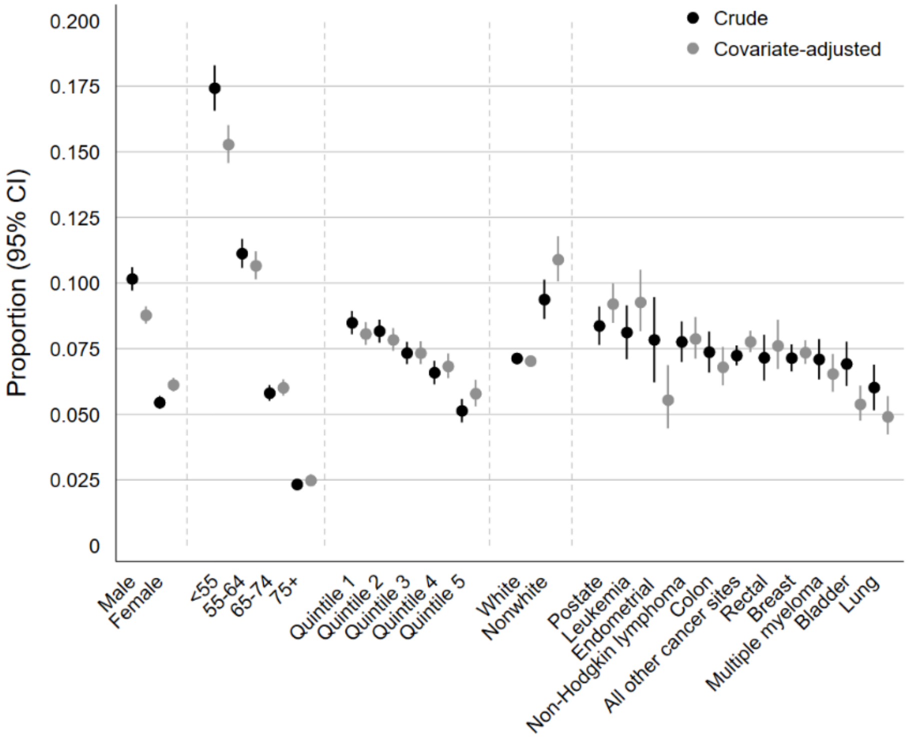


Table 2. Associations between response mode and reporting a satisfied experience of cancer care (scores 9 or 10 to question 59): frequency and percentage and related crude and adjusted odds ratios of reporting a satisfied experience (N=60,921).

Variable	Frequency and percentage of reporting a satisfied experience		Univariable logistic regression models		Multivariable logistic regression model	
	n (%)	Total, N	OR[a] (95% CI)	*P* value[b]	aOR[c] (95% CI)	*P* value[b]
Response mode				<.001		<.001
Postal	35,809 (63.58)	56,318	1		1	
Online	3082 (66.96)	4603	1.16 (1.09-1.24)		1.24 (1.16-1.32)	
Sex				.01		<.001
Male	18,006 (64.36)	27,976	1		1	
Female	20,885 (63.39)	32,945	0.96 (0.93-0.99)		0.89 (0.85-0.92)	
Age group (years)				<.001		<.001
<55	5551 (59.73)	9294	0.77 (0.73-0.81)		0.75 (0.71-0.79)	
55-64	7828 (62.18)	12,590	0.85 (0.81-0.89)		0.84 (0.81-0.88)	
65-74	14,286 (65.88)	21,686	1		1	
≥75	11,226 (64.70)	17,351	0.95 (0.91-0.99)		0.97 (0.93-1.01)	
IMD score				<.001		<.001
Quintile 1 (least deprived)	9575 (64.82)	14,772	1		1	
Quintile 2	9326 (65.10)	14,326	1.01 (0.97-1.06)		1.02 (0.97-1.07)	
Quintile 3	8613 (64.41)	13,373	0.98 (0.94-1.03)		1.01 (0.96-1.06)	
Quintile 4	6513 (61.65)	10,565	0.87 (0.83-0.92)		0.92 (0.87-0.97)	
Quintile 5 (most deprived)	4864 (61.69)	7885	0.87 (0.83-0.93)		0.95 (0.89-1.00)	
Ethnic group				<.001		<.001
White	36,343 (64.80)	56,088	1		1	
Nonwhite	2548 (52.72)	4833	0.61 (0.57-0.64)		0.63 (0.59-0.67)	
Cancer site[d]				<.001		<.001
Non-Hodgkin lymphoma	3025 (70.25)	4306	1.36 (1.23-1.50)		1.38 (1.25-1.53)	
Leukemia	1604 (69.86)	2296	1.34 (1.19-1.50)		1.36 (1.21-1.53)	
Breast	8257 (65.95)	12,521	1.12 (1.03-1.21)		1.30 (1.19-1.42)	
Endometrial	880 (65.67)	1340	1.10 (0.96-1.26)		1.21 (1.06-1.39)	
Colon	2800 (64.35)	4351	1.04 (0.94-1.15)		1.05 (0.95-1.16)	
Rectal	1898 (63.44)	2992	1		1	
Multiple myeloma	2788 (63.31)	4404	0.99 (0.90-1.10)		0.99 (0.90-1.09)	
Other	9634 (61.39)	15,692	0.92 (0.85-0.99)		0.96 (0.89-1.05)	
Lung	2002 (61.45)	3258	0.92 (0.83-1.02)		0.92 (0.83-1.02)	
Bladder	2600 (61.19)	4249	0.91 (0.83-1.00)		0.86 (0.78-0.95)	
Prostate	3403 (61.74)	5512	0.93 (0.85-1.02)		0.86 (0.78-0.94)	
Total	38,891 (63.84)	60,921				

[a] Unadjusted odds ratios (ORs) of reporting a satisfied experience from a series of univariable logistic regression models, conditional on each of the response mode, patient characteristic and cancer site variables considered.

[b] *P* values from joint Wald tests.

[c] Adjusted odds ratios (aORs) of reporting a satisfied experience from a multivariable logistic regression model, conditional on response mode and all patient characteristic and cancer site variables considered.

[d] Responders with renal and thyroid cancers (grouped into the “other” category) and melanoma skin cancer excluded due to their ethnic group being suppressed or missing.

Figure 2. Associations between response mode and reporting a satisfied experience of cancer care (scores 9 or 10 to question 59): adjusted odds ratios of reporting a satisfied experience. Ref: reference category.

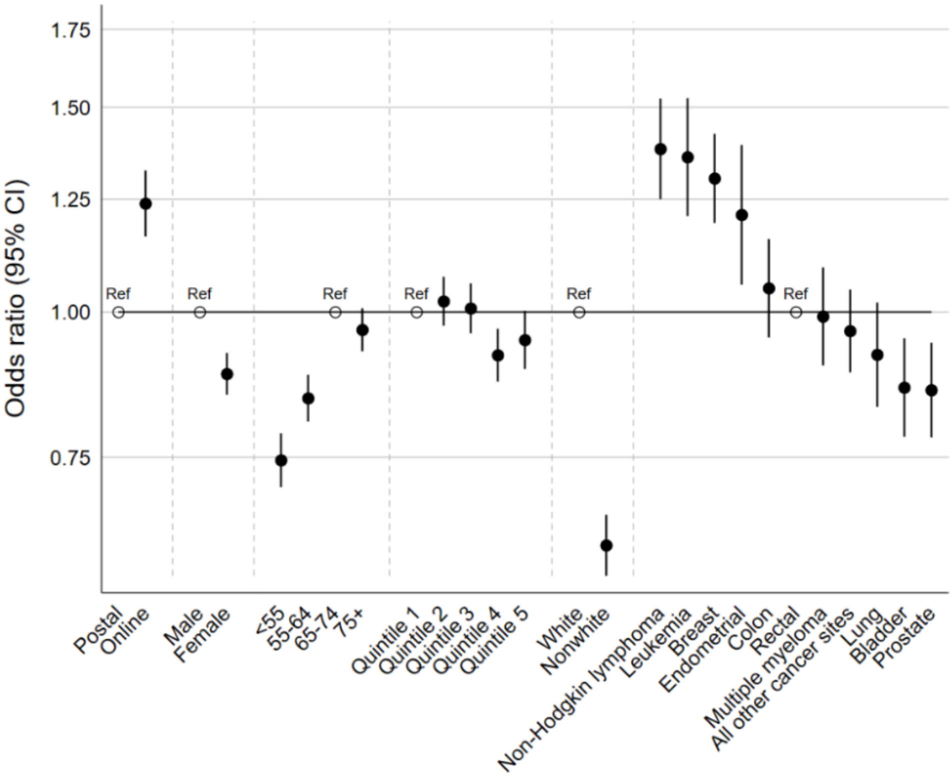
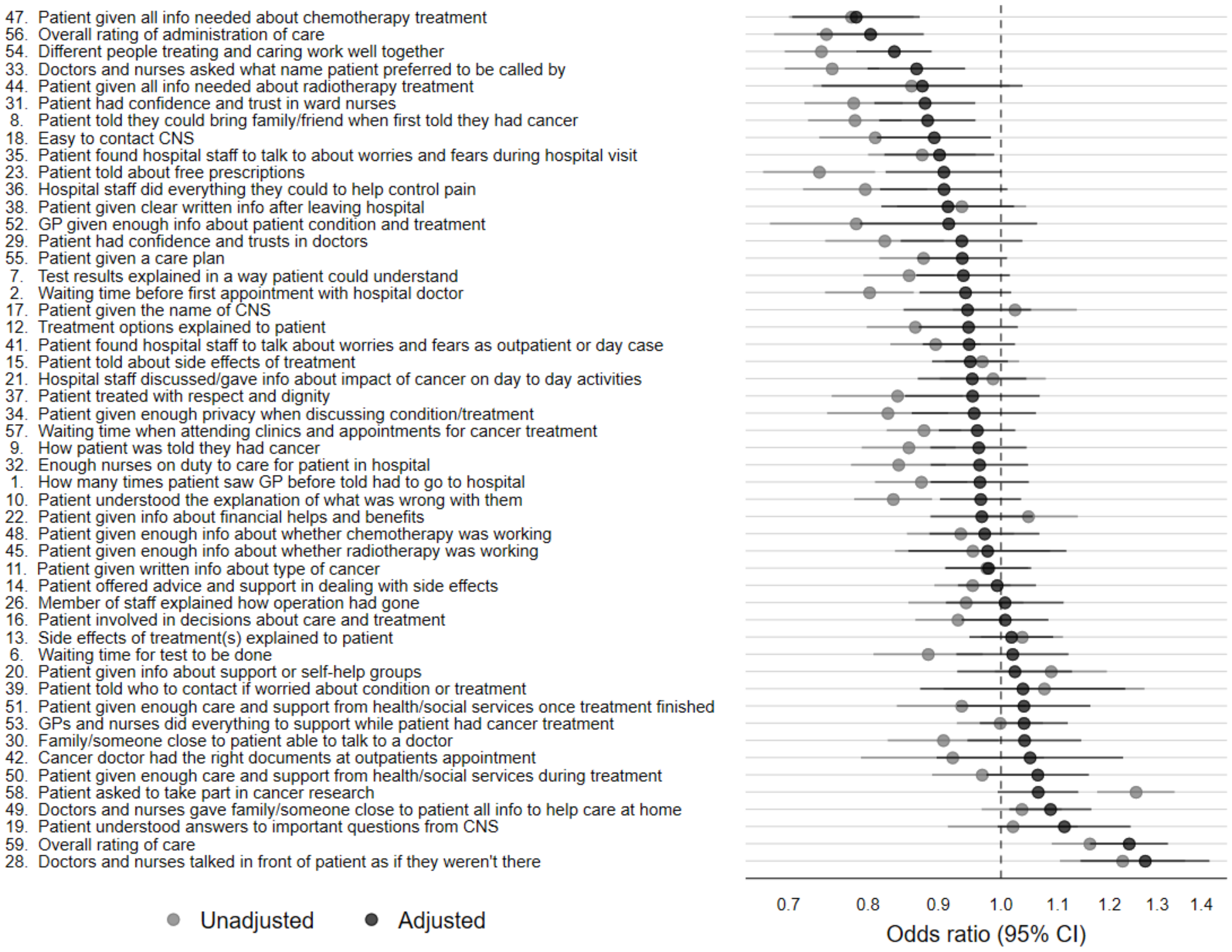


Figure 3. Associations between response mode and reported experience: adjusted odds ratios of reporting a satisfied/positive experience for online versus postal responders, by question. CNS: clinical nurse specialist; GP: general practitioner.



Multimedia Appendix 4 presents the variation in observed *P* values for the association between online response mode and reported experience for all 50 questions (49 aspect-specific questions plus the question addressing overall satisfaction) after adjustment for patient characteristic and cancer site variables. The observed variation is compared with what might be expected under the null hypothesis of no association (straight line). If there were no true associations, then two to three questions would be expected to have a *P* value less than .05 (dashed line) by chance alone, and the observed distribution would follow the straight line. The distribution of *P* values across these questions suggests that it is unlikely that online response mode was unrelated to reported experience.

Supplementary Analyses

When considering the association between response mode and overall satisfaction with cancer care, defining an alternative cut-off point for overall satisfaction (scores 8 to 10) yielded broadly comparable results to that using the original cut-off point (scores 9 or 10) (Multimedia Appendix 5). For this item, there was also strong evidence of an interaction between response mode and sex, as well as between response mode and ethnic group (*P* value for each interaction <.001). In contrast, there was no evidence for an interaction between response mode and any of the age group, social deprivation, or cancer site variables. Therefore, although for both response modes women were less likely than men to report a satisfied experience of care, this difference by sex was stronger for female online responders (OR 0.76, 95% CI 0.67-0.86 versus men) compared with female postal responders (OR 0.89, 95% CI 0.86-0.93 versus men). Furthermore, although nonwhite postal responders were less likely to report a satisfied experience of their care (OR 0.60, 95% CI 0.56-0.64 versus white ethnic group), there was no evidence for such a difference among nonwhite online responders (OR 0.94, 95% CI 0.78-1.14 versus white ethnic group) (Multimedia Appendix 6).

Discussion

Summary of Findings

Using data from a major English experience survey of patients recently treated with cancer, we found that a notable minority (approximately 1 in 14) of all participants responded to the survey online, with male, younger, least deprived, and nonwhite patients being more likely to do so. However, when examining associations between response mode and aspects of cancer patient experience, we observed statistically nonsignificant associations, with point estimates generally indicative of less frequent positive ratings for online responders. There were some exceptions; we found three questions with evidence that online responders were more likely to respond positively, particularly the question on overall satisfaction with care.

Strengths and Limitations

This study was based on the analysis of a large nationwide sample of responders, which allowed us to examine potential interactions, beyond main effect analyses, and we were able to adequately adjust for a range of patient characteristics (namely age group, sex, deprivation, ethnic group, and cancer site) of known relevance to patient experience surveys [15-17]. An important limitation is that we were unable to directly examine how reported experience related to actual care experience. As such, we could not establish whether the online response mode affected how people reported their experience of care, or whether patients who would have provided lower or higher ratings of care (regardless of response mode) were more likely to respond online. The challenge of drawing appropriate inferences in respect of this research question is further complicated by the fact that the associations between response mode and patient experience that we examined were heterogeneous in their presence, direction, and size. Another limitation of our analysis is that we had no means of examining potential differences by response mode in usability (eg, the time taken to complete and post or submit the survey questionnaire).

To aid interpretation of findings, we conducted a post hoc comparison of how each question appeared in the postal and online questionnaires. For this, we used the online demo questionnaire of the 2017 survey provided by the survey provider (Quality Health), which was identical in format to that of the 2015 survey, and the published postal questionnaire of the 2015 survey. Allowing for the difference in the medium, we found that the presentation of items was identical between the postal and online questionnaires, except for the question regarding overall satisfaction, which was presented slightly differently. In particular, the anchoring text for this question on the postal version covered scores 8 to 10 for “very good”, whereas it only covered 10 on the online version, with similar differences at the opposite end of the scale (Multimedia Appendix 7). This discrepancy in the appearance of the question might contribute to explaining the difference in overall rating of satisfaction between postal and online responders. However, we urge caution in this interpretation because this item is not particularly unusual in its association with online response mode, considering the distribution of associations observed for the other items (Figure 3).

Findings in Relation to Other Evidence and Implications

We were not aware of any relevant literature on the use of an online response option specifically in cancer patient experience surveys. However, a previous German study that examined health behaviors using data from a population-based longitudinal panel reported an overall equivalence in responses obtained from the Web-only response mode arm compared with a mixed mode (paper or Web) arm [18].

Although not the principal focus of our inquiry, we confirmed a previously identified variation in the association between sociodemographic characteristics and cancer site with rated satisfaction or experience of cancer care [15,19]. The findings of a higher probability of online response by male, younger, and less deprived patients might be expected, but nonwhite patients were also more likely to respond online. This observation contrasts with the patterns of variation by ethnic group in use of the internet among members of the general population [20]. Further research to help understand this ethnic variation is needed.

In survey research, it is generally important to consider whether response mode ought to be adjusted for when examining sociodemographic inequalities, trends over time, or when considering organizational comparisons [21]. The answer to this research question depends on the direction of causality between response mode and reported patient experience. We advocate the need for high-definition experimental studies in small, yet adequately powered, subsamples of responders who will be otherwise matched for all their characteristics except response mode. For example, Elliott et al [10] conducted a randomized controlled trial to examine the influence of survey response mode on experience ratings in the context of the Consumer Assessments of Healthcare Providers and Systems Hospital Survey. The results suggested the need for adjusting for survey response mode in the calculation of hospital scores. Unlike other forms of randomized controlled trials that are associated with substantial practical and ethical barriers, such trials are relatively easy to conduct in the context of survey research.

Conclusions

We described the sociodemographic and cancer site predictors of the online response option in a major national survey in England, and examined potential associations of response mode with rated satisfaction and experience of cancer care. The findings highlighted that online and postal responders differed in their patient characteristics, with less evidence for variation between online and postal responders in terms of experience ratings. Whether the association between response mode and satisfaction with care is causal in the context of cancer patient experience surveys needs to be examined experimentally.

Acknowledgments

This work is supported by Macmillan Cancer Support grant 5995414 for which GAA and GL are joint principal investigators. GL is supported by a Cancer Research UK Advanced Clinician Scientist Fellowship Award (C18081/A18180). We thank the UK Data Archive for access to the anonymous survey data (UKDA Study Number SN 8163 for CPES 2015), the Department of Health as the depositor and principal investigator of the survey, and Quality Health as the data collector. We are also grateful to

Quality Health for enabling access to the online demo questionnaire of CPES 2017. We thank all patients who participated in the survey.

Authors' Contributions

All authors made substantial contributions to the conception and design of the study, analysis and interpretation of data, drafting the article and revising it critically for important intellectual content, and final approval of the version to be submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of sociodemographic variables among patients with suppressed ethnic group.

[PDF File (Adobe PDF File), 22KB - jmir_v21i5e11855_app1.pdf]

Multimedia Appendix 2

Analysis sample derivation.

[PNG File, 29KB - jmir_v21i5e11855_app2.png]

Multimedia Appendix 3

Associations between response mode and reported experience: question-specific sample size, percentage of a satisfied/positive experience, and odds ratio of reporting a satisfied/positive experience for online versus postal responders, adjusted for patient characteristic and cancer site variables, by question. CNS: clinical nurse specialist; GP: general practitioner.

[PDF File (Adobe PDF File), 50KB - jmir_v21i5e11855_app3.pdf]

Multimedia Appendix 4

Associations between response mode and reported experience: variation in observed P values for the association between online response mode and reported experience for 50 examined questions.

[PNG File, 62KB - jmir_v21i5e11855_app4.png]

Multimedia Appendix 5

Supplementary analyses: adjusted odds ratios of reporting a satisfied experience of cancer care (scores 8 to 10 to question 59). Ref: reference category.

[PNG File, 81KB - jmir_v21i5e11855_app5.png]

Multimedia Appendix 6

Supplementary analyses: adjusted odds ratios of reporting a satisfied experience of cancer care (scores 9 or 10 to question 59) with the inclusion of pairwise interactions of response mode with sex and ethnic group (N=60,921).

[PDF File (Adobe PDF File), 29KB - jmir_v21i5e11855_app6.pdf]

Multimedia Appendix 7

Associations between response mode and reported experience: comparison of the presentation of question 59 on overall satisfaction with cancer care in the postal and online questionnaires.

[PDF File (Adobe PDF File), 43KB - jmir_v21i5e11855_app7.pdf]

References

1. Burt J, Campbell J, Abel G, Aboulghate A, Ahmed F, Asprey A, et al. Improving patient experience in primary care: a multimethod programme of research on the measurement and improvement of patient experience. Programme Grants Appl Res 2017;5(9):1 [FREE Full text] [doi: 10.3310/pgfar05090] [Medline: 28654227]
2. American Society of Clinical Oncology, European Society for Medical Oncology. ASCO-ESMO consensus statement on quality cancer care. Ann Oncol 2006 Jul;17(7):1063-1064. [doi: 10.1093/annonc/mdl152] [Medline: 16751232]

3. Department of Health. The NHS Outcomes Framework 2013/14. London: Department of Health; 2012 Nov 13. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213055/121109-NHS-Outcomes-Framework-2013-14.pdf [accessed 2018-05-08] [WebCite Cache ID 6zGOFHErD]
4. Independent Cancer Taskforce. Achieving World-Class Cancer Outcomes: A Strategy for England 2015-2020. 2015. URL: http://www.cancerresearchuk.org/sites/default/files/achieving_world-class_cancer_outcomes_-_a_strategy_for_england_2015-2020.pdf [accessed 2018-05-08] [WebCite Cache ID 6zGOWN2g4]
5. Quality Health. National Cancer Patient Experience Survey. URL: http://www.ncpes.co.uk/ [accessed 2018-05-08] [WebCite Cache ID 6zGObrCQ1]
6. Jenkinson C, Coulter A, Bruster S. The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries. Int J Qual Health Care 2002;14(5):353-358. [Medline: 12389801]
7. Paddison CA, Abel GA, Roland MO, Elliott MN, Lyratzopoulos G, Campbell JL. Drivers of overall satisfaction with primary care: evidence from the English General Practice Patient Survey. Health Expect 2015;18(5):1081-1092 [FREE Full text] [doi: 10.1111/hex.12081] [Medline: 23721257]
8. Elliott MN, Kanouse DE, Edwards CA, Hilborne LH. Components of care vary in importance for overall patient-reported experience by type of hospitalization. Med Care 2009;47(8):842-849. [doi: 10.1097/MLR.0b013e318197b22a] [Medline: 19584764]
9. Abel GA, Saunders CL, Lyratzopoulos G. Post-sampling mortality and non-response patterns in the English Cancer Patient Experience Survey: Implications for epidemiological studies based on surveys of cancer patients. Cancer Epidemiol 2016;41:34-41 [FREE Full text] [doi: 10.1016/j.canep.2015.12.010] [Medline: 26797675]
10. Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. Health Serv Res 2009;44(2):501-518 [FREE Full text] [doi: 10.1111/j.1475-6773.2008.00914.x] [Medline: 19317857]
11. Quality Health, NHS England. UK Data Service. 2017. National Cancer Patient Experience Survey, 2015. SN: 8163 URL: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8163 [accessed 2018-05-08] [WebCite Cache ID 77O2AnFhj]
12. The National Archives. Communities and neighbourhoods: indices of deprivation 2007 URL: http://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/ [accessed 2018-05-08] [WebCite Cache ID 6zGOthxq1]
13. World Health Organization. ICD-10 Version:2016. 2016. International Statistical Classification of Diseases and Related Health Problems 10th Revision URL: https://icd.who.int/browse10/2016/en [accessed 2018-05-08] [WebCite Cache ID 6zGP3Tx2y]
14. Quality Health. National Cancer Patient Experience Survey. 2015 Reports URL: http://www.ncpes.co.uk/index.php/reports/2015-reports [accessed 2018-05-08] [WebCite Cache ID 6zGP6XcAo]
15. Saunders CL, Abel GA, Lyratzopoulos G. Inequalities in reported cancer patient experience by socio-demographic characteristic and cancer site: evidence from respondents to the English Cancer Patient Experience Survey. Eur J Cancer Care (Engl) 2015;24(1):85-98 [FREE Full text] [doi: 10.1111/ecc.12267] [Medline: 25327713]
16. Lyratzopoulos G, Elliott M, Barbiere JM, Henderson A, Staetsky L, Paddison C, et al. Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. BMJ Qual Saf 2012;21(1):21-29 [FREE Full text] [doi: 10.1136/bmjqs-2011-000088] [Medline: 21900695]
17. Paddison C, Elliott M, Parker R, Staetsky L, Lyratzopoulos G, Campbell JL, et al. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. BMJ Qual Saf 2012 Aug;21(8):634-640 [FREE Full text] [doi: 10.1136/bmjqs-2011-000737] [Medline: 22626735]
18. Rübsamen N, Akmatov MK, Castell S, Karch A, Mikolajczyk RT. Comparison of response patterns in different survey designs: a longitudinal panel with mixed-mode and online-only design. Emerg Themes Epidemiol 2017;14:4 [FREE Full text] [doi: 10.1186/s12982-017-0058-2] [Medline: 28344629]
19. El Turabi A, Abel GA, Roland M, Lyratzopoulos G. Variation in reported experience of involvement in cancer treatment decision making: evidence from the National Cancer Patient Experience Survey. Br J Cancer 2013;109(3):780-787 [FREE Full text] [doi: 10.1038/bjc.2013.316] [Medline: 23807170]
20. Office for National Statistics. GOV.UK. 2018. Ethnicity facts and figures: Internet use URL: https://www.ethnicity-facts-figures.service.gov.uk/culture-and-community/digital/internet-use/latest [accessed 2018-05-08] [WebCite Cache ID 6zGPBtI2O]
21. Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD. Patterns of unit and item nonresponse in the CAHPS Hospital Survey. Health Serv Res 2005;40(6):2096-2119 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00476.x] [Medline: 16316440]

Abbreviations

CI: confidence interval
CPES: Cancer Patient Experience Survey

IMD: Index of Multiple Deprivation
NHS: National Health Service
OR: odds ratio

Edited by G Eysenbach; submitted 07.08.18; peer-reviewed by T Freeman, J Campbell; comments to author 08.11.18; revised version received 02.01.19; accepted 21.01.19; published 02.05.19.

Please cite as:
Pham TM, Abel GA, Gomez-Cano M, Lyratzopoulos G
Predictors of Postal or Online Response Mode and Associations With Patient Experience and Satisfaction in the English Cancer Patient Experience Survey
J Med Internet Res 2019;21(5):e11855
URL: https://www.jmir.org/2019/5/e11855/
doi:10.2196/11855
PMID:31045503

Original Paper

Paper Versus Digital Data Collection for Road Safety Risk Factors: Reliability Comparative Analysis From Three Cities in Low- and Middle-Income Countries

Amber Mehmood[1], MBBS, MPH, FCPS; Niloufer Taber[1], MPH, MPA; Abdulgafoor M Bachani[1], PhD, MHS; Shivam Gupta[1], MD, PhD, MPH; Nino Paichadze[2], MD, MPH; Adnan A Hyder[2], MD, MPH, PhD

[1]International Injury Research Unit, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Milken Institute School of Public Health, The George Washington University, Washington, DC, United States

Corresponding Author:
Amber Mehmood, MBBS, MPH, FCPS
International Injury Research Unit
Department of International Health
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
Suite E-8644
Baltimore, MD, 21205
United States
Phone: 1 443 287 2763
Email: amehmoo2@jhu.edu

Abstract

Background: Rapid advances in mobile technologies and applications and the continued growth in digital network coverage have the potential to transform data collection in low- and middle-income countries. A common perception is that digital data collection (DDC) is faster and quickly adaptable.

Objective: The objective of this study was to test whether DDC is faster and more adaptable in a roadside environment. We conducted a reliability study comparing digital versus paper data collection in 3 cities in Ghana, Vietnam, and Indonesia observing road safety risk factors in real time.

Methods: Roadside observation of helmet use among motorcycle passengers, seat belt use among 4-wheeler passengers, and speeding was conducted in Accra, Ghana; Ho Chi Minh City (HCMC), Vietnam; and Bandung, Indonesia. Two independent data collection teams were deployed to the same sites on the same dates and times, one using a paper-based data collection tool and the other using a digital tool. All research assistants were trained on paper-based data collection and DDC. A head-to-head analysis was conducted to compare the volume of observations, as well as the prevalence of each risk factor. Correlations (*r*) for continuous variables and kappa for categorical variables are reported with their level of statistical significance.

Results: In Accra, there were 119 observation periods (90-min each) identical by date, time, and location during the helmet and seat belt use risk factor data collection and 118 identical periods observing speeding prevalence. In Bandung, there were 150 observation periods common to digital and paper data collection methods, whereas in HCMC, there were 77 matching observation periods for helmet use, 82 for seat belt use, and 84 for speeding. Data collectors using paper tools were more productive than their DDC counterparts during the study. The highest mean volume per session was recorded for speeding, with Bandung recording over 1000 vehicles on paper (paper: mean 1092 [SD 435]; digital: mean 807 [SD 261]); whereas the lowest volume per session was from HCMC for seat belts (paper: mean 52 [SD 28]; digital: mean 62 [SD 30]). Accra and Bandung showed good-to-high correlation for all 3 risk factors (*r*=0.52 to 0.96), with higher reliability in speeding and helmet use over seat belt use; HCMC showed high reliability for speeding (*r*=0.99) but lower reliability for helmet and seat belt use (*r*=0.08 to 0.32). The reported prevalence of risk factors was comparable in all cities regardless of the data collection method.

Conclusions: DDC was convenient and reliable during roadside observational data collection. There was some site-related variability in implementing DDC methods, and generally the productivity was higher using the more familiar paper-based method. Even with low correlations between digital and paper data collection methods, the overall reported population prevalence was similar for all risk factors.

(J Med Internet Res 2019;21(5):e13222) doi:10.2196/13222

KEYWORDS

information technology; public health informatics; mHealth; risk factors; population surveillance

Introduction

Background

Road traffic injuries (RTIs) are the ninth leading cause of death worldwide, primarily affecting the young and productive age groups. In addition to causing over 1.35 million deaths each year, road traffic crashes are responsible for 50 million nonfatal injuries [1]. Although the rapid economic growth seen in many low- and middle-income countries (LMICs) has raised living standards and led to a reduction in many diseases of poverty, increased motorization without an increase in traffic enforcement or improvements in road environments has led to a rise in motor vehicle crashes [2-5]. Consequently, LMICs share a larger proportion of deaths and disability, and it is estimated that 966 to 1160 per 100,000 population disability-adjusted life years are lost because of RTIs [6].

To tackle this international problem, the Global Plan for the Decade of Action for Road Safety 2011-2020 recommended a range of road safety measures, including improvement in road user behaviors as an important pillar [7]. These recommendations focus on the development and implementation of comprehensive programs and strategies to positively affect seat belt and child restraint use, correct helmet wearing, speeding, and drunk driving. The Bloomberg Initiative for Global Road Safety (BIGRS), a consortium of international partners funded by Bloomberg Philanthropies, promotes the adoption of internationally recognized best practices to improve these 4 risk factors in 10 selected cities from across 9 LMICs [8-10]. As part of the project, the Johns Hopkins International Injury Research Unit conducts semiannual observational studies with local partners to measure the prevalence of these 4 RTI risk factors over time: helmet use, seat belt use, speeding, and drunk driving. These data provide evidence to inform program development, as well as monitoring and evaluation of interventions in the project.

Roadside Observational Studies for Risk Factor Monitoring

Paper-based data collection has been the standard method for primary roadside observational studies, and until recent years, the only method [11,12]. Paper-based data collection methods have certain advantages that make them easily adaptable. Paper-based tools are more flexible, immediately deployable, and do not require any specialized training other than the content of questionnaire. Writing on paper forms is easier especially for short and concise questionnaires. However, as other researchers have observed, legibility related errors are frequent, storage costs can be prohibitive, deployment and tracking of surveys are challenging, and double data entry is expensive and time-consuming [13,14]. Sometimes, when using paper data collection methods, other forms of data, including photographs or Global Positioning System (GPS) coordinates, require separate equipment and careful tracking to link external data to the correct observation. Moreover, in administered questionnaires, there may be variability among surveyors in adhering to complex eligibility requirements or logical dependencies (skip patterns) across questions [15].

Recently, to facilitate real-time data collection in a roadside environment, mobile health (mHealth) tools were developed for population-level observational studies on 3 road safety risk factors, speeding, helmet use, and seat belt use, as a part of the BIGRS project. Electronic (digital) methods of data collection have merged the process of data collection and data entry, potentially saving costs and time [16]. The diversity of mHealth apps has generated immense interest among researchers to test innovative ideas, especially where accuracy and standardized data collection is required. Transitioning from traditional paper-based methodology into digital data collection (DDC) supports rapid aggregation and analysis of a large amount of data by avoiding the costs and time lag of data entry [15]. In addition, DDC also facilitates remote monitoring of the data collection process and can improve data quality by introducing standardized responses, skip patterns, logic checks, and automatic calculations [17-19]. However, these advantages of DDC must be compared with the productivity and reliability of widely accepted and established paper-based data collection method.

Rationale and Study Objectives

Although DDC is rapidly replacing traditional paper-based methods, the majority of studies or commentary about DDC method do not use multiple methods to establish its reliability by independent data collectors on the same targets and are usually done in the context of household surveys rather than observational studies [14,20,21]. Owing to lack of comparative studies, there is little evidence about the productivity, reliability, and efficiency of large volume data collection using mobile devices in a highly dynamic roadside environment [20]. The term *productivity* refers to sheer quantity, and in the context of observational studies, the amount of data collected per session, real time in the field, distinguishing it from *efficiency*, which is used in the context of quality of data that might include creating output in less time, using fewer resources, or spending less money. In the context of project management, efficient output may have different dimensions, including human and material cost, effort, turnaround time, etc. Reliability of a new tool or method indicates its ability to produce same or consistent results when compared with a reference or standard method. This study aimed to assess productivity and reliability of DDC by comparing simultaneously collected paper and digital data in terms of volume of observations, overall measurement of prevalence of road safety risk factors, and interobserver agreement about a busy roadside environment. The paper then discusses the impact of study findings in guiding the choice between paper and digital methods of data collection in different contexts.

Methods

Setting

The study was conducted in 3 different cities where roadside observational studies were taking place: Accra, Ghana; Bandung, Indonesia; and Ho Chi Minh City (HCMC), Vietnam. These cities were selected out of the 10 participating in the BIGRS project, based on the willingness of the local partners to switch from paper to digital format. Simultaneous paper-based data collection and DDC were conducted in Accra during March 2017, in HCMC during April 2017, and in Bandung during August 2017, as part of a routine semiannual data collection schedule.

The BIGRS Project team used KoBoToolbox data collection software and its KoBoCollect Android smartphone application, which was developed by the Harvard Humanitarian Initiative as an open source suite of tools for data collection and analysis [22]. The digitization process included programming digital forms to be downloaded on mobile app, using Android tablets for data collection and uploading information to the secure cloud server [23]. DDC forms for each risk factor were based on the same tools used in paper-based data collection method. To maintain standardization and quality, all the digital forms were the same, though the forms were available in English for Accra and were bilingual for HCMC (Vietnamese and English) and Bandung (Indonesian and English).

In all 3 cities, local data collectors were hired and trained over 2 days to familiarize them with the study protocols, as well as the Android environment and the KoBoCollect app, including mobile data entry process, saving digital forms, and uploading data to the server. All data collectors were trained on both digital and paper data collection methods. This training session was followed by a mandatory hands-on practice for both digital and paper data collectors. Similarly, supervisors and data managers were trained to manage field site data collection, monitor data upload and server activity, and download data from the server. Any issues with data entry, saving, and uploading were identified and resolved during training and practice, before the roadside observational studies.

Data Collection Protocol

Within each city, locations for observation were selected using stratified randomization to ensure that all major road types and city administrative divisions were represented. Standardized observation methods were employed across all observation sites. At each location, observations were done by 2 independently working teams on the same date and times with one team using paper-based forms and the other used digital forms. Each team consisted of 1 observer who viewed vehicles and conveyed the information to the data recorder who marked the presence of risk factors and demographic data for each vehicle as applicable. Data collectors were randomly rotated among teams and between digital and paper data collection methods throughout the study to avoid individual data collector competence influencing the productivity or reliability of data collection method. Each data collector's schedule was randomly varied by date, time, location, data collection partner, and data collection method (paper or digital). Observations at each location were done during both weekdays and weekends and both rush hours and off-peak hours.

For speeding assessments, the sites were carefully chosen to avoid junctions or intersections, or areas where vehicles were slowing down because of construction or road blocks, as well as entrances to parking lots, gas stations, malls, or shopping centers. For observations on seat belts and child restraint use, junctions, intersections, or entrances of gas stations and rest areas where vehicles travel at reduced speeds were selected to facilitate close observations and ensure accuracy. The protocol required that only vehicles traveling in 1 direction were observed. Starting the observation with the vehicle closest to the curb or roadside also allowed data collectors to observe as many vehicles as possible with accuracy in a high-volume traffic flow.

At the beginning of the session at each site, the team filled a site description form using their respective method (digital or paper). These forms captured data about the road and traffic environment at each time, date, and location, including the traffic volume during a 15-min period, the weather, and the presence and nature of law enforcement, including the presence and placement of police and/or cameras for enforcement. Speeding observations also captured the posted speed limit and the existence of various environmental traffic calming measures, such as speed bumps. Data captured about the vehicle included the vehicle type (sedan, sport utility vehicle, truck, etc), vehicle ownership (private, commercial, government, etc), and for speeding observations, the actual speed of the vehicle in kilometer per hour (km/h). Finally, information on the vehicle's occupants was captured during assessments of seat belt and helmet use, including each occupant's gender, estimated age group, position within the vehicle, and use of safety equipment.

Consistent procedures and definitions were maintained between methods and across observation sites to ensure comparability of results. The metric for comparisons was volume of observations and prevalence of risk factor per 90-min session. We did not use an a priori number of observations and track the time to accomplish them. The helmet use risk factor was defined as wearing a strapped, standard helmet (not a cap helmet). Seat belt use was defined as wearing a buckled seat belt, or using a proper child restraint, on a single vehicle occupant. Speeding was defined as any speed in excess of the posted speed limit in kilometer per hour and levels of overspeed categorized in 5 or 10 km/h intervals.

Statistical Analysis

Observation periods in digital and paper formats were matched to each other by date, time, and location; digital observation periods without corresponding paper observation periods were not included in the analysis, and vice versa. The few sessions that did not match were because of issues with logistics, staffing, or equipment.

A head-to-head comparison of digital and paper data collection was conducted to assess the productivity during each session matched by date, time, and location. After pooling all sessions, the mean number of observations were compared between paper and digital methods. In addition, 2 sample tests of proportions were used to compare overall helmet use and seat belt use prevalence, pooling across all observation sessions. Furthermore, chi-square tests of independence were used to evaluate whether the numbers of vehicles in different categories of overspeeding varied between digital and paper data collection methods, again pooling across all observation sessions [24]. This was done to assess whether the larger picture of traffic safety was the same between digital and paper data collection methods, despite any differences between individual digital and paper observation sessions.

The reliability of a method could be measured in terms of inter- or intraobserver variations for which *r*, interclass correlations, and kappa are appropriate statistics. In this study interrater reliability between digital and paper data collection methods were assessed using Pearson correlations for continuous measures and kappa for roadside environment data recorded as categorical variables, with the observation session used as the rating object [25,26]. Pearson correlation is appropriate for proportions when the majority of proportions are not close to 0 or 1. Although Spearman rank correlation is often used for proportions, we were interested in exact values rather than rank ordering and therefore Pearson correlation was more appropriate.

The risk factors were not pooled across the cities to better appreciate the contextual differences in the productivity and reliability of city teams.

Templates in Microsoft Excel were used for data entry from paper formats and all statistical analyses were conducted in STATA SE version 15.1 software package [27]. Ethical approval was obtained from the Institutional Review Board of Johns Hopkins Bloomberg School of Public Health, United States.

Results

Productivity in Digital and Paper Data Collection Methods

In Accra, there were 119 helmet use observation sessions matched exactly by date, time, and location between digital and paper data collection methods; 119 matched seat belt use observation sessions and 118 matched sessions observing speeding. In Bandung, each risk factor had 150 matched observation sessions. In HCMC, the numbers of matched sessions were somewhat lower, with 77 matched helmet use observation sessions, 82 matched seat belt use sessions, and 84 sessions matched for speeding observations. This lower number was because of both the fewer number of sessions conducted in digital and paper data collection methods and mismatches by date, time, or location between digital and paper observation sessions. As the number of sessions, and hence the total sample size for each risk factor varied by city, the mean volume of observations per session was used for comparison.

In addition to the number of sessions, the number of observations made per session varied among the 3 cities and across risk factors, with Bandung generally having higher productivity per session, followed by Accra, and last by HCMC (Table 1). The number of observations made per session was lower among research assistants conducting DDC as compared with paper data collection, with correlations between digital and paper for the same date, time, and location ranging from 0.23 to 0.95 across cities and risk factors (Table 1).

To assess the impact these differences between digital and paper data collections in productivity and overall sample size may have had on the precision of our estimates, we calculated the level of precision for the current digital data sample size (Table 2). We also calculated the sample sizes needed to estimate proportions to achieve a CI half-width of 0.01 and 0.005, based on the DDC proportion. With 1 exception, all digital and paper sample sizes were able to provide estimates within 1 percentage point, and in 6 out of 9 cases, the sample size was large enough to estimate proportions to within a half percentage point, thus eliminating any risk of sample size affecting the overall prevalence of risk factors. We did not find any risk factor in the cities where the paper sample size was able to estimate precision to one or one-half percentage point, but not the DDC.

Table 1. Volume of observations: reliability between digital and paper observations.

Risk factor and city	Digital observation, mean (SD)	Paper observation, mean (SD)	Correlation value (*r*)	*P* value
Helmet				
Accra	181.81 (84.87)	196.86 (91.53)	0.95	<.001
Bandung	353.97 (104.03)	509.81 (151.97)	0.56	<.001
Ho Chi Minh City	210.45 (86.29)	249.17 (107.32)	0.23	.04
Seat belt				
Accra	200.47 (71.20)	258.31 (96.73)	0.73	<.001
Bandung	199.78 (71.35)	245.23 (88.06)	0.52	<.001
Ho Chi Minh City	62.40 (30.09)	51.56 (28.45)	0.32	.003
Speeding				
Accra	305.32 (90.42)	331.65 (94.83)	0.84	<.001
Bandung	807.49 (261.23)	1092.08 (435.45)	0.78	<.001
Ho Chi Minh City	228.27 (109.80)	225.24 (131.98)	0.77	<.001

Table 2. Level of precision and sample size requirements.

Risk factor and city	Digital observation: existing sample size	Paper observation: existing sample size	Digital observation: current level of precision	Sample size required for estimation within 1 percentage point (0.01)	Sample size required for estimation within 0.5 percentage point (0.005)
Helmet					
Accra	28,719	30,983	0.005	8655	34,618
Bandung	71,846	101,197	0.003	8378	33,512
Ho Chi Minh City	20,842	26,419	0.006	7524	30,094
Seat belt					
Accra	55,983	58,024	0.004	9553	38,211
Bandung	50,391	57,758	0.004	8573	34,292
Ho Chi Minh City	8205	5981	0.011	9567	38,268
Speeding					
Accra	36,028	39,135	0.004	7132	28,526
Bandung	121,123	163,812	0.002	4572	18,285
Ho Chi Minh City	19,175	18,920	0.002	613	2452

Interobserver Agreement between Digital and Paper Data Collection Methods

There were some discrepancies between how digital and paper research assistants recorded the presence of police and camera enforcement at each site. The values of calculated kappa statistics ranged from just under 0.51 up to 1.00, indicating moderate to perfect agreement. In Accra, all speeding observation sessions recorded that there was no police presence and no camera enforcement; although there was 100% agreement between digital and paper data collection methods, without any variation, kappa is undefined (Table 3).

Table 3. Interrater agreement between digital and paper data collection methods on law enforcement and environmental deterrents.

Risk factor and city	Kappa value	*P* value
Helmet[a]		
Accra	0.51	<.001
Bandung	0.72	<.001
Ho Chi Minh City	0.50	<.001
Seat belt[a]		
Accra	0.93	<.001
Bandung	0.92	<.001
Ho Chi Minh City	0.59	<.001
Speeding[a]		
Accra	N/A[b]	N/A[b]
Bandung	0.68	<.001
Ho Chi Minh City	1.00	<.001
Speeding[c]		
Accra	0.94	<.001
Bandung	0.78	<.001
Ho Chi Minh City	0.95	<.001

[a] Comparing police presence, camera enforcement, or both.

[b] N/A: not applicable; no sites were observed to have police or camera enforcement through digital or paper data collection; although there was 100% agreement between digital and paper data collection, without any variation, kappa is undefined.

[c] Comparing environmental speed deterrents, including speed bumps, cross walks, and stop signs.

Prevalence of Risk Factors and Reliability between Digital and Paper Data Collection Methods

The prevalence of each of the 3 risk factors were assessed with moderate to high levels of reliability between digital and paper data collection methods. Accra showed the highest levels of reliability overall. For example, on average 63% to 66% of motorcycle occupants per session were observed to wear helmets in both digital and paper data collection methods, for a correlation of r=0.94 (P<.001; Table 4). Reliability between digital and paper data collection methods was similar for seat belt use and speeding; the correlation for seat belt use was 0.76 (P<.001) and 0.97 (P<.001) in speeding. Bandung also had similar levels of correlation between digital and paper data collection per observation session, with excellent reliability in helmet use and speeding observations (r=0.89, P<.001 and r=0.95, P<.001, respectively), and very good correlation between data collection methods in seat belt use observations (r=0.70, P<.001). HCMC had lower and nonsignificant correlations between digital and paper data collection methods in helmet and seat belt use observations (r=0.11, P=.36 and r=0.08,P=.46, respectively). Interestingly, speeding correlation in HCMC was almost perfect (r=0.999, P<.001).

For helmet use and seat belt use risk factor assessments, reliability was also assessed within subgroups of road users, by gender, estimated age group, and occupant's position within the vehicle. When further breaking down helmet and seat belt use by occupant's role, driver or passenger, the correlations followed similar patterns as the overall figures, with Accra and Bandung showing very high levels of correlation, followed by HCMC (Figures 1 and 2). Across all 3 cities, observations made on drivers showed higher levels of reliability than observations made on passengers.

In each city, the largest proportion of observed motorcycle occupants were males over the age of 18 years, with motorcycle occupants being almost exclusively adult males in Accra (98% of all occupants in both digital and paper data collection methods; Figure 3). Assessments of the seat belt use risk factor considered finer divisions of estimated age, as younger and older children should use different child restraints, rather than seat belts alone (Tables 5 and 6). Furthermore, the age groups and genders were estimated on best guess by data collectors, rather than exact ages and genders. However, as the gender of children is often difficult to assess, we have pooled the genders in the under 5 years and 6 to 11 years age categories.

Table 4. Prevalence of risk factors: overall proportions and 2 sample tests of proportions.

Risk factor and city	Digital proportion (SD)	Paper proportion (SD)	*P* value
Helmet			
Accra	0.66 (0.47)	0.63 (0.48)	<.001
Bandung	0.68 (0.47)	0.70 (0.46)	<.001
Ho Chi Minh City	0.73 (0.44)	0.73 (0.45)	.18
Seat belt			
Accra	0.46 (0.50)	0.44 (0.50)	<.001
Bandung	0.63 (0.48)	0.66 (0.47)	<.001
Ho Chi Minh City	0.53 (0.50)	0.59 (0.49)	<.001
Speeding			
Accra	0.75 (0.43)	0.77 (0.42)	<.001
Bandung	0.14 (0.34)	0.12 (0.33)	<.001
Ho Chi Minh City	0.02 (0.12)	0.02 (0.12)	.65

Figure 1. Prevalence of correct helmet use by occupant role: reliability between digital and paper observations.

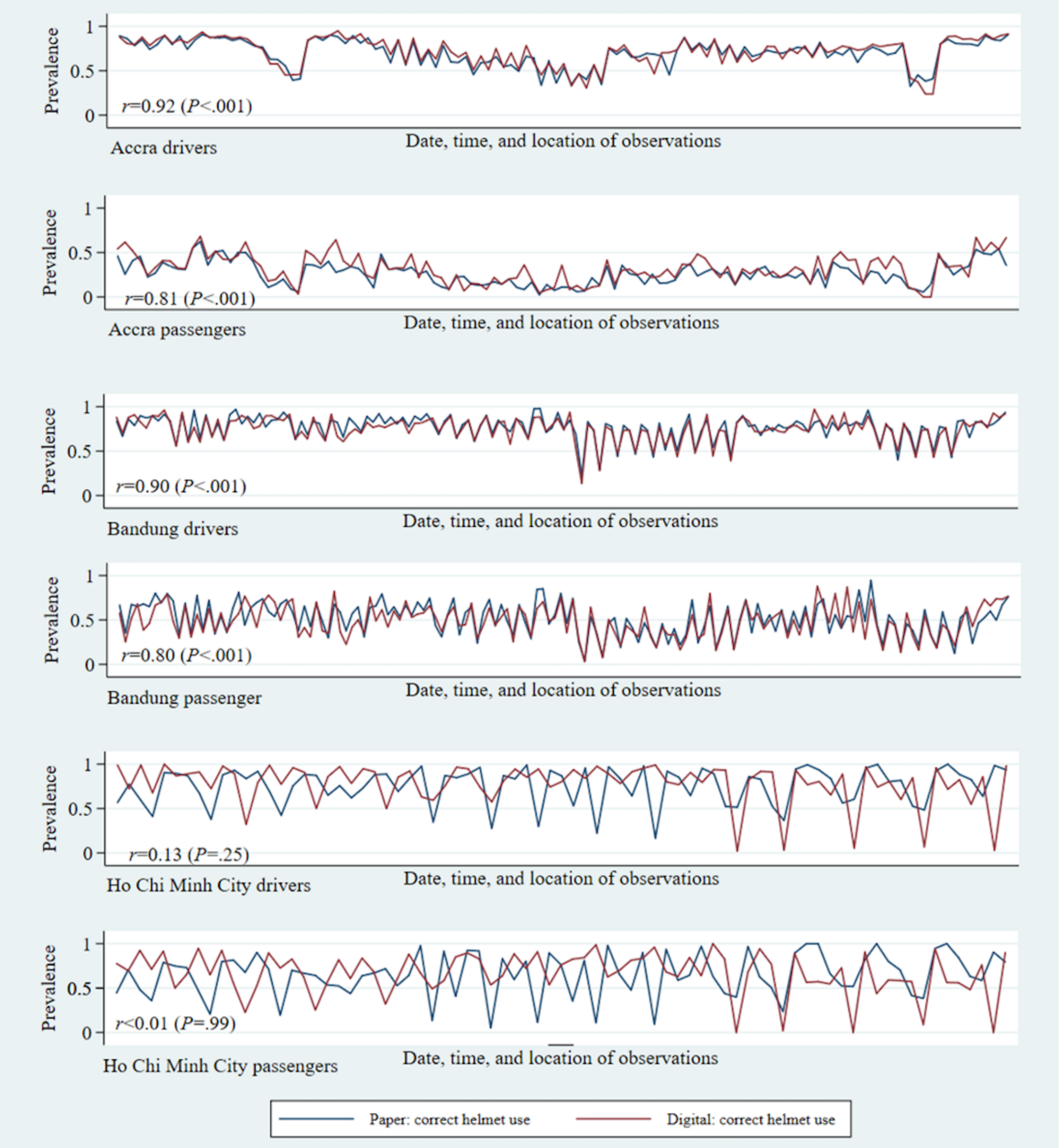


Figure 2. Prevalence of seat belt use by occupant role: reliability between digital and paper observations.

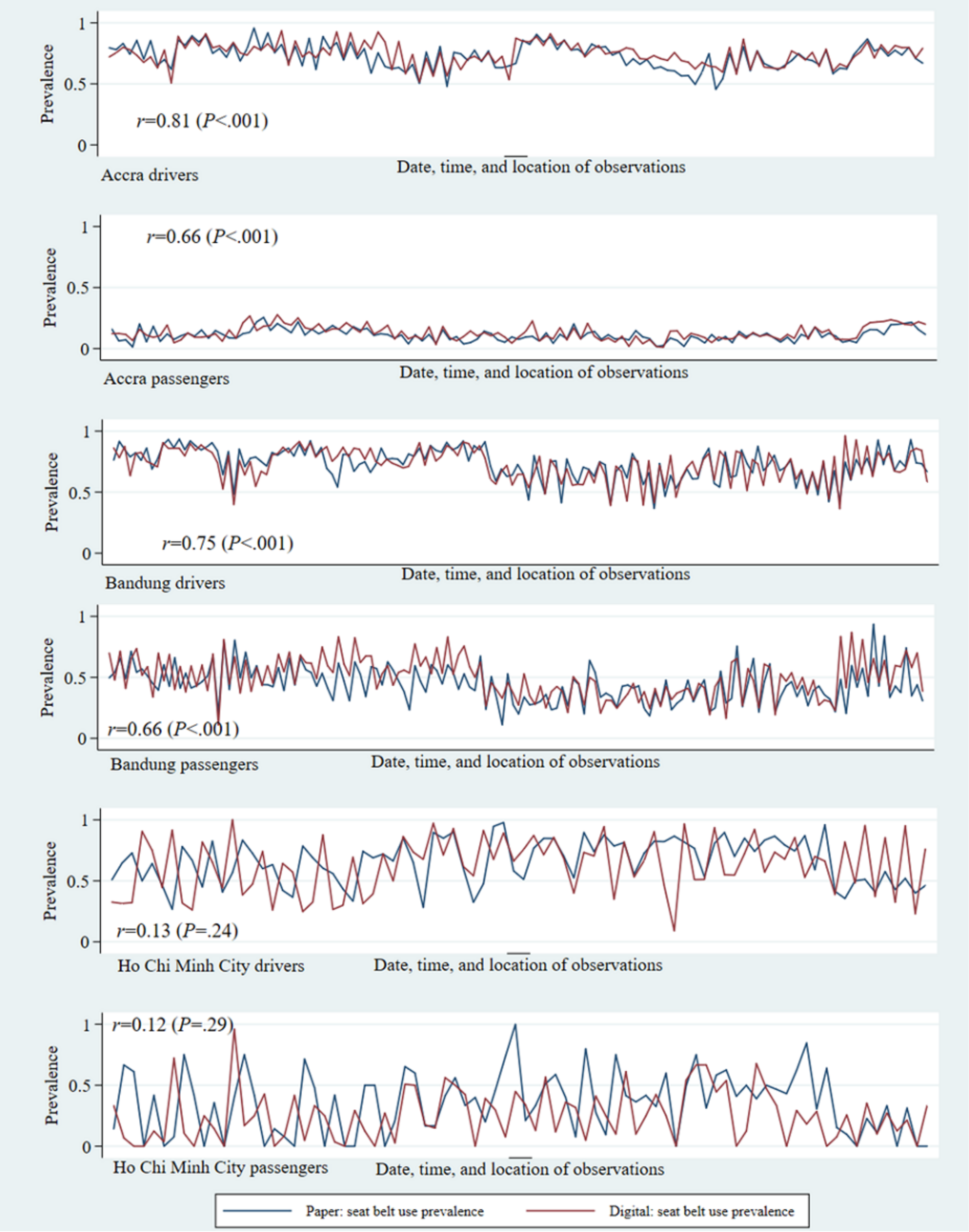


Figure 3. Motorcycle occupants observed by age-sex group.

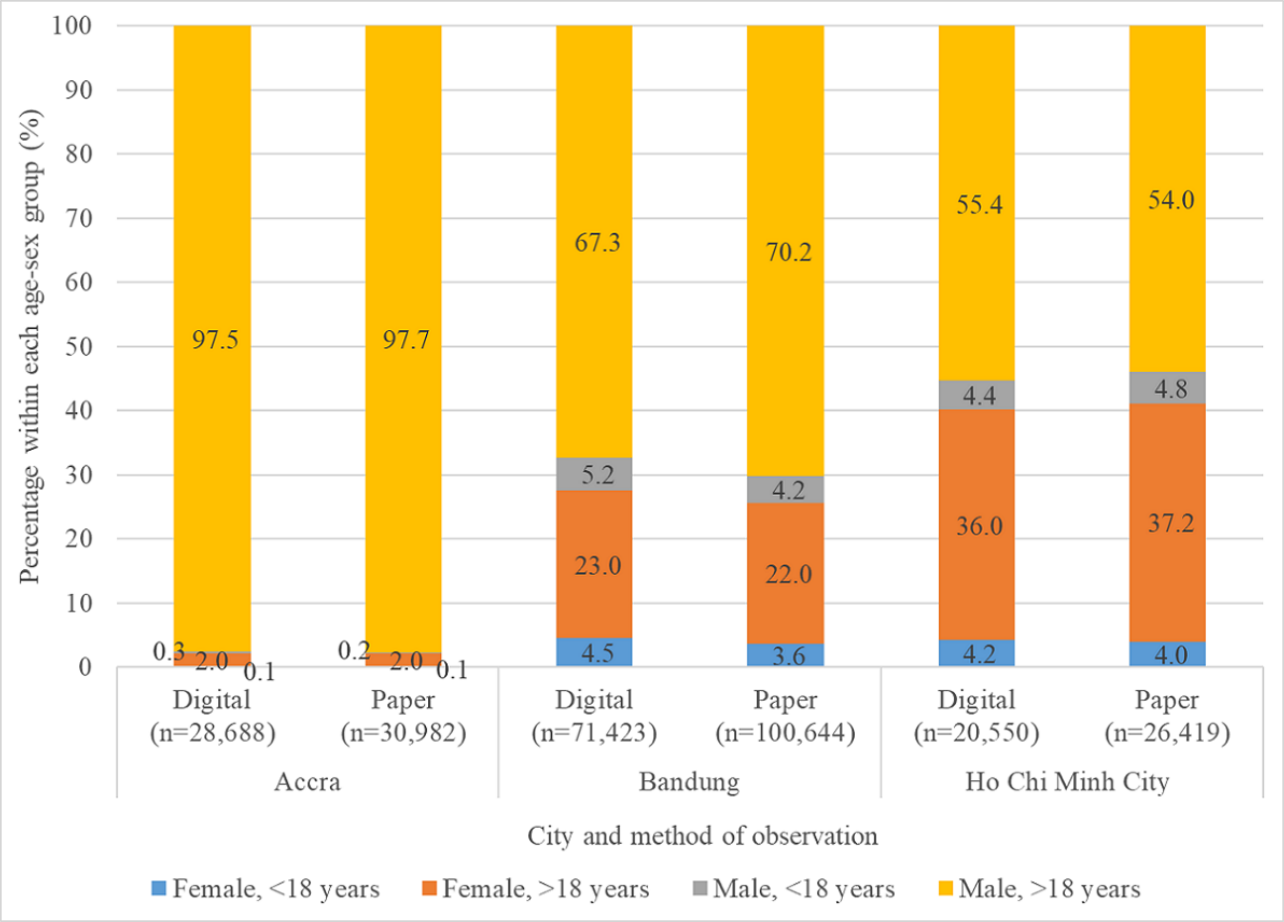


Table 5. Prevalence of helmet use by age-sex groups: reliability between digital and paper observations.

Risk factor and city	Digital observation proportion	Paper observation proportion	Correlation value (*r*)	*P* value
Female, <18 years				
Accra	0.00	0.13	N/A[a]	N/A
Bandung	0.21	0.20	0.35	<.001
Ho Chi Minh City	0.44	0.50	−0.14	.23
Female, >18 years				
Accra	0.26	0.25	0.33	<.001
Bandung	0.61	0.63	0.79	<.001
Ho Chi Minh City	0.70	0.72	0.17	.13
Male, <18 years				
Accra	0.22	0.10	0.27	.14
Bandung	0.35	0.32	0.37	<.001
Ho Chi Minh City	0.51	0.48	0.15	.20
Male, >18 years				
Accra	0.67	0.64	0.94	<.001
Bandung	0.76	0.78	0.87	<.001
Ho Chi Minh City	0.80	0.77	0.09	.42

[a] N/A: not applicable.

Table 6. Prevalence of seat belt use by age-sex groups: reliability between digital and paper observations.

Risk factor and city	Digital observation proportion	Paper observation proportion	Correlation value (*r*)	*P* value
Both genders, <5 years				
Accra	0.13	0.14	0.63	<.001
Bandung	0.07	0.07	−0.004	.97
Ho Chi Minh City	0.00	0.00	N/A[a]	N/A
Both genders, 5-11 years				
Accra	0.11	0.05	0.23	.23
Bandung	0.17	0.10	0.22	.02
Ho Chi Minh City	0.10	0.00	N/A	N/A
Female, 12-17 years				
Accra	0.10	0.17	−0.19	.33
Bandung	0.33	0.30	0.16	.23
Ho Chi Minh City	0.27	0.05	−0.04	.84
Female, 18-24 years				
Accra	0.13	0.11	0.13	.40
Bandung	0.64	0.51	0.06	.47
Ho Chi Minh City	0.16	0.10	0.29	.01
Female, 25-59 years				
Accra	0.30	0.26	0.70	<.001
Bandung	0.68	0.61	0.58	<.001
Ho Chi Minh City	0.23	0.21	0.21	.047
Female, >60 years				
Accra	0.28	0.21	0.09	.57
Bandung	0.57	0.54	0.16	.17
Ho Chi Minh City	0.00	0.00	N/A	N/A
Male, 12-17 years				
Accra	0.10	0.10	0.29	.09
Bandung	0.49	0.38	0.18	.11
Ho Chi Minh City	0.32	0.16	0.01	.96
Male, 18-24 years				
Accra	0.18	0.16	0.33	.01
Bandung	0.67	0.60	0.22	.01
Ho Chi Minh City	0.22	0.22	0.23	.05
Male, 25-59 years				
Accra	0.54	0.52	0.61	<.001
Bandung	0.67	0.67	0.70	<.001
Ho Chi Minh City	0.49	0.52	0.17	.10
Male, >60 years				
Accra	0.47	0.39	0.30	.01
Bandung	0.69	0.61	0.18	.06
Ho Chi Minh City	0.27	0.12	0.18	.57

[a] N/A: not applicable.

Among occupants of 4-wheeled vehicles, between two-thirds to three-fourths were adult males aged between 25 and 59 years (Figure 4). Generally, age-sex groups with low representation in the datasets, especially children, had lower reliability between digital and paper data collection methods, as did more narrowly defined age groups. For example, the proportion of 4-wheeler occupants that were females aged between 25 and 59 years was approximately 19% in Accra, 18% in Bandung, and 8% to 9% in HCMC; in those same cities, females aged between 12 and 17 years comprised 1% or less of 4-wheeler occupants, across digital and paper data collection methods (Figure 4). The correlation for seat belt use among females aged between 25 and 59 years was 0.73 in Accra, 0.58 in Bandung, and 0.04 in HCMC; the correlation for females aged between 12 and 17 years was 0.16 in Bandung, and was negative in Accra and HCMC.

Pooling across all observation sessions of helmets and seat belts that could be matched between the 2 modalities, the digital and paper data collection methods, resulted in very similar risk factor prevalence, within only a few percentage points of each other except that of HCMC (Figures 5 and 6). Speeding observations also demonstrated overall consistency between the 2 formats, with HCMC demonstrating very high correlation between digital and paper data collection (Figure 7). In Bandung and Accra, the Pearson chi-square tests of independence showed statistically significant difference in the proportion of vehicles in different categories of overspeeding, although the actual percentages fall between 1 to 2 points from each other. Across cities, the prevalence of helmet use, seat belt use, and speeding and category of overspeeding was largely similar between digital and paper data collection methods (Table 4).

Figure 4. Four-wheeler occupants observed by age-sex group.

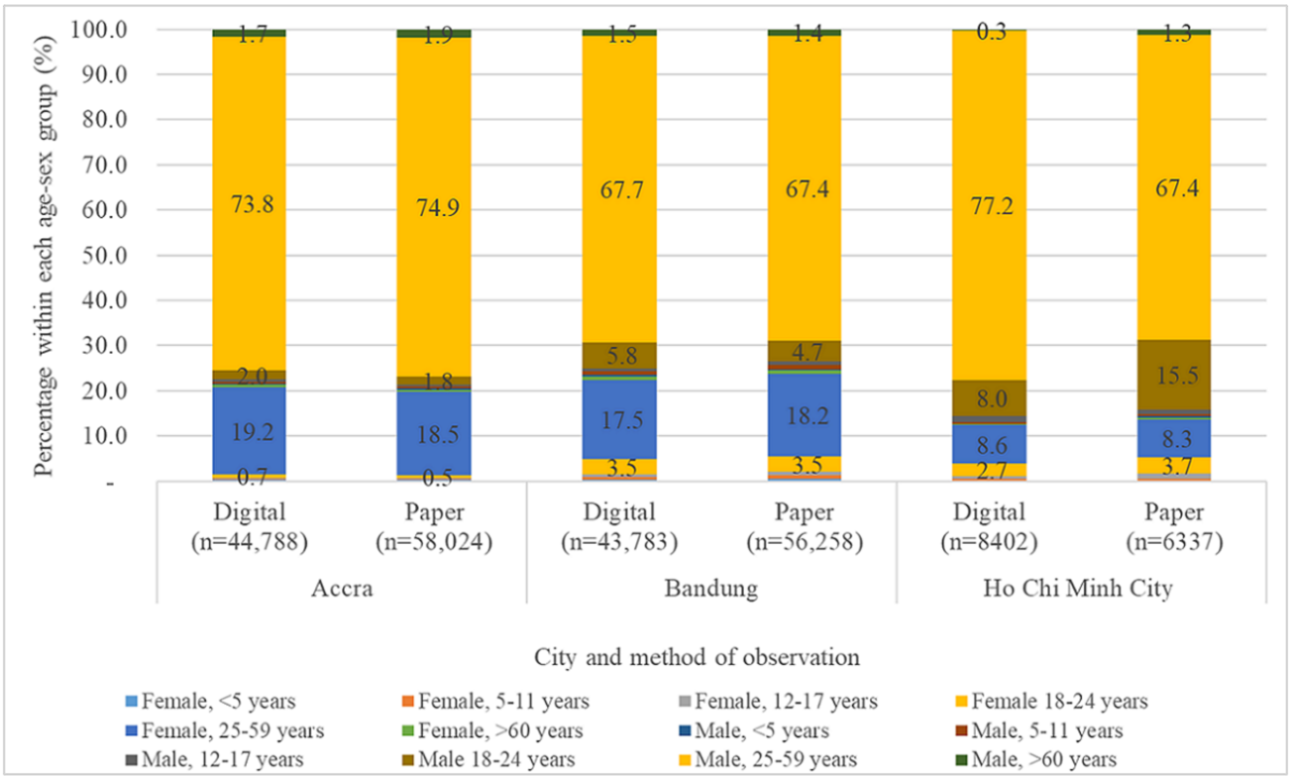


Figure 5. Prevalence of correct helmet use: reliability between digital and paper observations.

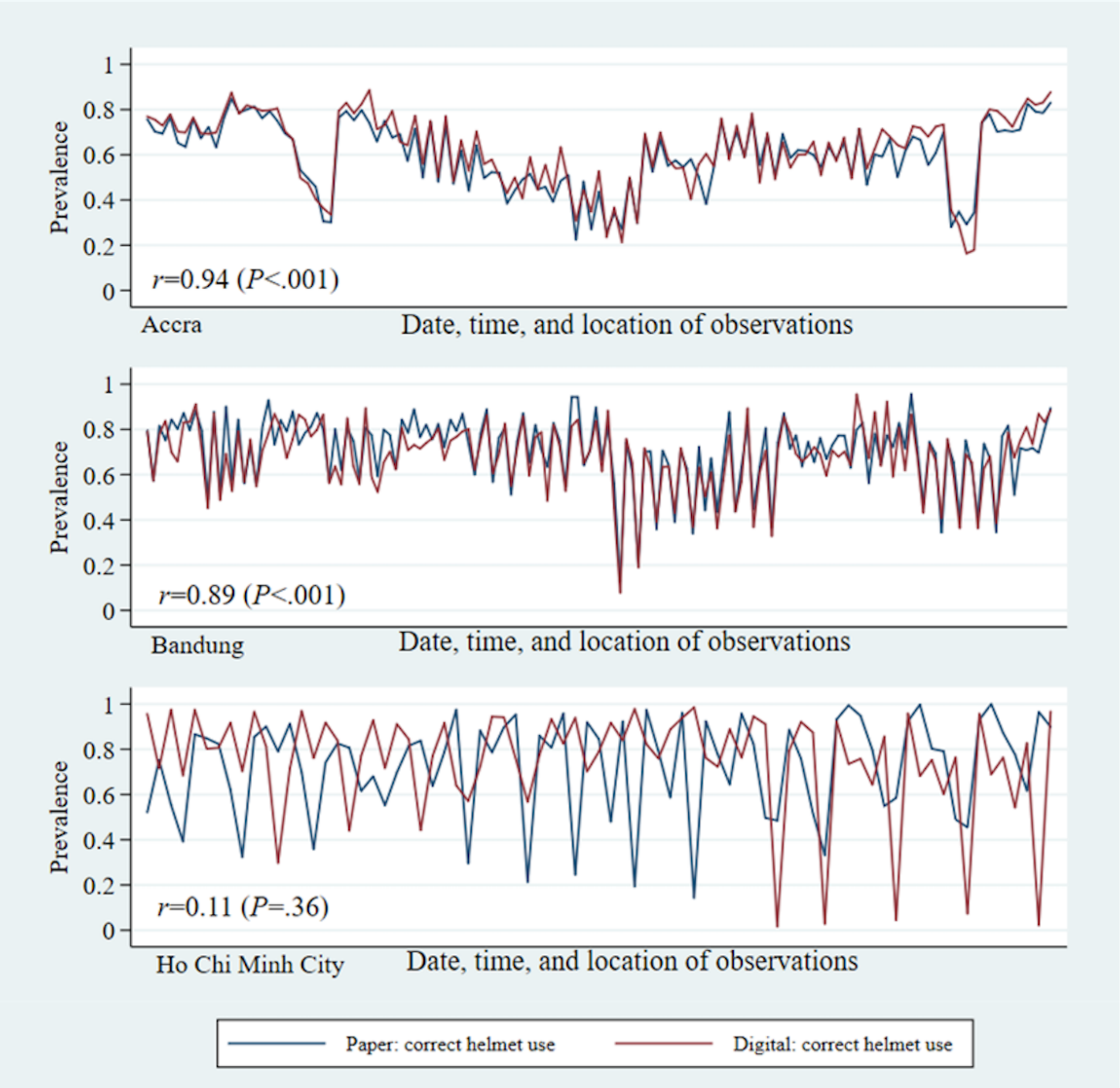


Figure 6. Prevalence of seat belt use: reliability between digital and paper observations.

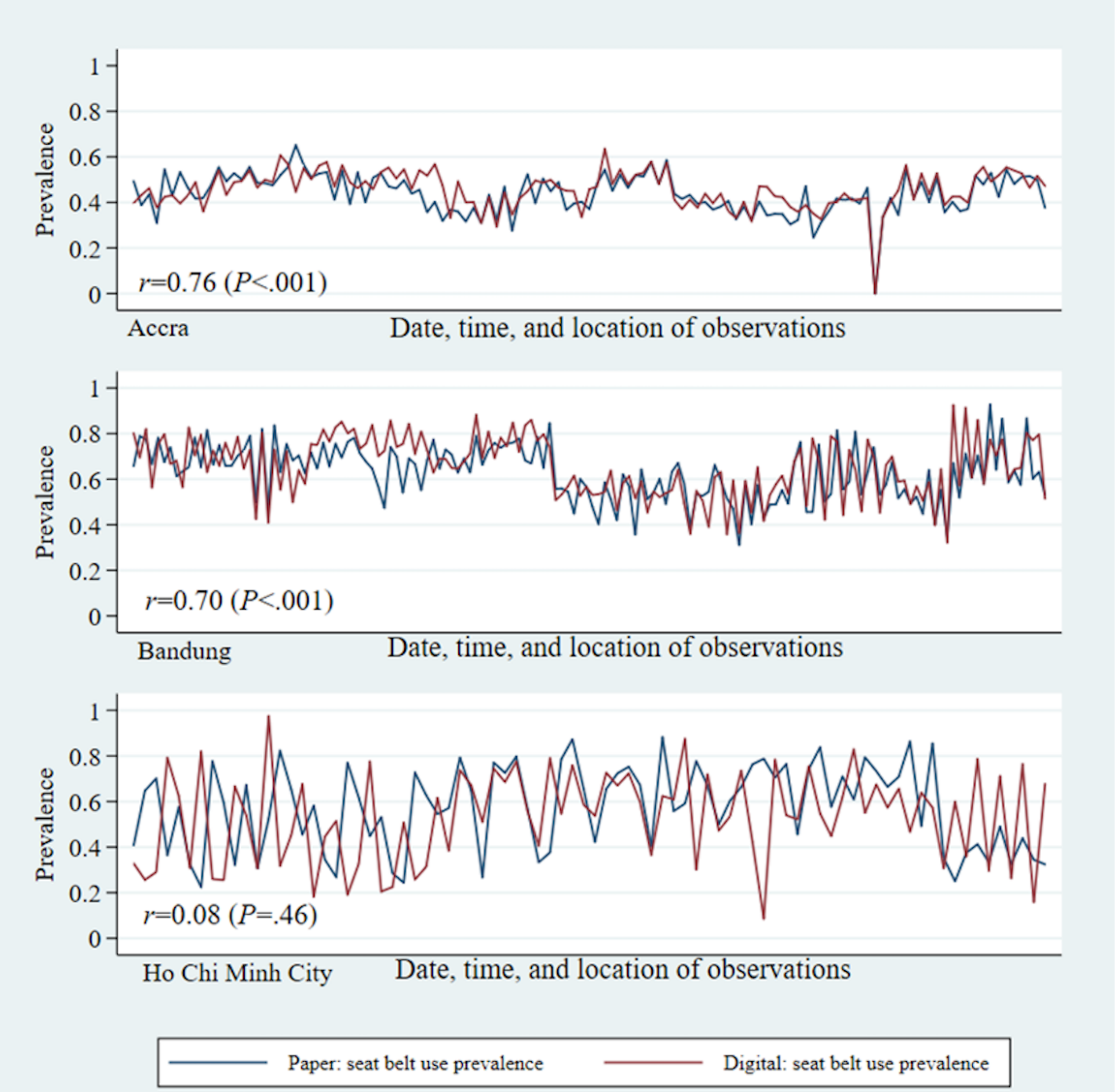
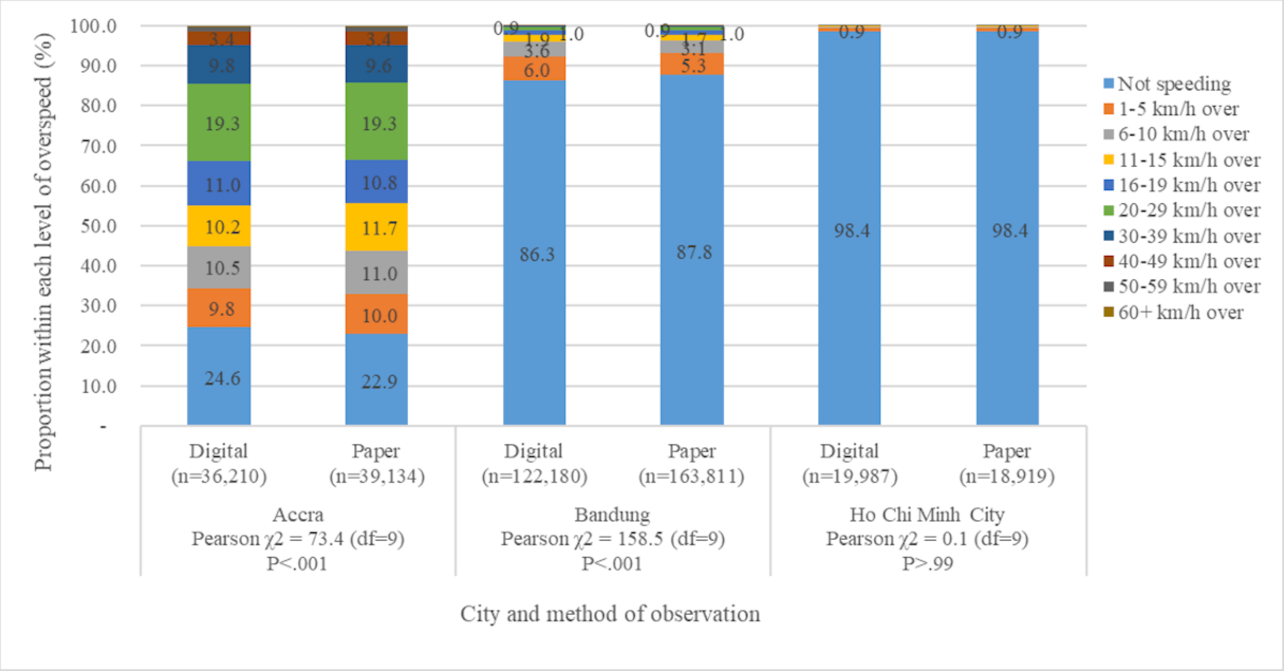


Figure 7. Prevalence of speeding by level of overspeeding and city: overall proportions and chi-square test of independence.



Discussion

Principal Findings

To our knowledge, this is the first study to evaluate the productivity and reliability of digital versus paper data collection in a roadside environment. This study also provides an illustration of different contexts and results from Accra, HCMC, and Bandung, using a standard methodology in all the sites. Each city had different number of observation sessions per risk factor, varying between 84 and 159 sessions depending on the risk factor and method. Considering only those digital and paper observation sessions matched by date, time, and location, the mean volumes of observations were statistically significantly higher for paper-based data than digitally collected data for all 3 risk factors in Accra and Bandung, and for helmet use in HCMC. However, larger sample size in paper data collection was not associated with meaningfully higher level of precision in forming prevalence estimates and same level of precision was achieved with relatively smaller sample size in DDC.

The differences in the number of observations per 90-min session between paper and digital methods varied by city, indicating that familiarity and dexterity varied by context. The data collection team in Accra showed the highest correlation and productivity across all risk factors. Bandung demonstrated high productivity with both data collection methods, with good reliability. In HCMC, the productivity of the DDC teams exceeded that of the paper data collection teams during the seat belt and speeding risk factor observational studies; however, the results showed moderate to low reliability in the prevalence of risk factors between the 2 methods. On inquiry, the reason behind this finding was the inadequate number of field staff, which led to the deviation from standard protocol. Although the protocol called for an observation team comprising separate observer and recorder, in HCMC the same person was observing and recording information. This resulted in overall low volumes and possibly negatively affected reliability.

There may be several reasons why the seat belt risk factor assessment showed lower reliability across the 3 cities, as compared with other risk factors. Although the reliability of correct helmet use was high in both Accra and Bandung, moderate reliability of seat belt observations in those cities might be offset by inability to capture accurate data on all the vehicle occupants. The seat belt use observations (in contrast to speeding observations) require the data collector to peer inside each vehicle; multiple research teams have reported difficulties with visibility into vehicles, particularly those with tinted windows [28]. In the helmet use observations, although the occupants of the motorcycle are visible, the data collector has to interpret whether a helmet is being worn correctly or not and may not have sufficient time to correctly assess all occupants. The helmet and seat belt use observations were more reliable for drivers than for passengers in both Accra and Bandung, and the reliability among the visible motorcycle passengers was higher than the less visible 4-wheeler passengers.

Similarly, in a moving vehicle it is sometimes difficult to document the age and sex of the occupants with certainty. This problem magnifies when occupants are young children or are wearing helmets, which could negatively impact the reliability of the observation. Furthermore, the seat belt observations required finer estimates of age group, which can lead to more chances for misclassification. However, because of the design of the observational study, the issue of misclassification by age or sex is minimized by the random assignments of the data collectors by date, time, location, team pairs, and digital versus paper data collection. These misclassifications are therefore randomly distributed across paper and digital format; however, despite random distribution, this likely also lowered the reliability when making comparisons by sex and age group.

Misclassification of sex and age grouping is a general limitation of all studies based on roadside observations [28].

Another area where different observations were noted was the level of agreement on presence or absence of law enforcement. The presence and nature of police and camera enforcement, as well as environmental speed deterrents, must be same regardless of the method of data collection. The differences between observations might be simply a matter of timing (eg, if traffic police were at the location during only part of the session) or understanding of the environment (speed cameras vs closed circuit cameras; functional vs inactive speed cameras). These differences could be avoided by better training of the research assistants.

Speeding prevalences in paper and digital methods were found to be similar in HCMC and within 2 percentage points in Accra and Bandung, reflecting overall estimates as well as individual category of overspeeding. These small observed differences, though not meaningful, were in most instances statistically significant likely because of our very large sample sizes of vehicles and vehicle occupants. For instance, the prevalence of correct helmet use in Bandung was 0.68 according to the digitally collected dataset and 0.70 in the data collected through paper format. Although we do not judge 68% to be meaningfully different from 70%, with 173,043 observations on motorcycle occupants, this difference is statistically significant at the <.001 level of significance. The most important finding of this study was that despite the differences between digital and paper data collection formats in the volume of observations and variations in reliability, the overall prevalence of each risk factor was comparable. This finding is important for 2 reasons. First, switching from paper to DDC may reduce the mean number of observations per session, but it does not translate into a different prevalence of risk factors. DDC provided the advantage of reducing turnaround time, by eliminating the need of double data entry and cleaning required in paper format, which often delayed data analysis and dissemination of results. Second, the reliability of prevalence estimates for each risk factor obtained through digital method would allow to switch to DDC for future rounds of data collection in suitable environments, without impeding or distorting prior analysis of time trends for each road safety risk factor.

Challenges

DDC was not completely error-free but was found to minimize data entry errors resulting from an extra data entry step [29]. Although automated skip patterns, mandatory fields, and logic checks support data completeness and accuracy, there were instances where these led to slow recording or incomplete information. For example, if research assistant initially recorded 4 car occupants but could only observe 3 as the vehicle moved on, the digital application was programmed to not allow the form to be uploaded without completing required information on all occupants. This issue was fixed by changing the required fields and making the form more flexible by adding a *nonobservable* option, in consideration of these extremely dynamic roadside environments. Misclassification error between genders and among age categories results in lower reliability between digital and paper methods, but as this error was random, it did not affect the risk factor prevalence in the whole samples.

Generally, DDC was well received in all 3 cities, but some challenges were identified by the DDC teams. First, the research assistants in Bandung found DDC to be tiring, especially in the upper back and neck areas because of prolonged rigid upper body position during information recording. This has been previously reported in other studies and this issue was resolved by limiting the number of sessions to 2 to 3 per day for each data collector [30]. Second, unstable network connections made it difficult for local teams to upload data, particularly in suburban road networks. Without being able to upload data and clear the tablet memory, the tablets slowed down, especially when research assistants had to conduct multiple back-to-back sessions. Third, the battery ran out quickly when the mobile network was used to upload data to the server; to tackle this issue, the data collectors were provided with backup batteries and power banks. There were at least 2 occasions where digital data were lost for the entire session; in one instance, the tablet malfunctioned and in the other, a research assistant ignored the prompt to save the completed forms after finishing the session. Fourth, research assistants reported that sometimes they recalled an error, such as misspecifying their location, only after uploading the data to the server; this recall error was handled by the data managers who corrected the error on the server. Fifth, although the data collection tool was uploaded bilingually in Bandung and HCMC, some research assistants recommended the use of visuals and photos for data entry as opposed to text-based drop-down menus. It was also recommended to have screens that could be scrolled down than swiped to improve the efficiency; to date, this function was not available in KoBoCollect app. It is important to note that most of the observed challenges could be addressed by training of field personnel, investing in good quality tablets and power backup, and further development of a user-friendly tablet interface.

Implications for the Choice of Data Collection Method

Overall, paper-based data collection was found to be more productive method for observational studies in roadside environment. A possible explanation for this finding is that writing on paper is easier or at least more familiar initially; typing using an onscreen keyboard might be slower in some circumstances, owing to the requirement of entering information on each individual vehicle or vehicle occupant separately. This could be initially challenging for the average data collector, particularly if they were not familiar with Android technology or had not used a mobile phone or tablet on a regular basis. There appeared to be 2 learning curves for data collectors when moving to an electronic format; not only must they develop familiarity with the data entry system, but also with the content of the survey form as displayed in an electronic format [29]. The learning curve to use an electronic data entry system is usually proportionate to the degree of complexity in the electronic format and length of time spent developing experience. Considering our observational forms were much shorter and less complex (relative to a typical household survey), the time advantages of DDC might be less pronounced using an initially unfamiliar technology.

In those circumstances where sheer productivity is not the central focus and precision of estimates could be maintained by comparatively smaller sample sizes, DDC may be preferred, as shown in this study. DDC provides the advantages of standardization; logic checks; immediate updating of questionnaire version without wasting previously printed material; automatic synchronization of metadata, pictures, and GPS coordinates with the correct survey; and both remote and real-time monitoring of data quality, as documented by other researchers as well [14,15]. DDC requires a one-time cost of mobile or tablet devices as well as the cost of setting up a server and designing the digital module. Other longer-term costs include maintenance of devices, data plans, or internet service. Therefore, in some environments, paper-based data collection might be more feasible. However, in settings and circumstances where these conditions could be met, DDC could cut the time to aggregate large datasets, reduce the cost related with printing, transporting, and storing paper questionnaires, double data entry, reconciliation through hard copy checks, and associated human resources. The relative cost and availability of human and material resource could also impact the choice of data collection method.

Limitations

This paper does not directly address the efficiency of digital versus paper data collection. As mentioned earlier, efficiency of a data collection system is contextually determined and may imply time and cost efficiency or could be tied to logistical feasibility in a given environment. This study did not collect information on differential cost of supplies, equipment, training, human resource, data collection, and management.

Conclusions

DDC provides a reliable and convenient means for conducting large volume roadside observational studies of behavioral risk factors and reducing the turnaround time from data collection to policy decisions. There is some site-related variability in implementing DDC, but the big-picture results are comparable with the paper-based approach. There are upfront costs associated with resources to program the digital applications and acquire the necessary equipment for digital data collection, but the benefits of automating future rounds of data collection with quality data may help in reducing turnaround time and thus prove beneficial in the long run.

Acknowledgments

The authors thank the Bloomberg Philanthropies for funding this project through the Bloomberg Initiative for Global Road Safety, grant number 119016. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Bloomberg Philanthropies or Johns Hopkins University.

The authors also acknowledge the support and contributions of their partners in all 3 cities, including Building and Road Research Institute, Kumasi, Ghana; Center for Injury Prevention Policy and Research, Hanoi University of Public Health, Vietnam; and Universitas Padjadjaran Bandung, Indonesia.

Authors' Contributions

AM conceptualized the study and developed all manuscript drafts. NT analyzed the data and prepared results; AB and SG led the data collection in Vietnam, Indonesia, and Ghana; NP and AM developed the DDC tool and participated in training and implementation of DDC; and AAH helped in study concept and provided oversight of the study and critical feedback in protocol development and manuscript drafts. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Global Status Report On Road Safety 2018. Geneva, Switzerland: World Health Organization; 2019. URL: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/ [accessed 2019-05-14] [WebCite Cache ID 78N1YFihs]
2. Garg N, Hyder AA. Exploring the relationship between development and road traffic injuries: a case study from India. Eur J Public Health 2006 Oct;16(5):487-491. [doi: 10.1093/eurpub/ckl031] [Medline: 16641159]
3. Moniruzzaman S, Andersson R. Economic development as a determinant of injury mortality - a longitudinal approach. Soc Sci Med 2008 Apr;66(8):1699-1708. [doi: 10.1016/j.socscimed.2007.12.020] [Medline: 18308440]
4. Yannis G, Papadimitriou E, Folla K. Effect of GDP changes on road traffic fatalities. Saf Sci 2014 Mar;63:42-49 [FREE Full text] [doi: 10.1016/j.ssci.2013.10.017]
5. Law TH. Factors associated with the relationship between non-fatal road injuries and economic growth. Transport Policy 2015 Aug;42:166-172. [doi: 10.1016/j.tranpol.2015.06.004]
6. Institute of Metrics and Evaluation, University of Washington. 2017. Global burden of disease: Road traffic injuries URL: https://vizhub.healthdata.org/gbd-compare/ [accessed 2018-12-03] [WebCite Cache ID 74Op28IT4]
7. Sminkey L. World Health Organization. Global Plan for the Decade of Action for Road Safety 2011-2020 URL: https://www.who.int/roadsafety/decade_of_action/plan/plan_english.pdf [accessed 2019-05-06] [WebCite Cache ID 78B21BfTH]

8. Hyder AA, Allen KA, Di Pietro G, Adriazola CA, Sobel R, Larson K, et al. Addressing the implementation gap in global road safety: exploring features of an effective response and introducing a 10-country program. Am J Public Health 2012 Jun;102(6):1061-1067. [doi: 10.2105/AJPH.2011.300563] [Medline: 22515864]
9. Peden M, Di Pietro G. Saving lives by improving road safety. Public Health 2017 Dec;144S:S3-S4 [FREE Full text] [doi: 10.1016/j.puhe.2016.12.038] [Medline: 28288728]
10. Bloomberg Philanthropies. Public Health: Road Safety URL: https://www.bloomberg.org/program/public-health/road-safety/ [accessed 2018-12-03] [WebCite Cache ID 74OpT0ekY]
11. Pérez-Núñez R, Chandran A, Híjar M, Celis A, Carmona-Lozano MS, Lunnen JC, et al. The use of seatbelts and child restraints in three Mexican cities. Int J Inj Contr Saf Promot 2013;20(4):385-393. [doi: 10.1080/17457300.2012.754477] [Medline: 23316999]
12. Mathur AK, Gupta S, Bandhu A. A baseline study on pattern of helmet use in the State of Rajasthan, India. J Health Manag 2017 Jul 23;19(3):417-434. [doi: 10.1177/0972063417717894]
13. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decis Mak 2009 Dec 23;9(1):51 [FREE Full text] [doi: 10.1186/1472-6947-9-51] [Medline: 20030813]
14. Caeyers B, Chalmers N, De Weerdt J. SSRN. 2011 Feb 08. A Comparison of CAPI and PAPI Through a Randomized Field Experiment URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1756224 [accessed 2018-12-22] [WebCite Cache ID 74qHmUhH2]
15. Tomkys E, Lombardini S. Going Digital: Using digital technology to conduct Oxfam's Effectiveness Reviews. Oxford, UK: Oxfam; Sep 2015:9.
16. Patnaik S, Brunskill E, Thies W. Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice. 2009 Apr 17 Presented at: International Conference on Information and Communication Technologies and Development (ICTD); 17-19 April, 2009; Doha, Qatar p. 74-84. [doi: 10.1109/ICTD.2009.5426700]
17. Schleyer TK, Forrest JL. Methods for the design and administration of web-based surveys. J Am Med Inform Assoc 2000;7(4):416-425 [FREE Full text] [Medline: 10887169]
18. Goldstein M, Benerjee R, Kilic T. The World Bank. 2012 Jul 24. Paper v Plastic Part I: The survey revolution is in progress URL: https://blogs.worldbank.org/impactevaluations/paper-v-plastic-part-i-the-survey-revolution-is-in-progress [accessed 2018-12-03] [WebCite Cache ID 74Os4k7c1]
19. Goldstein M, Banerjee R, Kilic T. The World Bank. 2012 Jul 25. Paper or Plastic? Part II: Approaching the survey revolution with caution URL: https://blogs.worldbank.org/impactevaluations/paper-or-plastic-part-ii-approaching-the-survey-revolution-with-caution [accessed 2018-12-03] [WebCite Cache ID 74OsU8d4p]
20. Fletcher LA, Erickson DJ, Toomey TL, Wagenaar AC. Handheld computers. A feasible alternative to paper forms for field data collection. Eval Rev 2003 Apr;27(2):165-178. [doi: 10.1177/0193841X02250527] [Medline: 12703341]
21. Lombardini S, Tomkys E. Going Digital: Using and sharing real-time data during fieldwork. Oxford, UK: Oxfam; Nov 2017:6-7.
22. KoBoToolbox at the Harvard Humanitarian Initiative. URL: https://www.kobotoolbox.org/ [accessed 2018-12-03] [WebCite Cache ID 74OqJZC3X]
23. Deniau C, Gaillard T, Mbagogo A, Réounodji F, Le Bel S. Using the KoBoCollect tool to analyze the socio-economic and socio-cultural aspects of commercial hunting and consumption of migratory waterbirds in the Lakes Chad and Fitri (Chad). 2017 Presented at: Conference proceedings of EFITA WCCA congress; July 6-7, 2017; Montpellier, France p. 87-88.
24. McHugh M. The chi-square test of independence. Biochem Med (Zagreb) 2013;23(2):143-149 [FREE Full text] [Medline: 23894860]
25. Agresti A, Coull B. Approximate is better than "exact" for interval estimation of binomial proportions. Am Stat 1998 May;52(2):119-126 [FREE Full text] [doi: 10.1080/00031305.1998.10480550]
26. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005 Mar;85(3):257-268. [Medline: 15733050]
27. StataCorp LLC. 2017. Stata Statistical Software: Release 15 URL: https://www.stata.com/ [accessed 2019-05-10] [WebCite Cache ID 78FzesEsJ]
28. Vecino-Ortiz AI, Bishai D, Chandran A, Bhalla K, Bachani AM, Gupta S, et al. Seatbelt wearing rates in middle income countries: a cross-country analysis. Accid Anal Prev 2014 Oct;71:115-119. [doi: 10.1016/j.aap.2014.04.020] [Medline: 24906165]
29. Cole E, Pisano ED, Clary GJ, Zeng D, Koomen M, Kuzmiak CM, et al. A comparative study of mobile electronic data entry systems for clinical trials data collection. Int J Med Inform 2006;75(10-11):722-729. [doi: 10.1016/j.ijmedinf.2005.10.007] [Medline: 16386460]
30. Toh SH, Coenen P, Howie EK, Straker LM. The associations of mobile touch screen device use with musculoskeletal symptoms and exposures: a systematic review. PLoS One 2017;12(8):e0181220 [FREE Full text] [doi: 10.1371/journal.pone.0181220] [Medline: 28787453]

Abbreviations

BIGRS: Bloomberg Initiative for Global Road Safety
DDC: digital data collection
GPS: Global Positioning System
HCMC: Ho Chi Minh City
LMIC: low- and middle-income country
mHealth: mobile health
RTI: road traffic injury

Edited by G Eysenbach; submitted 21.12.18; peer-reviewed by S Shah, K Goniewicz; comments to author 21.02.19; revised version received 06.03.19; accepted 08.04.19; published 28.05.19.

Please cite as:
Mehmood A, Taber N, Bachani AM, Gupta S, Paichadze N, Hyder AA
Paper Versus Digital Data Collection for Road Safety Risk Factors: Reliability Comparative Analysis From Three Cities in Low- and Middle-Income Countries
J Med Internet Res 2019;21(5):e13222
URL: http://www.jmir.org/2019/5/e13222/
doi:10.2196/13222
PMID:31140431

Original Paper

Profiles of a Health Information–Seeking Population and the Current Digital Divide: Cross-Sectional Analysis of the 2015-2016 California Health Interview Survey

Hena Naz Din[1,2], MPH; Corinne McDaniels-Davidson[2,3,4], MPH, PhD; Jesse Nodora[4,5], DrPH; Hala Madanat[2,6], PhD

[1]University of California, San Diego, La Jolla, CA, United States
[2]School of Public Health, San Diego State University, San Diego, CA, United States
[3]Institute for Public Health, San Diego State University, San Diego, CA, United States
[4]Moores Cancer Center, University of California, San Diego, La Jolla, CA, United States
[5]Department of Family Medicine and Public Health, School of Medicine, University of California, San Diego, La Jolla, CA, United States
[6]Institute for Behavioral and Community Health, San Diego State University, San Diego, CA, United States

Corresponding Author:
Hena Naz Din, MPH
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 (858) 534 2230
Email: hena.n.din@gmail.com

Abstract

Background: Internet use for health information is important, given the rise of electronic health (eHealth) that integrates technology into health care. Despite the perceived widespread use of the internet, a persistent “digital divide” exists in which many individuals have ready access to the internet and others do not. To date, most published reports have compared characteristics of internet users seeking health information vs nonusers. However, there is little understanding of the differences between internet users seeking health information online and users who do not seek such information online. Understanding these differences could enable targeted outreach for health interventions and promotion of eHealth technologies.

Objective: This study aims to assess population-level characteristics associated with different types of internet use, particularly for seeking online health information.

Methods: The 2015-2016 California Health Interview Survey datasets were used for this study. Internet use was classified as never used the internet (*Never use*), ever used the internet but not to search for health information in the last 12 months (*Use not for health*), and ever used the internet and have used it to search for health information in the last 12 months (*Use for health*). Weighted multinomial logistic regression was used to assess sociodemographic and health characteristics associated with types of internet use. Findings are reported as odds ratios (ORs) with 95% CIs.

Results: Among 42,087 participants (weighted sample of 29,236,426), 19% reported *Never Use* of the internet, 27.9% reported *Use not for health*, and 53.1% reported *Use for health*. Compared to *Never Use* individuals, *Use for health* individuals were more likely to be younger (OR: 0.1, 95% CI 0.1-0.2 for ≥60 years vs <60 years), female (OR: 1.6, 95% CI 1.3-1.9 compared to males), and non-Hispanic white (OR: 0.54, 95% CI 0.4-0.7 for Latinos and OR: 0.2, 95% CI 0.2-0.4 for African Americans) and have a higher socioeconomic status (>400% of Federal Poverty Guidelines; OR: 1.3, 95% CI 1.4-2.4). Overall, characteristics for the *Use not for health* and *Use for health* groups were similar, except for those with lower levels of education and respondents not having visited a physician in the last year. For these two characteristics, the *Use not for health* group was more similar to the *Never Use* group.

Conclusions: Our findings indicate that a digital divide characterized by sociodemographic and health information exists across three types of users. Our results are in line with those of previous studies on the divide, specifically with regard to disparities in use and access related to age, race/ethnicity, and socioeconomic status. Disparities in online health-seeking behavior may reflect existing disparities in health care access extending into a new era of health technology. These findings support the need for interventions to target internet access and health literacy among *Never Use* and *Use not for health* groups.

(J Med Internet Res 2019;21(5):e11931) doi:10.2196/11931

KEYWORDS

internet; digital divide; health information; patient portals; online health information seeking; eHealth

Introduction

The internet has become a widespread tool of social communication, economic opportunity, and health care information and access [1]. Since 2000, use of the internet has increased nationally from 52% to 89% [2]. In the realm of health and well-being, access to and use of the internet are central to the rise of electronic health (eHealth), which is the integration of technology into health care to improve and facilitate wellness and health maintenance [3-5]. The Affordable Care Act, in particular, has expanded the integration of the internet in health care by funding and incentivizing health information technology in the form of patient portals and electronic health records [6]. Although these advances are intended to improve patient wellness, a digital divide exists in which certain individuals have ready access to and can use the internet, whereas others do not [7]. This divide has the potential to exacerbate existing health disparities among vulnerable populations.

Although the digital divide was historically characterized by access to the internet, more recently, the divide has highlighted differences in use and skills [4]. Use of the internet for health information has real-world benefit, including application of information learned online to manage and monitor one’s health [4,8,9]. However, internet access and use vary by individual characteristics including race and ethnicity. Prior studies report that individuals of older age, lower income, male gender, and rural locality tend to be less likely to report internet access or use compared to their counterparts [7,10]. Latinos, African Americans, and individuals with low English proficiency are more likely to report no internet use compared to non-Hispanic Whites and English-proficient individuals [7,11]. Assessing characteristics of those who do and those who do not use the internet for health information may be important within the context of the digital divide, because it has implications on the potential impact of eHealth. Specifically, groups least likely to report internet use are also less likely to report use of patient portals and access online health information [12,13].

Published reports about internet use, specifically for seeking health information, have primarily compared individuals who do not use the internet in any capacity to those who use the internet specifically for seeking health information [7,10]. This type of dichotomy tends to focus primarily on internet use in the context of access (ie, those who have access to the internet for seeking health information are users and those who do not are nonusers). This overlooks a third group that exists—individuals who use the internet but not to seek health information. To our knowledge, only one other study explored characteristics of these individuals in comparison to internet users who seek health information online [7]. However, characterization of groups by internet use should include nonusers of the internet as well as other types of users. By delineating group characteristics according to the type of use, findings may reveal new opportunities for intervention.

Therefore, this study aims to assess population-level characteristics associated with internet use, particularly for seeking health information online. Sociodemographic and health characteristics will be explored among those who report no use of the internet, those who use the internet for seeking health information, and those who use the internet but not for seeking health information. Varied profiles of use and nonuse may indicate areas of the digital divide that are changing and may benefit from intervention and targeted resources.

Methods

Data Source and Participants

Data from the 2015-2016 California Health Interview Survey (CHIS) were used for this study. CHIS is a statewide study conducted by the University of California Los Angeles Center for Health Policy and Research in collaboration with the California Department of Public Health and the Department of Health Care Services. Data are collected annually through surveys administered by random-digit dialing to both landline phones and cellphones. CHIS data yield a representative sample of the noninstitutionalized population of California to explore health, wellness, and access by race and region across the state.

This study used publicly available data files that allow for estimation of state-level outcomes. The 2015-2016 CHIS data files are provided separately for each year and require pooling to aggregate data. The first wave of data was collected from May 2015 to February 2016 and the second, from January 2016 to December 2016. Participants were recruited through landline and cell phone sampling. All individuals who participated in the adult (age≥18 years) 2015-2016 surveys were included in this study.

Types of Internet Use

The CHIS assesses types of internet use by asking participants “Have you ever used the Internet?” and “In the past 12 months, did you use the Internet to look for health or medical information?” [14]. Responses to questions were “yes” or “no.” Only those who responded “yes” to having used the internet were asked about searching for health or medical information online. Internet use could be on any computer or electronic device and encompassed a variety of activities like emailing and searching the Web. Health and medical information was broadly defined as searching for health and wellness advice (eg, nutrition and physical activity), disease symptoms, health plans, and other related topics [14]. These questions were combined to create the main outcomes of this study: never used the internet (*Never use*), ever used the internet but have not used it to search for health information in the last 12 months (*Use not for health*), and ever used the internet and have used it to search for health information in the last 12 months (*Use for health*).

Measures

Participant sociodemographic and health information characteristics were included to assess association with the types of internet use. Variables were selected based on previous literature indicating an association of internet access with type of use and online health seeking. Sociodemographic variables included age, gender, marital status, education, primary language used at home, nativity, household size, federal poverty level, and race/ethnicity. Age was categorized as <60 years and ≥60 years based on past research, which highlights a disparity in internet use around this age [10]. Race and ethnicity are reported in CHIS based on the Office of Management and Budget classification and included six categories: Hispanic, Non-Hispanic (NH) white, African-American (NH), Asian (NH), American Indian/Alaskan Native (NH), and Other races. Rural and urban locality is a CHIS-generated variable based on zip code data.

Health information included the type of health insurance (private, public, or no insurance), presence of usual care (“yes” was classified as access to a doctor or clinic/health center and “no” was classified as no reported usual care, emergency department listed as usual care, or no one place used for usual care), current health status (excellent, very good, good, fair, or poor), and visit to a medical provider in the last 12 months (“yes” or “no”).

Weighting of Data

As recommended by the CHIS, this study utilized weights for replication methods. The CHIS uses a complex sampling design that involves oversampling by geographic area and group characteristics (eg, race/ethnicity). Weights are thus needed to adjust for the different probabilities of selection and to most accurately reflect the population of California. Design characteristics that require weighted variables include multiple people interviewed in one household (maximum 3), oversampling from certain geographic areas, and random-digit phone dialing for participant recruitment. Weights are designed for the jackknife method of variance and bias estimation.

Statistical Analyses

Descriptive statistics on internet use and variables of interest were performed using design-adjusted or weighted frequencies. Correlations between variables were assessed to avoid multi-collinearity; closely identified variables (rho≤0.5) were reduced to one variable for inclusion in the final model [15]. Bivariate analyses were conducted between each variable and internet use in order to determine inclusion into the final model at a significance of *P*<.05. To evaluate the association between the variables of interest and internet use, weighted multinomial logistic regression analysis was conducted to model internet use outcomes, using the *NeverUse* group assigned as the reference group. Findings are reported as odds ratios (OR) with 95% CIs. The difference of association between *Use not for health* and *Use for health* groups was assessed by evaluating overlapping CIs from the weighted multinomial logistic regression. Evaluating CIs is a conservative method of assessing significant differences as compared to other post-hoc analyses/comparisons of groups [16,17]. Assessment of CIs focused on CIs that did not overlap, thus indicating a significant difference between groups [16,17]. All analyses were conducted using SAS OnDemand for Academics (SAS Institute Inc, Cary, NC).

Results

Study Population

The CHIS sampled 21,034 individuals in 2015 and 21,055 in 2016. This resulted in a population-weighted sample of 14,541,326 from 2015 and 14,695,100 from 2016. The final population-weighted sample comprised 29,236,426 individuals.

Prevalence of Internet Use

Never use of the internet was reported by 83.8% (n=33,856) of the study sample (weighted: 84.2%, n=24,508,603). Among those, 65.6% (n=22,195; weighted: 65.4%, n=16,027,194) reported using the internet to search for health information in the last 12 months. In addition, 19% (n=7959, weighted: 15.8%, n=4,599,254) reported *Never Use* of the internet, 27.9% (n=11,641; weighted: 29.1%, n=8,473,075) reported *Use not for health*, and 53.1% (n=22,194; weighted: 55.1%, n=16,026,161) reported *Use for health.*

Internet Use by Population Characteristics

The distribution of population characteristics across levels of internet use is reported in Table 1. Comparison of proportions showed a significant difference (*P*<.05) within each characteristic across all categories of internet use. A greater percentage of older adults, Hispanics, and individuals with public health insurance were found in the *Never Use* group (Table 1). Public and private health insurance was split similarly among the *Use not for health* group (42.7% vs 45.3%). A majority of the *Never Use* group reported good or fair health (30.3% vs 34.9%), while very good and good health were more common among *Use not for health* (29.4% vs 32.9%) and *Use for health* (34.2 vs 29.8) groups (Table 1).

Weighted regression showed that compared that to *Never Use* individuals, *Use for health* individuals were more likely to be younger (OR: 0.1, 95% CI 0.1-0.2 for ≥60 years vs <60 years), female (OR: 1.6, 95% CI 1.3-1.9 compared to males), and non-Hispanic white (OR: 0.5, 95% CI 0.4-0.7 for Latinos and OR: 0.2, 95% CI 0.2-0.4 for African Americans), and have a higher socioeconomic status (≥400% of Federal Poverty Guidelines; OR: 1.3, 95% CI 1.4-2.4; Table 2).

Living in an urban location was significant for *Use for Health* compared to *Never Use* (OR: 0.7 95% CI 0.6-0.9 for rural vs urban); however, there was no significant difference in geographic locality between *Never Use* and *Use not for health.Use for health* individuals were also more likely to be employed (OR: 0.5, 95% CI 0.4-0.7 for unemployed vs employed) and privately insured (OR: 1.7, 95% CI 1.4-2.2 compared to *Never Use*; Table 2).

Table 1. Distribution of population characteristics according to type of internet use (weighted population values) from the 2015-2016 California Health Interview Survey.

Characteristic	*Never use* [a]	*Use not for health* [a]	*Use for health* [a]
Age (years), n (%)			
<60	2,057,498 (44.7)	6,640,214 (78.4)	12,907,147 (80.5)
≥60	2,541,756 (55.3)	1,832,861 (21.6)	3,119,014 (19.5)
Gender, n (%)			
Male	2,137,475 (46.5)	4,845,123 (57.2)	7,223,238 (45.1)
Female	2,461,779 (53.5)	3,627,952 (42.8)	8,802,923 (54.9)
Race[b] , n (%)			
Non-Hispanic white	1,038,704 (22.6)	3,069,101 (36.2)	8,058,396 (50.3)
Hispanic	2,650,416 (57.6)	3,489,369 (41.2)	4,164,014 (26)
African American	271,686 (5.9)	628,620 (7.4)	736,405 (4.6)
American Indian/Alaskan Native	25,902 (0.56)	51,779 (0.61)	56,660 (0.35)
Asian	561,019 (12.2)	1,037,029 (12.2)	2,548,049 (15.9)
Other	51,528 (1.1)	197,176 (2.3)	462,638 (2.9)
Household size			
Reponses obtained, n (%)	4,599,254 (15.8)	8,473,075 (29.1)	16,026,161 (55.1)
Mean (SEM)	3.20 (0.06)	3.43 (0.04)	3.15 (0.02)
Employment[c] , n (%)			
Employed	2,837,920 (61.7)	5,748,514 (67.8)	11,358,968 (70.9)
Unemployed	1,761,334 (38.3)	2,724,560 (32.2)	4,667,193 (29.1)
% Federal Poverty Level, n (%)			
0-138	2,339,046 (50.9)	2,561,345 (30.2)	3,653,178 (22.8)
139-200	687,038 (14.9)	1,018,756 (12.0)	1,562,263 (9.7)
201-400	938,716 (20.4)	2,158,134 (25.5)	3,683,057 (23.0)
≥400	634,454 (13.8)	2,734,840 (32.3)	7,127,664 (44.5)
Geographic location[d] , n (%)			
Urban	4,096,249 (89.1)	7,593,284 (89.6)	14,582,763 (91)
Rural	503,005 (10.9)	879,791 (10.4)	1,443,398 (9.0)
Place of birth, n (%)			
The United States	1,702,387 (37.0)	5,407,011 (63.8)	12,069,324 (75.3)
Outside the United States	2,896,866 (63)	3,066,064 (36.2)	3,956,837 (24.7)
General health, n (%)			
Excellent	453,312 (9.8)	1,580,799 (18.7)	3,265,295 (20.4)
Very good	647,263 (14.1)	2,494,385 (29.4)	5,481,351 (34.2)
Good	1,392,301 (30.3)	2,785,481 (32.9)	4,777,245 (29.8)
Fair	1,606,978 (34.9)	1,331,452 (15.7)	2,020,028 (12.6)
Poor	500,499 (10.9)	280,958 (3.3)	482,242 (3.0)
Insurance type[e] , n (%)			
Public	3,216,169 (69.9)	3,619,481 (42.7)	5,316,902 (33.2)
Private/employer	798,882 (17.4)	3,835,796 (45.3)	9,498,872 (59.3)
Uninsured	584,203 (12.7)	1,017,798 (12.0)	1,210,387 (7.6)

Characteristic	*Never use* [a]	*Use not for health* [a]	*Use for health* [a]
Education level , n (%)			
Less than high school	2,440,420 (53.1)	1,575,314 (18.6)	962,512 (6.0)
High school	1,104,174 (24.0)	2,368,448 (28.0)	2,900,243 (18.1)
Some college[f]	575,168 (12.5)	2,184,604 (25.8)	4,115,447 (25.7)
Bachelor of Arts/Science	360,689 (7.8)	1,600,450 (18.9)	5,062,836 (31.6)
Master’s degree or higher	118,802 (2.6)	744,259 (8.8)	2,985,123 (18.6)
Usual source of care[g] , n (%)			
Yes	3,710,563 (80.7)	6,889,486 (81.3)	13,892,613 (86.7)
No	888,691 (19.3)	1,583,588 (18.7)	2,133,548 (13.3)
Visited physician in the last 12 months , n (%)			
Yes	3,673,726 (79.9)	6,125,379 (72.3)	13,663,846 (85.3)
No	925,528 (20.1)	2,347,696 (27.7)	2,362,315 (14.7)
Marital status , n (%)			
Married	2,321,581 (50.5)	3,865,368 (45.6)	7,733,475 (48.3)
Other[h]	1,715,317 (37.3)	2,078,225 (24.5)	3,377,259 (21.1)
Never married	562,356 (12.2)	2,529,483 (29.9)	4,915,426 (30.7)
Language at home , n (%)			
English	1,464,336 (31.8)	4,388,202 (51.8)	9,998,790 (62.4)
Spanish	1,405,725 (30.6)	876,315 (10.3)	587,631 (3.7)
Asian languages[i]	231,906 (5.04)	182,780 (2.2)	267,543 (1.7)
Other	56,752 (1.2)	126,799 (1.5)	227,592 (1.4)
Multilingual	1,440,535 (31.3)	2,898,979 (34.2)	4,944,605 (30.9)

[a] *Never use*: never used the internet; *Use not for health*: ever used the internet but not for seeking health information in the last 12 months; *Use for health*: used internet ever and for health information in the last 12 months.

[b] Based on the classification by the Office of Management and Budget [18].

[c] “Employed” includes those with full- and part-time employment; “Unemployed” includes those looking for work and those not looking for work.

[d] Determined by zip codes.

[e] Public: categorized as only Medicare, only Medicaid, combination of the two, and the combination of one with a classified *other* insurance. Private: categorized as employment-based insurance and privately purchased insurance.

[f] “Some college” includes vocational school, Associates of Arts, Associates of Science, and some years of college.

[g] Usual care includes doctor or clinic/health center. No usual care includes no reported usual care, emergency department listed as usual care, or no one place used for usual care.

[h] “Other” includes widowed, separated, divorced, or living with partner.

[i] Asian languages include Chinese, Vietnamese, and Korean.

Table 2. Association between population characteristics and type of internet use (weighted population values). Values are presented as multivariable-adjusted odds ratio and 95% CI (Reference-*Never use*).

Characteristic	*Use not for health* [a]	*Use for health* [a]
Age (years)		
<60 years	1	1
≥60 years	0.2 (0.1-0.2)	0.1 (0.1-0.2)
Gender[b]		
Male	1	1
Female	0.9 (0.8-1.1)	1.6 (1.3-1.9)
Race		
Non-Hispanic white	1	1
Hispanic	0.79 (0.58-1.06)	0.54 (0.40-0.72)
African American[b]	0.64 (0.48-0.84)	0.29 (0.21-0.39)
American Indian/Alaskan Native	0.77 (0.43-1.40)	0.42 (0.24-0.72)
Asian	0.70 (0.48-1.03)	0.69 (0.46-1.03)
Other	0.89 (0.51-1.56)	0.79 (0.44-1.40)
Number of individuals in a household, mean (SEM)	1.1 (1.0-1.2)	1.1 (1.0-1.1)
Employment		
Employed	1	1
Unemployed	0.6 (0.5-0.7)	0.5 (0.4-0.7)
% Federal Poverty Guidelines		
0-138	1	1
139-200	1.1 (0.8-1.4)	1.0 (0.8-1.3)
201-400	1.5 (1.1-1.9)	1.3 (1.04-1.7)
>400	1.9 (1.4-2.4)	1.8 (1.4-2.4)
Geographic location		
Urban	1	1
Rural	0.9 (0.7-1.1)	0.7 (0.6-0.9)
Place of birth		
The United States	1	1
Outside the United States	0.6 (0.4-0.7)	0.4 (0.3-0.5)
General health		
Excellent	1	1
Very good	1.1 (0.8-1.4)	1.1 (0.8-1.5)
Good	0.9 (0.7-1.2)	0.9 (0.7-1.2)
Fair	0.6 (0.4-0.9)	0.7 (0.5-1.0)
Poor	0.5 (0.4-0.7)	0.6 (0.4-0.9)
Insurance type		
Public	1	1
Private/employer	1.4 (1.1-1.7)	1.7 (1.4-2.2)
Uninsured	1.1 (0.8-1.5)	1.4 (0.9-1.9)
Education level		
Less than high school	1	1
High school	2.3 (1.8-3.1)	3.5 (2.6-4.6)

Characteristic	*Use not for health* [a]	*Use for health* [a]
Some college[b]	3.9 (2.8-5.3)	8.2 (5.9-11.2)
Bachelor of Arts/Science[b]	4.5 (3.2-6.3)	15.3 (10.8-21.7)
Master’s degree or higher[b]	7.1 (4.5-11.4)	32.1 (19.6-52.4)
Usual source of care		
Yes	1	1
No	0.9 (0.7-1.1)	0.9 (0.7-1.2)
Visited physician in the last 12 months[b]		
Yes	1	1
No	1.2 (1.0-1.6)	0.6 (0.5-0.8)
Marital status		
Married	1	1
Other	0.9 (0.8-1.1)	0.9 (0.7-1.1)
Never married	1.6 (1.2-2.1)	1.8 (1.3-2.4)
Language at home		
English	1	1
Spanish	0.6 (0.4-0.8)	0.5 (0.3-0.7)
Asian languages	0.8 (0.4-1.7)	0.5 (0.2-1.1)
Other	0.7 (0.3-1.6)	0.5 (0.2-1.2)
Multilingual	0.9 (0.7-1.2)	1.0 (0.7-1.3)

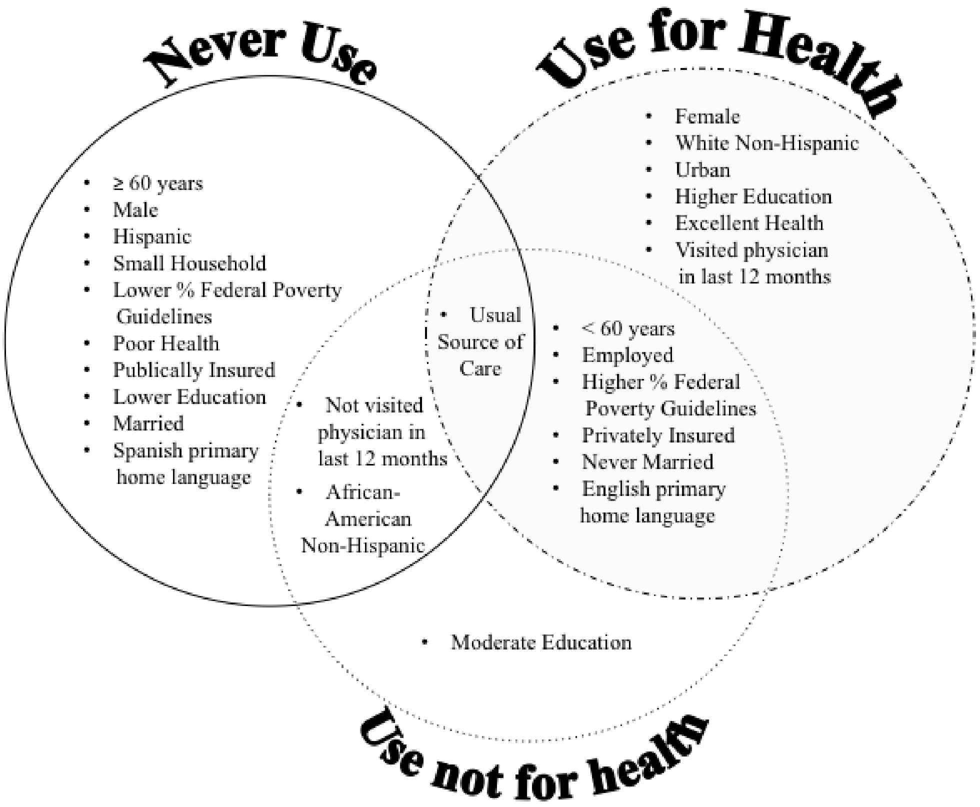
[a] *Use not for health*: ever used the internet but not for seeking health information in the last 12 months; *Use forhealth*: used internet ever and for health information in the last 12 months.

[b] Significant difference across *Use not for health* and *Use forhealth* groups assessed by 95% CI overlap.

The *Use not for health* group had similar characteristics as the *Use for health* group, except in the likelihood of being female, likelihood of being African American (non-Hispanic), levels of education, and likelihood of not having visited a physician in the last year. Compared to *Use for health* individuals, *Use not for health* individuals were significantly less likely to be female but significantly more likely to be African American (NH). Compared to *Never Use* individuals, more *Use not for health* individuals had college level or higher education (OR: 4.5, 95% CI 3.2-6.3 for Bachelor of Arts/Bachelor of Science and OR: 7.1, 95% CI 4.5-11.4 for master’s degree or higher). These CIs did not overlap when compared to *Use for health* (OR: 15.3, 95% CI 10.8-21.7 for Bachelor of Arts/Bachelor of Science and OR: 32.1, 95% CI 19.6-52.4 for master’s degree or higher), indicating a larger proportion of people with higher education among the *Use for health* group. Not visiting a physician in the last year was not significantly different between the *Never Use* and *Use not for health* groups (OR: 1.2, 95% CI 1.0-1.6) but was different when compared to the *Use for health* group (OR: 0.6, 95% CI 0.5-0.8).

Overall, the *Use not for health* and *Use for health* groups shared more similarities in sociodemographic characteristics than the *Never Use* group (Figure 1). Only the usual source of care did not significantly differ between the groups.

Figure 1. Characteristics of groups by types of internet use. Never use: never used the internet; Use not for health: ever used the internet but not for seeking health information in the last 12 months; Use for health: used internet ever and for health information in the last 12 months.



Discussion

Principal Results

This study assessed population characteristics associated with types of internet use for seeking health information among a population representative sample in California. The study adds to the literature by updating the assessment of the digital divide across three categories of use.

Limitations

Strengths of this study include use of a population representative sample of California and assessment of internet use, specifically for seeking health information online. The size of the sample and subsequent weighting allowed for nuanced evaluation of small but key groups within California like racial minority and rural residing groups. Limitations of the study include lack of specificity in terms of what type of health information respondents sought online and with what frequency respondents sought this information. Additionally, CHIS does not report on other aspects considered important for internet use including health literacy and technical skills to navigate online searching. As a population-based study specific to California, the CHIS results are not generalizable to other states. Additionally, CHIS is based on self-report and may be subject to recall bias.

Comparison to Prior Work

Compared to the study by Nguyen et al [7], the proportion of Californians who are internet users and who searched for health information has grown from 81.5% to 84.2% and from 64.5% to 65.4%, respectively [7]. This study shows that compared to *Use for health* individuals, *Never Use* individuals tend to be older, male, Hispanic, or African American (vs NH white); be of a lower socioeconomic status; and report poorer health (compared to excellent health). These distinctions mirror what is already understood about the digital divide, specifically with regard to age disparities (older adults vs adults) and racial disparities (African American/Hispanic vs NH white) [7,19]. Although general internet use is growing among older adults (67%) and racial/ethnic minorities (87%), these groups still report less internet use and access [2]. Older adults are a vulnerable group, given their high health needs, and may benefit from internet use for health. Barriers to internet use among older adults include technical difficulties, confusion with the amount of information online, disability in terms of psychomotor function, health literacy, costs, and distrust in internet sources [10,20]. Of note, distrust and low health literacy (eg, understanding health and medical information) have been significantly associated with less internet use for seeking health information even when older adults report using the internet [10,20].

Among African American and Hispanic individuals, internet access, defined by desktop, laptop, or handheld computer (ie, smartphone) ownership or broadband subscription, is low (63.5% and 69.6%, respectively) as compared to NH white individuals (78.8%) [21]. Smartphones have helped increase access to the internet for minority groups. Compared to only 9% of NH white individuals, 22% of Hispanic and 15% of African American individuals are smartphone-only internet users [22]. Among these groups, activities such as health information seeking are more likely to be performed on smartphones than on a traditional computer. Among Hispanic individuals, nonusers tend to be less proficient in English, and Spanish is the primary language in which media is consumed [11]. Language and literacy barriers are significant, as searching, reading, and comprehending health information online can be complex. Use of a patient portal may be more complicated than searching for health information online, as the former involves logging in, password creation, and more specific medical information. Thus, language and literacy are significant barriers to internet use generally and in the context of eHealth.

Unique to our study is the classification of *Use not for health* and comparing this group to both *Never Use* and *Use for health* groups. Given the saturation of internet use in our sample (>80% reported internet use), we were able to subcategorize this group into use but not for health and use for health seeking. Differences in educational attainment, gender, race (specifically of African American race), and physician visits in the last year differentiate the *Use not for health* group from the *Use for health* group. Nguyen et al [7] reported similar findings when only characterizing internet users who search for health information online. Specifically, they found that individuals who use the internet to search for health information were more likely to have seen a physician in the past year, to have higher educational attainment, to be of an ethnic minority (ie, African American or Latino), and to be female [7]. Not having visited a physician in the last 12 months may explain the low use of the internet for seeking health information, because it may indicate no serious diagnosis and thus no concerns or interests in seeking health information online. A serious diagnosis, or risk of one, is associated with health information seeking, as evidenced by a national US sample in which health information seeking was more prevalent among cancer survivors (69.8%) and individuals with a family history of cancer (51.2%) than those who had no history of cancer (29.6%) [23]. In the context of eHealth and health technology integration, it is important to explore why individuals who have access to the internet in some capacity do not use it for health-seeking activities (eg, searching for health information).

Types of internet use may indicate the extent to which eHealth expansion and integrated online health tools may be missing key populations. Health systems and policy makers may consider the characteristics of populations identified in this study by *Never Use* and *Use not for health* groups, because these populations may be lacking in the intended benefit of patient portals and eHealth resources. In particular, if *Never Use* and *Use not for health* groups do not frequent medical providers, accessing these groups may require other techniques besides, for example, designating a patient portal during a clinic visit. Types of use may also reflect existing disparities in health care access continuing into a new era of health technology. As evidenced by a study of a representative sample in the Dutch population, disparities in internet usage reflected real-life disparities experienced within the Dutch population according to gender, age, and level of education [24]. Careful consideration and purposeful design of health and technology integration may have the potential to eliminate disparities [25]. Patient portals have already been shown to improve patient outcomes and engagement and reduce health care costs, especially among patients with chronic conditions [26]. However, this only applies to patients who are able to access, use, and comprehend the information being shared through these portals. Healthy People 2020 has incorporated internet access and use into its objectives to target these disparities [27]. Particularly relevant are objectives to increase the proportion of online health information seekers who report easy access to information to increase health literacy and the proportion of health-related websites that are simple and usable [27].

Conclusions

Advances in health care and management have promoted internet integration and its use to support health and wellness. Although our study found a high prevalence of internet use across the study population, findings suggest that a digital divide continues to exist. The widest divide still remains between nonusers and users, and a lesser divide exists among users who search for health information and users who do not. Disparities identified in both internet use and health information seeking reflect a lack of health equity in a new era of technological advances in society and health care. The internet is one important tool for the development of an empowered patient and individual. Underutilization of the internet as a tool of health information leaves behind vulnerable populations and may have an adverse impact on health care for these individuals. Future studies may explore specific barriers to internet use and online health information seeking among *Never Users* and *Use but not for health* to inform and shape targeted and tailored interventions. Targeting interventions and educational materials to nonusers and users of the internet may improve internet and health literacy and support the ultimate goal of developing a more informed and empowered individual.

Acknowledgments

We would like to acknowledge Drs. Elena Martinez, Greg Norman, and Eileen Pitpitan for their scientific advice and guidance. This work was supported by the SDSU/UCSD Comprehensive Cancer Center Partnership (CA132379 and CA132384).

Conflicts of Interest

None declared.

References

1. JAGM VD. Digital divide: impact of access. The International Encyclopedia of Media Effects 2017:1-11. [doi: 10.1002/9781118783764.wbieme0043]
2. Pew Research Center: Internet, Science & Technology. 2018. Internet/broadband fact sheet URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2019-04-29] [WebCite Cache ID 71eNMok3L]
3. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
4. 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]
5. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21th Century. Washington, DC: The National Academies Press; 2001:1-6.
6. Centers for Disease Control and Prevention. 2018. Public Health and Promoting Interoperability Programs (formerly, known as Electronic Health Records Meaningful Use) URL: https://www.cdc.gov/ehrmeaningfuluse/introduction.html [accessed 2019-04-29] [WebCite Cache ID 71eNcePLQ]
7. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the Internet as a resource for health information: Results from a California population-based study. Int J Med Inform 2017 Dec;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
8. Lyles CR, Sarkar U. Health literacy, vulnerable patients, and health information technology use: where do we go from here? J Gen Intern Med 2015 Mar;30(3):271-272 [FREE Full text] [doi: 10.1007/s11606-014-3166-5] [Medline: 25588688]
9. 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;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
10. Levy H, Janke AT, Langa KM. Health literacy and the digital divide among older Americans. J Gen Intern Med 2015;30(3):284-289 [FREE Full text] [doi: 10.1007/s11606-014-3069-5] [Medline: 25387437]
11. Gonzalez M, Sanders-Jackson A, Emory J. Online Health Information-Seeking Behavior and Confidence in Filling Out Online Forms Among Latinos: A Cross-Sectional Analysis of the California Health Interview Survey, 2011-2012. J Med Internet Res 2016 Dec 04;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]
12. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The Digital Divide and Patient Portals: Internet Access Explained Differences in Patient Portal Use for Secure Messaging by Age, Race, and Income. Med Care 2016 Dec;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
13. Cardona-Grau D, Bush R, Din H, Richardson A, Kuelbs C, Chiang G. MP76-14 Patient portal usage in pediatric urology: is it meaningful use to everyone? Journal of Urology 2017;197(4S):1. [doi: 10.1016/j.juro.2017.02.2142]
14. UCLA Center for Health Policy Research. 2017. Get CHIS Data (CHIS 2016 Adult Survey) URL: http://healthpolicy.ucla.edu/chis/data/Pages/GetCHISData.aspx [accessed 2019-04-29] [WebCite Cache ID 71eNl7z4z]
15. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiology (Sunnyvale) 2016 Apr;6(2):277 [FREE Full text] [doi: 10.4172/2161-1165.1000227] [Medline: 27274911]
16. Cumming G. A method for testing a prespecified subgroup in clinical trials. Stat Med 2009;28:205-220. [doi: 10.1002/sim]
17. Schenker N, Gentleman J. On judging the significance of differences by examining the overlap between confidence intervals. Am Stat 2001;55(3):182-186.
18. Explanation of Data Standards for Race, Ethnicity, Sex, Primary Language, and Disability. URL: https://minorityhealth.hhs.gov/assets/pdf/checked/1/Explanation_of_Draft_Standards.pdf [accessed 2019-04-29] [WebCite Cache ID 77zrZ00lP]
19. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
20. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011 Feb 16;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
21. United States Census Bureau. 2017. The digital divide: by internet, computer, race & hispanic origin URL: https://www.census.gov/library/visualizations/2017/comm/internet.html [accessed 2019-04-29] [WebCite Cache ID 71eNtkLAa]
22. Perrin A. Pew Research Center. 2018. Smartphones help blacks, Hispanics bridge some-but not all-digital gaps with whites URL: http://www.pewresearch.org/fact-tank/2017/08/31/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ [accessed 2019-04-29] [WebCite Cache ID 71eNyHRJS]
23. Finney RLJ, Agunwamba AA, Wilson P, Chawla N, Vieux S, Blanch-Hartigan D, et al. Cancer-Related Information Seeking Among Cancer Survivors: Trends Over a Decade (2003-2013). J Cancer Educ 2016 Jun;31(2):348-357. [doi: 10.1007/s13187-015-0802-7] [Medline: 25712202]
24. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. New Media & Society 2013;16(3):507-526. [doi: 10.1177/1461444813487959]

25. López L, Green AR, Tan-McGrory A, King R, Betancourt JR. Bridging the digital divide in health care: the role of health information technology in addressing racial and ethnic disparities. Jt Comm J Qual Patient Saf 2011;37(10):437-445. [Medline: 22013816]
26. Sorondo B, Allen A, Fathima S, Bayleran J, Sabbagh I. Patient Portal as a Tool for Enhancing Patient Experience and Improving Quality of Care in Primary Care Practices. EGEMS (Wash DC) 2016;4(1):1262 [FREE Full text] [doi: 10.13063/2327-9214.1262] [Medline: 28203611]
27. Health communication and health information technology. 2018. Healthy People 2020 URL: https://www.healthypeople.gov/2020/topics-objectives/topic/health-communication-and-health-information-technology/objectives [accessed 2018-08-11] [WebCite Cache ID 71eO589gT]

Abbreviations

CHIS: California Health Interview Survey
eHealth: electronic health
NH: Non-Hispanic
OR: odds ratio

Edited by G Eysenbach; submitted 14.08.18; peer-reviewed by C Almario; comments to author 14.10.18; revised version received 14.03.19; accepted 31.03.19; published 14.05.19.

Please cite as:
Din HN, McDaniels-Davidson C, Nodora J, Madanat H
Profiles of a Health Information–Seeking Population and the Current Digital Divide: Cross-Sectional Analysis of the 2015-2016 California Health Interview Survey
J Med Internet Res 2019;21(5):e11931
URL: https://www.jmir.org/2019/5/e11931/
doi:10.2196/11931
PMID:31094350

Original Paper

Assessment of Use and Preferences Regarding Internet-Based Health Care Delivery: Cross-Sectional Questionnaire Study

Georgios Paslakis[1,2], MD; Josefine Fischer-Jacobs[3], MSc; Lars Pape[4,5], MD; Mario Schiffer[6,7], MD; Raoul Gertges[5]; Uwe Tegtbur[5,8], MD; Tanja Zimmermann[3], PhD; Mariel Nöhre[3,5], MD; Martina de Zwaan[3,5], MD

[1]University Health Network, Toronto General Hospital, Toronto, ON, Canada
[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[3]Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School, Hannover, Germany
[4]Department of Pediatric, Liver and Metabolic Diseases, Hannover Medical School, Hannover, Germany
[5]Project Kidney Transplantation 360°, Hannover Medical School, Hannover, Germany
[6]Department of Nephrology and Hypertension, Hannover Medical School, Hannover, Germany
[7]Department of Nephrology, University Hospital, Friedrich-Alexander University, Erlangen, Germany
[8]Department of Sports Medicine, Hannover Medical School, Hannover, Germany

Corresponding Author:
Martina de Zwaan, MD
Department of Psychosomatic Medicine and Psychotherapy
Hannover Medical School
Carl-Neuberg-Straße. 1
Hannover,
Germany
Phone: 49 49511532 ext 6570
Fax: 49 511 5323190
Email: dezwaan.martina@mh-hannover.de

Abstract

Background: There has been an incremental increase in the use of technology in health care delivery. Feasibility, acceptability, and efficacy of interventions based on internet technologies are supported by a growing body of evidence.

Objective: The aim of this study was to investigate use and preferences in the general adult population in Germany for remote, internet-based interaction (eg, email, videoconferencing, electronic medical records, apps).

Methods: A nationwide cross-sectional questionnaire survey in adults that was representative in terms of age, sex and educational level was carried out.

Results: A total of 22.16% (538/2428) of survey participants reported not using the internet for work or private use. The nonuser phenotype can be described as being older, having lower educational and income status, and living in less populated areas. The majority of participants within the cohort of internet users reported that they would not consider using electronic medical records (973/1849, 52.62%), apps (988/1854, 53.29%), or emails to report symptoms (1040/1838, 56.58%); teleconference with one (1185/1852, 63.98%) or more experts (1239/1853, 66.86%); or participate in video psychotherapy (1476/1853, 79.65%) for the purpose of medical consultation or treatment. Older age and lower educational level were the most robust predictors of assumed future denial of use.

Conclusions: Our results point toward low use and preference rates among the general population for the use of telemedicine. It also seems that those who might benefit from telemedical interventions the most, are, in fact, those who are most hesitating. These low use and preference rates of eHealth should be considered prior to designing and providing future telemedical care, supporting the need for easy-to-use, data secure solutions.

(J Med Internet Res 2019;21(5):e12416) doi:10.2196/12416

KEYWORDS

telemedicine; health care delivery; internet; eHealth; teleconference; electronic medical records; online psychotherapy

Introduction

Background

With technological innovations permeating all aspects of life, there has been an incremental increase in the use of technology in health care delivery. Feasibility, acceptability, and efficacy in terms of symptom improvement or improvements in quality of life using interventions based on internet technologies are supported by a growing body of evidence [1-4]. Patient access to online tools implemented to address chronic conditions (diabetes mellitus, asthma, etc) has improved patient self-care [5-7], with a main advantage being self-care delivery in low- and middle-income countries [8-10]. In order to best exploit the advantages while simultaneously avoiding the pitfalls, the US government has issued a mandate for the appropriate use of health care technology [11].

Beside the implementation of internet-based interventions used by patients on their own independent of any patient-physician interaction, the actual (mostly real time, depending on definition) practice of medicine through the internet involving a physician-patient interaction, known as telemedicine, presents an opportunity to revolutionize health care delivery on a global scale [12]. While early work has focused on the delivery of telemedicine services on an organizational level (eg, clinic-to-clinic teleneurology or telestroke [13]), more recently telemedicine has expanded to the delivery of health care services directly to a patient's home. Frequently called direct-to-consumer care, this kind of telemedicine, in which a patient interacts with a physician or another health care specialist via email or videoconferencing, for example, is growing rapidly [14,15].

In many countries, direct-to-consumer telehealth companies offer patients with minor illnesses around-the-clock access to a physician. In Germany, health insurance companies offer telephone or guided internet support for specific illnesses such as depression or tinnitus, while an increasing financial grant support is dedicated to the investigation of the potential benefits of different types of telemedicine services [16]. In other European countries, for example Estonia, electronic medical records have already been implemented. Patient-accessible electronic health records are thought to increase patient involvement in their own health care matters as access to information increases [17]. Still, direct-to-consumer telemedicine requires patients to use potentially unfamiliar technology, and data security aspects may restrict its use in view of privacy and liability concerns. It is therefore important to examine the perspective of the population toward telemedicine prior to the implementation of definite solutions. On the provider side, costs for infrastructure and the mobilization of resources (eg, need for trainings) are some of the potential limiting factors for the implementation of telemedicine services. The economic advantages of such internet-based services remain unclear; also, even though remote consultation or treatment may be less expensive than personal visits, new use may not decrease overall health care spending [18].

Need for the Study

In order to facilitate a safe, timely, efficient, effective, and equitable delivery of internet-based health care, assessment of needs of all stakeholders involved is crucial [19]. Most studies so far have intended to shed light on the provider point of view, mostly on an institutional level, in developed and developing countries ([20], systematic review in [21]), while the recipient point of view has been scarcely examined, mostly in patient cohorts with very specific needs living with chronic conditions [22,23]. Early efforts to implement the electronic health card and corresponding telematic infrastructure are in progress in Germany, and some health insurance companies have established electronic medical records, but German use and preferences regarding internet-based health care delivery has not been assessed. Therefore, in a hypothesis-free manner, the aim of this study was to investigate use and preferences of the general adult population in Germany by means of a questionnaire about remote, internet-based interaction (email, videoconferencing, electronic medical record, apps) with medical professionals. The goal was to gather data and stratify them according to sociodemographic variables and in the light of benefit-related aspects, thus serving as a basis for the planning and implementation of future telemedical solutions.

Methods

Recruitment

Participants in the study were part of a larger cross-sectional survey on physical and mental well-being in a random sample of German residents aged 14 years and older (range 14 to 91 years). The demographic consulting company USUMA (Unabhängiger Service für Umfragen, Methoden und Analysen) assisted with sampling and data collection. The procedure was designed to yield a nationwide sample representative in terms of age, sex, and educational level over the fieldwork period from May to July 2018. Sociodemographic data were collected in person by trained interviewers. In addition, participants returned a battery of self-report questionnaires including the telemedicine questions. This study was part of a larger survey assessing different issues in the general population for research purposes in Germany. For the purposes of this study, we assessed adults only. Thus, only participants aged 18 years and older were included in the analyses.

Data Acquisition

In Germany, no directory is generally available containing the addresses of all private households or individuals that can be used by market research agencies as a sampling frame. The data collected by the local authorities are only available for surveys considered to be of major public interest.

The consortium Arbeitsgemeinschaft ADM-Stichproben closes this gap by providing a sampling frame, the ADM Sampling System for Face-to-Face Surveys, to member agencies. The demographic consulting company, USUMA, supporting this study is a member agency and has access to this sampling system [24]. This frame allows representative face-to-face samples to be drawn for all households in Germany and for all people living in those households. The main statistical data are provided on a detailed level for this population.

The ADM Sampling System is organized as an area sample comprising all populated areas in Germany, organized by state, county, and community with the statistical areas within communities described by public data and the geographical data taken from traffic navigation systems. Taken together, the area sample consists of about 53,000 areas, each containing a minimum of 350 and an average of 700 private households. All areas were first regionally stratified resulting in approximately 1500 strata. Next, 128 nonoverlapping nets were randomly extracted containing a total of 258 areas across Germany. These 258 areas were drawn proportionally to the distribution of private households. Since the sampling is done randomly in three steps (first step: stratified drawing of a sample point system after random allocation, second step: random-walk household selection procedure, third step: Kish-Selection-Grid method for randomly selecting the target person within the household), this method for face-to-face surveys is based entirely on random sampling and fully meets the scientific requirements regarding randomization based on statistical theory [24].

The participation rate was 47.3%, taking into account all refusals to participate as well as interviews that failed to take place due to respondent illness or being otherwise unavailable during the fieldwork. All participants provided their written informed consent in accordance with the Helsinki declaration. The study was approved by the ethics committee of the University of Leipzig.

The following sociodemographic data were assessed: Sex (male and female), age (distinguished according to groups: 18-24, 25-34, 35-44, 45-54, 55-65, >65 years), educational level (<12 and ≥12 years of education), monthly income (0 to <1000, 1000-2500, and ≥2500 euros/month), population size (<5000, 5000-50,000, and ≥50,000 residents).

Participants were first asked about their internet use, email use, and use of videoconference technologies (eg, Skype) in general. Participants were asked to choose between never, rarely (sporadic), frequently (on single days of the week), regularly (on most days of the weeks), and daily (every day) for internet use and yes or no for email and videoconference use.

Participants who reported at least sporadic use of the internet (internet users) were then asked a series of 7 pairs of questions about the medical context of consultation or treatment. Questions were constructed and chosen from a larger pool by physicians and psychologists at the Department of Psychosomatic Medicine and Psychotherapy at Hannover Medical School who have previously been involved in internet-based studies [16,25-27]. Affirmative (yes) answers about the use of internet-based health care delivery were considered to reflect preferences, while no answers to the same questions were considered indicative of reluctance. All questions were short, target-oriented, and simple:

1. Would you use/have you used email to schedule visits with your physician?
2. Would you use/have you used email to report symptoms to your physician?
3. Would you use/have you used videoconferencing with your physician?
4. Would you use/have you used videoconferencing with more than one physician (eg, general practitioner and specialist) at the same time?
5. Would you use/have you used videoconferencing for psychotherapy?
6. Would you use/have you used electronic medical records you can access at any time to see your exam results and leave messages?
7. Would you use/have you used an app that offers personalized information about your condition and recommends exercises and support?

Statistical Analyses

Statistical analyses were performed using SPSS Statistics for Windows version 25.0 (IBM Corp). Analyses of variance or *t* tests for comparisons between groups were performed appropriately with sociodemographic data as between-subject factors or independent variables, respectively. In order to examine the predictive value of the independent variables, binary logistic regressions and multiple linear regressions were performed for dependent variables consisting of two or more than two categories, respectively. The level of significance was set at $P \leq .05$, but a Bonferroni correction for multiple testing was performed according to the number of independent variables in each hypothesis testing.

Results

Cohort

A total of 2516 individuals participated in the study. Of those, 77 were excluded for being younger than 18 years. Thus, data from a total of 2439 adults were analyzed. This cohort consisted of 45.10% (1100/2439) males and 54.90% (1339/2439) females. Mean age was 49.04 (SD 16.87) years, and 40.01% (976/2439) of participants were aged older than 55 years. Almost 4 out of 5 participants (1926/2439, 78.97%) had less than 12 years of education, with 27.71% (662/2439) belonging to a low-income group and the majority (1487/2439, 62.24%) having an average income of 1000 to <2500 euros per month. More details on sociodemographics of the cohort are shown in Table 1.

Table 1. Sociodemographics of study participants as a cohort.

Variables	Survey participants, n (%)
Sex (n=2439)	
Male	1100 (45.10)
Female	1339 (54.90)
Age in years (n=2439)	
18-24	196 (8.04)
25-34	372 (15.25)
35-44	423 (17.34)
45-54	472 (19.36)
55-64	483 (19.80)
65+	493 (20.21)
Education, years (n=2439)	
<12	1926 (78.97)
≥12	513 (21.03)
Income (euros/month) (n=2389)	
0 to <1000	662 (27.71)
1000 to <2500	1487 (62.24)
≥2500	240 (10.05)
Population (n=2439)	
<5000	351 (14.39)
5000 to <50,000	1028 (42.15)
≥50,000	1060 (43.46)

Internet, Email, and Videoconference Use in General

When asked if they used the internet in general for work or for private purposes, 2428 responded (11 participants with missing data). Of these, 22.16% (538/2428) denied using the internet (nonusers), while 77.84% (1890/2428) said they used the internet for work or in private (internet users); 43.2% use the internet on a daily basis.

Among the 2428 respondents, there was no difference in the frequency of internet use in general by sex (P=.30). However, age group revealed a significant main effect ($F_{5,2422}$=189.55, P<.001), with linearly declining internet use frequency parallel to increasing age. Similarly, a significant main effect was found for educational level (P<.001), with lower rates of internet use frequency in individuals with less than 12 years compared to those with 12 years or more of education. In fact, there was a significant interaction between age group and educational level ($F_{5,2416}$=10.32, P<.001) (Figure 1). Higher income corresponded to higher internet use frequency ($F_{2,2375}$=25.33, P<.001; Games-Howell post hoc: highest income group > middle-income group = low-income group) (data not shown). Finally, there was also a significant main effect for population size ($F_{2,2425}$=11.80, P<.001; Games-Howell post hoc: small communities/cities < middle size cities < big cities), with higher rates of internet use in larger cities (data not shown).

In a multiple linear regression to predict internet use frequency based on sex, age, educational level, income, and population size, a significant regression equation was found ($F_{5,2372}$=213.71, P<.001), with an R^2 of .31. Age, educational level, income (all P<.001), and population size (P=.001) were significant predictors of internet use frequency and remained significant even after Bonferroni correction with the new level of significance set at .05/5=.01. However, sex was not a significant predictor of internet use frequency (Table 2).

Figure 1. Internet use depending on age and educational level in a survey of adults from the general population. A significant interaction between age and educational level is found.

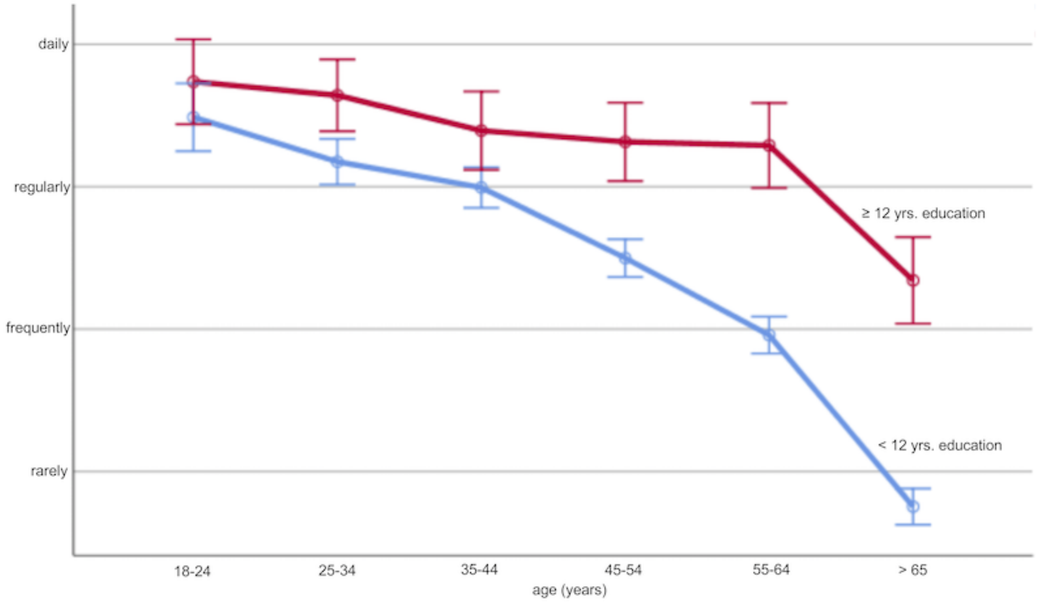


Table 2. Linear regression analysis to predict internet use based on sex, age, income, population size, and educational level.

Variable	Nonstandardized coefficient		Standardized coefficient beta	*t*-value	*P* value
	Regression coefficient B (SE[a])	95% CI			
Constant	4.09 (.18)	3.74 to 4.44		22.81	<.001
Sex	.11 (.06)	–0.002 to 0.23	.03	1.93	.054
Age	–.50 (.02)	–0.54 to –0.47	–.49	–28.17	<.001
Education	.63 (.07)	0.49 to 0.77	.16	8.94	<.001
Income	.39 (.05)	0.29 to 0.49	.14	7.66	<.001
Population	.14 (.04)	0.06 to 0.22	.06	3.45	.001

[a] SE: standard error.

Internet Users

Comparing internet users and nonusers, no differences in sex distribution or population size of the community or city of origin were found. On the other hand, age distribution, educational level, and income differed (eg, more than half of nonusers were aged 65 years and older, most of them had an educational level of less than 12 years, and there more nonusers belonging to the low-income group compared to internet users). More details are shown in Table 3.

Within the group of internet users, 6.85% (129/1884, 6 missing) stated they did not use email, while 93.15% (1755/1884) did. Within the same group of internet users, the majority (58.71%, 1089/1855, 35 missing) did not use videoconference programs, and 41.29% (766/1855) were familiar with their use.

Internet Users Versus Nonusers

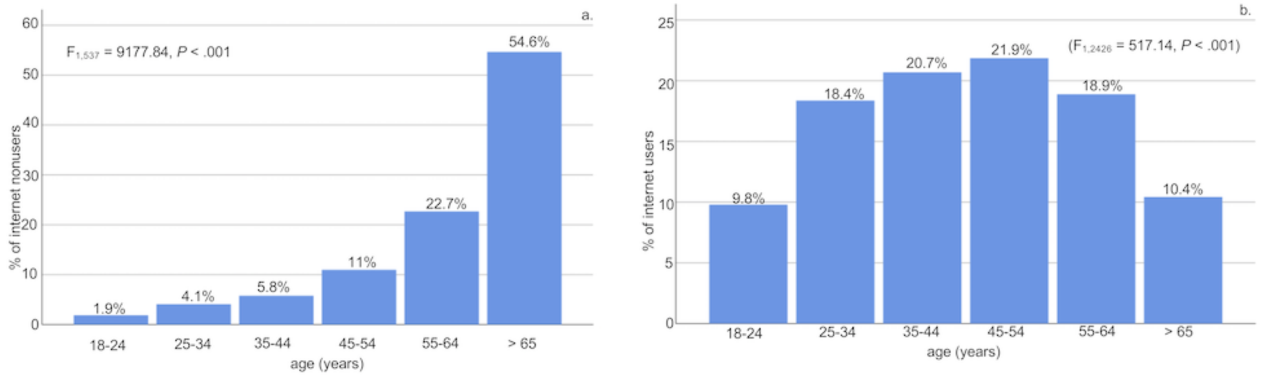
Within the group of internet users, age had a significant main effect ($F_{1,1889}$=10,664.20, *P*<.001) and educational level was significantly different with regard to the total years of education (*P*<.001). Similarly, age had a significant main effect ($F_{1,537}$=9177.84, *P*<.001) and educational level was significantly different with regard to the total years of education (*P*<.001) within the group of nonusers. These results are displayed in Figures 2 and 3.

Table 3. Sociodemographic characteristics of internet users and nonusers.

Variables	Internet users, n (%)	Internet nonusers, n (%)	df[a]	Chi-square	*P* value
Sex (n=1890)			**1**	**0.6**	**.44**
Male	861 (45.56)	235 (43.68)			
Female	1029 (54.44)	303 (56.32)			
Age in years (n=1890)			**5**	**576.17**	**<.001**
18-24	185 (9.79)	10 (1.86)			
25-34	347 (18.36)	22 (4.09)			
35-44	391 (20.69)	31 (5.76)			
45-54	413 (21.85)	59 (10.97)			
55-64	357 (18.89)	122 (22.67)			
65+	197 (10.42)	294 (54.65)			
Education in years (n=1890)			**1**	**101.17**	**<.001**
<12	1410 (74.60)	509 (94.61)			
≥12	480 (25.40)	29 (5.39)			
Income in euros per month (n=1845)			**2**	**28.9**	**<.001**
0 to <1000	486 (26.34)	169 (31.70)			
1000 to <2500	1141 (61.84)	342 (64.17)			
≥2500	218 (11.82)	22 (4.13)			
Population (n=1890)			**2**	**3.33**	**.19**
<5000	261 (13.81)	88 (16.35)			
5000 to <50,000	790 (41.80)	231 (42.94)			
≥50,000	839 (44.39)	219 (40.71)			

[a] df: degree of freedom.

Figure 2. Distribution of age groups in noninternet users (a) compared to internet users (b). Within each of the groups users versus nonusers, age groups had a significant main effect. Age groups also differed between the groups.



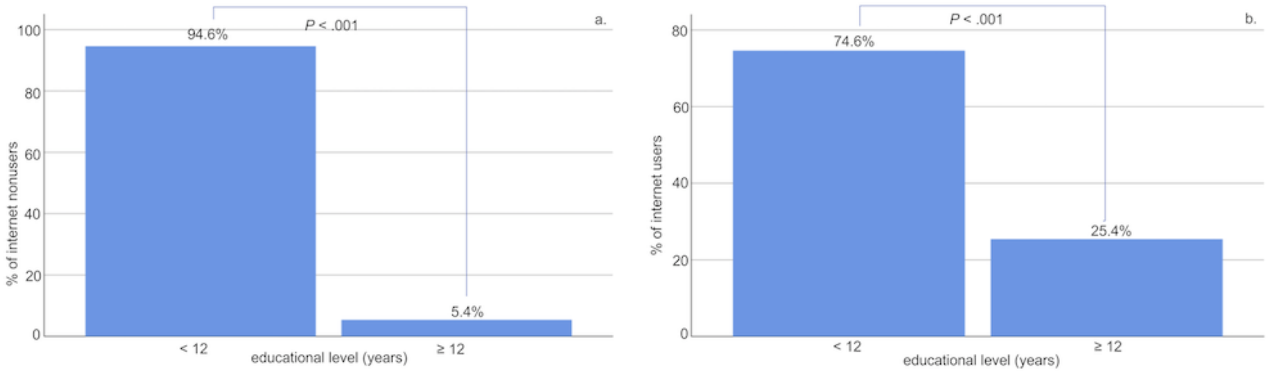


Figure 3. Comparisons of educational levels in noninternet users (a) compared to internet users (b). Within each of the groups users versus nonusers, educational level was significantly different with regard to the total years of education. Educational level also differed between the groups.

Telemedicine Within the Medical Consultation and Treatment Context

The 1890 internet users were asked whether they would consider using email, videoconferencing, electronic patient charts, or apps within a medical context of consultation or treatment in the future and whether they have already made use of any of these options.

The majority of participants answered that they would take advantage of using email for scheduling medical visits (1257/1890, 66.51%), although the majority of participants have not used email for this purpose so far (1247/1890, 66.00%). In contrast, most participants would not use email for reporting symptoms to their physicians (1040/1890, 55.03%), and few participants (64/1890, 3.39%) have done this in the past.

The majority of internet users also proved to be reluctant to use videoconferencing for consultations with their physician (1185/1836, 64.54%), and only 0.76% (14/1836) reported having had experience with it. Similarly, consultation with more than one physician at once by means of videoconference would not be considered by most internet users (1239/1836, 67.48%). Only 1.35% (19/1836) had actual experience with psychotherapy using videoconference technology, and most users reported low preference rates for online video-psychotherapy (377/1836, 20.53%). Electronic patient charts were not familiar to most users (1547/1836, 84.26%), but a slight majority would consider electronic patient charts as a future option (973/1836, 53.00%). Almost half of internet users would use a medical app (866/1836, 47.17%); however, just 1.85% (34/1836) have already done so.

In Table 4, more details about the results concerning the willingness of adult internet users to consider the use of internet technologies within a medical context of consultation or treatment as well as their actual experiences with technologies of this kind within this specific context are displayed.

Next, binary logistic regression analyses were performed for the use of internet technology within a medical context based on sex, age, educational level, income, population size, and previous experience with the respective online tool under investigation. Due to the insertion of 6 independent variables in the regression equation, the new level of significance was set at .05/6=.008.

Regarding email use for scheduling future medical visits, a significant regression model was found, χ^2_{12}=296.4, *P*<.001. Age, educational level, and previous experience with email for this purpose were significant predictors (*P*<.001) as was population size (*P*=.007). Sex and income (following Bonferroni correction) were not significant predictors.

For using email for reporting symptoms, the logistic regression model turned out to be significant, χ^2_{12}=137.5, *P*<.001. Age, educational level, and previous email use for reporting symptoms were significant predictors (all *P*<.001). Sex, income, and population size were not significant predictors at the new level of statistical significance.

For videoconference use to communicate with a physician, the regression model was significant, χ^2_{12}=62.8, *P*<.001. Age (*P*=.03), educational level (*P*<.001), and income (*P*=.005) were significant predictors whereas sex, population size, and previous experience with teleconference were not.

The regression equation for the possibility of videoconference use with more than one physician at once turned out to be significant χ^2_{12}=77.0, *P*<.001. Age (*P*=.001), educational level (*P*<.001), and previous experience with group teleconference in a medical context (*P*=.008) were significant predictors. Sex, income and population size were not significant predictors.

The regression model for prediction of psychotherapy to be considered by means of videoconference also proved to be significant χ^2_{12}=60.1, *P*<.001. Here, only educational level (*P*=.002) and previous experience with online video psychotherapy (*P*<.001) were significant predictors.

Preferences regarding the use of electronic patient files were statistically predicted by age (*P*=.001), educational level (*P*<.001) as well as previous use of electronic patient files (*P*=.006) in a significant regression equation χ^2_{12}=77.0, *P*<.001.

Table 4. Willingness of adult internet users (n=1836) to consider the use of internet technologies within a medical context of consultation or treatment and their actual experiences with technologies of this kind within this specific context.

Variable	Yes, n (%)	No, n (%)
Would use email to schedule visits	1257 (68.46)	579 (31.54)
Have used email to schedule visits	358 (22.31)	1247 (77.69)
Would use email to report symptoms	798 (43.42)	1040 (56.58)
Have used email to report symptoms	64 (4.04)	1519 (95.96)
Would use videoconferencing with their physician	667 (36.02)	1185 (63.98)
Have used videoconferencing with their physician	14 (0.89)	1561 (99.11)
Would use videoconferencing with more than 1 physician at the same time	614 (33.14)	1239 (66.86)
Have used videoconferencing with more than 1 physician at the same time	16 (1.02)	1559 (98.98)
Would use videoconferencing for psychotherapy	377 (20.35)	1476 (79.65)
Have used videoconferencing for psychotherapy	19 (1.21)	1554 (98.79)
Would use electronic medical records	876 (47.38)	973 (52.62)
Have used electronic medical records	22 (1.40)	1547 (98.60)
Would use apps	866 (46.71)	988 (53.29)
Have used apps	34 (2.20)	1508 (97.80)

Finally, considering app use within a medical context was once more statistically predicted by age and educational level (both $P<.001$) as well as previous app experience ($P=.001$) in a significant regression model $\chi^2_{12}=90.5$, $P<.001$.

In all models, younger age, higher educational level, higher income, and previous experience with the internet-based interventions in question were associated with higher rates of preferences for future use of internet-based solutions. All regression models can be found in Multimedia Appendix 1.

Discussion

Internet-based solutions to provide health care consultation or even treatments have been shown to improve patient activation and engagement [28-32] and thus improve outcomes [33-35]. However, as most scientific studies have examined clinical cohorts with very specific needs (eg, patients suffering from chronic diseases), data on use and future preferences in the general population regarding internet-based solutions are scarce [21,36]. Obviously, personal preferences may differ depending on the actual individual need for medical consultation or treatment. The assessment of personal preferences and especially stratification according to specific variables may be important prior to designing and providing internet-based medical consultation or treatment approaches, the purpose being a consistent or repeated rather than a singular or just transient use.

Principal Findings

Even in a country as technologically developed as Germany, it was quite a surprising finding that 22.5% of survey participants reported they did not use the internet for work or private use. In contrast, in a recent survey conducted in 2017 by the Federal Statistical Office (Destatis) in Germany, internet use was above 90% in all age groups except age 65 and older, in which the respective percentage was 55% [37]. We explain this obvious discrepancy with the way the questions were asked; while the Federal Statistical Office asked about lifetime internet use, our questionnaire addressed current ongoing use. According to our results, those individuals with a low preference for the implementation of internet-based health care delivery can be described as being older and having lower educational and income status compared to those individuals who expressed a preference and were younger, more educated, and had a higher income. Since age, educational level, and living in remote (rural) areas are associated with lower utilization rates of medical services and thus with higher morbidity and mortality rates [38-40], it becomes obvious that designing and providing future internet-based telemedical solutions aimed at reaching a wide-ranging and region-wide number of recipients should take the above nonuser phenotype into serious consideration. It would therefore be necessary to develop easy-to-use solutions (eg, tablet-based) suitable for those individuals who are not familiar with technology but would still like to try following simple instructions. For those without access to either the internet or computers and tablets, further solutions are needed (eg, internet hot spots, supply of tablets to the elderly or similar).

In addition, and also rather surprisingly, the majority of participants within the subcohort of internet users reported that they would not consider the use of email, videoconference, video psychotherapy, electronic medical records, or apps for the purpose of medical consultation or treatment, with the exception of using emails to schedule medical visits. This finding may point toward the fact that internet use has not been associated with the idea of a tool helping to address medical issues and needs within the same context by the general population—not even by its very own regular consumers. There is a mismatch between everyday use of internet-based technologies (email, apps, skype, etc) and willingness to use the exact same technologies for the sake of delivering health care. This is important to consider prior to implementing internet-based health care delivery solutions in Germany. Many individuals may not be able to anticipate how such systems would function with success. Data security may also be an issue; a recent market research study in Germany revealed that 95.2% of all respondents indicated fearing a possibly fraudulent use of their personal data on the internet [41]. With older age and lower educational level in our study being the most robust predictors of assumed future use, it seems as those who are supposed to benefit from telemedical interventions the most, are, in fact, those who might not use it. Thus, the development of simple, secure systems as described above might be able to overcome these obstacles.

This is a study examining use and preferences regarding a variety of internet-based technologies for health care delivery in a large representative sample in Germany. A limited number of similar studies have been performed in convenience samples, with the associated risks of sampling error and lack of representativeness inherent in convenience samples. Apolinário-Hagen et al [42] performed a Web-based study in a convenience sample (N=646) examining general preferences with reference to internet-based therapy and found similar rates of interesting in psychotherapy by means of videoconference (22.8% compared to 20.3% in our cohort). Interestingly, attachment avoidance and stress were associated with preference to internet-based interventions, and those individuals who were well aware of internet-based approaches showed a higher preference for therapist-guided internet treatment. The authors highlight the importance of increasing public knowledge about internet interventions in order to promote acceptance and uptake [42]. In another convenience sample in German and Austrian citizens (N=496), Hoerbst et al [43] examined knowledge, expectations, fears, and barriers toward electronic health records. In contrast to our findings of 47.4% of respondents being open to the idea of electronic medical records, they found more than 80% of respondents being supportive of the idea, although data protection issues were the major concerns. Our results showing low interest rates among the general population for internet-based technologies for telemedical purposes are in accordance with the results by Jenssen et al [44], in a US national survey, showing that despite regular use of new digital technology, few participants would consider the use of these tools for communicating with their physicians. In another US nationwide survey assessing participant preferences for telemedicine, Welch et al [45] found that 41% percent of survey participants felt it to be unimportant if their current health care professional offered telemedicine, only 15% would consider changing to a new professional who would provide telemedical care, and 56% of participants felt it was important to have an established relationship with a physician before considering telemedical visits. White race and higher education and relatively younger age and higher income have been described as predictors for internet use and eHealth literacy among older adults (aged 55 years and older) who were patients at clinics serving low-income populations [46]; policy makers are asked to consider such findings prior to the implementation of eHealth programs in order to meet the needs of people living different realities.

In our study, a significantly higher proportion of the population reported interest in using new telemedical technologies compared to the small minority that reported actual previous experience. Prior use was dependent on whether such technology-based solutions were provided by their practitioners and whether the need for such a use had actually existed in the past. This fact could be interpreted as insufficient exposure to the respective technologies so far; in other words, specific offers might increase specific demand. If recipients of telemedical care solutions were to experience short-term advantages of such systems (eg, being able to schedule appointments online at any time of the day, being able to see examination results and have easy access to further personal medical data) in an uncomplicated manner combined with maximum security of the systems, higher levels of actual use would be expected. Countries such as Estonia have shown that user-friendly internet-based health solutions are accepted by the majority of the population. It may therefore be expected that the implementation of simple-to-use, data-secure systems in Germany might lead to much higher use rates than shown here.

Another survey in Germany investigated preferences for online interventions in psychiatry and psychotherapy between health care professionals and nonprofessionals and found that, overall, nonprofessionals were more skeptical in their ratings than professionals [47]. More favorable eHealth utilization rates by European general practitioners were associated with younger age of the practitioners, female sex, and specific working conditions (eg, self-employment) in the study by Torrent-Sellens et al [48]. Although assessment of preferences of the medical staff involved was not part of our survey, aspects such as changes in roles and responsibilities and need for additional resources, reimbursement, and training may be crucial for successful implementation of new health care delivery approaches [49] and should be targets of investigation in further studies.

Blended treatments—the use of online tools (eg, self-help) combined with face-to-face psychotherapy—are already part of clinical routines aimed at increasing the impact of psychotherapy [50]. Nevertheless, computerized treatments and mobile phone apps for mental health problems seem to be negatively viewed and the likelihood of their future use to be rated low, indicating that policy makers need to improve the public perception of such options in order to facilitate their dissemination [51]. In a recent study in Austrian psychotherapists, both recipients of internet-based psychotherapy and their providers associated the new modes of treatment with more disadvantages and risks [52]. Again, assessment of barriers to and facilitators of use for remote psychotherapy is necessary to develop tailored, feasible, and acceptable practice designs for all stakeholders involved.

Strengths and Limitations

Strengths of this study clearly include the large number of participants and representativeness of the cohort. However, our survey also has limitations. First, we asked the general population to express their opinions on the use of email, videoconferencing, apps, and electronic patient files, all of which are internet-based instruments intended to deliver health care.

The assessment of differences between these instruments was not part of our protocol, which aimed at assessing global preferences for a variety of new technologies that may provide future meaningful contributions to improve health care delivery in the general population. In addition, the reasons for or against considering the use of internet-based telemedicine interventions were not assessed (eg, fear of dataveillance, needs remaining unmet) and may only be assumed. Other predictors should also be investigated in future studies (eg, ethnic/cultural background, strength of the physician-patient relationship); in addition, reasearch has shown that a health-related information seeking personality was a significant positive predictor of the willingness to undergo online treatment [53]. Furthermore, we did not use a validated questionnaire but a series of questions generated by experts in psychosomatic medicine. Presumably, although not explicitly assessed, a minority of the participants has been offered the use of the technologies under investigation within a medical context, as such systems are not yet fully developed or implemented in Germany in routine practice.

Conclusion

Among internet users, the majority reported that they would not consider using internet technology for the purposes of medical consultation or treatment. Internet-based health care approaches using the right tools aimed at the right recipients may prove to be of great benefit. Our results emphasize the importance of developing solutions able to convince the general population to make proper use of and benefit from the potential telemedicine has to offer. As a putative practicable future way to increase acceptance and promote implementation of internet-based approaches in routine care, it is crucial to provide continuous, large-scale information on newer technological possibilities in health care delivery and ensure that concerns (eg, data security) are adequately addressed. In addition, policy makers should be focus on outreach to the elderly and those with lower income and educational level, as it is those groups who are more likely to report lower use of and preferences for internet-based health care delivery. Such an approach includes easy-to-use solutions and facilitation of access to internet-based services.

Acknowledgments

This study is supported in part by a grant from the Federal Joint Committee of the Federal Republic of Germany (number 01NVF16009).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Logistic regression analyses for the prediction of attitudes toward internet-based health care delivery based on 6 predictors: sex, age, educational level, income, population size, and previous experience.

[DOCX File, 68KB - jmir_v21i5e12416_app1.docx]

References

1. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
2. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
3. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
4. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
5. Al-Durra M, Torio M, Cafazzo JA. The use of behavior change theory in Internet-based asthma self-management interventions: a systematic review. J Med Internet Res 2015;17(4):e89 [FREE Full text] [doi: 10.2196/jmir.4110] [Medline: 25835564]
6. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized hypertension management using patient-generated health data integrated with electronic health records (EMPOWER-H): six-month pre-post study. J Med Internet Res 2017 Sep 19;19(9):e311 [FREE Full text] [doi: 10.2196/jmir.7831] [Medline: 28928111]
7. Saslow LR, Mason AE, Kim S, Goldman V, Ploutz-Snyder R, Bayandorian H, et al. An online intervention comparing a very low-carbohydrate ketogenic diet and lifestyle recommendations versus a plate method diet in overweight individuals

with type 2 diabetes: a randomized controlled trial. J Med Internet Res 2017 Feb 13;19(2):e36 [FREE Full text] [doi: 10.2196/jmir.5806] [Medline: 28193599]
8. Abaza H, Marschollek M. mHealth application areas and technology combinations. A comparison of literature from high and low/middle income countries. Methods Inf Med 2017 Aug 08;56(7):e105-e122. [doi: 10.3414/ME17-05-0003] [Medline: 28925418]
9. Hurt K, Walker RJ, Campbell JA, Egede LE. mHealth interventions in low and middle-income countries: a systematic review. Glob J Health Sci 2016 Dec 01;8(9):54429 [FREE Full text] [doi: 10.5539/gjhs.v8n9p183] [Medline: 27157176]
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. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
12. Combi C, Pozzani G, Pozzi G. Telemedicine for developing countries: a survey and some design issues. Appl Clin Inform 2016 Nov 02;7(4):1025-1050 [FREE Full text] [doi: 10.4338/ACI-2016-06-R-0089] [Medline: 27803948]
13. Dorsey ER, Glidden AM, Holloway MR, Birbeck GL, Schwamm LH. Teleneurology and mobile technologies: the future of neurological care. Nat Rev Neurol 2018 May;14(5):285-297. [doi: 10.1038/nrneurol.2018.31] [Medline: 29623949]
14. Dash J, Haller DM, Sommer J, Junod PN. Use of email, cell phone and text message between patients and primary-care physicians: cross-sectional study in a French-speaking part of Switzerland. BMC Health Serv Res 2016 Dec 05;16(1):549 [FREE Full text] [doi: 10.1186/s12913-016-1776-9] [Medline: 27716256]
15. Fletcher TL, Hogan JB, Keegan F, Davis ML, Wassef M, Day S, et al. Recent advances in delivering mental health treatment via video to home. Curr Psychiatry Rep 2018 Jul 21;20(8):56. [doi: 10.1007/s11920-018-0922-y] [Medline: 30032337]
16. Pape L, de Zwaan M, Tegtbur U, Feldhaus F, Wolff JK, Schiffer L, et al. The KTx360°-study: a multicenter, multisectoral, multimodal, telemedicine-based follow-up care model to improve care and reduce health-care costs after kidney transplantation in children and adults. BMC Health Serv Res 2017 Aug 23;17(1):587 [FREE Full text] [doi: 10.1186/s12913-017-2545-0] [Medline: 28830408]
17. Wass S, Vimarlund V, Ros A. Exploring patients' perceptions of accessing electronic health records: innovation in healthcare. Health Informatics J 2019 Mar;25(1):203-215. [doi: 10.1177/1460458217704258] [Medline: 28457195]
18. de la Torre-Diez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
19. Blumenthal D. Stimulating the adoption of health information technology. N Engl J Med 2009 Apr 9;360(15):1477-1479. [doi: 10.1056/NEJMp0901592] [Medline: 19321856]
20. Strauss AT, Martinez DA, Garcia-Arce A, Taylor S, Mateja C, Fabri PJ, et al. A user needs assessment to inform health information exchange design and implementation. BMC Med Inform Decis Mak 2015 Oct 12;15:81 [FREE Full text] [doi: 10.1186/s12911-015-0207-x] [Medline: 26459258]
21. Yusif S, Hafeez-Baig A, Soar J. e-Health readiness assessment factors and measuring tools: a systematic review. Int J Med Inform 2017 Dec;107:56-64. [doi: 10.1016/j.ijmedinf.2017.08.006] [Medline: 29029692]
22. Duplaga M. A cross-sectional study assessing determinants of the attitude to the introduction of eHealth services among patients suffering from chronic conditions. BMC Med Inform Decis Mak 2015 Apr 19;15:33 [FREE Full text] [doi: 10.1186/s12911-015-0157-3] [Medline: 25927312]
23. Sun N, Rau PP. The acceptance of personal health devices among patients with chronic conditions. Int J Med Inform 2015 Apr;84(4):288-297. [doi: 10.1016/j.ijmedinf.2015.01.002] [Medline: 25655783]
24. ADM sampling system. URL: https://www.adm-ev.de/en/services/the-adm-sampling-system/ [accessed 2018-12-20] [WebCite Cache ID 74nwxFvQN]
25. Teufel M, Schäffeler N, de Zwaan M, Graap H, Zipfel S, Giel KE. Internet use among patients with psychosomatic disorders: what are the health-related demands and needs? J Health Psychol 2011 Oct;16(7):1120-1126. [doi: 10.1177/1359105311401671] [Medline: 21459918]
26. de Zwaan M, Herpertz S, Zipfel S, Tuschen-Caffier B, Friederich H, Schmidt F, et al. INTERBED: internet-based guided self-help for overweight and obese patients with full or subsyndromal binge eating disorder. A multicenter randomized controlled trial. Trials 2012 Nov 21;13:220 [FREE Full text] [doi: 10.1186/1745-6215-13-220] [Medline: 23171536]
27. de Zwaan M, Herpertz S, Zipfel S, Svaldi J, Friederich H, Schmidt F, et al. Effect of internet-based guided self-help vs individual face-to-face treatment on full or subsyndromal binge eating disorder in overweight or obese patients: the INTERBED randomized clinical trial. JAMA Psychiatry 2017 Oct 01;74(10):987-995. [doi: 10.1001/jamapsychiatry.2017.2150] [Medline: 28768334]
28. Haas K, Martin A, Park KT. Text message intervention (TEACH) improves quality of life and patient activation in celiac disease: a randomized clinical trial. J Pediatr 2017 Jun;185:62-67 [FREE Full text] [doi: 10.1016/j.jpeds.2017.02.062] [Medline: 28343658]
29. John ME, Samson-Akpan PE, Etowa JB, Akpabio II, John EE. Enhancing self-care, adjustment and engagement through mobile phones in youth with HIV. Int Nurs Rev 2016 Dec;63(4):555-561. [doi: 10.1111/inr.12313] [Medline: 27653013]

30. Knoerl R, Lee D, Yang J, Bridges C, Kanzawa-Lee G, Lita SG, et al. Examining the impact of a web-based intervention to promote patient activation in chemotherapy-induced peripheral neuropathy assessment and management. J Cancer Educ 2017 Mar 06:1. [doi: 10.1007/s13187-017-1200-0] [Medline: 28265863]
31. Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving hypertension control and patient engagement using digital tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]
32. Solomon M, Wagner SL, Goes J. Effects of a Web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]
33. Naparstek J, Wing RR, Xu X, Leahey TM. Internet-delivered obesity treatment improves symptoms of and risk for depression. Obesity (Silver Spring) 2017 Dec;25(4):671-675 [FREE Full text] [doi: 10.1002/oby.21773] [Medline: 28224713]
34. Peters AE, Keeley EC. Patient engagement following acute myocardial infarction and its influence on outcomes. Am J Cardiol 2017 Nov 01;120(9):1467-1471. [doi: 10.1016/j.amjcard.2017.07.037] [Medline: 28843395]
35. 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]
36. Lee JL, Choudhry NK, Wu AW, Matlin OS, Brennan TA, Shrank WH. Patient use of email, facebook, and physician websites to communicate with physicians: a national online survey of retail pharmacy users. J Gen Intern Med 2015 Jun 24;31(1):45-51. [doi: 10.1007/s11606-015-3374-7] [Medline: 26105675]
37. Destatis.: Federal Statistical Office of Germany URL: https://www.destatis.de/DE/Publikationen/Qualitaetsberichte/EinkommenKonsumLebensbedingungen/Informationsgesellschaft/Informationstechnikhaushalte05.html [accessed 2018-10-02] [WebCite Cache ID 72rfGQOW9]
38. Goldman D, Smith JP. The increasing value of education to health. Soc Sci Med 2011 May;72(10):1728-1737 [FREE Full text] [doi: 10.1016/j.socscimed.2011.02.047] [Medline: 21555176]
39. Montez JK, Berkman LF. Trends in the educational gradient of mortality among US adults aged 45 to 84 years: bringing regional context into the explanation. Am J Public Health 2014 Jan;104(1):e82-e90 [FREE Full text] [doi: 10.2105/AJPH.2013.301526] [Medline: 24228659]
40. van Gaans D, Dent E. Issues of accessibility to health services by older Australians: a review. Public Health Rev 2018;39:20 [FREE Full text] [doi: 10.1186/s40985-018-0097-4] [Medline: 30027001]
41. [Furcht vor Datenmissbrauch ist weit verbreitet Die Studie]. URL: https://www.presseportal.de/pm/80428/3236590 [accessed 2018-10-02] [WebCite Cache ID 72rc2Ednv]
42. 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]
43. Hoerbst A, Kohl CD, Knaup P, Ammenwerth E. Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens. Int J Med Inform 2010 Feb;79(2):81-89. [doi: 10.1016/j.ijmedinf.2009.11.002] [Medline: 20031482]
44. Jenssen BP, Mitra N, Shah A, Wan F, Grande D. Using digital technology to engage and communicate with patients: a survey of patient attitudes. J Gen Intern Med 2016 Jan;31(1):85-92. [doi: 10.1007/s11606-015-3517-x] [Medline: 26385117]
45. Welch BM, Harvey J, O'Connell NS, McElligott JT. Patient preferences for direct-to-consumer telemedicine services: a nationwide survey. BMC Health Serv Res 2017 Nov 28;17(1):784 [FREE Full text] [doi: 10.1186/s12913-017-2744-8] [Medline: 29183372]
46. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2018 Oct 24:733464818807468. [doi: 10.1177/0733464818807468] [Medline: 30353776]
47. Schulze N, Reuter SC, Kuchler I, Reinke B, Hinkelmann L, Stöckigt S, et al. Differences in attitudes toward online interventions in psychiatry and psychotherapy between health care professionals and nonprofessionals: a survey. Telemed J E Health 2018 Nov 09:1. [doi: 10.1089/tmj.2018.0225] [Medline: 30412450]
48. Torrent-Sellens J, Díaz-Chao A, Soler-Ramos I, Saigí-Rubió F. Modeling and predicting outcomes of eHealth usage by European physicians: multidimensional approach from a survey of 9196 general practitioners. J Med Internet Res 2018 Oct 22;20(10):e279 [FREE Full text] [doi: 10.2196/jmir.9253] [Medline: 30348628]
49. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
50. Zwerenz R, Becker J, Knickenberg RJ, Siepmann M, Hagen K, Beutel ME. Online self-help as an add-on to inpatient psychotherapy: efficacy of a new blended treatment approach. Psychother Psychosom 2017;86(6):341-350. [doi: 10.1159/000481177] [Medline: 29131090]
51. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health—attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]

52. Schuster R, Pokorny R, Berger T, Topooco N, Laireiter A. The advantages and disadvantages of online and blended therapy: survey study amongst licensed psychotherapists in Austria. J Med Internet Res 2018 Dec 18;20(12):e11007 [FREE Full text] [doi: 10.2196/11007] [Medline: 30563817]
53. Roettl J, Bidmon S, Terlutter R. What predicts patients' willingness to undergo online treatment and pay for online treatment? Results from a web-based survey to investigate the changing patient-physician relationship. J Med Internet Res 2016 Feb 04;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]

Edited by G Eysenbach; submitted 05.10.18; peer-reviewed by A Zeleke, M Duplaga, S Sawesi, S Zheng; comments to author 29.11.18; revised version received 14.01.19; accepted 08.04.19; published 16.05.19.

Please cite as:
Paslakis G, Fischer-Jacobs J, Pape L, Schiffer M, Gertges R, Tegtbur U, Zimmermann T, Nöhre M, de Zwaan M
Assessment of Use and Preferences Regarding Internet-Based Health Care Delivery: Cross-Sectional Questionnaire Study
J Med Internet Res 2019;21(5):e12416
URL: http://www.jmir.org/2019/5/e12416/
doi:10.2196/12416
PMID:31099338

Original Paper

Inequalities in the Use of eHealth Between Socioeconomic Groups Among Patients With Type 1 and Type 2 Diabetes: Cross-Sectional Study

Anne Helen Hansen[1,2], BSW, MD, PhD; Meghan Bradway[3,4], MBA; Jan Broz[5], MD; Tor Claudi[6], MD; Øystein Henriksen[7], BSW, PhD; Silje C Wangberg[8], PsyD, PhD; Eirik Årsand[3,9], MSc, PhD

[1]Centre for Quality Improvement and Development, University Hospital of North Norway, Tromsø, Norway
[2]Department of Community Medicine, Faculty of Health Sciences, UiT–The Arctic University of Norway, Tromsø, Norway
[3]Norwegian Centre for eHealth Research, University Hospital of North Norway, Tromsø, Norway
[4]Department of Clinical Medicine, University Hospital of North Norway, Tromsø, Norway
[5]Department of Internal Medicine, Second Faculty of Medicine, Charles University, Prague, Czech Republic
[6]Department of Medicine, Nordland Hospital, Bodø, Norway
[7]Faculty of Social Sciences, Nord University, Bodø, Norway
[8]Department of Health and Care Sciences, Faculty of Health Sciences, UiT–The Arctic University of Norway, Narvik, Norway
[9]Department of Clinical Medicine, UiT–The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Anne Helen Hansen, BSW, MD, PhD
Centre for Quality Improvement and Development
University Hospital of North Norway
PO Box 35
Tromsø, 9038
Norway
Phone: 47 91619655
Email: anne.helen.hansen@unn.no

Abstract

Background: The prevalence of diabetes and the use of electronic health (eHealth) are increasing. People with diabetes need frequent monitoring and follow-up of health parameters, and eHealth services can be highly valuable. However, little is known about the use of eHealth in different socioeconomic groups among people with diabetes.

Objective: The aim of this study was to investigate the use of 4 different eHealth platforms (apps, search engines, video services, and social media sites) and the association with socioeconomic status (SES) among people diagnosed with type 1 and type 2 diabetes mellitus (T1D and T2D, respectively).

Methods: We used email survey data from 1250 members of the Norwegian Diabetes Association (aged 18-89 years), collected in 2018. Eligible for analyses were the 1063 respondents having T1D (n=523) and T2D (n=545). 5 respondents reported having both diabetes types and thus entered into both groups. Using descriptive statistics, we estimated the use of the different types of eHealth. By logistic regressions, we studied the associations between the use of these types of eHealth and SES (education and household income), adjusted for gender, age, and self-rated health.

Results: We found that 87.0% (447/514) of people with T1D and 77.7% (421/542) of people with T2D had used 1 or more forms of eHealth sometimes or often during the previous year. The proportion of people using search engines was the largest in both diagnostic groups, followed by apps, social media, and video services. We found a strong association between a high level of education and the use of search engines, whereas there were no educational differences for the use of apps, social media, or video services. In both diagnostic groups, high income was associated with the use of apps. In people with T1D, lower income was associated with the use of video services.

Conclusions: This paper indicates a digital divide among people with diabetes in Norway, with consequences that may contribute to sustaining and shaping inequalities in health outcomes. The strong relationship between higher education and the use of search engines, along with the finding that the use of apps, social media, and video services was not associated with education, indicates that adequate communication strategies for audiences with varying education levels should be a focus in future efforts to reduce inequalities in health outcomes.

(J Med Internet Res 2019;21(5):e13615) doi:10.2196/13615

KEYWORDS

inequalities; eHealth; internet; health care utilization; cross-sectional study; diabetes mellitus, type 1; diabetes mellitus, type 2; education; income; Norway

Introduction

Increasing Prevalence of Diabetes

The prevalence of diabetes is increasing worldwide; it is expected to rise to 642 million cases in 2040 [1]. Global prevalence in adults is estimated at 8.8% [1]. Around 245,000 people have been diagnosed with diabetes in Norway, of whom around 28,000 (11.4%) have type 1 diabetes (T1D) [2]. Most patients do not reach the combined national treatment targets for prevention of complications [3-5].

Increasing Use of Electronic Health

The World Health Organization defines electronic health (eHealth) as “the use of information and communication technologies for health” [6]. The use of eHealth has increased significantly over the past decades, and around 80% of the general population in the United States and Europe conducts health-related searches on the Web [7-10]. It has recently been reported that 87% of people with T1D in Norway used eHealth in 1 or more forms, and 84% of people with T1D had used search engines sometimes or often during the previous year [11]. Most Norwegian households (98%) have internet access [12], 96% of the population aged 16 to 79 years has used the internet during the previous 3 months, and 90% of this population uses the internet every day [13]. Social media is used by 80% of the Norwegian population [13].

Socioeconomic Inequalities in the Prevalence, Morbidity, and Mortality of Diabetes

Despite a relatively high average standard of living, all European countries still have substantial inequalities in health outcomes among socioeconomic groups, as affluent groups have better somatic and mental health and lower mortality than disadvantaged groups [14]. Relative health differences between the highest and lowest socioeconomic groups in Norway even rank among the highest in Europe [14]. The inverse association between socioeconomic status (SES) and the prevalence, morbidity, and mortality of diabetes is well documented [15,16]. In addition, there is evidence of worse health care for diabetes patients with low SES [15,17]. Health care services and individuals’ abilities to take advantage of them are considered parts of the explanations for inequalities in health, even in universal health care systems [17,18].

Socioeconomic Inequalities in the Use of Electronic Health

One might assume that communication inequalities can contribute to inequalities in health, as it is well known that new interventions and treatments reach people in higher socioeconomic groups first [19-21]. Research consistently indicates that women, younger people, and people with middle and high SES are more likely to seek health information and advice from the internet [22-25]. In addition, both long-term illness and good health are reported to be positively associated with eHealth use [8,22]. However, in the case of diabetes, there is evidence that there is no immediate benefit from health technology implementation in lower-education groups, in contrast to medium- and especially higher-education groups, with possible consequences regarding health outcomes [24]. Wangberg et al found that SES is related to differential use of eHealth, as people with higher education use eHealth tools that more likely influence health behaviors [26]. Given access to the internet, the digital health divide implies that some people are less likely to use the internet for health purposes, as well as to benefit from eHealth resources [10,25].

Despite recent reports of a decrease in gender and health disparities in the use of eHealth, persistent predictors of less use of eHealth seem to be higher age (75 years and older), lower education (lower than high school), and (very) low income [9].

Diabetes and Electronic Health

Type 2 diabetes (T2D) is partly caused, maintained, and deteriorated by preventable risk factors, such as physical inactivity, unhealthy diet, obesity, and smoking. In daily life, individuals with diabetes are in charge of managing their disease and self-management, and empowerment is essential in care and prevention of complications of T1D and T2D diabetes. Recent systematic reviews have shown that eHealth can play a positive role in this regard [27-29]. It is a core political ambition to equalize social inequalities in health, as reflected in the Norwegian Public Health Act [30]. To our knowledge, no one has studied the relationship among SES groups in the form of education and household income, and different forms of eHealth used by people with diabetes in Norway. As there is consistent evidence of socioeconomic inequalities in health, as well as inequalities in the distribution of diabetes and in the use of eHealth, studies of possible gaps in information seeking among people with diabetes deserve close attention. This is equally relevant in relation to the importance of information in the follow-up of this prevalent and lifelong chronic disease.

Objective

The aim of this study was to investigate the use of different eHealth platforms among people with diabetes (T1D and T2D) and investigate whether the use of eHealth was associated with SES. Specifically, we tested whether the use of apps, search engines (such as Google), video services (such as YouTube), and social media (such as Facebook) was associated with education and household income, adjusted for gender, age, and self-rated health.

Methods

Data

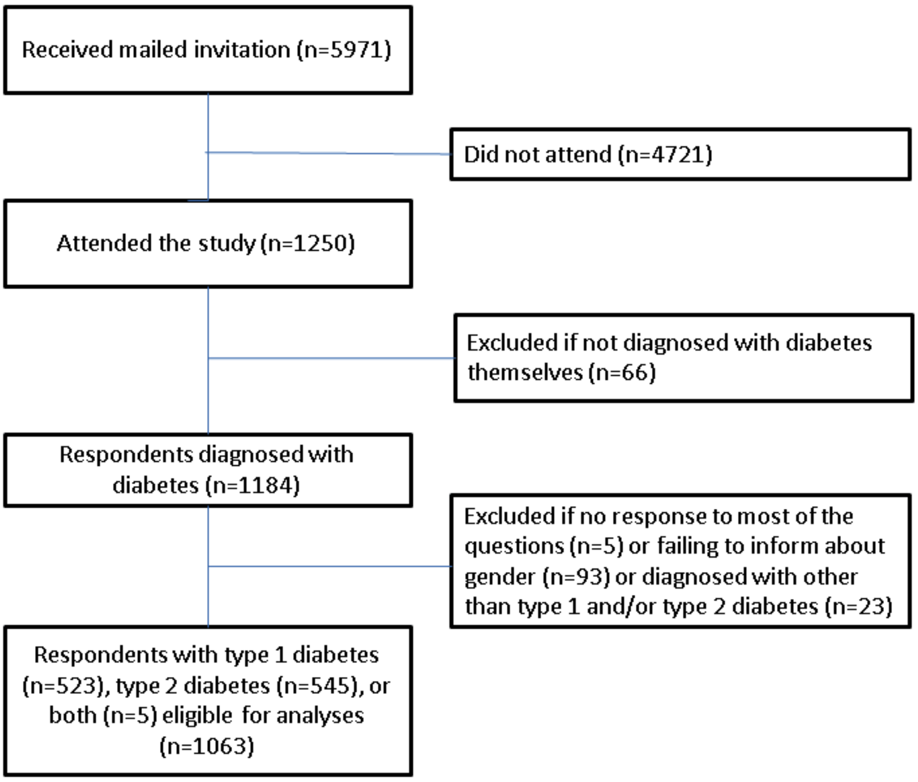
The current cross-sectional study is a part of the DIAcare project [31], investigating relations between eHealth use and the use of provider-based health care services. The project has previously published 3 papers, partly using the same dataset and methodology [11,32,33]. Initially, as described in our protocol paper [31], we planned to use data from the seventh Tromsø Study, conducted in 2015-2016. However, these data were not available to us because of an agreement with another researcher regarding exclusive rights to decide about the collected eHealth data for 3 years. Consequently, we developed a tailored questionnaire on the basis of the specific objectives of this study [31], using relevant questions from other published surveys on health seeking behavior [34,35].

Email survey data were obtained in January and February 2018 from members of The Norwegian Diabetes Association (NDA). By December 31, 2017, the patient organization had 33,908 members, 53% were women and 47% were men. Around 30% of the members have T1D [36]. The Norwegian Centre for Research Data (NSD) Web Survey distributed the invitations to a randomly selected sample of 5971 individuals (about 18% of all members). We distributed information about the study purpose and what participation would entail together with the invitation. The questionnaire (Multimedia Appendix 1) included questions about demographic and socioeconomic characteristics, health status, including specific questions about duration, severity and treatment of the individuals' diabetes, and use of and experiences with eHealth. Before data collection, 2 people diagnosed with diabetes and 2 experts from our research group (EÅ and AHH) reviewed and tested the questionnaire several times. Nonrespondents were given 1 reminder, submitted by email 15 days after the first request.

Participants

It was not possible for the same respondent to fill in the questionnaire more than once. From a total of 1250 participants, we first excluded those who had not been diagnosed with diabetes themselves (n=66). This group comprised 61 family members, 4 health personnel (2 overlapping), and 3 others. We also excluded participants who failed to respond to most of the questions (n=5) and those who did not give information about gender (n=93). Finally, as we had decided to investigate T1D and T2D in this part of the study, participants with other diabetes types were excluded (n=23). The final sample included 1063 respondents (Figure 1). Of these, 523 reported having T1D, and 545 reported having T2D. A total of 5 of these were overlapping, interpreted as double diabetes [37,38].

Figure 1. Flow chart of the study population.



Variables

The 4 dependent variables were the use of apps (for mobile phone or tablet computer), search engines (such as Google), social media (such as Facebook), and video services (such as YouTube) for health purposes during the previous 12 months. For an easier interpretation of logistic regressions, and in line with previous research [34], these eHealth variables were dichotomized by merging the original 4 answering options into "never or once" and "sometimes or often."

Education and household income were the key independent variables. Education was categorized into low (primary/part of secondary school), middle (completed secondary school), high (college/university<4 years), and highest education (college/university 4 years or more). The response options for household income were merged into 3 groups, labeled low (NOK 350,000 or less), medium (NOK 351,000-750,000), and high (NOK 751,000 or more) income. Adjustment independent variables were gender, age, and self-rated health. We grouped age in 20-year intervals. Response options for self-rated health were excellent, good, fair, bad, and very bad. The bad and very bad categories were merged because of low numbers in the very bad category (4 respondents). The 2 categories of the dichotomous response time variable were early and late respondents.

Analyses

Data were analyzed by descriptive statistics and logistic regressions. Correlations were determined using Spearman correlation coefficients. As T1D and T2D are different diseases, and as the samples differed substantially, particularly regarding gender, age, and duration of the participants' diabetes, we decided to stratify analyses by diabetes type. For each diagnostic category, we constructed 1 multivariable regression model for each of the dependent variables. The independent variables (gender, age, education, household income, and self-rated health) were introduced collectively into the models. We performed similar analyses after excluding the 5 participants reporting double diabetes. Owing to a relatively low response rate, we compared respondents who did not respond initially but eventually consented (late respondents) with early respondents, assuming that the late respondents were more similar to nonrespondents [39]. This was done by subsequently introducing the response time variable into the regression models. We used 95% CIs throughout the study. Stata, version 14.2 (StataCorp LLC), was used for all analyses.

Ethics

Ethics approval was not required, which was confirmed by the Regional Committee for Medical and Health Research Ethics. This study has been approved by the Data Protection Officer (Personvernombudet) at the University Hospital of North-Norway (ref 2017/6579). The NSD resource center received no information about the participants other than their email addresses.

Results

Participation

In total, 1250 persons aged 18 to 89 years participated, constituting a minimum response rate of 20.9% (Figure 1). Eligible for analysis were the 523 participants with T1D and the 545 with T2D, which made a total of 1063 participants. A total of 5 of them reported having both diabetes types, and thus they entered into both groups (Figure 1) [37].

Sample Characteristics

Mean age was 54.8 (95% CI 53.9-55.7) years; it was 47.0 (95% CI 45.7-48.3) years for those with T1D and 62.3 (95% CI 61.4-63.2) years for those with T2D. Median age was 57 years; it was 48 years (range 18-89) for T1D and 63 years (range 22-89) for T2D. Mean disease duration was 17.9 (95% CI 17.1-18.7) years; it was 23.2 (95% CI 21.9-24.6) years for T1D and 12.7 (95% CI 12.0-13.5) years for T2D, whereas median disease duration was 15 years; it was 22 years (range 0-75) for T1D and 10 years (range 0-68) for T2D.

Those who were married, working, and had good self-rated regulation of diabetes and good self-rated health made up the largest groups in both the T1D and T2D samples. For all other characteristics, the largest groups differed between the 2 diagnoses (Table 1).

Among participants with T1D, women (281/523, 53.7%), people aged 40 to 59 years (223/523, 42.6%),

people with high education (152/480, 31.7%), high household income (238/467, 51.0%), and a diabetes duration of at least 30 years (187/522, 35.8%) made up the largest groups (Table 1). In contrast, the largest groups among participants with T2D were men (342/545, 62.7%), persons aged 60 years and over (356/545, 62.7%), persons with medium education (165/510, 32.4%), medium household income (253/488, 51.9%), and persons with a diabetes duration of less than 10 years (209/542, 38.6%; Table 1). As expected, the duration of diabetes differed substantially between the 2 diagnostic groups (Table 1).

In the T1D group, 87.0% (447/514) of the participants had used 1 or more forms of eHealth sometimes or often during the previous year. The corresponding proportion for those with T2D was 77.7% (421/542). The proportion who used search engines was the largest among those with T1D, as well as those with T2D (84.0% vs 73.0%, respectively), followed by apps (55.5% vs 50.8%, respectively), social media (45.2% vs 31.2%, respectively), and video services (23.2% vs 12.3%, respectively; Table 2).

Table 1. Sample characteristics of participants.

Characteristic	Total sample, n (%)	Type 1 diabetes, n (%)	Type 2 diabetes, n (%)
Gender	**N=1063**	**n=523**	**n=545**
Female	483 (45.43)	281 (53.7)	203 (37.3)
Male	580 (54.57)	242 (46.3)	342 (62.7)
Age	**N=1063**	**n=523**	**n=545**
18-39	192 (18.06)	178 (34.0)	14 (2.6)
40-59	394 (37.07)	223 (42.6)	175 (32.1)
60+	477 (44.87)	122 (23.4)	356 (65.3)
Marital status	**N=778**	**n=380**	**n=401**
Single	55 (7.1)	42 (11.0)	13 (3.2)
Married/cohabitant	723 (92.9)	338 (89.0)	388 (96.8)
Main daily activity	**N=984**	**n=481**	**n=507**
Working (full-time or part-time)	519 (52.7)	308 (64.0)	215 (42.4)
Pensioner old age	265 (26.9)	65 (13.5)	200 (39.5)
Pensioner disability	132 (13.4)	53 (11.0)	79 (15.6)
Pupil/student	35 (3.6)	35 (7.3)	0 (0.0)
Other	33 (3.4)	20 (4.2)	13 (2.5)
Education	**N=986**	**n=480**	**n=510**
Low (primary/part of secondary school)	107 (10.9)	39 (8.1)	68 (13.3)
Medium (completed secondary school)	302 (30.6)	139 (29.)	165 (32.4)
High (college/university <4 years)	300 (30.4)	152 (31.7)	150 (29.4)
Highest (college/university, 4 years or more)	277 (28.1)	150 (31.2)	127 (24.9)
Household income	**N=951**	**n=467**	**n=488**
Low (<NOK 350,000)	114 (12.0)	66 (14.1)	48 (9.8)
Medium (NOK 351,000-750,000)	413 (43.4)	163 (34.9)	253 (51.9)
High (NOK 751,000 or more)	424 (44.6)	238 (51.0)	187 (38.3)
Duration of diabetes	**N=1059**	**n=522**	**n=542**
<10 years	333 (31.45)	127 (24.3)	209 (38.6)
10-19 years	308 (29.08)	107 (20.5)	202 (37.3)
20-29 years	201 (18.98)	101 (19.4)	100 (18.5)
30 years and over	217 (20.49)	187 (35.8)	31 (5.7)
Self-rated regulation of diabetes	**N=1054**	**n=520**	**n=539**
Excellent	245 (23.25)	101 (19.4)	146 (27.1)
Good	577 (54.74)	292 (56.2)	287 (53.2)
Fair	193 (18.31)	103 (19.8)	91 (16.9)
Bad/very bad	39 (3.70)	24 (4.6)	15 (2.8)
Self-rated health	**N=1054**	**n=521**	**n=538**
Excellent	155 (14.71)	93 (17.9)	63 (11.7)
Good	542 (51.42)	269 (51.6)	275 (51.1)
Fair	261 (24.76)	113 (21.7)	150 (27.9)
Bad/very bad	96 (9.11)	46 (8.8)	50 (9.3)

Table 2. Proportion of participants using different kinds of electronic health (eHealth) "sometimes" or "often" during the previous 12 months.

Type of eHealth	Total sample			Type 1 diabetes[a]			Type 2 diabetes[a]		
	n/N	%	95% CI	n/N	%	95% CI	n/N	%	95% CI
One or more forms of eHealth	864/1051	82.21	79.8-84.4	*447/514*	*87.0*	*83.8-89.6*	*421/542*	*77.7*	*74.0-81.0*
Apps	556/1048	53.05	50.0-56.1	285/514	55.5	51.1-59.7	274/539	50.8	46.6-55.1
Search engines	821/1048	78.34	75.7-80.7	*431/513*	*84.0*	*80.6-86.9*	*394/540*	*73.0*	*69.0-76.6*
Social media	399/1050	38.00	35.1-41.0	*232/513*	*45.2*	*40.9-49.6*	*169/541*	*31.2*	*27.5-35.3*
Video services	183/1037	17.65	15.4-20.1	*118/506*	*23.3*	*19.8-27.2*	*66/536*	*12.3*	*9.8-15.4*

[a] Statistically significant differences between T1D and T2D are marked in italics.

Positive Association Between Higher Education and the Use of Search Engines

We found a strong association between higher education and the use of search engines. In people with T1D, the odds were more than 3 times higher for the high education group (odds ratio, OR 3.26, 95% CI 1.34-7.96) and almost 6 times higher for the highest education group (OR 5.78, 95% CI 2.14-15.57) compared with the low education group (Table 3). Among those with T2D, the odds were more than doubled for those with high education (OR 2.17, 95% CI 1.11-4.26) and more than 3 times as high for those with the highest education (OR 3.30, 95% CI 1.58-6.89) compared with the low education group (Table 4). We found no educational differences for the use of apps, social media, or video services in any of the diagnostic categories (Tables 3 and 4).

Type 1 Diabetes Patients With Higher Income Are Less Likely to Use Video Services

In people with T1D, we found that the middle- and high-income groups had lower odds of using video services (OR 0.51, 95% CI 0.26-0.99 and OR 0.50, 95% CI 0.26-0.98, respectively) compared with the low-income group (Table 3).

Positive Association Between Higher Income and the Use of Apps

Among people with T1D, as well as T2D, the high income group more likely used apps, compared with the low income group (OR 3.05, 95% CI 1.63-5.71 and OR 2.06, 95% CI 1.02-4.19, respectively; Tables 3 and 4). We found no associations between household income and the use of the other eHealth types in people with T2D (Table 4).

Gender and Age

Men with T1D and T2D were less likely to use social media (OR 0.50, 95% CI 0.34-0.74 and OR 0.62, 95% CI 0.41-0.94, respectively), and men with T2D were less likely to use search engines (OR 0.54, 95% CI 0.34-0.87), compared with women (Tables 3 and 4).

In people with T1D, higher age was inversely associated with the use of apps and search engines (Table 3). We found no statistically significant age differences in the use of eHealth among people with T2D (Table 4).

Self-Rated Health

In many of the T1D groups that reported fair or bad/very bad health, the odds of using eHealth (except for the use of apps) were significantly higher than in the excellent health group. This was most apparent regarding the use of social media, where the odds of use were almost 3 times higher among those in bad/very bad health compared with those in excellent health (OR 2.96, 95% CI 1.34-6.54; Table 3). In the fair health group, the use of search engines was more than twice as likely as in the excellent health group (OR 2.37, 95% CI 1.03-5.46). The group with bad/very bad health was more likely to use video services (OR 2.62, 95% CI 1.15-5.97), compared with the excellent health group (Table 3). Among people with T2D reporting bad/very bad health, the odds of using search engines was significantly higher than among those in excellent health (OR 2.73, 95% CI 1.02-7.31).

There were no strong correlations (defined as rho>0.5) among the independent variables in any of the models. The strongest correlations were found for education and household income in both the T1D and T2D models (rho 0.2757 and 0.2555, respectively). Performing the regression analyses after excluding the 5 participants with “double diabetes” did not alter the results.

Table 3. Associations of using electronic health "sometimes" or "often" in people with type 1 diabetes.

Characteristics	Apps (n=467), OR[a] (95% CI)[b]	Search engines (n=466), OR (95% CI)[b]	Social media (n=466), OR (95% CI)[b]	Video services (n=463), OR (95% CI)[b]
Gender				
Female[c]	1.00	1.00	1.00	1.00
Male	1.40 (0.95-2.06)	0.71 (0.41-1.21)	*0.50 (0.34-0.74)*	1.43 (0.91-2.24)
Age				
18-39 years[c]	1.00	1.00	1.00	1.00
40-59 years	*0.52 (0.33-0.84)*	*0.34 (0.16-0.74)*	1.20 (0.76-1.89)	0.84 (0.50-1.40)
60+ years	*0.45 (0.26-0.77)*	*0.23 (0.10-0.52)*	0.72 (0.42-1.23)	0.59 (0.31-1.12)
Education				
Low education[c] (primary/part of secondary school)	1.00	1.00	1.00	1.00
Medium education (completed secondary school)	0.84 (0.39-1.80)	1.64 (0.70-3.82)	1.03 (0.48-2.20)	0.89 (0.36-2.16)
High education (college/university, <4 years)	0.81 (0.38-1.73)	*3.26 (1.34-7.96)*	0.95 (0.44-2.03)	1.10 (0.45-2.68)
Highest education (college/university, 4 years or more)	0.84 (0.39-1.82)	*5.78 (2.14-15.57)*	0.79 (0.37-1.71)	0.97 (0.39-2.40)
Household income				
Low income[c] (NOK <350,000)	1.00	1.00	1.00	1.00
Medium income (NOK 351,000-750,000)	1.26 (0.68-2.33)	1.11 (0.49-2.50)	1.00 (0.54-1.86)	*0.51 (0.26-0.99)*
High income (NOK 751,000 or more)	*3.05 (1.63-5.71)*	1.60 (0.68-3.78)	1.55 (0.83-2.89)	*0.50 (0.26-0.98)*
Self-rated health				
Excellent health[c]	1.00	1.00	1.00	1.00
Good health	1.31 (0.78-2.20)	1.88 (0.95-3.71)	*1.99 (1.15-3.45)*	0.96 (0.52-1.78)
Fair health	1.68 (0.91-3.12)	*2.37 (1.03-5.46)*	*2.46 (1.30-4.64)*	1.08 (0.53-2.23)
Bad/very bad health	1.96 (0.89-4.32)	2.13 (0.69-6.60)	*2.96 (1.34-6.54)*	*2.62 (1.15-5.97)*

[a] OR: odds ratio.

[b] Statistically significant findings are marked in italics.

[c] Reference groups.

Table 4. Associations of using electronic health "sometimes" or "often" in people with type 2 diabetes.

Characteristics	Apps (n=478), OR[a] (95% CI)[b]	Search engines (n=479), OR (95% CI)[b]	Social media (n=480), OR (95% CI)[b]	Video services (n=476), OR (95% CI)[b]
Gender				
Female[c]	1.00	1.00	1.00	1.00
Male	0.92 (0.63-1.36)	*0.54 (0.34-0.87)*	*0.62 (0.41-0.94)*	0.72 (0.40-1.28)
Age				
18-39 years[c]	1.00	1.00	1.00	1.00
40-59 years	0.83 (0.23-2.96)	1.17 (0.21-6.37)	1.11 (0.30-4.12)	0.88 (0.16-4.70)
60+ years	0.83 (0.24-2.91)	0.86 (0.16-4.54)	0.77 (0.21-2.82)	0.58 (0.11-3.06)
Education				
Low education[c] (primary/part of secondary school)	1.00	1.00	1.00	1.00
Medium education (completed secondary school)	1.29 (0.71-2.35)	1.74 (0.92-3.29)	1.06 (0.56-2.04)	0.69 (0.27-1.74)
High education (college/university, <4 years)	1.42 (0.76-2.63)	*2.17 (1.11-4.26)*	1.05 (0.54-2.05)	1.18 (0.48-2.91)
Highest education (college/university, 4 years or more)	1.28 (0.68-2.43)	*3.30 (1.58-6.89)*	0.84 (0.42-1.69)	0.91 (0.35-2.38)
Household income				
Low income[c] (NOK <350,000)	1.00	1.00	1.00	1.00
Medium income (NOK 351,000-750,000)	1.66 (0.84-3.26)	1.11 (0.52-2.33)	1.20 (0.58-2.50)	1.13 (0.39-3.21)
High income (NOK 751,000 or more)	*2.06 (1.02-4.19)*	2.25 (0.99-5.11)	1.15 (0.53-2.46)	0.92 (0.31-2.78)
Self-rated health				
Excellent health[c]	1	1	1	1
Good health	1.11 (0.62-1.97)	1.42 (0.74-2.74)	1.08 (0.57-2.05)	1.11 (0.43-2.89)
Fair health	1.20 (0.64-2.23)	1.56 (0.77-3.19)	1.37 (0.69-2.70)	1.75 (0.66-4.64)
Bad/very bad health	1.44 (0.66-3.18)	*2.73 (1.02-7.31)*	1.06 (0.45-2.51)	1.13 (0.31-4.06)

[a] OR: odds ratio.

[b] Statistically significant findings are marked in italics.

[c] Reference groups.

Discussion

Principal Findings

We found that 87.0% (447/514) of people with T1D and 77.7% (421/542) of people with T2D had used 1 or more forms of eHealth sometimes or often during the previous year. In both diagnostic groups, the proportion using search engines was the largest, followed by apps, social media, and video services. Those with higher levels of education had higher odds of using search engines, whereas we found no educational differences for the use of apps, social media, or video services. In both diagnostic groups, those with high income more likely used apps, whereas T1D patients with medium and high income had lower odds of using video services. Men in both diagnostic groups used social media less than women did. In people with T1D, higher age was inversely associated with the use of apps and search engines, whereas those in poorer health had higher odds of using eHealth, particularly social media. There was no association between self-rated health and the use of apps. Among people with T2D, those with bad/very bad health had significantly higher odds of using search engines compared with those in excellent health.

Overall Use of Electronic Health Among People With Diabetes

We revealed a high overall use of eHealth, and search engines were the most commonly used platforms. This conforms with the study by Hong et al, which reported that seeking health information on the Web has been the most typical health-related internet use in the general US population of older adults (55 years and over), increasing from 57% in 2003 to 80% in 2011 [9]. Considering the high mean age in this study's sample, this study's rates for a Norwegian diabetes population can be seen as a confirmation of this trend, and for T1D, they can be seen as an extension of it. This is also in line with previous research, indicating that long-term illness and good health, which characterize this study's sample, are related to increased use of eHealth [22]. The high overall use of eHealth suggests increased efforts in providing high-quality electronic information and services tailored for people with diabetes.

Differences According to Education

There was a strong association between higher education and the use of search engines in both diagnostic groups. This is in line with most other studies of education and Web-based health information seeking in general and disease-specific populations [10,25,40]. The finding might be explained by higher-educated groups' capabilities and experiences of seeking out, finding, understanding, and making sense of health- and disease-related information [25,40]. It might also reflect educational differences in engagement with health, health care systems, and health care activities [25]. This might be reinforced by information tailored for people with higher education more than for people with lower education, hindering future searches among those who experienced that they did not fully understand what was found [25]. Notably, education was not significantly associated with the use of apps, social media, or video services in this study. Kontos et al even found lower levels of education to be associated with increased use of social media for health purposes [25], which might indicate that social media information is experienced as more accessible and useful in groups with lower education. The peer-to-peer interactions and social and emotional support provided by social media may be of significance in this regard [41]. Regarding the use of apps, our finding of no association with education supports other recently conducted studies [42,43]. This might indicate benefit from communication through apps, social media, and video services, regardless of education, whereas higher-educated people might additionally experience benefit from information through literature and other texts, whether internet-based or not. When targeting people with lower education, one might thus consider providing information through apps, videos, or social media. This is important, as it might have consequences for engagement in healthy lifestyle behavior and the ability to achieve better health outcomes.

Differences According to Household Income

Health apps may be used for a wide variety of health purposes, including self-management and control of diabetes. In both diagnostic categories, the high-income group had higher odds of using apps than the low income group. The association was stronger for T1D than for T2D (Tables 3 and 4). In contrast, an Australian study recently reported no association between SES and the use of apps for T1D and T2D [42]. This is interesting even if findings are not directly comparable. A possible explanation of our finding might be the costs of downloading some of the apps. As the use of apps is still increasing, another possible explanation might be that novel solutions and treatments reach people in higher socioeconomic groups first [19-21]. We found that T1D patients with medium and high income had lower odds of using video services compared with the low-income group. This might indicate that people with low income are more likely to benefit from video-based information regarding health issues. However, this association was not found for T2D. A possible explanation might be the higher age in this group and less use of video services compared with the T1D group (Tables 1 and 2). In a general US population, higher household income was significantly associated with seeking health information on the Web in 2003, 2005, and 2011-12 [9]. This is only partly supported by this study's results, as we did not find any significant association between household income and the use of search engines or social media.

Differences According to Gender

The only significant gender differences in this study were that men in both diagnostic categories used social media less than women and that men with T2D used search engines less than women. Previous research reports a general trend that women have a higher engagement in health issues, eHealth, and social media, and women often act as a liaison for their family [9,25], which might underpin and explain these findings. Nevertheless, we found no gender differences for the use of apps and video services. Previous research has produced conflicting results regarding gender and the use of apps [43,44], and evidence is scarce about predictors for the use of video services. The small gender differences in this study (Tables 3 and 4) are in line with the trend reported by Hong et al that the digital health divide between genders narrowed from 2003 and was no longer significant in 2011 [9].

Differences According to Age

Higher age was inversely associated with the use of apps and search engines in people with T1D. Others have specifically described an association between younger age and app use among people with diabetes [42]. In line with these results, previous research consistently reports that older people use eHealth less than younger people, both in general populations and elderly populations, as well as in populations with chronic disease [10,11,22,25]. Another possible explanation is that people with T1D are likely to have a longer disease duration the older they are, as incidence of T1D decreases with increasing age. With their greater experience in living with diabetes, they may not use apps or search engines as much as if they were more recently diagnosed. This is consistent with the duration of diabetes described in Table 2. In addition, the lower use of apps and search engines among older people with T1D might partly be because of less use of mobile devices in general, as well as lower education in older age groups [13,45]. Age-sensitive information design for elderly people could be an area for development. However, in the T2D group, we found no age-related differences in the use of eHealth. The association regarding app and search engine use described above for the T1D group may be consistent with the lack of association with age among those with T2D. As T2D incidence is higher with increasing age, more of the T2D sample is at any given age likely to be more recently diagnosed than the veterans of T1D at their comparable ages, thus explaining that eHealth use does not differ by age in the T2D group. Hong et al described a narrowing of the age divide in the general US population from 2003 to 2011 [9]. This study's findings suggest a disappearance of the age divide in the T2D group, as well as regarding the use of social media and video services in the T1D group. Our findings are not surprising, as the use of eHealth is increasing rapidly among elderly people in the western world [10,13,46], and the elderly population is gradually transforming from "digital immigrants" (having to learn and acquire digital activity as adults) to “digital natives” (having grown up with digital technology) [25]. Thus, the inverse association between age and eHealth use reported in previous research might not be sustained in the future [13,46] .

Differences According to Self-Rated Health

Previous research has produced conflicting results regarding the relationship between health status and the use of eHealth [47]. A striking finding in this study was that those reporting bad/very bad health were almost 3 times as likely to use social media compared with those in excellent health. Access of shared information through social media might give valuable fellowship, along with clinical and emotional support for people with T1D, particularly in periods of poorer health [41]. In people with T2D, we found a positive association between bad/very bad health and the use of search engines, which is in line with the illness behavior model, stating that people in poorer health more likely seek disease-related information on the Web [23]. We found no significant associations between self-rated health and app use (Tables 3 and 4). Others have found an association in terms of better outcomes of health parameters among app users, stating that the use of apps contributes to better disease management and health outcomes [42].

Limitations and Strengths

Limitations and strengths have been discussed in detail in our first study in this project [11]. One limitation is the low estimated participation rate. However, response rate must not be confused with response quality [39]. Older people dominated among the late respondents compared with the early respondents [11]. As late respondents might be more similar to nonrespondents [39], seniors might be underrepresented in this study.

Distribution of the questionnaire by email is another limitation, which excluded those who do not use the internet. As 98% of Norwegian households have internet access, we do not think this affected our results significantly [11]. It is well known that women, healthier persons, higher socioeconomic groups, and middle-aged people are more likely to participate in surveys [11]. This suggests that women, people around 40 to 80 years, people in better health, and higher socioeconomic groups might be overrepresented in this study. As different factors might pull the tendency in different directions or level each other out, it is not possible to judge the magnitude or direction of a possible nonresponse bias. The low response rate is in itself not an indication of low representativeness [48]. We suggest that nonresponse bias posed a limited threat to this study’s validity; however, generalization must be made with caution. We investigated socioeconomic differences in the form of self-reported education and self-reported household income. Education was measured at the individual level, whereas household income was measured at family level, which is a limitation. The answers regarding household income might be less accurate than those regarding one’s own education. We thus consider education a more solid measure than household income. However, we think that both these aspects of SES, despite limitations, offer a broader picture than one of them might do alone.

Recall bias might have occurred for all aspects of this questionnaire study. Other relevant limitations explored in the first study were the validity of self-reported data, the cross-sectional study design, and interest in the subject studied [11]. Finally, we cannot exclude the possibility of unmeasured confounders of the reported associations. This study also has some strengths, which are similar to the strengths discussed in the first paper in this series [11]. The most important strength is the focus on a scarcely investigated research field. Other strengths are the detailed questionnaire specifically tailored to people with diabetes, the recruitment of participants from all of Norway, the inclusion of a wide age span of participants, and the opportunity to analyze the data shortly after they were collected. Finally, the collection of data in cooperation with the NDA enabled us to develop an excellent user participation, with a large and important group of health care users.

Conclusions

Overall, this study indicates a digital divide among people with diabetes in Norway, with consequences that may contribute to shaping inequalities in health outcomes. We want to highlight the strong relationship between higher education and the use of search engines, along with the finding that educational level was not associated with differences in the use of apps, social media, and video services. We also revealed that the use of video services was more likely in lower income groups. Collectively, our findings suggest that information through apps, social media, and video services might be a good choice when targeting lower educational groups. In society’s effort to reduce inequalities in health outcomes, clinicians and health care leaders should be aware of these inequalities in eHealth use to design adequate health communication strategies for different target groups, particularly according to educational level. More research is needed to confirm our findings.

Acknowledgments

The authors wish to thank the NDA for their cooperation in the conduct and performance of this study. Without their generous contribution, this study could not have been realized. The authors also wish to thank the Northern Norway Health Authorities and The University Hospital of North Norway, who provided funding for this research and approved the changes regarding data collection that had to be made to complete the study.

Authors' Contributions

All authors contributed to the design and conduct of the study. AHH drafted the manuscript. All authors contributed with improvements and critical revisions and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey questionnaire.

[PDF File (Adobe PDF File), 62KB - jmir_v21i5e13615_app1.pdf]

References

1. Ogurtsova K, da RFJD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
2. Norwegian Institute of Public Health. [Public Health in Norway 2018] URL: https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2018/helsetilstanden-i-norge-2018.pdf [accessed 2018-06-16] [WebCite Cache ID 70DeT070T]
3. Jenssen TG, Tonstad S, Claudi T, Midthjell K, Cooper J. The gap between guidelines and practice in the treatment of type 2 diabetes A nationwide survey in Norway. Diabetes Res Clin Pract 2008 May;80(2):314-320. [doi: 10.1016/j.diabres.2007.12.025] [Medline: 18279994]
4. Mouland G. [Diabetes in general practice--were treatment goals reached?]. Tidsskr Nor Laegeforen 2014 Jan 28;134(2):168-172 [FREE Full text] [doi: 10.4045/tidsskr.13.0375] [Medline: 24477150]
5. Cooper JG, Claudi T, Thordarson HB, Løvaas KF, Carlsen S, Sandberg S, et al. Treatment of type 1 diabetes in the specialist health service--data from the Norwegian Diabetes Register for Adults. Tidsskr Nor Laegeforen 2013 Nov 12;133(21):2257-2262 [FREE Full text] [doi: 10.4045/tidsskr.13.0153] [Medline: 24226332]
6. World Health Organization. eHealth URL: http://www.who.int/ehealth/en/ [accessed 2018-06-16] [WebCite Cache ID 70DfcHSuE]
7. Wangberg SC, Andreassen HK, Prokosch H, Santana SMV, Sørensen T, Chronaki CE. Relations between internet use, socio-economic status (SES), social support and subjective health. Health Promot Int 2008 Mar;23(1):70-77 [FREE Full text] [doi: 10.1093/heapro/dam039] [Medline: 18083686]
8. Fox S, Duggan M. Pew Internet. Washington DC: Pew Research Center; 2013. Health Online URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2018-06-16] [WebCite Cache ID 70Dh4ruWf]
9. Hong YA, Cho J. Has the digital health divide widened? Trends of health-related internet use among older adults from 2003 to 2011. J Gerontol B Psychol Sci Soc Sci 2017 Sep 1;72(5):856-863. [doi: 10.1093/geronb/gbw100] [Medline: 27558403]
10. Sørensen T, Andreassen HK, Wangberg SC. Norwegian Centre for eHealth Research. 2014. [Project report eHealth in Norway 2013] URL: https://ehealthresearch.no/files/documents/Prosjektrapporter/NST-rapport_2014-02_e-helse_i_Norge_2013.pdf [accessed 2018-06-16] [WebCite Cache ID 70DhrW5OT]
11. 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 7;20(11):e11322 [FREE Full text] [doi: 10.2196/11322] [Medline: 30404766]
12. [Media news]. [Share with internet access] URL: http://medienorge.uib.no/statistikk/medium/ikt/347 [accessed 2019-05-17] [WebCite Cache ID 78RHFww6P]
13. Statistics Norway. [Nine out of ten surf the net every day] URL: https://www.ssb.no/teknologi-og-innovasjon/artikler-og-publikasjoner/ni-av-ti-surfer-pa-nettet-hver-dag [accessed 2019-01-31] [WebCite Cache ID 75pxg0TzO]
14. Mackenbach JP. Nordic paradox, Southern miracle, Eastern disaster: persistence of inequalities in mortality in Europe. Eur J Public Health 2017 Dec 1;27(suppl_4):14-17. [doi: 10.1093/eurpub/ckx160] [Medline: 29028239]
15. Grintsova O, Maier W, Mielck A. Inequalities in health care among patients with type 2 diabetes by individual socio-economic status (SES) and regional deprivation: a systematic literature review. Int J Equity Health 2014 Jun 2;13:43 [FREE Full text] [doi: 10.1186/1475-9276-13-43] [Medline: 24889694]
16. Espelt A, Borrell C, Roskam AJ, Rodríguez-Sanz M, Stirbu I, Dalmau-Bueno A, et al. Socioeconomic inequalities in diabetes mellitus across Europe at the beginning of the 21st century. Diabetologia 2008 Nov;51(11):1971-1979. [doi: 10.1007/s00125-008-1146-1] [Medline: 18779946]
17. Sortsø C, Lauridsen J, Emneus M, Green A, Jensen PB. Socioeconomic inequality of diabetes patients' health care utilization in Denmark. Health Econ Rev 2017 Dec;7(1):21 [FREE Full text] [doi: 10.1186/s13561-017-0155-5] [Medline: 28550486]
18. Starfield B. Pathways of influence on equity in health. Soc Sci Med 2007 Apr;64(7):1355-1362. [doi: 10.1016/j.socscimed.2006.11.027] [Medline: 17208343]
19. Whitehead M. CIEE-FLACSO. Copenhagen: World Health Organisastion (WHO) Regional Office for Europe; 1990. The concept and principles of equity and health URL: http://salud.ciee.flacso.org.ar/flacso/optativas/equity_and_health.pdf [accessed 2019-01-31] [WebCite Cache ID 75pyZvn3a]

20. Bell AV. "I think about Oprah": social class differences in sources of health information. Qual Health Res 2014 Apr;24(4):506-516. [doi: 10.1177/1049732314524637] [Medline: 24623661]
21. Victora CG, Fenn B, Bryce J, Kirkwood BR. Co-coverage of preventive interventions and implications for child-survival strategies: evidence from national surveys. Lancet 2005;366(9495):1460-1466. [doi: 10.1016/S0140-6736(05)67599-X] [Medline: 16243091]
22. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
23. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2015 Aug 10:804-814. [doi: 10.1177/1460458215595851] [Medline: 26261218]
24. Buysse HE, de Moor GJ, de Maeseneer J. Introducing a telemonitoring platform for diabetic patients in primary care: will it increase the socio-digital divide? Prim Care Diabetes 2013 Jul;7(2):119-127. [doi: 10.1016/j.pcd.2012.10.085] [Medline: 23273770]
25. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
26. Wangberg SC, Sørensen T, Andreassen HK. Using the internet to support exercise and diet: a stratified Norwegian survey. Med 2 0 2015;4(2):e3 [FREE Full text] [doi: 10.2196/med20.4116] [Medline: 26310277]
27. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. Technology-assisted weight loss interventions in primary care: a systematic review. J Gen Intern Med 2015 Jan;30(1):107-117. [doi: 10.1007/s11606-014-2987-6] [Medline: 25134692]
28. Bashshur RL, Shannon GW, Smith BR, Woodward MA. The empirical evidence for the telemedicine intervention in diabetes management. Telemed J E Health 2015 May;21(5):321-354 [FREE Full text] [doi: 10.1089/tmj.2015.0029] [Medline: 25806910]
29. 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]
30. Ministry of Health and Care Services. The Norwegian Public Health Act URL: https://www.regjeringen.no/globalassets/upload/hod/hoeringer-fha_fos/123.pdf [accessed 2019-01-31] [WebCite Cache ID 75rass382]
31. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen O, Wangberg SC, et al. The use of eHealth and provider-based health services by patients with diabetes mellitus: protocol for a cross-sectional study. JMIR Res Protoc 2016 Oct 31;5(4):e207 [FREE Full text] [doi: 10.2196/resprot.6529] [Medline: 27799136]
32. Hansen AH, Claudi T, Årsand E. Associations between the use of eHealth and out-of-hours services in people with type 1 diabetes: cross-sectional study. J Med Internet Res 2019 Mar 21;21(3):e13465 [FREE Full text] [doi: 10.2196/13465] [Medline: 30896437]
33. Hansen AH, Claudi T, Årsand E. Use of electronic health and its impact on doctor-visiting decisions among people with diabetes: cross-sectional study. J Med Internet Res 2019 Apr 26;21(4):e13678 [FREE Full text] [doi: 10.2196/13678] [Medline: 31025957]
34. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
35. The Tromsø Study Website. URL: http://tromsoundersokelsen.uit.no/tromso/ [accessed 2018-06-16] [WebCite Cache ID 70Dkah0hU]
36. The Norwegian Diabetes Association. A˚rsberetning 2017 URL: https://issuu.com/knutjarle/docs/a_rsberetning_2017_ [accessed 2018-06-16] [WebCite Cache ID 70DkKIFTA]
37. Merger SR, Kerner W, Stadler M, Zeyfang A, Jehle P, Müller-Korbsch M, DPV Initiative, German BMBF Competence Network Diabetes Mellitus. Prevalence and comorbidities of double diabetes. Diabetes Res Clin Pract 2016 Sep;119:48-56. [doi: 10.1016/j.diabres.2016.06.003] [Medline: 27449710]
38. Cleland SJ, Fisher BM, Colhoun HM, Sattar N, Petrie JR. Insulin resistance in type 1 diabetes: what is 'double diabetes' and what are the risks? Diabetologia 2013 Jul;56(7):1462-1470 [FREE Full text] [doi: 10.1007/s00125-013-2904-2] [Medline: 23613085]
39. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. J Am Med Assoc 2012 May 2;307(17):1805-1806. [doi: 10.1001/jama.2012.3532] [Medline: 22550194]
40. Lee C, Ramírez AS, Lewis N, Gray SW, Hornik RC. Looking beyond the Internet: examining socioeconomic inequalities in cancer information seeking among cancer patients. Health Commun 2012;27(8):806-817 [FREE Full text] [doi: 10.1080/10410236.2011.647621] [Medline: 22356137]
41. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
42. Trawley S, Baptista S, Browne JL, Pouwer F, Speight J. The use of mobile applications among adults with type 1 and type 2 diabetes: results from the Second MILES-Australia (MILES-2) study. Diabetes Technol Ther 2017 Dec;19(12):730-738. [doi: 10.1089/dia.2017.0235] [Medline: 29028442]

43. Ernsting C, Dombrowski SU, Oedekoven M, Kanzler M, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
44. Elavsky S, Smahel D, Machackova H. Who are mobile app users from healthy lifestyle websites? Analysis of patterns of app use and user characteristics. Transl Behav Med 2017 Dec;7(4):891-901 [FREE Full text] [doi: 10.1007/s13142-017-0525-x] [Medline: 28929368]
45. Statistics Norway. Level of education in the population URL: https://www.ssb.no/utniv/ [accessed 2019-02-01] [WebCite Cache ID 75rcSjTbT]
46. Bujnowska-Fedak MM. Trends in the use of the internet for health purposes in Poland. BMC Public Health 2015;15:194 [FREE Full text] [doi: 10.1186/s12889-015-1473-3] [Medline: 25886280]
47. Hale TM. Is there such a thing as an online health lifestyle? Examining the relationship between social status, Internet access, and health behaviors. Information, Communication & Society 2013 May;16(4):501-518. [doi: 10.1080/1369118X.2013.777759]
48. Halbesleben JR, Whitman MV. Evaluating survey quality in health services research: a decision framework for assessing nonresponse bias. Health Serv Res 2013 Jun;48(3):913-930 [FREE Full text] [doi: 10.1111/1475-6773.12002] [Medline: 23046097]

Abbreviations

eHealth: electronic health
NDA: Norwegian Diabetes Association
NSD: Norwegian Centre for Research Data
OR: odds ratio
SES: socioeconomic status
T1D: type 1 diabetes
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 04.02.19; peer-reviewed by Z Huang, L Nelson, S Oser, S Li, E Da Silva; comments to author 02.03.19; revised version received 28.03.19; accepted 05.05.19; published 29.05.19.

Please cite as:
Hansen AH, Bradway M, Broz J, Claudi T, Henriksen Ø, Wangberg SC, Årsand E
Inequalities in the Use of eHealth Between Socioeconomic Groups Among Patients With Type 1 and Type 2 Diabetes: Cross-Sectional Study
J Med Internet Res 2019;21(5):e13615
URL: http://www.jmir.org/2019/5/e13615/
doi:10.2196/13615
PMID:31144669

Original Paper

Health-Related Internet Use in Hard-to-Reach Populations: Empirical Findings From a Survey in a Remote and Mountainous Province in China

HongMin Li[1], PhD; Jin Xu[2,3], PhD; Lingui Li[4], PhD; Qingyue Meng[2,3], PhD

[1]National Health Commission Key Laboratory of Health Economics and Policy Research, School of Health Care Management, Shandong University, Ji'nan, China

[2]China Center for Health Development Studies, Peking University, Beijing, China

[3]School of Public Health, Peking University, Beijing, China

[4]Zhejiang Chinese Medical Univeristy, Hangzhou, China

Corresponding Author:
Qingyue Meng, PhD
China Center for Health Development Studies
Peking University
38 Xue Yuan Road, Haidian District
Beijing, 100191
China
Phone: 86 18668926330
Fax: 86 053188382693
Email: qmeng@bjmu.edu.cn

Abstract

Background: The expanding use of the internet contributes to more effective searches for health-related information and opens up opportunities for direct Web-based communication with health care professionals. However, little is known about how users' characteristics on the demand side influence health-related internet use, especially in remote and rural areas within developing countries. The absence of accurate estimates of users' characteristics and their impact on adaptations of health care services in developing countries constrains focused policy-centered discussions and the design of appropriate policies.

Objective: The aim of this study was to assess the prevalence of health-related internet use and to identify its determinants in a remote province in China.

Methods: We conducted a cross-sectional survey in June and July of 2018 in Ningxia, located in northwestern China. Rural households were selected by multistage random sampling, and households' key members were interviewed face-to-face at the respondents' home. Dependent variables were whether the households use Web health services or not. Independent variables were chosen based on the Andersen behavioral model. Sociodemographic characteristics were compared between households that used health-related Web services with nonusers. We applied logistic regression models to evaluate multivariate associations between respondents' characteristics and their usage of Web-based health services and obtained odds ratios with 95% CI.

Results: A total of 1354 respondents from rural households were interviewed, of whom 707 (52.22%) were men. The mean age of the respondents was 44.54 years (SD 10.22). Almost half of the surveyed households (640/1354, 47.27%) reported using 1 or more Web-based health care services, whereas 37.8% (502/1354) reported using the internet to obtain health-related information, 15.51% (210/1354) used the internet to communicate with professionals about health issues, and 7.24% (98/1354) had engaged in Web-based consultations in the last year. After controlling for potential confounders, households engaged in health-related internet use were found to be wealthier, have higher health demands, and have less geographic access to high-quality health care compared with other households.

Conclusions: The internet has become a major health information resource in rural Ningxia. Social structures, family enabling factors, health needs, and characteristics relating to health care access were significant predictors of households' health-related internet use in rural and remote areas in China. Those who belong to older age groups, have low income, and whose education levels do not extend beyond primary school education are significantly less likely to use Web-based health care services and to benefit from Web-based health care programs. A need for continued collaborative efforts involving multiple stakeholders, including communities, Web-based and other health care providers, family members, and the government is needed.

(J Med Internet Res 2019;21(5):e12693) doi:10.2196/12693

KEYWORDS

eHealth; rural population; cross-sectional survey; China

Introduction

Key Challenges in China's Health Care System

The progress of the health care reform, initiated in China in 2009, has led to significant changes in the country's health care system. China's achievements in strengthening its health care system have been remarkable, as evidenced by the substantial scale of inputs, efforts, and social expectations [1]. Nevertheless, there are challenges in the accessibility and quality of health care and the equitable allocation of associated resources. A particular challenge is the disparity in health care across regions, rural and urban areas, and different segments of the population [2]. These disparities, including inequitable and inadequate allocation of resources, critically impact the health care system and constitute major barriers that affect health care accessibility in rural and remote areas [3].

The health service delivery system in China is divided into rural and urban parts. The urban part is set up with community-level services as the grassroots level, with collaboration between community health institutions and city hospitals. The rural health system is set up with county-level hospitals as the backbone, with township- and country-level clinics as the grassroots level. At higher levels, large general hospitals at the city, provincial, and national levels provide acute and emergency care and focus on severe and complicated diseases. However, the current health resource allocation is mostly concentrated on large general hospitals, which leads to disparities between the urban and rural areas [4]. Geographically, disparities still exist among eastern, central, and western China [2].

The rural areas in western China have special geography and economy characteristics; therefore, those residents have poor access to health services and are hard-to-reach populations. "Hard-to-reach" is a term used to describe populations that are difficult to reach or involve in research or public health programs because of their physical and geographical location (eg, in mountains, forests, or deserts) or their social and economic situation [5].

The Internet and Health Care Accessibility

Expansion of internet use contributes to more effective searches for health-related information and creates opportunities for direct communication with health care professionals through the internet. Health-related internet use can potentially reduce overall medical costs [6] by eliminating the need for diagnostic testing as well as travel expenses incurred by rural patients who repeatedly visit urban hospitals. Moreover, by encouraging users to have regular medical checkups, Web-based health care services may lead to a reduction in the number of emergency cases [6]. Electronic health (eHealth), defined as the use of emerging information and communications technology (ICT), and especially the internet for health-related issues, has become an important supplement to traditional health care resources [7].

Health care can be delivered to geographically remote locations via the internet, which can potentially play an important role in facilitating access to medical information and service delivery in rural and remote areas for hard-to-reach populations. The internet provides access to quality health care for those who may be underserved by the traditional medical system [8,9], thus addressing disparities in health care [10]. Moreover, studies have shown that internet-based health care training, education, and research programs targeting health care providers in rural and remote areas have led to increased communication among clinicians, resulting in more efficient, higher quality, and less costly care in these areas [11,12].

The Internet and Health Care in China

With the rapid development of the internet and ICT in China, 802 million people were reported to have access to the internet in 2018, of which 26.3% were from rural areas [13]. The government is actively promoting integrated provision of internet and health care facilities. The Internet Plus health care strategy was proposed in 2015 by the National Development and Reform Commission and the National Health and Family Planning Commission (renamed the National Health Commission) to improve access to health care in rural and remote areas through ICT. Extensive investments of social capital have been made in parallel in Web-based health services available on the internet through the development of consultations, hospital facilities, and medicines as well as in health management. However, the effects and consequences of large investments of capital in the emerging Web-based health care market on the performance and accessibility of health care remain unclear.

Disparities in health care across geographical, socioeconomic, racial, and ethnic groups can only be reduced through effective eHealth designs and practices that are responsive to public demands [10]. Evidence of the demanding characteristics and determinants are needed to improve efficiency and effectiveness of policies and investments. However, the limited availability of demand-related information has resulted in uncertainty regarding the performance of eHealth programs and their adoption.

Research on Electronic Health Usage

Previous studies [14-17] on the internet and health care, which have mainly focused on developed countries, have contributed to a better understanding of their characteristics and impacts on adaptations of health care services. They have shown that people who use the internet for health-related purposes are more likely to be young and female, with high levels of income and education. Moreover, factors such as health status and attitudes toward the internet have been found to influence Web health-related activities.

As per the World Health Organization, in a low-resource setting, ICT can be a key component for providing universal health coverage [18]. However, studies conducted in developing countries [19-22] have only focused on internet hospitals, the market for Web-based consultations, and the government's role in the provision of eHealth. Their findings were based on secondary data and Web-based resources. In contrast, less is known about how users' characteristics on the demand side influence health-related internet use, especially for hard-to-reach populations living in remote areas in developing countries. The absence of accurate estimates of users' characteristics and their impact on adaptation of health care services in developing countries limits focused policy-centered discussions and the design of appropriate policies.

For populations facing barriers to health care access, the internet can be a particularly appealing source of health information and professional advice. It is important to know how internet-based health care services have been used among such hard-to-reach populations. Previous studies were mostly conducted in cities, and a few studies were conducted among hard-to-reach populations. Furthermore, previous studies lacked a concept framework in their methods.

Objectives

The aims of this study were to assess the prevalence of health-related internet use, to identify its determinants in rural areas in a western province in China, and to make policy recommendations for Web-based health care development in hard-to-reach populations. We sought to answer the following research questions: (1) What are the characteristics of users of Web-based health care services in remote areas? (2) What are the determinants of health-related internet use among the hard-to-reach populations?

Methods

Study Site

Ningxia, located in northwestern China, is the representative region for remote areas in China. Of the 6.68 million inhabitants of Ningxia, 45% were living in rural areas at the end of 2016. In 2016, the number of medical and technical personnel per 1000 inhabitants was 3.9 in rural Ningxia, compared to the national number of 6.1. Furthermore, the number of beds was 3.33 per 1000 people in this region compared to the national number of 5.37 [23]. In addition, Ningxia has a variety of complex landforms—mountains, plateaus, plains, hills, and valleys—that affect the population's geographic accessibility to health care, especially in rural areas.

Study Design and Data

We conducted a cross-sectional survey in June and July 2018 in Ningxia. Of the 22 counties and districts in Ningxia, 4 were selected by considering the socioeconomic factors and environmental and health care resources (Table 1). Of the selected counties and districts, 46 rural villages were chosen for the survey. Within each village, about 30 households were randomly selected. The respondents within the selected households had to meet the following inclusion criteria: age of 18-65 years and possessing the right to make decisions related to the use of health care services within their families. Only one respondent in a family was selected and responsible for the family's information on the Web-based health-related service usage. Face-to-face interviews were conducted at the respondents' homes, and the mean time required for completing the survey was around 30 min.

Measurements

The Andersen Behavioral Model of Health Services Use is a conceptual framework widely used to investigate demands for health care [24]. The model enhances our understanding of the reasons individuals use health services and promotes equitable access to health care services. It proposes determinants to explain why people use health care services, such as an individual's predisposition to use services, factors that enable or impede use, and people's need for care. Even if the model has been widely applied to the general health service field, few studies apply it to the Web health care context [25]. Web-based health care is an innovation used in health care services; this model can help develop a comprehensive adoption model and explain why people use the internet to obtain health information and services.

The dependent variables in our analysis were the respondents' internet search for health-related information, Web-based communication with health professionals, and Web-based consultation. A Web-based consultation takes place on professional medical platforms or websites, and the process of consultation includes history taking, diagnosis, and intervention advice. Web-based communication is more casual: It can occur on social apps between patients and professionals, and the context of communication is health related. Use of the Web-based health care services provided was judged based on whether the respondents use any of the three abovementioned services. Table 2 presents the definitions and response options for the selected variables.

Table 1. Economic development and human resources for health in the selected counties or districts.

Counties/districts	Disposable income (Renminbi, in rural areas)	Number of doctors per 1000 people
Tongxin	6710.69	0.84
Xiji	6857.12	1.16
Huinong	10995.46	3.12
Qingtongxia	11200.00	1.24
Ningxia	9118.69	2.14

Table 2. Definitions and response options for the selected dependent and independent variables.

Variables	Options
Dependent variables	
Use of Web-based health care services[a]	Yes/no
Searching the internet for health-related information[a]	Yes/no
Web-based communication with health professionals[a,b]	Yes/no
Web-based consultation[a,b]	Yes/no
Independent variables	
Social structure factors	
Education levels	No formal education, primary school, junior middle school, high school, college and above
Occupation	Farmer, informally or formally employed
Ethnic groups	Han, Hui, other
Family enabling factors	
Household annual income	Renminbi
Health knowledge scores[c]	Range: 1-5
Health care accessibility factors	
Distances to county hospital	Kilometers
Distances to city hospital	Kilometers
Health need characteristics	
Does the household have chronic patient(s)?[d]	Yes/no
Did the household have outpatient(s) in the last month?	Yes/no
Did the household have inpatient(s) in last 12 months?	Yes/no
Does the household have older persons (age≥60 years)?	Yes/no
Does the household have kids (age≤5 years)?	Yes/no
Other confounding factors	
Gender	Female/male
Age	Number of years
Family members	Number

[a] Internet use for health purposes was measured as usage by all family members for addressing their health issues over the last 12 months.

[b] A Web-based consultation must include history taking, diagnosis, and intervention advice and can take place on the internet platforms of professional medical services. Web-based communication is more casual; it can occur on social apps between patients and professionals, and the context of communication is health related.

[c] Five questions on health-related knowledge were posed, entailing true/false responses.

[d] Chronic diseases were hypertension, dyslipidemia, diabetes or high blood sugar, cancer or malignant tumor, cardiovascular disease, stroke, chronic lung diseases, liver disease, stomach or other digestive disease, arthritis or rheumatism, asthma, emotional or psychiatric problems, and memory-related disease.

Data Analysis

We first performed a descriptive analysis to obtain a basic sociodemographic profile of respondents and their characteristics. The sociodemographic characteristics of households that used Web-based health-related services were then compared with those that did not use these services. We performed a univariate analysis for each of the variables considered by using an appropriate statistical test (Chi-square test for discrete variables and *t* test for continuous variables).

We applied logistic regression models to evaluate multivariate associations between respondents' characteristics and their usage of Web-based health services and obtained odds ratios (ORs) with 95% CIs. Regression analysis was performed according to the following equation:

$$\Pr[y_i = 1|SS, FE, HA, HN, O] = F(\beta_0 + \beta_1 SS_i + \beta_2 FE_i + \beta_3 HA_i + \beta_4 HN_i + \beta_5 O_i)$$

Where y is an indicator of health-related internet use; SS, FE, HA, HN, and O are measures of social structure, family enabling factors, health care accessibility, health care need, and other confounding factors, respectively; and F denotes the logistic cumulative distribution function.

Regression analysis was performed on the sample as a whole and for each individual Web-based health-related activity, namely, searches conducted for health-related information, communication with professionals, and Web-based consultations. In addition, we performed regression analysis for households that availed of any of these three Web-based activities to obtain health care assistance. A *P* value <.05 was considered statistically significant. Stata version 14.0 for Microsoft Windows (Stata Corp, College Station, TX) was used for the statistical analysis.

Results

Study Population

Descriptive statistics for the study sample are presented in Table 3. A total of 1354 respondents from rural households were interviewed, of whom 707 (52.22%) were men. The mean age of the respondents was 44.54 years (SD 10.22). More than three-quarters of the respondents (1038/1354, 76.66%) were able to use the internet. In this study, those who were able to use the internet were considered to be able to use social apps such as WeChat to contact with people or search information on their mobile phone or computer with an internet connection.

An examination of the factors related to the social structure revealed that more than 74.15% (1004/1354) of the respondents were farmers, most of whom lacked an education beyond junior middle school (1176/1354, 86.86%). The majority of the respondents (917/1354, 67.73%) were of Han ethnicity. In 2017, the mean household annual income was ¥39,150.17 (or US $5725.63), and the mean score for health-related knowledge was 3.57 (range: 0-5). Almost half of the selected households (619/1354, 45.72%) had members who were chronic patients, 37.81% (512/1354) of households had used outpatient services in the last 1 month, and 35.86% (486/1354) of households had used inpatient services in the last 12 months. The mean distances of the surveyed households from county- and city-based hospitals were 21.39 km (SD 30.04) and 69.19 km (SD 42.74), respectively.

Overall Health-Related Internet Use

Almost half of the surveyed households (640/1354, 47.27%) reported using one or more Web-based health care services, 37.8% (502/1354) reported using the internet to obtain health-related information, 15.51% (210/1354) used the internet to communicate with professionals about health issues, and 7.24% (98/1354) had engaged in Web-based consultations in the last year (Table 4).

We conducted a univariate analysis of differences among households that had used and those that had not used Web-based services for all of the variables. As shown in Table 2, the overall utilization of Web-based services was associated with the respondents’ positions within the social structure, family enabling factors, family members’ health status, and geographic access to health care. Specifically, respondents with the following characteristics were more likely to use Web-based health care services: higher education levels, higher health-related knowledge scores (3.79 vs 3.57), higher household incomes (¥50,514 vs ¥28,963 or US $7399 vs US $4242), higher health demands, and living nearer to county- or city-based hospitals (Table 5).

Determinants of Health-Related Internet Use

Table 6 shows the determinants of health-related internet use. The ORs from the logistic regression analysis were summarized after controlling for potential confounders. For the category of factors related to the social structure, households with higher levels of education and formal or informal jobs were more likely to use Web-based health services. There was no difference in the use of Web-based health services among respondents of Han and Hui ethnicities. For the category of family enabling factors, respondents with higher health-related knowledge scores were more likely to seek Web-based communication (OR 1.43) or Web-based consultations (OR 1.40) with health care professionals. High-income households were more likely to search for health-related information (OR 1.25) and communicate through the internet on health issues (OR 1.44). Findings for the health need characteristics revealed that households with higher health demands were more likely to use Web-based health services. In particular, households that included outpatients over the last 1 month were more likely to search for Web-based health-related information (OR 1.38), perform Web-based communication (OR 1.99), and schedule Web-based consultation (OR 2.00). Findings on the geographic factors related to health care access revealed that greater distances to city-based hospitals were more likely to prompt members of households to search for health-related information (OR 1.27), perform Web-based communication with professionals (OR 1.60), and schedule Web-based consultations (OR 1.80). Younger respondents were more likely to opt for health-related internet use than older respondents.

Table 3. Characteristics of the study sample (N=1354).

Variable	Value
Social structure, n (%)	
Education level	
No formal education	306 (22.60)
Primary school	403 (29.76)
Junior school	472 (34.86)
High school	116 (8.57)
College and above	57 (4.20)
Occupation	
Farmers	1004 (74.15)
Others	350 (25.85)
Ethnic groups	
Han	917 (67.73)
Hui	437 (32.27)
Family enabling factors, mean (SD)	
Health knowledge	3.57 (1.18)
Family income (Renminbi)	39,150.17 (51,196.3)
Health need factors, n (%)	
Members of age <5 years	
Yes	394 (29.10)
No	960 (70.90)
Members of age >60 years	
Yes	406 (29.99)
No	960 (70.90)
Members with chronic conditions	
Yes	619 (45.72)
No	735 (54.28)
Outpatient clinic visit in last 1 month	
Yes	512 (37.81)
No	842 (62.19)
Hospitalization in last 12 months	
Yes	486 (35.89)
No	868 (64.11)
Health accessibility, mean (SD)	
Distance to county hospital (km)	21.39 (30.04)
Distance to city hospital (km)	69.19 (42.74)
Gender, n (%)	
Female	647 (47.78)
Male	707 (52.22)
Age (years), mean (SD)	44.54 (10.22)
Family size, mean (SD)	4.43 (1.60)

Table 4. The prevalence of health-related internet use (N=1354).

Variables	n (%)
Use of Web-based health care services	640 (47.27)
Searching internet for health-related information	502 (37.8)
Web-based communication with health professionals	210 (15.51)
Web-based consultation	98 (7.24)

Table 5. Characteristics of the study sample in relation to Web-based health-related activities.

Indicators/variables	Overall use of Web-based services (n=640)	*P* value	Web-based information use (n=502)	*P* value	Web-based communication (n=210)	*P* value	Web-based consultation (n=98)	*P* value
Social structure								
Education level, n (%)[a]		<.001		<.001		<.001		<.001
No formal education	66 (21.57)		34 (11.11)		16 (5.23)		6 (1.96)	
Primary school	161 (39.95)		112 (27.79)		54 (13.40)		20 (4.96)	
Junior school	278 (58.90)		232 (49.15)		79 (16.74)		38 (8.05)	
High school	84 (72.41)		74 (63.79)		33 (28.45)		17 (14.66)	
College and above	51 (89.47)		50 (87.72)		28 (49.12)		17 (29.82)	
Occupation, n (%)[a]		<.001		<.001		<.001		<.001
Farming	407 (40.54)		302 (30.08)		111 (11.06)		48 (4.78)	
Others	233 (66.57)		200 (57.14)		99 (28.29)		50 (14.29)	
Ethnic groups, n (%)[a]		.94		.54		.02		.09
Han	434 (47.33)		345 (37.62)		128 (13.96)		59 (6.43)	
Hui	206 (47.14)		157 (35.93)		82 (18.76)		39 (8.92)	
Family enabling factors, mean (SD)								
Health knowledge score[b]	3.79 (1.17)	<.001	3.84 (1.17)	<.001	4.05 (1.10)	<.001	4.12 (1.04)	<.001
Family income (Renminbi)[b]	50514.77 (64409.1)	<.001	53223.2 (67844.9)	<.001	57890.55 (43781.19)	<.001	55173.62 (59225.8)	.001
Health need factors, n (%)								
Members of age <5 years[a]		.19		.03		.07		.02
Yes	197 (50)		163 (41.37)		72 (18.24)		38 (9.64)	
No	443 (46.15)		339 (35.31)		138 (14.37)		60 (6.25)	
Members of age >60 years[a]		.54		.58		.86		.08
Yes	197 (48.52)		155 (38.18)		64 (15.76)		37 (9.11)	
No	443 (46.73)		347 (36.6)		146 (15.40)		61 (6.43)	
Members with chronic conditions[a]		.01		.46		.13		.08
Yes	315 (50.89)		236 (38.13)		106 (17.12)		53 (8.56)	
No	325 (44.22)		266 (36.19)		104 (14.15)		45 (6.12)	
Outpatient clinic visit in last 1 month[a]		<.001		.02		<.001		.001
Yes	276 (53.91)		209 (40.82)		108 (21.09)		53 (10.35)	
No	364 (43.23)		293 (34.80)		102 (12.11)		45 (5.34)	
Hospitalization in last 12 months[a]		<.001		.08		.004		.05
Yes	263 (54.12)		195 (40.12)		94 (19.34)		44 (9.05)	
No	377 (43.43)		307 (33.37)		116 (13.36)		54 (6.22)	
Health accessibility								
Distance to county hospital (km), mean (SD)[b]	17.42 (18.13)	<.001	17.54 (19.62)	<.001	18.06 (30.04)	.08	19.37 (26.68)	.49
Distance to city hospital (km), mean (SD)[b]	65.22 (44.34)	.001	64.09 (43.62)	<.001	72.83 (47.47)	.17	76.27 (44.35)	.04
Gender[a]		.89		.66		.06		.31
Female, n (%)	307 (47.45)		236 (36.48)		88 (13.60)		42 (6.49)	

Indicators/variables	Overall use of Web-based services (n=640)	*P* value	Web-based information use (n=502)	*P* value	Web-based communication (n=210)	*P* value	Web-based consultation (n=98)	*P* value
Male, n (%)	333 (47.10)		266 (37.62)		122 (17.26)		56 (7.92)	
Age (years), mean (SD)[b]	42.22 (10.56)	<.001	41.03 (10.47)	<.001	41.01 (10.88)	<.001	39.03 (11.41)	<.001
Family size, mean (SD)[b]	4.43 (1.56)	.89	4.42 (1.66)	.87	4.52 (1.49)	.33	4.41 (1.60)	.23

[a] Pearson Chi-squared test was performed for the univariable analysis.

[b] *t* test was performed.

Table 6. Multivariate associations of respondents' and households' characteristics and Web-based health-related activities.

Variables	Web-based service use (n=640), OR[a] (95% CI)	P value	Web-based information use (n=502), OR (95% CI)	P value	Web-based communication (n=210), OR (95% CI)	P value	Web-based consultation (n=98), OR (95% CI)	P value
Social structure								
Education (reference: no formal education)[a]								
Primary school	2.16 (1.51-3.10)	<.001	2.55 (1.65-3.94)	<.001	2.41 (1.32-4.42)	.004	2.19 (0.84-5.65)	.10
Junior school	4.05 (2.79-5.88)	<.001	5.55 (3.59-8.58)	<.001	2.72 (1.47-5.02)	.001	3.32 (1.31-8.43)	.01
High school	6.67 (3.88-11.48)	<.001	9.18 (5.21-16.18)	<.001	4.25 (2.07-8.71)	<.001	5.11 (1.80-14.49)	.002
College and above	15.26 (5.87-39.64)	<.001	25.90 (10.20-65.74)	<.001	8.27 (3.53-19.35)	<.001	9.89 (3.15-30.99)	<.001
Occupation (reference: farming)[b]								
Employed	1.27 (0.93-1.74)	.12	1.19 (0.87-1.64)	.25	1.72 (1.17-2.52)	.006	1.71 (1.01-2.90)	.04
Ethnic group (reference: Han)[b]								
Hui	0.83 (0.61-1.15)	.27	0.75 (0.53-1.04)	.09	1.21 (0.80-1.84)	.36	0.97 (0.54-1.74)	.93
Health need factors								
Members of age <5 years (reference: No)[b]								
Yes	0.91 (0.67-1.22)	.54	0.97 (0.71-1.33)	.86	0.97 (0.66-1.43)	.89	1.11 (0.66-1.86)	.68
Members with chronic conditions (reference: No)[b]								
Yes	1.35 (1.05-1.75)	.02	1.19 (0.91-1.56)	.18	1.13 (0.80-1.59)	.47	1.33 (0.84-2.13)	.22
Outpatient clinic visit in last 1 month (reference: No)[b]								
Yes	1.57 (1.21-2.04)	.001	1.38 (1.05-1.81)	.01	1.99 (1.42-2.78)	<.001	2.00 (1.27-3.16)	.003
Inpatient clinic visit in last 12 months (reference: No)[b]								
Yes	1.63 (1.25-2.12)	<.001	1.30 (0.99-1.71)	.05	1.51 (1.07-2.12)	.01	1.28 (0.80-2.03)	.29
Family enabling factors								
Family income	1.26 (1.10-1.44)	<.001	1.25 (1.09-1.44)	.002	1.44 (1.18-1.75)	<.001	1.19 (0.938-1.52)	.14
Health knowledge score	1.13 (1.01-1.26)	.02	1.08 (0.96-1.21)	.19	1.43 (1.22-1.67)	<.001	1.40 (1.12-1.75)	.002
Health accessibility								
Distance to county hospital	0.61 (0.50-0.77)	<.001	0.68 (0.54-0.85)	.001	0.75 (0.56-1.01)	.06	0.81 (0.54-1.23)	.33
Distance to city hospital	1.26 (1.01-1.55)	.03	1.27 (1.02-1.58)	.03	1.60 (1.19-2.14)	.002	1.80 (1.18-2.76)	.006
Age	0.97 (0.97-0.99)	.006	0.96 (0.95-0.98)	<.001	1.00 (0.98-1.02)	.74	0.98 (0.96-1.01)	.34
Family size	0.99 (0.90-1.08)	.76	0.98 (0.90-1.08)	.76	0.98 (0.87-1.10)	.81	0.98 (0.84-1.15)	.86

[a] OR: odds ratio.

[b] Reference category set to a value of 1.

Discussion

Principal Findings

Our findings indicate that in rural and remote areas, almost half (640/1354, 47.27%) of the surveyed households had used the internet to obtain health-related information or Web-based treatment over the last year. Households engaged in health-related internet use were found to be richer, have higher health demands, and have less geographic access to high-quality health care compared with other households.

As previous studies reported [16,17,26,27], the most common way to use the internet for health is to search for health information. The internet has become a major health information resource in rural Ningxia, although the prevalence is less than that in developed countries [16,17,26,27]. At the same time, the internet provides an alternative for households in rural and remote areas to contact professionals about health issues.

Social structures, family enabling factors, health needs, and characteristics related to health care access were significant predictors of households' health-related internet use in rural and remote areas in China. A more favorable position within the social structure increased the likelihood of Web-based use of health services, as higher education levels or employment potentially increased knowledge and proficiency about internet use as well as eHealth literacy [28]. Family enabling factors, especially household incomes, were found to be an important determinant of Web-based health care use. This finding is consistent with those of other studies [15,17]. As expected, higher household incomes increased the probability of respondents' use of Web-based health-related services. Higher health demands also increased the likelihood of health-related internet use. Households with either inpatients or outpatients were more likely to use the internet to search for health-related information. Our findings related to the final predictor of health access supported a geographic digital divide, as the likelihood of health-related internet use significantly increased among respondents located at greater distances from city-based hospitals. This finding is consistent with that of a previous study [29], which showed that the use of Web-based health services is associated with limited access to health care. The internet may offer a low-cost source of health information and could help meet the heightened demand for health-related information among those facing barriers of access to care. Before the potential confounders were controlled, people who lived nearer to county- or city-based hospitals were more likely to use Web-based health care services. However, after controlling for potential confounders, the residents who live farther from the city hospitals were found to be more likely to use Web-based health care services, which confirms that residents in low-resource settings are more likely to ask for help from the internet. In addition, residents who lived nearer to the county hospitals were found to be more likely to use Web-based health care services even after controlling for the variables. A possible explanation is that the service provision by county hospitals in western regions is still limited and needs further research and deep exploration.

Our findings also showed that different types of Web-based health care users demonstrate similar as well as contrasting characteristics. Although younger participants were more likely to search for health-related information on the internet, the age factor did not significantly influence Web-based communication with doctors or Web-based consultations. Our findings further indicated that farmers were more likely to search for health-related information on the internet than to conduct other Web-based health-related activities. This finding could be attributed to the fact that farmers have lived in rural areas for a long time and have had few opportunities to work in cities, which could reduce the likelihood of their participation in Web-based communication and consultations.

Although health-related knowledge scores had no significant impact on searches for information on health-related issues, participants whose scores were higher were more likely to conduct Web-based consultations and communication. The presence of extensive health-related information on the internet highlights the importance of health literacy. Health literacy was independently related to health knowledge [30]. Respondents with low levels of health literacy are unable to evaluate health information available on the internet, owing to which, they are misled by unsubstantiated information obtained from the internet [31].

The internet may also play a major role in future health care delivery. Although previous studies [15-17] have examined the characteristics of internet users who engage in Web-based health-related activities, this study is one of the first to examine the characteristics of households within rural and remote areas that include different types of health-related internet users.

In summary, the internet is an important alternative to obtain health information and services for rural and remote residents in China. The findings of this study confirm the results of previous studies [15-17]: Socioeconomically disadvantaged individuals (according to age, income, and education levels) are less likely to access health-related information and engage in Web-based consultations. Those who belong to older age groups, have low incomes, and have education levels less than primary school are significantly less likely to use Web-based health care services and benefit from related health care programs.

Implications for Policies, Practices, and Future Studies

Our findings indicate a need for continued collaborative efforts involving multiple stakeholders, including communities, Web-based and other health care providers, family members, and the government. Web-based health-related services and programs should focus on the provision of training for older adults, lower-income households, and households located in remote areas to reduce disparities in health care. Health education programs conducted by communities and health care institutions could integrate internet skills development with health education when targeting the abovementioned key population segments.

More specifically, populations should be educated on ways to acquire high-quality and useful health-related information and services from internet. Health knowledge and promotion programs should ensure that education efforts take into account a population's health literacy skills. Health educators need to be aware of their target population's health literacy skills and adjust educational interventions accordingly. Considering the characteristics of populations in rural and remote areas, researchers and providers of Web-based health-related services should develop simple and engaging methods for providing information and treatment.

From the family perspective, younger family members and those with higher education levels can provide meaningful assistance for older family members to enhance their health-related internet use. Regulations, legal restrictions, and rules should be clarified and strictly implemented in relation to Web-based consultations and the dissemination of health-related information to strengthen the management of marketing activities associated with Web-based health care delivery.

Further studies should examine the effects of internet usage on the delivery of health care services using longitudinal or experimental research designs and focusing especially on remote and rural areas within developing regions. The association between chronic conditions and the characteristics of specific Web-based health services is another important area of inquiry. Studies should attempt to identify the characteristics of acquired health-related knowledge that may be associated with the likelihood of increased use of health care facilities from the internet, as the conditions of knowledge acquisition may help resolve the barriers to Web-based communication and consultation. Furthermore, intervention studies are needed to examine what works for people with low health literacy skills, to improve their health knowledge and evaluate whether increased knowledge or health literacy results in improved health outcomes.

Limitations

The findings of this study provide important insights into the use of internet-based health care services among populations located in remote and rural areas and contribute to the evidence base on the delivery of Web-based health care services from the demand perspective. However, there were several limitations to our analysis. First, as this was a cross-sectional study, meaningful differences could only be considered as correlational, not causal. Second, a comprehensive spectrum of health-related Web-based activities such as Web-based purchases of medicines was not considered. Moreover, the study site (northwestern China) is not representative of other regions within the country.

Despite these limitations, these results have significant implications for future research and the development of computer/internet training for inhabitants of rural and remote areas.

Conclusions

The internet has become a major health information resource in rural Ningxia. Social structures, family enabling factors, health needs, and characteristics related to health care access were significant predictors of households' health-related internet use in rural and remote areas in China. People who belong to older age groups, have low incomes, and have education levels below primary school are significantly less likely to use Web-based health care services and benefit from Web-based health care programs. There is a need for continued collaborative efforts involving multiple stakeholders, including communities, Web-based and other health care providers, family members, and the government.

Acknowledgments

The authors would like to thank Wei Yaqing, Li Qizhe, Zhang Feng, Li Ji, Xu Ning, Yang Kun, Zhang Yajun, Zhu Mengjie, and Yan Fan from the Ningxia Medical University for their field work in this study.

Conflicts of Interest

None declared.

References

1. Li X, Lu J, Hu S, Cheng KK, De Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 9;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
2. Meng Q, Xu L, Zhang Y, Qian J, Cai M, Xin Y, et al. Trends in access to health services and financial protection in China between 2003 and 2011: a cross-sectional study. Lancet 2012 Mar 3;379(9818):805-814. [doi: 10.1016/S0140-6736(12)60278-5] [Medline: 22386034]
3. Liu W, Liu Y, Twum P, Li S. National equity of health resource allocation in China: data from 2009 to 2013. Int J Equity Health 2016 Apr 19;15:68 [FREE Full text] [doi: 10.1186/s12939-016-0357-1] [Medline: 27093960]
4. Meng Q, Yang H, Chen W, Sun QX. People's Republic of China health system review. Health Syst Trans 2015;5(7):2-3 [FREE Full text] [doi: 10.1787/health_glance_ap-2014-en]
5. Cortis N. Overlooked and under-served? Promoting service use and engagement among 'hard-to-reach' populations. Int J Soc Welf 2011;21(4):351 [FREE Full text] [doi: 10.1111/j.1468-2397.2011.00825.x]
6. Alkhaldi B, Sahama T, Huxley C, Gajanayake R. Barriers to implementing eHealth: a multi-dimensional perspective. Stud Health Technol Inform 2014;205:875-879. [Medline: 25160313]
7. Morris ME, Aguilera A. Mobile, social, and wearable computing and the evolution of psychological practice. Prof Psychol Res Pr 2012 Dec;43(6):622-626 [FREE Full text] [doi: 10.1037/a0029041] [Medline: 25587207]
8. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]
9. Wildenbos GA, Peute LW, Jaspers MWM. A framework for evaluating mHealth tools for older patients on usability. Stud Health Technol Inform 2015;210:783-787. [Medline: 25991261]
10. Gibbons MC. EHealth Solutions for Healthcare Disparities. New York: Springer-Verlag; 2018.
11. Alicata D, Schroepfer A, Unten T, Agoha R, Helm S, Fukuda M, et al. Telemental health training, team building, and workforce development in cultural context: the Hawaii experience. J Child Adolesc Psychopharmacol 2016 Apr;26(3):260-265. [doi: 10.1089/cap.2015.0036] [Medline: 26466015]

12. Committee on Pediatric Workforce, Marcin JP, Rimsza ME, Moskowitz WB. The use of telemedicine to address access and physician workforce shortages. Pediatrics 2015 Jul;136(1):202-209 [FREE Full text] [doi: 10.1542/peds.2015-1253] [Medline: 26122802]
13. Central Cyberspace Affairs Commission. The 42nd China statistical report on internet development URL: http://cac.gov.cn/wxb_pdf/CNNIC42.pdf [accessed 2019-05-15] [WebCite Cache ID 78OCPuyNa]
14. Bradford WD, Kleit A, Krousel-Wood MA, Re RM. Comparing willingness to pay for telemedicine across a chronic heart failure and hypertension population. Telemed J E Health 2005 Aug;11(4):430-438. [doi: 10.1089/tmj.2005.11.430] [Medline: 16149888]
15. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
16. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
17. Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. J Med Internet Res 2011;13(4):e110 [FREE Full text] [doi: 10.2196/jmir.1796] [Medline: 22169526]
18. Global Diffusion of eHealth: Making Universal Health Coverage Achievable, Report of the Third Global Survey on eHealth. Washington, DC: World Health Organization; 2016. URL: https://www.who.int/goe/publications/global_diffusion/en/ [accessed 2019-05-21] [WebCite Cache ID 78Xbi2wtp]
19. Li J, Zhang Y, Ma L, Liu X. The impact of the internet on health consultation market concentration: an econometric analysis of secondary data. J Med Internet Res 2016 Oct 28;18(10):e276 [FREE Full text] [doi: 10.2196/jmir.6423] [Medline: 27793793]
20. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017 Jul 4;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
21. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Dec;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
22. Alwan K, Awoke T, Tilahun B. Knowledge and utilization of computers among health professionals in a developing country: a cross-sectional study. JMIR Hum Factors 2015 Mar 26;2(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.4184] [Medline: 27025996]
23. China Knowledge Network. Pecking Union Medical College Press URL: http://kns.cnki.net/kns/detail/detail.aspx?FileName=N2017120225000003&DbName=CSYD2018
24. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
25. Euehun L, Semi H. Determinants of adoption of mobile health services. Online Inf Rev 2015;39(4):556-573 [FREE Full text] [doi: 10.1108/OIR-01-2015-0007]
26. Fox S, Duggan M. Health Online. Washington, DC: Pew Internet & American Life Project; 2013. URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-05-21] [WebCite Cache ID 78Xbi3Emg]
27. Newhouse N, Lupiáñez-Villanueva F, Codagnone C, Atherton H. Patient use of email for health care communication purposes across 14 European countries: an analysis of users according to demographic and health-related factors. J Med Internet Res 2015;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
28. 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;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
29. Bhandari N, Shi Y, Jung K. Seeking health information online: does limited healthcare access matter? J Am Med Inform Assoc 2014;21(6):1113-1117 [FREE Full text] [doi: 10.1136/amiajnl-2013-002350] [Medline: 24948558]
30. Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Edu Counsel 2003 Nov;51(3):267-275 [FREE Full text] [doi: 10.1016/S0738-3991(02)00239-2]
31. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

Abbreviations

eHealth: electronic health
ICT: information and communications technology
OR: odds ratio

Edited by G Eysenbach; submitted 16.11.18; peer-reviewed by S Paige, X Liu, Y Liu; comments to author 03.04.19; revised version received 29.04.19; accepted 07.05.19; published 31.05.19.

Please cite as:
Li H, Xu J, Li L, Meng Q
Health-Related Internet Use in Hard-to-Reach Populations: Empirical Findings From a Survey in a Remote and Mountainous Province in China
J Med Internet Res 2019;21(5):e12693
URL: http://www.jmir.org/2019/5/e12693/
doi:10.2196/12693
PMID:31152526

Original Paper

Evaluating a Video-Based, Personalized Webpage in Genitourinary Oncology Clinical Trials: A Phase 2 Randomized Trial

Rana McKay[1*], MD; Hannah Mills[2*], BSN; Lillian Werner[2], MS; Atish Choudhury[2], MD, PhD; Toni Choueiri[2], MD; Susanna Jacobus[2], MS; Amanda Pace[2], BSN; Laura Polacek[2], BSc; Mark Pomerantz[2], MD; Judith Prisby[2], BSN; Christopher Sweeney[2], MBBS; Meghara Walsh[2], BSN; Mary-Ellen Taplin[2], MD

[1]University of California San Diego, La Jolla, CA, United States
[2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Mary-Ellen Taplin, MD
Department of Medical Oncology
Dana-Farber Cancer Institute
450 Brookline Avenue, Dana 1230
Boston, MA, 02215
United States
Phone: 1 (617) 582 7221
Fax: 1 (617) 632 2165
Email: Mary_Taplin@dfci.harvard.edu

Abstract

Background: The pace of drug discovery and approvals has led to expanding treatments for cancer patients. Although extensive research exists regarding barriers to enrollment in oncology clinical trials, there are limited studies evaluating processes to optimize patient education, oral anticancer therapy administration, and adherence for patients enrolled in clinical trials. In this study, we assess the feasibility of a video-based, personalized webpage for patients enrolled in genitourinary oncology clinical trials involving 1 or more oral anticancer therapy.

Objective: The primary objective of this trial was to assess the differences in the number of patient-initiated violations in the intervention arm compared with a control arm over 4 treatment cycles. Secondary objectives included patient satisfaction, frequently asked questions by patients on the intervention arm, patient-initiated calls to study team members, and patient-reported stress levels.

Methods: Eligible patients enrolling on a therapeutic clinical trial for a genitourinary malignancy were randomized 2:1 to the intervention arm or control arm. Patients randomized to the intervention arm received access to a video-based, personalized webpage, which included videos of patients' own clinic encounters with their providers, instructional videos on medication administration and side effects, and electronic versions of educational documents.

Results: A total of 99 patients were enrolled (89 were evaluable; 66 completed 4 cycles). In total, 71% (40/56) of patients in the intervention arm had 1 or more patient-initiated violation compared with 70% (23/33) in the control arm. There was no difference in the total number of violations across 4 cycles between the 2 arms (estimate=−0.0939, 95% CI−0.6295 to 0.4418, *P* value=.73). Median baseline satisfaction scores for the intervention and control arms were 72 and 73, respectively, indicating high levels of patient satisfaction in both arms. Median baseline patient-reported stress levels were 10 and 13 for the intervention and control arms, respectively, indicating low stress levels in both arms at baseline.

Conclusions: This study is among the first to evaluate a video-based, personalized webpage that provides patients with educational videos and video recordings of clinical trial appointments. Despite not meeting the primary endpoint of reduced patient-initiated violations, this study demonstrates the feasibility of a video-based, personalized webpage in clinical trials. Future research assessing this tool might be better suited for realms outside of clinical trials and might consider the use of an endpoint that assesses patient-reported outcomes directly. A major limitation of this study was the lack of prior data for estimating the null hypothesis in this population.

(J Med Internet Res 2019;21(5):e12044) doi:10.2196/12044

KEYWORDS

cancer; prostatic neoplasms; kidney neoplasms; clinical trial; instructional films and videos; education

Introduction

Background

Over the past several years, the Food and Drug Administration (FDA) has approved several new therapies for the treatment of genitourinary malignancies [1], and this is largely because of patient enrollment in clinical trials [2]. Between June of 2017 and May of 2018, the FDA approved 58 oncology drugs and granted new indications for previously approved drugs, 6 of which were indicated for genitourinary malignancies [3]. Improvement of patient outcomes hinges on well-designed clinical trials [4]. Therefore, it is essential for investigators to optimize clinical trial processes and develop strategies to improve the patient experience on therapeutic clinical trials. Researchers have identified several oncology clinical trial obstacles, which include low patient enrollment, underrepresentation of minorities [5], and data collection burden on the clinical trials system [6].

Given that approximately 25% to 30% of the oncology drug pipeline involves oral anticancer therapy [7], researchers can anticipate new and evolving barriers to oncology clinical trial conduct. Self-administered oral therapies create the added responsibility of ensuring out-of-hospital drug dosing and monitoring. The transition to more oral anticancer therapy has resulted in increased challenges related to patient adherence to prescribed treatment regimens [8]. Adherence is defined as the degree to which patients follow recommendations for day-to-day treatment with respect to the timing, dosing, and frequency of oral anticancer therapy [9]. Nonadherence can result from improper timing of doses, missed doses, incorrect amount, and erratic dosing schedules. Nonadherence has been shown to affect treatment efficacy, leading to worsened survival, resistance, and treatment failure [10-12].

Adherence interventions for patients in clinical trials have typically focused on providing oral anticancer therapy education at the time of clinical trial informed consent or initial investigational drug prescription. Experience demonstrates that trials involving oral anticancer therapy necessitate additional support for patients to optimally manage treatment outside the secure and supportive clinic environment [13].

Before the rise of oral chemotherapy, patients typically received their treatments in the clinic where intravenous drugs could be administered in a safe, protected environment. Now, patients can take their oral chemotherapy at home without direct observation. The level to which adherence is an issue in oncology clinical trials is not clearly defined, and the degree of additional support patients require to properly self-administer oral anticancer therapy has not been quantified. It is known, however, that the strict requirements of a clinical trial (eg, frequent clinic visits, pharmacokinetic blood draws, and medication diary completion) can cause patients to feel overwhelmed and confused. Research has shown that patients can have misconceptions about aspects of research, such as the risk of side effects, trial aims, and the likelihood of personal benefit [14].

Prior Research

Research looking at the use of Web-based interventions and video content to improve oral medication adherence remains largely untapped, but data do suggest that Web-based interventions can effectively increase medication adherence among chronically ill patients, patients undergoing smoking cessation, patients initiating HIV treatment, and in other clinical scenarios [15-17]. Research has also shown that patients, with their increasing aptitude for technology, benefit from innovative approaches that go beyond the written informed consent document, such as interactive Web-based tools that can be used to engage patients and improve understanding of their disease and treatment [18-21]. A systematic review and meta-analysis of studies assessing Web-based interventions compared with usual care or other decision aids within the realm of screening for prostate cancer found that of the 4 studies that used knowledge as an outcome, patients assigned to the Web-based intervention group had higher average knowledge scores than those assigned to usual care [22]. Of the 2 studies that compared Web-based interventions with video decision aids, patients' average knowledge was higher in the group of patients using video.

Phase 3 studies have shown that video content that is tailored to the cancer patient compared with standardized text can effectively improve patient knowledge related to clinical trials and reduce attitudinal barriers [18]. Furthermore, video content has been shown not only to increase knowledge but also to increase patient satisfaction as well [23,24]. Advantages of video and Web-based modes of communication are that they can be stored, shared, and accessed repeatedly by patients and their families. On the basis of current research implications and the evidence supporting video and Web-based interventions to improve knowledge, medication adherence, stress, and satisfaction, we believe creative technology and personalization of communication will improve the patient experience and reinforce the patient care plan.

This Study

In this randomized phase 2 study, we assessed the use of a video-based, personalized webpage via the information-sharing platform, Postwire, to help patients navigate their participation in genitourinary oncology clinical trials involving 1 or more oral anticancer therapy. We used a novel and objective endpoint, patient-initiated protocol violations, for this study to quantitatively evaluate whether a personalized, video-based webpage might help clinical trial patients better adhere to the strict clinical trial guidelines. Patient-initiated protocol violations can be objectively measured and reflect patient misunderstanding regarding medication administration and other protocol procedures. We hypothesized that patients receiving the video-based intervention would have less patient-initiated protocol violations given improved understanding about mediation administration and trial procedures, which would be provided through the video-based, personalized webpage compared with a control arm. In addition, we hypothesized improved patient satisfaction and reduced stress.

Methods

Study Design

This was a randomized, phase 2 clinical trial assessing a video-based, personalized webpage as a complement to standard patient education for clinical trial patients. The institutional review board (IRB) at the Dana-Farber Cancer Institute (DFCI) approved this intervention and determined that this randomized, phase 2 trial did not meet the requirements for registration to a WHO-accredited trial registry as it was an ancillary, noninvasive trial.

The Postwire application was chosen as it is Health Insurance Portability and Accountability Act (HIPAA)–compliant and uses secure sockets layer technology to prevent hacking and ensure privacy. All patients provided written informed consent. This study was pilot-tested before the initiation of patient recruitment and enrollment to optimize workflow and minimize technology malfunctions. A safety run-in of 9 weeks took place to carefully monitor patients and protocol procedures, and adjustments were made as necessary to optimize workflow and study conduct.

The study consisted of 2 arms: the video-based intervention arm and the control arm. Patients were randomized 2:1 to the intervention arm or control arm and were stratified by therapeutic or *parent* clinical trial protocol type (targeted, hormone, or combination therapy). Patients randomized to the control arm received standard clinical trial oral anticancer therapy education and care without a video-based personalized webpage. Patients enrolled in the intervention arm had access to a video-based personalized webpage in addition to standard of care educational materials.

Intervention Arm

Patients assigned to the intervention arm were introduced to the Postwire interface prior to cycle 1 day 1. A designated clinical research coordinator used a tablet to show the patient how to navigate the personalized webpage using a template page. The study team obtained the patient's email address, and within 24 hours of the patient's cycle 1 day 1 clinic visit, a secure email containing a link to the patient's personalized webpage was sent. The next 3 clinic visits with the patient's oncology provider were recorded in real time and uploaded to the patient's personalized webpage. This webpage included videos of the patient's own clinic encounters, their parent clinical trial informed consent, emergency contact information, oral anticancer therapy dosing guidelines, and medication diaries. When new or revised content was uploaded to a patient's personalized webpage, a notification email was sent. In addition to the video recordings of patient's encounters with their providers, the webpage was equipped with templated videos of the patients' own nurses describing how to self-administer oral anticancer therapy (Multimedia Appendix 1), what side effects to expect versus side effects requiring an emergency room visit or a call to the study team, and how to complete study-required documents, such as a medication diary. Patients had the ability to share their webpage access with caregivers (referred to as *patient designees*).

The personalized webpage was available to patients on the intervention arm until 6 full cycles had been completed. Once data collection was complete, the webpage was deleted. If a patient withdrew from the parent, therapeutic clinical trial, or this ancillary trial, the webpage was deactivated. All patients on the intervention arm received paper-based instructional documents and were encouraged to contact the study team with any questions. This intervention was a complement to standard practice rather than a replacement. Patients randomized to the control arm received the same educational content in paper form and were educated by their treating oncologists and care teams according to standard practice.

Study Calendar

This ancillary study's treatment cycles matched the patient's parent clinical trial treatment cycles, which ranged from 21 to 42 days except for 1 trial which had 90-day treatment cycles after the first 28 days. Patients remained on the intervention for 6 cycles (4 cycles of the intervention and 2 cycles of follow-up), withdrawal of consent, or until meeting parent trial discontinuation criteria.

Patient Population

Patients with genitourinary malignancies enrolling in a clinical trial involving 1 or more oral anticancer therapy were eligible if they had consented but not yet started on 1 of several selected parent clinical trials. When a patient was consenting to 1 of the selected parent clinical trials, the provider would discuss this ancillary trial and assess patient interest. Eligible patients were English-speaking and had adequate internet competency and use, as determined by a 3-point scale eligibility questionnaire. Patients' level of functioning and ability to perform activities of daily living as well as physical ability were recorded using the Eastern Cooperative Oncology Group (ECOG) Scale of Performance Status and included in the baseline characteristics. Support was not available to develop videos in multiple languages. Patients who were interested and eligible were consented at the same time they were consented to the parent clinical trial. If the provider was unable to collect consent to this ancillary trial on the day of parent trial consent, it was possible to obtain consent and receive registration and randomization status before the initial cycle 1 day 1 clinic visit. The IRB deemed this trial low risk and determined it acceptable for clinical research coordinators trained on the study to consent patients, if delegated by the principal investigator.

Study Objectives

The primary objective of this trial was to assess the differences in the number of patient-initiated violations in the intervention arm compared with the control arm over 4 treatment cycles. Patient-initiated protocol violations reflect incidents of patient misunderstanding that could potentially be improved upon with the personalized, video-based intervention. Although patient-initiated protocol violations could be measured quantitatively and objectively, the degree to which patient-initiated protocol violations affect the clinical trials process and the rate at which they occur was not known at baseline. This was the rationale for choosing this novel endpoint. Secondary objectives included assessment of patient satisfaction as measured by the Functional Assessment of Chronic Illness Therapy-Treatment Satisfaction-Patient Satisfaction (FACIT-TS-PS) [25], patient perceived stress as measured by the Perceived Stress Scale-10 Item (PSS-10) [26], frequency of webpage use, and number of patient-initiated phone calls to providers.

Patient-Initiated Violations

Patient-initiated violations were defined as events that deviated from instructions detailed in the clinical protocol and including inappropriate drug dosing and operational or safety-related events. All instances of violations were recorded by cycle. To give equal weight to the 11 violations, a given violation, no matter how often it occurred within a cycle, counted once per cycle. For example, a patient would have 2 patient-initiated violations if he missed 2 doses (counted as 1 violation) and self-administered 3 doses at the wrong time (counted as 1 violation). We chose to calculate violations as such because each violation reflected 1 episode of patient misunderstanding. Consequently, a patient completing 4 cycles would have a count ranging from 0 to 44 patient-initiated violations.

Frequency of Webpage Usage

Postwire allowed for internal tracking of user access, and individual patient and patient designee access was recorded. A designated clinical research coordinator documented individual webpage accession on a weekly basis and inserted the data into the electronic data capture system.

Patient Satisfaction and Stress Measures

The FACIT-TS-PS and PSS-10 were collected at each day 1 visit for 6 cycles. The FACIT-TS-PS was used as it is an expansion of the Functional Assessment of Cancer Therapy scale, used specifically in the cancer population. FACIT questionnaires have been robustly validated and address many realms of psychology including physical well-being, social and family well-being, emotional well-being, and functional well-being [27]. The FACIT questionnaire is appropriate for use in patients with any form of cancer, and it has been used and validated in other chronic illnesses, too. The questionnaire can be tailored to include the most relevant questions. Finally, administration time for any 1 assessment is usually less than 15 min.

The PSS-10 predicts both objective biological markers of stress and increased risk for disease among persons with higher perceived stress levels. The PSS-10, which has been externally validated, was chosen because it is one of the most widely used assessments to measure perceived stress and allows patients to appraise what aspects of their lives are stressful [26]. In addition, each question regarding perceived stress is specific to the past month, which is how often most patients on this trial were seen in the clinic. Each question is scored 0 to 4, on a 40-point scale, with scores over 20 representing above average stress. The FACIT-TS-PS contains the following subscales: physician communication, treatment staff communication, technical competence, nurse communication, and confidence and trust. For this study, we did not use the technical competence subscale. On the remaining 4 subscales, there are 23 items, scored 0 to 3, and 1 overall question scored from 0 to 4. This tool is scored on a 73-point scale, with higher scores representing high levels of satisfaction and confidence.

Frequently Asked Questions in the Intervention Arm

Questions asked during the filmed patient encounters were retrospectively reviewed and placed into predefined categories by a dedicated clinical research coordinator assigned to this trial. Questions were collected to assess patient clinical trial–related educational needs for future development of educational materials.

Patient-Initiated Provider Calls

Standard practice among the DFCI genitourinary clinical trial study team is to document any outgoing or incoming phone calls in the patient electronic medical record. In the event that incoming or outgoing calls were not documented, a clinical research coordinator contacted providers whose patients were participating on this ancillary study for additional phone data. This information was entered into the electronic data capture system unique to this study.

Statistical Design

To determine an appropriate sample size, the number of cumulative patient-initiated violations over 4 treatment cycles was simulated under a Poisson distribution, accounting for various proportions of patients without any patient-initiated violations. The intervention and control arms were compared with a 1-sided Wilcoxon rank-sum test 3000 times. On the basis of these simulations, the initial sample size was estimated at 75 patients, 25 patients in the control arm and 50 patients in the intervention arm, resulting in 93% power (1-sided alpha .10) to detect a 1:3 rate of patient-initiated violations in the intervention arm to the control arm equivalent to a rate ratio of 33%. This assumed 50% (12.5/25) of the control arm patients would have 1 or more patient-initiated violation or violations. Power was reduced to 80% if fewer patients experienced 1 or more patient-initiated violation or violations (33%).

Given early dropout of patients from their parent clinical trials (causing subsequent removal from this ancillary study), the sample size was expanded from 75 to 99 patients (66 intervention: 33 control) to enable well-powered analyses based on the original hypothesis. Patient dropout was because of our patient population’s aggressive and advanced disease states. A total of 61 out of 89 (69%) patients were enrolled in parent trials related to metastatic cancers of the genitourinary system (Table 1). Given patient replacement was not permitted on this study, the sample size was expanded to enable well-powered analyses based on the original hypothesis.

Statistical Analysis

On the basis of the statistical design, the primary analysis was to compare the intervention and control arms using the Wilcoxon rank-sum test (1-sided alpha=.10). The primary endpoint was also evaluated as a violation rate (total numbers of violations per cycle), and the comparison between 2 arms was analyzed using a Poisson model. With the Poisson model, patients who dropped off before 4 cycles contributed to the endpoint for as long as they remained in the study. A log-linked model was applied for rate as a function of the predictor variable. By including the log of the treatment cycles into the model, treatment cycles became the denominator of log of the rate of total number of violations. Pearson chi-square was used to estimate the dispersion parameter to account for overdispersion in the model. Standard errors of regression coefficients were adjusted as well. As a sensitivity analysis, exact Wilcoxon rank-sum test adjusting for ties was used to assess whether total number of violations differed by arms in patients who completed all 4 cycles. Descriptive statistics were used to summarize secondary endpoints by arm and cycle.

Results

Baseline Patient Characteristics

A total of 99 patients were enrolled from September 2014 to November 2016 (Multimedia Appendix 2). In total, 90% (89/99) of patients were evaluable, of whom 63% (56/89) were in the intervention arm and 37% (33/89) were in the control arm. A total of 74% (66/89) of evaluable patients completed 4 treatment cycles. Reasons for early study discontinuation (23/89, 26%) included disease progression (17/23, 74%), adverse events on primary study therapy (2/23, 9%), withdrawal (1/23, 4%), and unknown (3/23, 13%). In total, 76% (68/89) of patients on this study had prostate cancer, and 24% (21/89) had renal cell cancer in the final analysis. Baseline and disease characteristics are detailed in Table 1.

Most patients had metastatic disease (61/89, 69%), were enrolled in parent trials of hormonal therapy (54/89, 61%), and had ECOG performance scores of 0 (70/89, 79%). The ECOG performance scale is routinely used to assess how patients' cancers are progressing holistically. A score of 0 means a patient is fully active, and a score of 5 means death [28]. In total, 64% (57/89) were *extremely confident* in their use of the internet, whereas the remaining patients (32/89, 36%) were *somewhat confident* in their use of the internet.

Patient-Initiated Violations

Of the 89 patients analyzed, 71% (40/56) of patients in the intervention arm had 1 or more patient-initiated violations compared with 70% (23/33) of patients in the control arm. Table 2 describes the total violations.

For the intervention arm, the range of violations was 0 to 11, with a median of 1 in patients who completed greater than or equal to 4 cycles. For the control arm, the range of violations was 0 to 7, with a median of 1 in patients who completed greater than or equal to 4 cycles. For all patients who completed greater than or equal to 4 cycles, the range of violations was 0 to 11, with a median of 1. In total, 71% (63/89) had 1 or more violations over the course of 4 cycles.

Improper dosing (, 2/89, 2%), improper concomitant medication administration (5/89, 6%), and wrong doses (5/89, 6%) were among the patient-initiated violations. The most common patient-initiated violations were procedural in nature and included failing to return unused medication for pharmacy drug accountability (16/89, 18%), failing to return medication diaries (10/89, 11%), and incomplete medication diaries (6/89, 7%). There was no difference in the total number of patient-initiated violations across 4 cycles between arms (estimate=−0.0939, 95% CI −0.6295 to 0.4418, *P* value=.73). As a sensitivity analysis, the total number of patient-initiated violations was summarized for patients who completed 4 cycles of the intervention. There was no difference in total patient-initiated violations between arms (*P*=.92) in the 66 patients who completed greater than or equal to 4 cycles.

Table 1. Baseline patient characteristics.

Analysis population baseline characteristics	Arm 1 (N=56), n (%)	Arm 2 (N=33), n (%)	All (N=89), n (%)
Disease stage			
Metastatic	39 (70)	22 (67)	61 (69)
Nonmetastatic	17 (30)	11 (33)	28 (32)
Disease type			
Renal cell carcinoma	12 (21)	9 (27)	21 (24)
Prostate adenocarcinoma	44 (79)	24 (73)	68 (76)
Parent protocol type			
Unknown	1 (2)	2 (6)	3 (3)
Combination therapy	10 (18)	5 (15)	15 (17)
Hormone therapy	35 (63)	19 (58)	54 (61)
Targeted therapy	10 (18)	7 (21)	17 (19)
Eastern Cooperative Oncology Group			
0	45 (80)	25 (76)	70 (79)
1	10 (18)	8 (24)	18 (20)
2	1 (2)	—[a]	1 (1)
Internet use confidence scale			
Extremely confident	38 (68)	19 (58)	57 (64)
Somewhat confident	18 (32)	14 (42)	32 (36)
Frequency of internet access			
1 time a week	1 (2)	1 (3)	2 (2)
2 times a week	1 (2)	1 (3)	2 (2)
5 or more times a week	54 (96)	31 (94)	85 (96)
Number of times per week internet is accessed			
1 time a week	1 (2)	1 (3)	2 (2)
2 times a week	—	1 (3)	1 (1)
3 times a week	—	1 (3)	1 (1)
4 times a week	1 (2)	—	1 (1)
5 or more times a week	54 (96)	30 (91)	84 (94)
All	56 (100)	33 (100)	89 (100)
Baseline characteristics in patients who completed greater than or equal to 4 cycles			
Parent protocol type			
Unknown	1 (2)	1 (4)	2 (3)
Combination therapy	5 (12)	1 (4)	6 (9)
Hormone therapy	34 (79)	16 (70)	50 (76)
Targeted therapy	3 (7)	5 (22)	8 (12)
Disease stage			
Metastatic	26 (60)	12 (52)	38 (58)
Nonmetastatic	17 (41)	11 (48)	28 (42)
All	43 (100)	23 (100)	66 (100)

[a] Not applicable.

Table 2. Summary of overall frequency and type of patient-initiated violations by arm.

Patient-initiated violations by arm	Arm 1 (N=56), n (%)	Arm 2 (N=33), n (%)	All (N=89), n (%)
Total patient-initiated violations in the analysis population			
0	16 (28)	10 (30)	26 (29)
1	15 (26)	9 (27)	24 (27)
2	12 (21)	4 (12)	16 (18)
3	9 (16)	4 (12)	13 (14)
4	1 (1)	3 (9)	4 (4)
5	—[a]	1 (3)	1 (1)
6	1 (1)	1 (3)	2 (2)
7	1 (1)	1 (3)	2 (2)
11	1 (1)	—	1 (1)
All	56 (100)	33 (100)	89 (100)
Summary of type of patient-initiated violations in the analysis population			
Incomplete medication diary			
1	5 (8)	1 (3)	6 (6)
3	2 (3)	2 (6)	4 (4)
4	—	1 (3)	1 (1)
Total	7 (12)	4 (12)	11 (12)
Failure to return medication diary to visit			
1	4 (7)	1 (3)	5 (5)
3	2 (3)	2 (6)	4 (4)
4	—	1 (3)	1 (1)
Total	6 (10)	4 (12)	10 (11)
Failure to return leftover medication to visit			
1	3 (5)	2 (6)	5 (5)
3	8 (14)	2 (6)	10 (11)
4	—	1 (3)	1 (1)
Total	11 (19)	5 (15)	16 (18)
Failure to notify study team of interim cycle adverse event			
1	3 (5)	2 (6)	5 (5)
3	5 (8)	—	5 (5)
4	2 (3)	—	2 (2)
Total	10 (17)	2 (6)	12 (13)
Failure to notify study team of interim emergency department visit			
1	1 (1)	—	1 (1)
3	1 (1)	—	1 (1)
Total	2 (3)	0 (0)	2 (2)
Number of improper doses			
6	—	1 (3)	1 (1)
12	1 (1)	—	1 (1)
Total	1 (1)	1 (3)	2 (2)
Number of improper self con-medication administration			
1	3 (5)	—	3 (3)

Patient-initiated violations by arm	Arm 1 (N=56), n (%)	Arm 2 (N=33), n (%)	All (N=89), n (%)
3	—	1 (3)	1 (1)
12	—	1 (3)	1 (1)
Total	3 (5)	2 (6)	5 (5)
Number of missed appointments			
1	—	1 (3)	1 (1)
5	1 (1)	—	1 (1)
12	—	1 (3)	1 (1)
Total	1 (1)	2 (6)	3 (3)
Number of missed doses			
1	3 (5)	2 (6)	5 (5)
3	5 (8)	3 (9)	8 (9)
6	3 (5)	—	3 (3)
15	1 (1)	—	1 (1)
Total	12 (21)	5 (15)	17 (19)
Number of wrong doses			
1	2 (3)	—	2 (2)
3	2 (3)	1 (3)	3 (3)
Total	4 (7)	1 (3)	5 (5)
Number of doses self-administered at the wrong time			
1	—	1 (3)	1 (1)
3	2 (3)	2 (6)	4 (4)
4	—	1 (3)	1 (1)
Total	2 (3)	4 (12)	6 (6)

[a] Not applicable.

Table 3. Webpage accession by cycle for patients in the intervention arm.

Cycle number and number of times webpage accessed	n (%)
1	
0	17 (30)
1	13 (23)
2	17 (30)
3	3 (5)
4	2 (3)
5	2 (3)
6	1 (1)
14	1 (1)
All	56 (100)
2	
0	29 (53)
1	15 (27)
2	6 (11)
3	3 (5)
4	1 (1)
All	54 (100)
3	
0	30 (66)
1	13 (28)
2	2 (4)
All	45 (100)
4	
0	26 (60)
1	11 (25)
2	3 (7)
3	2 (4)
4	1 (2)
All	43 (100)

Frequency of Webpage Usage

Table 3 describes webpage accession by patients and their designees.

In total, 70% (39/56) of patients assigned to the intervention arm (including 9 designees) accessed the video-based, personalized webpage and its subcontents during cycle 1, 45% (25/56) during cycle 2, 27% (15/56) during cycle 3, and 25% (14/56) during cycle 4. During the 2 follow-up cycles, 5 patients accessed the webpage during cycle 5, and 1 patient accessed the webpage during cycle 6.

Patient Satisfaction and Stress Measures

The median baseline satisfaction score for the intervention arm was 72 and 73 for the control arm (Table 4). Following baseline, the median satisfaction score for both arms at each time point was 73. The median baseline stress level for the intervention arm was 10 and 13 for the control arm (Table 5). Median stress levels for the intervention arm at cycles 2 to 6 were 9, 8, 8, 9, and 7, respectively. Median stress levels for the control arm at cycles 2 to 6 were 11.5, 10.5, 10, 11, and 9.5, respectively. Median stress score for both arms at the end of study (following 6 cycles) was 7.

Table 4. Patient satisfaction scores.

Cycle and arm	Minimum	Median	Maximum
1			
1 (N=37)	53	72	73
2 (N=26)	55	73	73
All (N=63)	53	72	73
2			
1 (N=50)	54	73	73
2 (N=29)	66	73	73
All (N=79)	54	73	73
3			
1 (N=46)	29	73	73
2 (N=26)	68	73	73
All (N=72)	29	73	73
4			
1 (N=34)	60	73	73
2 (N=22)	55	73	73
All (N=56)	55	73	73
5			
1 (N=38)	48	73	73
2 (N=18)	67	73	73
All (N=56)	48	73	73
6			
1 (N=37)	49	73	73
2 (N=18)	68	73	73
All (N=55)	49	73	73
End of study			
1 (N=39)	53	73	73
2 (N=19)	72	73	73
All (N=58)	53	73	73

Table 5. Patient perceived stress scores.

Cycle and arm	Minimum	Median	Maximum
1			
1 (N=54)	0	10	25
2 (N=31)	0	13	28
All (N=85)	0	11	28
2			
1 (N=53)	0	9	24
2 (N=30)	0	11.5	21
All (N-83)	0	11	24
3			
1 (N=40)	0	8	27
2 (N=24)	0	10.5	20
All (N=64)	0	9	27
4			
1 (N=47)	0	8	25
2 (N=27)	0	10	23
All (N=74)	0	8.5	25
5			
1 (N=41)	0	9	25
2 (N=21)	0	11	22
All (N=62)	0	9.5	25
6			
1 (N=41)	0	7	25
2 (N=20)	0	9.5	22
All (N=61)	0	8	25
End of study			
1 (N=43)	0	7	24
2 (N=21)	0	7	24
All (N=64)	0	7	24

Table 6. Number of questions asked during the video encounters, according to type of question and cycle.

Cycle and question type	Median number of questions asked by patient (interquartile range)	Maximum
1 (n=56)		
Drug administration	2 (1-4)	10
Drug handling	0 (0-1)	4
Research team contact	0.5 (0-1)	3
Scheduling	1 (0.5-2)	8
Side effects	1 (0-3.5)	8
Webpage	0 (0-0)	2
2 (n=52)		
Drug administration	0 (0-1)	7
Drug handling	0 (0-0)	4
Research team contact	0 (0-0)	6
Scheduling	0 (0-1)	3
Side effects	0 (0-1)	5
Webpage	0 (0-0)	2
3 (n=44)		
Drug administration	0 (0-2)	4
Drug handling	0 (0-0)	5
Research team contact	0 (0-0)	1
Scheduling	0 (0-1)	7
Side effects	1 (0-2)	7
Webpage	0 (0-0)	1
4 (n=43)[a]		
Drug administration	0 (0-1)	4
Drug handling	0 (0-0)	4
Research team contact	0 (0-0)	1
Scheduling	0 (0-1)	3
Side effects	1 (0-2)	7
Webpage	0 (0-0)	2

[a] Data missing on 1 patient.

Frequently Asked Questions in the Intervention Arm

The most frequently asked questions by patients during the video encounters were tabulated for individuals in the intervention arm. Most questions asked were related to drug administration, drug side effects, and scheduling (Table 6).

Patient-Initiated Provider Calls

Overall, 21% (19/89) of patients in this study called an oncologist or nurse practitioner (Table 7) at least once (range 1-5), and 73% (65/89) of patients on study called a research nurse at least once (range 1-12).

Table 7. Patient-initiated calls to providers.

Cycle and call type	Arm 1, n (%)	Arm 2, n (%)
1		
Calls to MD[a] or NP[b]		
1	2 (4)	2 (6)
Calls to RN[c]		
1	12 (21)	6 18)
2	7 (13)	2 (6)
3	1 (2)	4 (12)
2		
Calls to MD or NP		
1	4 (8)	1 (3)
Calls to RN		
1	10 (20)	9 (30)
2	5 (10)	2 (7)
3	0 (0)	1 (3)
4	3 (6)	0 (0)
5	0 (0)	2 (7)
3		
Calls to MD or NP		
1	1 (2)	0 (0)
2	1 (2)	0 (0)
Calls to RN		
1	11 (24)	4 (15)
2	3 (7)	3 (11)
3	1 (2)	0 (0)
4	2 (4)	0 (0)
4		
Calls to MD or NP		
1	1 (2)	0 (0)
Calls to RN		
1	8 (19)	3 (13)
2	1 (2)	0 (0)

[a] Medical Doctor.

[b] Nurse Practitioner.

[c] Registered Nurse.

Discussion

Principal Findings

In this study, we evaluate a video-based, personalized webpage as a complement to routine clinical trial patient education. Our trial did not prove that the video-based, personalized webpage intervention reduced patient-initiated violations. This result might in part be related to the low total number of patient-initiated violations in both arms. Nearly one-third of patients had no violations, and 27% (24/89) had only 1 violation over the course of 4 cycles. We suspect that given the sophistication needed to seek care at an urban tertiary cancer center and enroll in, not 1 but 2 clinical trials [29], our patient population was perhaps more engaged in their cancer care than the general cancer population, contributing to the low number of patient-initiated violations observed in this trial. Such an intervention might be of more use in a less engaged patient population to aid in adherence of oral anticancer therapy prescribed in a standard of care setting.

Due to the exclusion criteria of the parent clinical trials and the lack of resources necessary to create materials for non-English-speaking patients, only English-speaking patients were included in this study. Research has touched upon the lack of minority representation in clinical trials. Video education has been shown to increase non-English-speaking patients' understanding of consent information compared with oral education and home visits and has helped with recruiting and retaining non-English speaking patients in clinical trials [30]. This video-based, personalized webpage might be of use in the recruitment of a greater number of minority patients in oncology clinical trials.

The FACIT-TS-PS and PSS-10 results show that patients were satisfied with their research teams and stress levels were low in both arms. We suspect that this is because patients who remained on study were deriving clinical benefit from their parent clinical trials, thus impacting their satisfaction and stress levels. Though generally patients were satisfied with their care teams, experienced low levels of stress, and were relatively compliant regarding patient-initiated violations, there remain opportunities for continued improvement of communication and education in the clinical trials process [8,19].

On the basis of patient feedback in the form of emails to the study team, patient caregivers appeared to derive benefit from the intervention. Future work can assess the effect of similar interventions on patient family members or caregivers as a means of supporting the patient and improving the clinical trial experience. Future research might also consider enabling webpage communication functions, allowing researchers to assess the utility of real-time, Web-based modes of communication over standard methods [31]. Video recording by use of iPads required trained personnel in the exam room for each encounter. This approach is not well suited for environments with limited staffing resources. Future studies should consider the use of remote video-recording to eliminate this need. Further development of this intervention should also explore mobile phone accessibility.

Limitations

To objectively quantify the impact of the video-based, personalized webpage, we selected an innovative primary endpoint for this study: patient-initiated violations. Developing the primary endpoint was challenging given the lack of clinical studies of this type. We chose patient-initiated violations because this endpoint was quantifiable and clinically meaningful. Furthermore, we thought that this novel primary endpoint would serve as a surrogate of patient knowledge of trial medication adherence, administration, and overall clinical trial procedures. Defining the null and alternative hypotheses at the onset of the study was also difficult, as there are no published examples of patient-initiated violations in similar patients. An alternative endpoint that has been used in other studies of video-based education delivery includes patient-reported outcomes of knowledge, attitude, and preparation for making decisions about clinical trials [18]. Such an endpoint should be considered in future research assessing this intervention.

The intervention might have broader impact in a less resourced setting or community practice to enhance clinical trial procedures and patient education as well as in settings outside the realm of clinical trials. This trial, which was conducted at a single institution with extensive experience in clinical trial operations, required patients to have access to the internet and some degree of confidence in internet use. These enrollment criteria might have biased the patient population to those with the most potential for understanding clinical trial procedures and excluded those who could have experienced the greatest benefit.

Comparison With Prior Work

Strategies to incorporate Web-based and mobile communication on clinical trials are underway, such as the Preparatory Education About Clinical Trials, a Web-based and interactive computer program delivering educational content to patients considering clinical trials, and a Web-based prostate cancer treatment decision aid assessing treatment satisfaction, decisional regret, and quality of life [18,21,32]. Advances in technology, such as video-based interventions and mobile apps, are expected to impact the delivery of oncology care over the next decade [33]. Advantages of e-technology (computer-assisted interventions and mobile phone apps) in clinical trials include improved efficiency, cost reduction, and fostering research and development [32]. These interventions have the potential to enhance patient understanding, improve patient enrollment, and streamline clinic operations [18].

Conclusions

As the population in general becomes more skilled with technology, we believe this type of intervention will be useful for patients and providers of varying disciplines. Although the primary endpoint of reduced patient-initiated violations over 4 cycles of treatment in the intervention arm was not met, this study demonstrates the feasibility of providing patients and caregivers with instant, at-home access to the discussions that take place between patients and their care teams. There was no difference between arms at the end of 4 cycles or following 2 cycles of follow-up. Nevertheless, our video-based personalized webpage offered patients an alternative to impersonal, lengthy, and written information as is currently provided in informed consent documents. Our intervention allowed patients to store, share, and access complex discussions related to their cancer care repeatedly over extended periods. With the ever-increasing role of technology in health care, the potential of a video-based, personalized webpage in oncology care includes optimization of oral anticancer therapy adherence, management of drug side effects, increase in patient safety, and improved trial operational quality and efficacy in a sustainable and translatable way that could benefit patients and providers to come.

Acknowledgments

The authors would like to thank the patients and caregivers who participated in this clinical trial and Dr Kate Burke for introducing our team to the Postwire webpage. The authors also acknowledge support from Fairweather Family Fund and Fat Boys Slim Sisters Fund (MET).

Conflicts of Interest

RM received research funding from Bayer and Pfizer and serves on the advisory board for Novartis and Janssen. MET received research funding from Medivation and Janssen. TC received research funding from AstraZeneca, Alexion, Bayer, Bristol Myers-Squibb/ER Squibb and sons LLC, Cerulean, Eisai, Foundation Medicine Inc, Exelixis, Ipsen, Tracon, Genentech, Roche, Roche Products Limited, F. Hoffman-La Roche, GlaxoSmithKline, Lilly, Merck, Novartis, Peloton, Pfizer, Prometheus Labs, Corvus, Calithera, Analysis Group, Sanofi/Aventis, and Takeda and serves in a consulting or advisory role for AstraZeneca, Alexion, Sanofi/Aventis, Bayer, Bristol Myers-Squibb/ER Squibb and sons LLC, Cerulean, Eisai, Foundation Medicine Inc, Exelixis, Genentech, Heron Therapeutics, Roche, GlaxoSmithKline, Merck, Novartis, Peloton, Pfizer, EMD Serono, Prometheus Labs, Corvus, Lilly, Ipsen, Up-to-Date, NCCN, and Analysis Group.

Multimedia Appendix 1

Your clinical trial research nurse explains how to take your doses of abiraterone acetate and prednisone each day.

[MOV File, 20MB - jmir_v21i5e12044_app1.MOV]

Multimedia Appendix 2

Consolidated Standards of Reporting Trials (CONSORT) diagram.

[PDF File (Adobe PDF File), 113KB - jmir_v21i5e12044_app2.pdf]

Multimedia Appendix 3

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2MB - jmir_v21i4e12044_app3.pdf]

References

1. Simon S. American Cancer Society. 2017. New Cancer Drug Approvals From 2017 URL: https://www.cancer.org/latest-news/new-cancer-drug-approvals-from-2017.html [accessed 2018-08-06] [WebCite Cache ID 71TI8BU0x]
2. Brown RF, Butow PN, Boyle F, Tattersall MH. Seeking informed consent to cancer clinical trials; evaluating the efficacy of doctor communication skills training. Psychooncology 2007 Jun;16(6):507-516. [doi: 10.1002/pon.1095] [Medline: 16986176]
3. ASCO Post. FDA Oncology Drug Approvals Granted Between June 2017 and May 16, 2018 URL: http://www.ascopost.com/issues/june-3-2018-narratives-special-issue/fda-oncology-drug-approvals-granted-between-june-2017-and-may-16-2018/ [accessed 2018-08-06] [WebCite Cache ID 71TIMepp5]
4. Goldberg RM, Wei L, Fernandez S. The evolution of clinical trials in oncology: defining who benefits from new drugs using innovative study designs. The Oncologist 2017;22(9):1015-1019. [Medline: 28620092]
5. Nass SJ, Moses HL, Mendelsohn J, editors. Physician and patient participation in cancer clinical trials. In: A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program. Washington, DC: National Academies Press (US); 2010.
6. Dwyer JT, editor. Regulatory issues. In: Implementing a National Cancer Clinical Trials System for the 21st Century: Second Workshop Summary. Washington, DC: National Academies (US); 2013.
7. Dobbels F, Van Damme-Lombaert R, Vanhaecke J, De Geest S. Growing pains: non-adherence with the immunosuppressive regimen in adolescent transplant recipients. Pediatr Transplant 2005 Jun;9(3):381-390. [doi: 10.1111/j.1399-3046.2005.00356.x] [Medline: 15910397]
8. Schneider SM, Hess K, Gosselin T. Interventions to promote adherence with oral agents. Semin Oncol Nurs 2011 May;27(2):133-141 [FREE Full text] [doi: 10.1016/j.soncn.2011.02.005] [Medline: 21514482]
9. Huang WC, Chen CY, Lin SJ, Chang CS. Medication adherence to oral anticancer drugs: systematic review. Expert Rev Anticancer Ther 2016;16(4):423-432. [doi: 10.1586/14737140.2016.1159515] [Medline: 26935964]
10. Gater A, Heron L, Abetz-Webb L, Coombs J, Simmons J, Guilhot F, et al. Adherence to oral tyrosine kinase inhibitor therapies in chronic myeloid leukemia. Leuk Res 2012 Jul;36(7):817-825. [doi: 10.1016/j.leukres.2012.01.021] [Medline: 22364811]

11. Ganesan P, Sagar TG, Dubashi B, Rajendranath R, Kannan K, Cyriac S, et al. Nonadherence to imatinib adversely affects event free survival in chronic phase chronic myeloid leukemia. Am J Hematol 2011 Jun;86(6):471-474 [FREE Full text] [doi: 10.1002/ajh.22019] [Medline: 21538468]
12. Hoffner B, Bauer-Wu S, Hitchcock-Bryan S, Powell M, Wolanski A, Joffe S. "Entering a Clinical Trial: Is it Right for You?": a randomized study of The Clinical Trials Video and its impact on the informed consent process. Cancer 2012 Apr 01;118(7):1877-1883 [FREE Full text] [doi: 10.1002/cncr.26438] [Medline: 22009665]
13. Bassan F, Peter F, Houbre B, Brennstuhl MJ, Costantini M, Speyer E, et al. Adherence to oral antineoplastic agents by cancer patients: definition and literature review. Eur J Cancer Care (Engl) 2014 Jan;23(1):22-35. [doi: 10.1111/ecc.12124] [Medline: 24079854]
14. Moorcraft SY, Marriott C, Peckitt C, Cunningham D, Chau I, Starling N, et al. Patients' willingness to participate in clinical trials and their views on aspects of cancer research: results of a prospective patient survey. Trials 2016 Jan 09;17:17 [FREE Full text] [doi: 10.1186/s13063-015-1105-3] [Medline: 26745891]
15. Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JC. Effects of eHealth interventions on medication adherence: a systematic review of the literature. J Med Internet Res 2011;13(4):e103 [FREE Full text] [doi: 10.2196/jmir.1738] [Medline: 22138112]
16. Graham AL, Papandonatos GD, Cha S, Erar B, Amato MS, Cobb NK, et al. Improving adherence to smoking cessation treatment: intervention effects in a web-based randomized trial. Nicotine Tob Res 2017 Mar 01;19(3):324-332 [FREE Full text] [doi: 10.1093/ntr/ntw282] [Medline: 27836982]
17. Neumann MS, Plant A, Margolis AD, Borkowf CB, Malotte CK, Rietmeijer CA, et al. Effects of a brief video intervention on treatment initiation and adherence among patients attending human immunodeficiency virus treatment clinics. PLoS One 2018;13(10):e0204599 [FREE Full text] [doi: 10.1371/journal.pone.0204599] [Medline: 30289884]
18. Meropol NJ, Wong Y, Albrecht T, Manne S, Miller SM, Flamm AL, et al. Randomized trial of a web-based intervention to address barriers to clinical trials. J Clin Oncol 2016 Feb 10;34(5):469-478 [FREE Full text] [doi: 10.1200/JCO.2015.63.2257] [Medline: 26700123]
19. Stacey D, Bennett CL, Barry MJ, Col NF, Eden KB, Holmes-Rovner M, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2011 Oct 05(10):CD001431. [doi: 10.1002/14651858.CD001431.pub3] [Medline: 21975733]
20. Song L, Rini C, Deal AM, Nielsen ME, Chang H, Kinneer P, et al. Improving couples' quality of life through a web-based prostate cancer education intervention. Oncol Nurs Forum 2015 Mar;42(2):183-192 [FREE Full text] [doi: 10.1188/15.ONF.183-192] [Medline: 25806885]
21. Cuypers M, Lamers RE, Kil PJ, van de Poll-Franse LV, de Vries M. Impact of a web-based treatment decision aid for early-stage prostate cancer on shared decision-making and health outcomes: study protocol for a randomized controlled trial. Trials 2015 May 27;16:231 [FREE Full text] [doi: 10.1186/s13063-015-0750-x] [Medline: 26012700]
22. Baptista S, Teles Sampaio E, Heleno B, Azevedo LF, Martins C. Web-based versus usual care and other formats of decision aids to support prostate cancer screening decisions: systematic review and meta-analysis. J Med Internet Res 2018 Jun 26;20(6):e228 [FREE Full text] [doi: 10.2196/jmir.9070] [Medline: 29945857]
23. Denny MC, Vahidy F, Vu KY, Sharrief AZ, Savitz SI. Video-based educational intervention associated with improved stroke literacy, self-efficacy, and patient satisfaction. PLoS One 2017;12(3):e0171952 [FREE Full text] [doi: 10.1371/journal.pone.0171952] [Medline: 28333925]
24. van Eck CF, Toor A, Banffy MB, Gambardella RA. Web-based education prior to outpatient orthopaedic surgery enhances early patient satisfaction scores: a prospective randomized controlled study. Orthop J Sports Med 2018 Jan;6(1):2325967117751418 [FREE Full text] [doi: 10.1177/2325967117751418] [Medline: 29399589]
25. Peipert JD, Beaumont JL, Bode R, Cella D, Garcia SF, Hahn EA. Development and validation of the functional assessment of chronic illness therapy treatment satisfaction (FACIT TS) measures. Qual Life Res 2014 Apr;23(3):815-824. [doi: 10.1007/s11136-013-0520-8] [Medline: 24062239]
26. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
27. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Health Qual Life Outcomes 2003 Dec 16;1:79 [FREE Full text] [doi: 10.1186/1477-7525-1-79] [Medline: 14678568]
28. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982 Dec;5(6):649-655. [Medline: 7165009]
29. Begg CB, Zelen M, Carbone PP, McFadden ET, Brodovsky H, Engstrom P, et al. Cooperative groups and community hospitals. Measurement of impact in the community hospitals. Cancer 1983 Nov 01;52(9):1760-1767. [Medline: 6616426]
30. Hughson J, Woodward-Kron R, Parker A, Hajek J, Bresin A, Knoch U, et al. A review of approaches to improve participation of culturally and linguistically diverse populations in clinical trials. Trials 2016 Dec 26;17(1):263 [FREE Full text] [doi: 10.1186/s13063-016-1384-3] [Medline: 27229153]

31. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. J Am Med Assoc 2017 Dec 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
32. Rosa C, Campbell AN, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. Contemp Clin Trials 2015 Nov;45(Pt A):41-54 [FREE Full text] [doi: 10.1016/j.cct.2015.07.007] [Medline: 26176884]
33. Aitken M, Kleinrock M, Simorellis A, Nass D. IQVIA Institute for Human Data. 2018. Global Oncology Trends 2018: Innovation, Expansion and Disruption URL: https://www.iqvia.com/institute/reports/global-oncology-trends-2018 [accessed 2018-08-06] [WebCite Cache ID 71TMJBe3u]

Abbreviations

DFCI: Dana-Farber Cancer Institute
ECOG: Eastern Cooperative Oncology Group
FACIT: Functional Assessment of Chronic Illness Therapy
FACIT-TS-PS: Functional Assessment of Chronic Illness Therapy-Treatment Satisfaction-Patient Satisfaction
FDA: Food and Drug Administration
IRB: institutional review board
PSS-10: Perceived Stress Scale-10 Item

Edited by G Eysenbach; submitted 27.08.18; peer-reviewed by E Arden-Close; comments to author 14.10.18; revised version received 08.12.18; accepted 31.12.18; published 02.05.19.

Please cite as:
McKay R, Mills H, Werner L, Choudhury A, Choueiri T, Jacobus S, Pace A, Polacek L, Pomerantz M, Prisby J, Sweeney C, Walsh M, Taplin ME
Evaluating a Video-Based, Personalized Webpage in Genitourinary Oncology Clinical Trials: A Phase 2 Randomized Trial
J Med Internet Res 2019;21(5):e12044
URL: https://www.jmir.org/2019/5/e12044/
doi:10.2196/12044
PMID:31045501

Review

Use of Commercial Off-The-Shelf Devices for the Detection of Manual Gestures in Surgery: Systematic Literature Review

Fernando Alvarez-Lopez[1,2*], MD; Marcelo Fabián Maina[3*], PhD; Francesc Saigí-Rubió[1*], PhD

[1]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
[2]Faculty of Health Sciences, Universidad de Manizales, Caldas, Colombia
[3]Faculty of Psychology and Education Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
[*]all authors contributed equally

Corresponding Author:
Francesc Saigí-Rubió, PhD
Faculty of Health Sciences
Universitat Oberta de Catalunya
Avinguda del Tibidabo 39-43
Barcelona, 08035
Spain
Phone: 34 933263622
Email: fsaigi@uoc.edu

Abstract

Background: The increasingly pervasive presence of technology in the operating room raises the need to study the interaction between the surgeon and computer system. A new generation of tools known as commercial off-the-shelf (COTS) devices enabling touchless gesture–based human-computer interaction is currently being explored as a solution in surgical environments.

Objective: The aim of this systematic literature review was to provide an account of the state of the art of COTS devices in the detection of manual gestures in surgery and to identify their use as a simulation tool for motor skills teaching in minimally invasive surgery (MIS).

Methods: For this systematic literature review, a search was conducted in PubMed, Excerpta Medica dataBASE, ScienceDirect, Espacenet, OpenGrey, and the Institute of Electrical and Electronics Engineers databases. Articles published between January 2000 and December 2017 on the use of COTS devices for gesture detection in surgical environments and in simulation for surgical skills learning in MIS were evaluated and selected.

Results: A total of 3180 studies were identified, 86 of which met the search selection criteria. Microsoft Kinect (Microsoft Corp) and the Leap Motion Controller (Leap Motion Inc) were the most widely used COTS devices. The most common intervention was image manipulation in surgical and interventional radiology environments, followed by interaction with virtual reality environments for educational or interventional purposes. The possibility of using this technology to develop portable low-cost simulators for skills learning in MIS was also examined. As most of the articles identified in this systematic review were proof-of-concept or prototype user testing and feasibility testing studies, we concluded that the field was still in the exploratory phase in areas requiring touchless manipulation within environments and settings that must adhere to asepsis and antisepsis protocols, such as angiography suites and operating rooms.

Conclusions: COTS devices applied to hand and instrument gesture–based interfaces in the field of simulation for skills learning and training in MIS could open up a promising field to achieve ubiquitous training and presurgical warm up.

(J Med Internet Res 2019;21(5):e11925) doi:10.2196/11925

KEYWORDS

minimally invasive surgery; user-computer interface; operating room; education, medical; computer-assisted surgery

Introduction

Background

The increasingly pervasive presence of technology in the operating room raises the need to study the interaction between the surgeon and computer system. In sterile environments, using the hand to operate a mouse, keyboard, or touchscreen is unacceptable as it alters the normal pace of surgery and breaks asepsis and antisepsis protocols [1-6]. Using a physical barrier between the surgeon's gloves and the interaction device [7], or the foot for manipulation, are not practical solutions either, as they do not allow fine interaction and carry risks of contamination [8]. Moreover, using a person to manipulate images in accordance with the surgeon's verbal instructions has proven difficult and is prone to giving rise to misunderstandings when the visualization of specific areas of the image are requested [9,10].

Early solutions to circumvent any contact between the surgeon and computer were based on voice recognition Automated Endoscopic System for Optimal Positioning (AESOP) and HERMES (Stryker Europe) [11,12], but these systems were impractical as they were difficult to use when performing complex tasks [13]. Natural user interfaces were first developed in the 1990s to enable interaction with the computer through natural human movements to manipulate radiological images in sterile surgical environments [14]. Gesture-based interfaces were another variant [15]. These enabled touchless manipulations to be performed and held great promise as a viable solution in the operating room and autopsy suites [10,16-19]. However, they could not be employed in sterile environments as they required some contact when gloves or position sensors were used [20-24].

Early attempts to use touchless gestures in minimally invasive surgery (MIS) involved hand and facial gestures [9,25]. Gesture recognition systems with Web and video cameras were later described [26,27] using the time-of-flight principle [28] and achieving interaction with the OsiriX viewer [17,29]. However, these systems were very expensive and inaccurate and required calibration and a complex setup, making them impractical for use in the operating room [30].

A new generation of tools known as commercial off-the-shelf (COTS) devices enabling touchless gesture–based human-computer interaction is currently being explored as a solution in surgical environments. The term COTS refers to a device that can be taken from a shelf, that is, sold over the counter. In addition to being low-cost, wireless, and ergonomic, they facilitate real-time interactivity and allow the user to point to and manipulate objects with 6 degrees of freedom [31]. Hansen et al described the use of the Wii Remote (Nintendo) for the intraoperative modification of resection planes in liver surgery [32], whereas Gallo et al used it for pointing to and manipulating 3-dimensional (3D) medical data in a number of ways [31,33-36]. However, intraoperative manipulation of the device required it to be wrapped in a sterile bag, thus eliminating the concept of contactless. In November 2010, the Microsoft Kinect (MK) 3D depth camera system (Microsoft Corp) was launched as a device for the Xbox 360 games console. The first descriptions of MK for medical use were in relation to physical and cognitive rehabilitation [37]. Subsequent experiences in this field showed that additional studies were required on issues such as effectiveness, commitment, and usability [38-40]. Its use in an operating room was first reported in 2011, at Sunnybrook Hospital in Toronto, when it was used to view magnetic resonance imaging and computed tomography scans, eventually giving rise to the GestSure system [13]. In 2012, the Leap Motion Controller (LMC; Leap Motion Inc) was launched, and in July 2013, the Myo armband (Thalmic Labs) was launched.

Construct validity [41,42], concurrent validity [43,44], and predictive validity [45,46] studies, as well as systematic reviews [47,48], have shown that simulation in virtual reality environments is an effective tool for motor skills learning in MIS. However, the high cost of virtual reality and augmented reality simulators calls for the development of new, portable low-cost solutions enabling ubiquitous learning. New COTS technologies that allow hand gestures and instrument movements to be detected open up an interesting field of exploration for the development and validation of new simulation models in virtual environments. One of the objectives of this systematic review was to recognize the existence of developments in this area.

Objectives

The aim of this systematic review was to provide an account of the state of the art of COTS devices in the detection of manual gestures in surgery and to identify their use as a simulation tool for motor skills teaching in MIS.

Methods

Article Retrieval

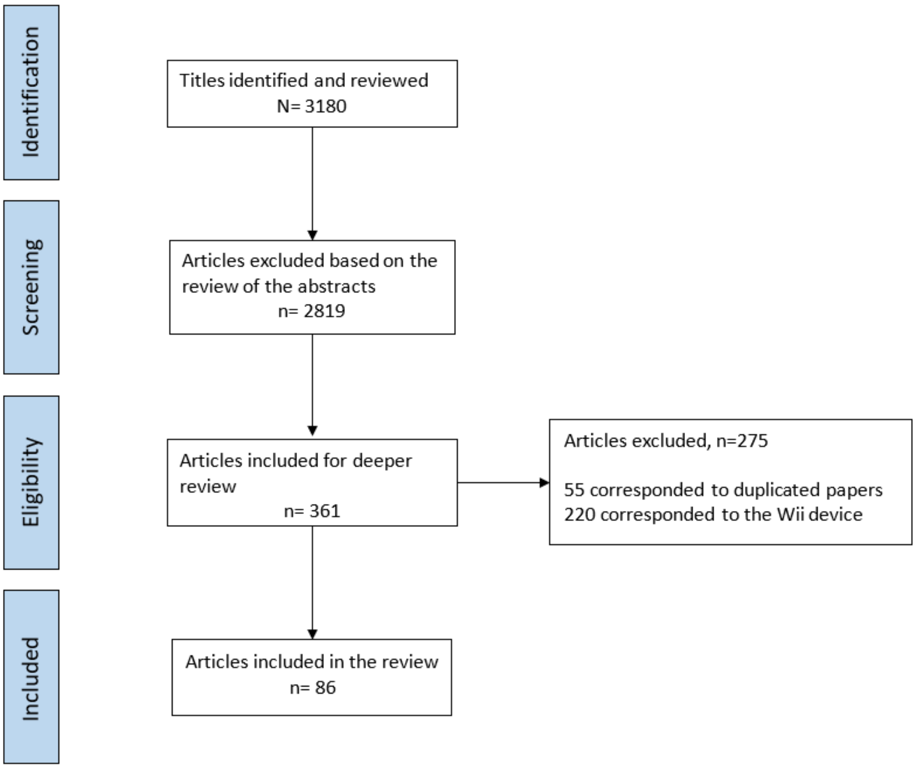
A search was conducted in the electronic databases PubMed, Excerpta Medica database (EMBASE), ScienceDirect, Espacenet, OpenGrey, and the Institute of Electrical and Electronics Engineers (IEEE) for articles published between January 2000 and December 2017, using combinations of the following Medical Subject Headings (MeSH) terms: *surgery*, *computer simulation*, *simulation training*, *laparoscopy*, *minimally invasive surgical procedures*, *robotic surgical procedures*, and *virtual reality*. The following were used as free terms: *commercial off-the-shelf*, *COTS*, *surgical education*, *surgical simulation*, *Wii*, *Microsoft Kinect*, *Xbox Kinect*, *Leap Motion*, *Leap Motion Controller*, *Myo armband*, and *gesture control*. The search strategy used a combination of MeSH terms and free terms. Boolean operators (AND and OR) were used to expand, exclude, or join keywords in the search. The devised strategy was applied first to PubMed and then to the remaining databases.

The search was limited to English-language publications and was complemented using the snowballing technique to identify relevant articles in the references of articles returned by our search [49]. A manual search was also conducted on the indices of the following publications: *Surgical Endoscopy*, *Surgical Innovation*, *Minimally Invasive Therapy and Allied Technologies*, the *Journal of Medical Internet Research*, and the *Journal of Surgical Education*. The snowballing search and the manual reviews enabled the retrieval of conference proceedings, letters to the editor, and simple concept descriptions. A MeaSurement Tool to Assess systematic Reviews (AMSTAR) [50] and Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [51] checklists were used to ensure the quality of the review. In total, 3 authors assessed the risk of bias. Disagreement on bias assessment and the interpretation of results was resolved by consensus discussions.

Study Selection

A total of 3180 studies were identified, and the abstracts were reviewed to determine whether they met the inclusion and exclusion criteria. The inclusion criteria were (1) original research articles, (2) proof-of-concept or prototype user testing and feasibility testing studies, (3) studies conducted in surgical environments (preoperative, intraoperative, or postoperative), and (4) studies carried out in real or simulated surgical settings. The exclusion criteria were (1) studies on COTS devices requiring hand contact, (2) studies conducted in nonsurgical clinical environments, and (3) studies on the technical description of devices that did not include criteria of clinical usability, feasibility, or acceptance as an outcome. Studies on COTS devices requiring hand contact (ie, Wii) were excluded from the analysis. After the first review of the titles and abstracts, 361 studies were selected, 220 of which corresponded to the Wii device and were therefore discarded. Of the 141 remaining articles, 55 were duplicate references. After reading the full texts of these studies, 86 were deemed to have met the search selection criteria. The search and selection processes are summarized in Figure 1.

Figure 1. Flow diagram of studies through the review.



We used a standardized form for data extraction, which included the following items: study, device on which the study was conducted, year of publication, aim, type of study, intervention, metrics, sample, and results and conclusions; clinical areas in which the study was conducted and types of surgical intervention (Tables 1-4) (see Multimedia Appendices 1-3 for the full Tables 1-3) and use of gesture-based COTS devices in surgery (Table 5). In total, 2 authors (FAL and MM) screened all the articles individually. Discrepancies were always resolved through discussion with the senior author (FSR) whenever necessary. All the data were analyzed qualitatively and quantitatively.

Results

Of the 86 articles identified, 43 (50%) were on MK, 31 (36%) were on the LMC, 2 compared MK with the LMC [77,113], 1 compared the LMC with the Myo armband [58], 1 compared MK with the LMC and the Myo armband [52], 6 were on web, video, or commercial cameras (7%), and 2 reviewed gesture interaction in general [59,65]. The data and detailed information on the studies reviewed are shown in Tables 1-3 (see Multimedia Appendices 1-3 for the full Tables 1-3). The results are organized by the type of COTS device used (Tables 1-3, see Multimedia Appendices 1-3 for the full Tables 1-3), by the type of surgical specialties in which COTS devices were used (Table 4), and by the type of use made of COTS devices in surgery, including simulation for motor skills learning (Table 5).

Table 1. Summary of included studies evaluating Microsoft Kinect.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[17]	To describe a system for the interactive exploration of medical images through a gesture-controlled interface using MK[a].	Proof-of-concept.	Manipulation of CT[b], MRI[c] and Positron emission tomography images.	Not described.	As the interface does not require direct contact or calibration, it is suitable for use in the operating room.
[99]	To explore the potential simplifications derived from using 3D[d] sensors in medical augmented reality applications by designing a low-cost system.	Proof-of-concept.	Augmented reality in Medicine.	Not described.	The concept is feasible but the whole process is still too time-consuming to be executed in real time.
[101]	To present an augmented reality magic mirror for anatomy teaching.	Proof-of-concept.	Augmented reality in Medicine. Anatomy education.	A hospital and a school.	The system can be used for educational purposes, to improve communication between doctor and patients. A possible use for anatomy teaching in surgery is not mentioned.
[5]	To evaluate the response time and usability (gestures and voice commands) compared with mouse and keyboard controls.	Prototype user testing and feasibility testing.	Manipulation of CT images.	2 radiologists and 8 forensic pathologists who recreated 12 images.	Users took 1.4 times longer to recreate an image with gesture control and rated the system 3.4 out of 5 for ease of use in comparison with the keyboard and mouse. The voice recognition system did not work properly.
[84]	To develop a system to allow the surgeon to interact with the standard PACS system during sterile surgical management of orthopedic patients.	Proof-of-concept.	Manipulation of radiological images in orthopedics.	Not described.	This is the first example of this technology being used to control digital X-rays in clinical practice.
[83]	To present a sterile method for the surgeon to manipulate images using touchless freehand gestures.	Experiment.	Manipulation of MRI images.	9 veterinary surgeons. 22 students.	The hypothesis that contextual information integrated with hand trajectory gesture information can significantly improve the overall recognition system performance was validated. The recognition accuracy was 98.7%
[76]	To evaluate an MK-based interaction system for manipulating imaging data using 'Magic Lens visualization.‘	Proof-of-concept in the operating room.	Manipulation of radiological images.	A laryngoplasty.	The surgeon can manipulate the preoperative information with the intraoperative video and the simulations to correctly place the implant.
[79]	To compare the accuracy and speed of interaction of MK with that of a mouse. To study the performance of the interaction methods in rotation tasks and localization of internal structures in a 3D dataset.	User testing.	Manipulation of radiological images.	15 users.	The gesture-based interface outperformed the traditional mouse with respect to time and accuracy in the orientation and rotation task. The mouse was superior in terms of accuracy of localization of internal structures. However, the gesture-based interface was found to have the fastest target localization time.
[74]	To develop a user-friendly touchless system for controlling the presentation of medical images based on hand gesture recognition in the operating room.	Proof-of-concept in the operating room.	Manipulation of radiological images in orthopedic surgery.	Not described.	The system does not require calibration and was adapted to the surgical environment following the principles of asepsis/antisepsis.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[30]	To present a touchless gesture interface that allows the surgeon to control medical images using hand gestures.	Proof-of-concept and prototype feasibility testing.	Manipulation of CT images.	Enucleation of 4 tumors in 3 urology patients.	First description in the literature of a gesture user interface using MK in the operating room in in-vivo surgery, showing that it is an efficient and low-cost solution.
[100]	To develop a low-cost augmented reality interface projected onto a mannequin simulator.	Proof-of-concept.	Augmented reality for education in Medicine.	A physical simulator, video projector, Wii Remote and MK.	The manipulations obtained using MK were similar to those described with the Wii.
[67]	To develop a version of a gesture-based system for controlling images.	Proof-of-concept.	Manipulation of MRI images.	Resection of a glioma.	Except for the scanning movement, each movement was recognized with great accuracy. The algorithm can be installed in the clinical area.
[128]	To use MK to operate an automated operating-room light system.	Prototype user testing.	Manipulation of operating room lights.	18 volunteers.	The gestures were easy to learn and the movement of the light beam was sufficiently precise.
[102]	To create a touchless head tracking system for an immersive virtual operating room.	Proof-of-concept.	Virtual reality for simulation and education in surgery.	A 3D virtual operating room with a virtual operating table.	Using MK, it was possible to implement a very accurate interactive tracking system regardless of the complexity of the virtual reality system.
[85]	To present a new prototype that allows the user to control the OsiriX system with finger gestures using a low-cost depth camera.	Proof-of-concept and prototype feasibility testing.	Manipulation of CT images.	4 forensic pathologists, 1 radiologist and 1 engineer.	On average, 4.5 min were required to learn to use the system. Participants rated the intuitiveness of the gestures with 3.8 out of 5 and control of the images with 3.8 out of 5. The low cost of the system makes it affordable for any potential user.
[104]	To present a new immersive surgical training system.	Proof-of-concept and prototype fidelity testing.	Virtual reality for education in surgery.	Cholecystectomy training on animal tissue blocks.	Initial feedback from the residents showed that the system is much more effective than the conventional videotaped system.
[60]	To test a speech and gesture-controlled interventional radiology system.	User testing.	Manipulation of CT and angiography images.	10 radiology residents used commands under different lighting conditions during 18 angiographies and 10 CT- guided punctures.	93% of commands were recognized successfully. Speech commands were less prone to errors than gesture commands. 60% of participants would use the application in their routine clinical practice.
[86]	To develop an image operation system for image manipulation using a motion sensor.	Proof-of-concept.	Manipulation of angiographic images.	Not described.	The system can be implemented as a useful tool in angiography for controlling image viewing using gestures in the operating room.
[19]	The working hypothesis is that contextual information such as the focus of attention, integrated with gestural information, can significantly improve overall system recognition performance compared with interfaces relying on gesture recognition alone.	Ethnographic study. Experiment. Survey.	Manipulation of MRI images.	10 veterinary surgeons. 20 volunteers.	The surgeon’s intention to perform a gesture can be accurately recognized by observing environmental cues (context). The hypothesis was validated by a drop in the false positive rate of gesture recognition from 20.76% to 2.33%. A significant rate of reduction of the mean task completion time indicated that the user operates the interface more efficiently with experience. The tracking algorithm occasionally failed in the presence of several people in the camera’s field of view.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[96]	To examine the functionality and usability of MK to complete the visualization of 3D anatomical images.	User testing. Survey.	Manipulation of anatomical images.	32 participants: Medical students, professors and anatomy laboratory staff.	MK users reached accuracy levels almost identical to those who used a mouse, and spent less time on performing the same tasks. MK showed potential as a device for interaction with medical images.
[103]	To examine usability for navigating through 3D medical images using MK compared with a traditional mouse.	User testing. Survey.	Manipulation of anatomical images. Education.	17 veterinary students.	Improvements should be made to MK before it can be implemented as a device for medical use. The preferred method was the mouse. MK has the potential to reduce time on the task.
[13]	To develop a prototype and to examine the feasibility of this new device to help bridge the sterility barrier and eliminate the time and space gap that exists between image review and visual correlation with real-time operative field anatomy.	Proof-of-concept and prototype feasibility testing.	Manipulation of CT and MRI images.	2 MIS[e] procedures and 4 open procedures performed by a surgeon.	The system worked well in a wide range of lighting conditions and procedures. There was an increase in the use of intraoperative image consultation. The gesture library was intuitive and easy to learn. Gestures were mastered within 10 min.
[61]	To investigate a solution for manipulating medical images using MK.	Proof-of-concept and prototype feasibility testing.	Manipulation of CT images.	29 radiologists (diagnostic and interventional).	The potential of the device to enhance image-guided treatment in an interventional radiology suite while maintaining a sterile surgical field was demonstrated. 69% of those surveyed believed that the device could be useful in the interventional radiology field.
[112]	To investigate the need for posture and position training during bronchoscopy using a tool called ETrack	Pilot study.	Analysis of the operator's movements during a bronchoscopy. Education.	Not described.	The results highlight the importance of posture during bronchoscopy and the need to implement a training module for the simulator.
[71]	To evaluate a new touchless, portable, low-cost 3D measurement system for objective breast assessment.	Concurrent validation study.	Calculation of breast implant volumes.	9 silicone implants of known volumes.	The implant volumes were calculated with an error margin of 10%. Reproducibility was satisfactory. The system was validated for clinical use.
[106]	To describe a gesture-controlled 3D teaching tool in which temporal bone anatomy is manipulated without using a mouse or keyboard. To provide a teaching tool for patient-specific anatomy.	Proof-of-concept.	Manipulation of anatomical images. Education.	0.15 mm slice thickness cadaveric temporal bone images.	The interactive 3D model developed seems promising as an educational tool.
[62]	To develop hand recognition software based on MK, linked to an interventional CT, to manipulate images.	Feasibility testing	Manipulation of CT images in surgery.	10 interventional radiology procedures. 1 operator.	Tested on 10 procedures, feasibility was 100%. The system also allowed information to be obtained without using the CT system interface or a third party, and without the loss of operator sterility.
[131]	To present a novel method for training intentional and nonintentional gesture recognition.	Experiment.	Performance of a simulated brain biopsy on a mannequin assisted by images manipulated using gestures.	19 subjects.	Continuous gesture recognition was successful 92.26% of the time with a reliability of 89.97%. Significant improvements in task completion time were obtained through the context integration effect.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[113]	To evaluate 2 contactless hand tracking systems, the LMC[f] and MK, for their potential to control surgical robots.	Experiment.	Manipulation of robots in surgery.	4 trained surgeons.	Neither system has the high level of accuracy and robustness that would be required for controlling medical robots.
[107]	To use a projector for visualization and to provide intuitive means for direct interaction with the information projected onto the surgical surface, using MK to capture the interaction zone and the surgeon's actions on a deformable surface.	Proof-of-concept.	Augmented reality in surgery.	Not described.	The system eliminates the need for the surgeon to look at a location other than the surgical field. It therefore removes distractions and enhances his or her performance. It not only provides the surgeon with medical data during the intervention, but also allows interaction with such information by using gestures.
[10]	To present an ethnographic study of a system based on MK developed to allow touchless control of medical images during vascular surgery. The study aims to go beyond demonstrating technical feasibility in order to understand the collaborative practices that emerge from its use in this context.	Ethnographic study.	Manipulation of radiological images.	Endovascular suite of a large hospital.	With touchless interaction, the visual resources were embedded and made meaningful in the collaborative practices of surgery. The importance of direct and dynamic control of the images by the clinicians in the context of talks and in the context of other artefact use is discussed.
[130]	To evaluate a system for manipulating an operating table using gestures.	Prototype user testing.	Manipulation of an operating table.	15 participants.	Major problems were encountered during gesture recognition and with obstruction by other people in the interaction area due to the size and layout of the operating room. The system cannot yet be integrated into a surgical environment.
[110]	To study the technical skills of colonoscopists using MK for motion analysis to develop a tool to guide colonoscopy education and to select discriminative motion patterns.	Construct validity study.	Analysis of the movements of the operator during a colonoscopy.	10 experienced and 11 novice endoscopists.	Certain types of metric can be used to discriminate between experienced and novice operators.
[72]	To develop a 3D surface imaging system and to assess the accuracy and repeatability on a female mannequin.	Interrater reliability study.	Measurement of the surface distances of the breast on a mannequin.	A female mannequin.	MK seems to be a useful and feasible system for capturing 3D images of the breast. There was agreement between the measurements obtained by the system and those taken manually with a measuring tape.
[105]	To present a new surgical training system.	Proof-of-concept.	Real-time immersive 3D surgical training. Education.	Not described.	Preliminary experiments show that this immersive training system is portable, effective and reliable.
[68]	To present the development and clinical testing of a device that enables intraoperative control of images with hand gestures during neurosurgical procedures.	Proof-of-concept. Initial clinical testing.	Manipulation of MRI images.	30 neurosurgical operations.	OPECT demonstrated high effectiveness, simplicity of use and precise recognition of the individual user profile. In all cases, surgeons were satisfied with the performance of the device.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[68]	To test whether an automatic motion analysis system could be used to explore if there is a correlation in scope movements and the level of experience of the surgeon performing the bronchoscopy.	Construct validity study. Prospective, comparative study.	Analysis of the operator's movements during a bronchoscopy. Education.	11 novice, 9 intermediate and 9 experienced bronchoscopy operators performed 3 procedures each on a bronchoscopy simulator.	The motion analysis system could discriminate between different levels of experience. Automatic feedback on correct movements during self-directed training on simulators might help new bronchoscopists learn how to handle the bronchoscope like an expert.
[77]	To compare 2 commercial motion sensors (MK and the LMC) to manipulate CT images, in terms of their utility, usability, speed, accuracy and user acceptance.	Two-strand sequential observational study. Qualitative and quantitative descriptive field study using a semi-structured questionnaire.	Manipulation of CT images.	42 participants: radiologists, surgeons and interventional radiologists.	Marginal to average acceptability of the 2 devices. MK was found to be more useful and easier to use, but the LMC was more accurate. Further research is required to establish the design specifications, installation guidelines and user training requirements to ensure successful implementation in clinical areas.
[57]	To develop an integrated and comprehensive operating room information system compatible with HL7 and DICOM (MediNav). A natural user interface is designed specifically for operating rooms based on MK.	Prototype user testing.	Users tested the application's various modules.	A prototype system is tested in a live operating room at an Iranian teaching hospital. 30 general surgeries.	The results of usability tests are promising, and indicate that integration of these systems into a complete solution is the key. Touchless natural user interfaces can help to collect and visualize medical information in a comprehensive manner.
[75]	To propose a novel system to visualize a surgical scene in augmented reality using the different sources of information provided by a C-arm and MK.	Prototype user testing.	Augmented reality in orthopedic surgery.	Simulations of 12 orthopedic procedures. 5 participating clinicians, 3 experienced surgeons, 2 fourth-year medical students.	The system showed promising results with respect to better surgical scene understanding and improved depth perception using augmented reality in simulated orthopedic surgery.
[114]	To explore 3D perception technologies in the operating room.	Ethnographic. Prototype testing.	Detection of the interaction between operating staff and the robot.	Not described.	The paper described a supervision system for the operating room that enables intention tracking. The system had low latency, good registration accuracy and high tracking reliability, which make it useful for workflow monitoring, tracking and avoiding collisions between medical robots and operating room staff.
[125]	To use MK and color markers to track the position of MIS instruments in real time.	Comparative study between MK and the SinaSim trainer.	Movement of the instrument to position its tip in 81 holes of a Plexiglas plate on 5 occasions.	1 user.	Although the new method had inferior accuracy compared with mechanical sensors, its low cost and portability make it a candidate for replacing traditional tracking methods.
[80]	To compare 3 different interaction modes for image manipulation in a surgery setting: 1) A gesture-controlled approach using MK; 2) verbal instructions to a third party; and 3) direct manipulation using a mouse.	Crossover randomized controlled trial with blocked randomization.	Interaction modes were direct manipulation using a mouse, verbal instructions given to a third party, and gesture-controlled manipulation using MK.	30 physicians and senior medical students	Under the premise that a mouse cannot be used directly during surgery, gesture-controlled approaches were shown to be superior to verbal instructions for image manipulation.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[121]	To evaluate the feasibility, validity, and reliability of the training system for motion parameter and ergonomic analyses between different experience levels of surgeons using the NDI Polaris System and MK camera.	Construct validity, concurrent validity and test-retest reliability. Prospective blinded study.	Tying of intra-corporeal MIS knots.	10 MIS novices, 10 intermediate level and 10 experts.	Validity and reliability of the self-developed sensor and expert model-based MIS training system ‘iSurgeon’ were established.
[73]	To analyze preoperative breast volume in patients with breast cancer in order to predict implant size for reconstruction.	Exploratory study.	MK was used to acquire 3D images of the patients’ breasts before surgery and after surgery.	10 patients.	This study showed the feasibility of using fast, simple and inexpensive 3D imaging technology for predicting implant size before surgery, although there were significant technical challenges in determining breast volume by surface imaging.
[52]	To evaluate the feasibility of using 3 different gesture control sensors (MK, the LMC and the Myo armband) to interact in a sterile manner with preoperative data as well as in settings of an integrated operating room during MIS.	Pilot user study.	2 hepatectomies and 2 partial nephrectomies on an experimental porcine model.	3 surgeons.	Natural user interfaces are feasible for directly interacting, in a more intuitive and sterile manner, with preoperative images and integrated operating room functionalities during MIS. The combination of the Myo armband and voice commands provided the most intuitive and accurate natural user interface.

[a] MK: Microsoft Kinect.

[b] CT: Computed Tomography.

[c] MRI: magnetic resonance imaging.

[d] 3D: 3-dimensional.

[e] MIS: minimally invasive surgery.

[f] LMC: Leap Motion Controller.

Table 2. Summary of included studies evaluating the Leap Motion Controller.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[63]	To evaluate the implementation of a low-cost device for touchless PACS control in an interventional radiology suite. To demonstrate that interaction with gestures can decrease the duration of the procedures, the risk of re-intervention, and improve technical performance.	Proof-of-concept and prototype feasibility testing.	Manipulation of images in interventional radiology.	Interventional radiology suite.	The LMC[a] is a feasible, portable and low-cost alternative to other touchless PACS interaction systems. A decrease in the need for re-intervention was reported, but no explanation was given of how it was measured.
[54]	To present the first experience of using new systems for image control in the operating room: the LMC and OsiriX.	Proof-of-concept.	Manipulation of CT[b] and MRI[c] images.	2 general surgeons, 1 urologist, 3 orthopedic surgeons and 2 surgeons	The average training time was 5 min. The system is very cost-effective, efficient and prevents contamination during surgery. First experience of using the LMC to control CT and MRI images during surgery.
[116]	To validate the possibility of performing precise telesurgical tasks by means of the LMC.	Comparative study of the Sigma.7 electro-mechanical device and the LMC.	Peg transferring task and answering a questionnaire. The success rate of peg transfers.	10 researchers.	The results allowed the authors to confirm that fine tracking of the hand could be performed with the LMC. The observed performance of the optical interface proved to be comparable with that of traditional electromechanical devices.
[87]	To describe a piece of software for image processing with OsiriX using finger gestures.	Proof-of-concept.	Manipulation of radiological images.	Not described.	It is possible to implement gesture control of medical devices with low-cost, minimal resources. The device is very sensitive to surface dirt and this affects performance. The device favors the occlusion phenomenon.
[113]	To evaluate 2 contactless hand tracking systems, the LMC and MK[d], for their potential to control surgical robots.	Experiment.	Manipulation of robots in surgery.	4 trained surgeons.	Neither system has the high level of accuracy and robustness that would be required for controlling medical robots.
[129]	To evaluate the LMC for simple 2-dimensional interaction and the action of entering a value.	Proof-of-concept and prototype testing.	Manipulation of medical information and operating room lights.	A 90-min conference on computer science and untrained users.	The user cases should be carefully classified and the most appropriate gestures for each application should be detected and implemented. Optimal lighting conditions for the LMC have still not been evaluated as unwanted light with deterioration of the IR light emitted may lead to a reduction in the recognition rate.
[81]	To compare the average time required by the conventional method using a mouse and an operating method with a finger-motion sensor.	Observational study.	Manipulation of angiographic images.	11 radiologists who observed a simulated clinical case.	After a practice time of 30 min, the average operation time by the finger method was significantly shorter than that by the mouse method.
[14]	To develop a workstation that allows intraoperative touchless control of diagnostic and surgical images in dentistry.	Prototype user testing.	Manipulation of radiological images.	2 surgeons. A case series of 11 dental surgery procedures.	The system performed very well. Its low cost favors its incorporation into clinical facilities of developing countries, reducing the number of staff required in operating rooms.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[88]	To propose an interface to control hand gestures and gestures with hand-held tools. In this approach, hand-held tools can become gesture devices that the user can use to control the images.	Prototype user testing.	Manipulation of ultrasound images.	12 participants.	Users were able to significantly improve their performance with practice.
[56]	To develop a software application for the manipulation of a 3D[e] pancreatic or liver tumor model by using CT and real-time elastography data.	Proof-of-concept.	Manipulation of CT and real-time elastography images.	15 patients with liver cancer and 10 patients with pancreatic cancer.	A 3D model of liver and pancreatic tumors was successfully implemented with a hands-free interaction device suitable for sterile environments and for aiding diagnostic or therapeutic interventions.
[117]	To present a new gesture recognition system for manipulating 2 surgical robots in a virtual simulator.	Proof-of-concept.	Manipulation of robots in surgery.	2 surgical robots in a virtual simulator.	The device provided satisfactory accuracy and speed. It requires a more complete Application Programming Interface.
[90]	To propose a web-based interface to retrieve medical images using gestures.	User testing. Pilot study.	Manipulation of radiological images.	2 users.	User feedback was positive. Users reported fatigue with prolonged use of gestures. Additional studies are required to validate the interface.
[64]	To describe the use of the LMC for image manipulation during hepatic transarterial chemoembolization and internal radiotherapy procedures.	Proof-of-concept.	Manipulation of images in interventional radiology.	Not described.	Gesture-based imaging control may lead to increased efficacy and safety with decreased radiation exposure during hepatic transarterial chemoembolization procedures.
[77]	To compare 2 commercial motion sensors (MK and the LMC) to manipulate CT images, in terms of their utility, usability, speed, accuracy and user acceptance.	Two-strand sequential observational study. Qualitative and quantitative descriptive field study using a semi-structured questionnaire.	Manipulation of CT images.	42 participants: radiologists, surgeons and interventional radiologists.	Marginal to average acceptability of the 2 devices. MK was found to be more useful and easier to use, but the LMC was more accurate. Further research is required to establish the design specifications, installation guidelines and user training requirements to ensure successful implementation in clinical areas.
[91]	To evaluate a new method for image manipulation using a motion sensor.	Observational study. User testing and proof-of-concept.	Manipulation of radiological images in dentistry.	14 students. 6 images.	Using the system, several processes can be performed quickly with finger movements. Using gestures was significantly superior to using a mouse in terms of time.
[92]	To develop a new system for manipulating images using a motion sensor.	Observational study.	Manipulation of radiological images in dentistry.	14 students. 25 images.	The operation time with the LMC was significantly shorter than with the conventional method using a mouse.
[108]	To design a virtual 3D online environment for motor skills learning in MIS[f] using exercises from the MISR-VR. The environment is designed in Unity, and the LMC is used as the device for interaction with the MIS forceps.	Letter to the editor.	None.	Not described	If it can be shown that 3D online environments mediated by natural user interfaces enable motor skills learning in MIS, a new field of research and development in the area of surgical simulation will be opened up.
[124]	Patent for accurate 3D instrument positioning.	Patent.	None.	Not described	Representing, on an output display, 3D positions and orientations of an instrument while medical procedures are being performed.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[69]	To describe the configuration for using the LMC in neurosurgery for image manipulation during a surgical procedure.	User testing.	Manipulation of images during a surgical procedure.	Resection of a meningioma and sarcoma surgery.	The learning curve only took 30 min. Although the main disadvantage was the lack of standardization of the gestures, the LMC is a low-cost, reliable and easily personalized device for controlling images in the surgical environment.
[109]	To develop skills in students and professionals using computer simulation technologies based on hand gesture capture systems.	User testing.	Description of the virtual environment.	Not described.	Simulation and new gesture recognition technologies open up new possibilities for the generation of computer-mediated procedures for medical training.
[93]	To present a gesture-controlled projection display that enables a direct and natural physician-machine interaction during CT-based interventions.	User testing (pilot and main).	8 tasks manipulating CT images.	12 participants (biomedical engineers, medical students and radiologists).	Gesture recognition is robust, although there is potential for improvement. The gesture training times are less than 10 min, but vary considerably between study participants.
[94]	To develop an anatomy learning system using the LMC.	User testing.	Manipulation of 220 anatomical images.	30 students and lecturers from an anatomy department.	The anatomy learning system using the LMC was successfully developed and it is suitable and acceptable as a support tool in an anatomy learning system.
[123]	To study the possibility of tracking laparoscopic instruments using the LMC in a box trainer.	Experiment.	3 static experiments and 1 dynamic experiment.	1 user.	The LMC had acceptable precision for tracking laparoscopic instruments in a box trainer.
[126]	To assess the potential of the LMC to track the movement of hands using MIS instruments.	Construct validity, concurrent validity. Comparative study with the InsTrac.	Passing a thread through pegs using the eoSim simulator.	3 experts and 10 novices.	The LMC is able to track the movement of hands using instruments in a MIS box simulator. Construct validity was demonstrated. Concurrent validity was only demonstrated for time and instrument path distance. A number of limitations to the tracking method used by LMC have been identified.
[118]	To explore the use of the LMC in endonasal pituitary surgery and to compare it with the Phantom Omni.	Comparative study between the LMC and the Phantom Omni.	16 resections of simulated pituitary gland tumors using a robot manipulated by the Phantom Omni and by the LMC.	3 neurosurgeons.	Users were able to achieve a very similar percentage of resection and procedure duration using the LMC.
[95]	To try to interact with medical images via a web browser using the LMC.	Prototype user testing.	Rotation, panning, scaling and selection of slices of a reconstructed 3D model based on CT or MRI.	1 user.	It is feasible to build this system and interaction can be carried out in real time.
[58]	To analyze the value of 2 gesture input modalities (the Myo armband and the LMC) versus 2 clinically established methods (task delegation and joystick control).	User study. Comparative study.	Simulating a diagnostic neuroradiological vascular treatment with 2 frequently used interaction tasks in an experimental operating room.	10 neuroradiologists	Novel input modalities have the potential to carry out single tasks more efficiently than clinically established methods.
[120]	To investigate the potential of a virtual reality simulator for the assessment of basic laparoscopic skills, based on the LMC	Face and construct validity.	3 basic tasks: camera navigation, instrument navigation, and two-handed operation.	2 groups of surgeons (28 experts and 21 novices).	This study provides evidence of the potential use of the LMC for assessing basic laparoscopic skills. The proposed system allows the dexterity of hand movements to be evaluated.

Study	Aim	Type of study	Intervention	Sample	Results/Conclusions
[52]	To evaluate the feasibility of using 3 different gesture control sensors (MK, the LMC and the Myo armband) to interact in a sterile manner with preoperative data as well as in settings of an integrated operating room during MIS.	Pilot user study.	2 hepatectomies and 2 partial nephrectomies on an experimental porcine model.	3 surgeons	Natural user interfaces are feasible for directly interacting, in a more intuitive and sterile manner, with preoperative images and integrated operating room functionalities during MIS. The combination of the Myo armband and voice commands provided the most intuitive and accurate natural user interface.
[127]	To evaluate the LMC as a tool for the objective measurement and assessment of surgical dexterity among users at different experience levels.	Construct validity study.	Surgical knot tying and manual transfer of objects.	11 participants.	The study showed 100% accuracy in discriminating between expert and novice performances.
[66]	To design an affordable and easily accessible endoscopic third ventriculostomy simulator based on the LMC, and to compare it with the NeuroTouch for its usability and training effectiveness.	Concurrent and construct validity study.	4 ellipsoid practice targeting tasks and 36 ventricle targeting tasks.	16 novice users and 2 expert neurosurgeons	An easy-access simulator was created, which has the potential to become a training tool and a surgical training assessment tool. This system can be used for planning procedures using patient datasets.
[119]	To present the LMC as a novel control device to manipulate the RAVEN-II robot.	Comparative study between the LMC and the electro-mechanical Sigma.7.	Comparison of peg manipulations during a training task with a contact-based device (Sigma.7).	3 operators.	With contactless control, manipulability is not as good as it is with contact-based control. Complete control of the surgical instruments is feasible. This work is promising for the development of future human-machine interfaces dedicated to robotic surgical training systems.
[98]	To evaluate the effect of using virtual reality surgery on the self-confidence and knowledge of surgical residents (the LMC and Oculus Rift).	Multisite, single-blind, parallel, randomized controlled trial.	The study group used the virtual reality surgery application. The control group used similar content in a standard presentation.	95 residents from 7 dental schools.	Immersive virtual reality experiences improve the knowledge and self-confidence of the surgical residents.
[97]	To develop and validate a novel training tool for Le Fort I osteotomy based on immersive virtual reality (the LMC and Oculus Rift).	Face and content validity.	A pre-intervention questionnaire to understand training needs and a postintervention feedback questionnaire.	7 consultant oral and maxillofacial surgeons.	The results confirmed the clinical applicability of virtual reality for delivering training in orthognathic surgery.
[70]	To investigate the feasibility and practicability of a low-cost multimodal head-mounted display system in neuroendoscopic surgery (the LMC and Oculus Rift).	Proof-of-concept in the operating room.	Ventriculocysto- cisternostomy. Ventriculostomy. Tumoral biopsy.	21 patients with ventricular diseases. 1 neurosurgeon.	The head-mounted display system is feasible, practical, helpful, and relatively cost efficient in neuroendoscopic surgery.

[a] LMC: Leap Motion Controller.

[b] CT: Computed Tomography.

[c] MRI: magnetic resonance imaging.

[d] 3D: 3-dimensional.

[e] MK: Microsoft Kinect.

[f] MIS: minimally invasive surgery.

Table 3. Summary of included studies evaluating other devices.

Study	Device	Aim	Type of study	Intervention	Results/Conclusions
[53]	Camera with Complementary Metal-Oxide-Semiconductor sensor	To propose an architecture for a real-time multimodal system to provide a touchless user interface in surgery.	Prototype user testing.	Gesture detection in computer-assisted surgery.	The preliminary results show good usability and rapid learning. The average time to click anywhere on the screen was less than 5 seconds. Lighting conditions affected the performance of the system. The surgeon showed strong interest in the system and satisfactorily assessed the use of gestures within the operating room.
[82]	Webcam	To describe a vision-based system that can interpret gestures in real time to manipulate objects within a medical data visualization environment.	Prototype user testing.	Manipulation of medical data (radiology images and selection of medical records) and movement of objects and windows on the screen.	The system implemented in a sterile environment demonstrated performance rates between 95% and 100%.
[27]	Canon VC-C4 color camera	To describe a vision-based gesture capture system that interprets gestures in real time to manipulate medical images.	Beta testing during a surgical procedure. Experiment.	A beta test of a system prototype was conducted during a live brain biopsy operation, where neurosurgeons were able to browse through MRI[a] images of the patient's brain using the sterile hand gesture interface.	Gesture recognition accuracy was 96%. For every repeat of trials, the task completion time decreased by 28% and the learning curve levelled off at the 10th attempt. The gestures were learned very quickly and there was a significant decrease in the number of excess gestures. Rotation accuracy was reasonable. The surgeons rated the system as easy to use, with a rapid response, and useful in the surgical environment.
[26]	Canon VC-C4 camera	To evaluate the Gestix system.	Prototype user testing.	Manipulation of MRI images during a neurosurgical biopsy.	The system setup time was 20 min. The surgeons found the Gestix system easy to use, with a rapid response, and easy to learn. The system does not require the use of wearable devices.
[59]	Interaction with gestures in general	Fieldwork focusing on work practices and interactions in an angiography suite and on understanding the collaborative work practices in terms of image production and use.	Ethnographic study of minimally invasive image-guided procedures within an interventional radiology department.	Manipulation of radiological images.	The paper discusses the implications of the findings in the work environment for touchless interaction technologies, and suggests that these will be of importance in considering new input techniques in other medical settings.
[115]	Commercial video camera	To describe the development of Gestonurse, a robotic system for surgical instruments.	Proof-of-concept.	Surgical instrumentation using a robot.	95% of gestures were recognized correctly. The system was only 0.83 seconds slower when compared with the performance of a human instrument handler.

Study	Device	Aim	Type of study	Intervention	Results/Conclusions
[65]	Touchless interaction systems in general	To understand and use common practices in the surgical setting from a proxemics point of view to uncover implications for the design of touchless interaction systems. The aim is to think of touchlessness in terms of its spatial properties. What does spatial separation imply for the introduction of the touchless control of medical images?	Ethnographic study.	Field observations of work practices in neurosurgery.	Alternative ideas, such as multiple cameras, are the kind of solution that these findings suggest. Such reflections and considerations can be revealed through careful analysis of the spatial organization of activity and proxemics of particular interaction mechanisms. However, it is very important to study current practice in order to speculate about new systems, because they in turn may alter practice.
[122]	Webcam	To present a system for tracking the movement of MIS[b] instruments based on an orthogonal webcam system installed in a physical simulator.	Experiment.	Recording the movements of the instrument within an imaginary cube.	The results showed a resolution of 0.616 mm on each axis of work, linearity and repeatability in motion tracking, as well as automatic detection of the 3D position of the tip of the surgical instruments with sufficient accuracy. The system is a low-cost and portable alternative to traditional instrument tracking devices.
[52]	MK, the LMC[c], the Myo armband and voice control	To evaluate the feasibility of using 3 different gesture control sensors (MK, the LMC and the Myo armband) to interact in a sterile manner with preoperative data as well as in settings of an integrated operating room during MIS.	Pilot user study.	2 hepatectomies and 2 partial nephrectomies on an experimental porcine model.	Natural user interfaces are feasible for directly interacting, in a more intuitive and sterile manner, with preoperative images and integrated operating room functionalities during MIS. The combination of the Myo armband and voice commands provided the most intuitive and accurate natural user interface.
[58]	The Myo armband and the LMC	To analyze the value of 2 gesture input modalities (the Myo armband and the LMC) versus 2 clinically established methods (task delegation and joystick control).	User study. Comparative study.	Simulating a diagnostic neuroradiological vascular treatment with 2 frequently used interaction tasks in an experimental operating room.	Novel input modalities have the potential to carry out single tasks more efficiently than clinically established methods.

[a] MRI: magnetic resonance imaging.

[b] MIS: minimally invasive surgery.

[c] LMC: Leap Motion Controller.

Table 4. Clinical areas and types of surgical intervention in which gesture-based commercial off-the-shelf devices were used.

Clinical areas	Types of surgical intervention	Studies
General surgery (N=7)	Intraoperative image control, image-guided minimally invasive surgery (adrenalectomy, pancreatectomy, liver resection, a Whipple procedure, as well as liver and pancreatic cancer and renal carcinoma resection), open and laparoscopic bile duct surgery, cholecystectomy, and hepatectomy and nephrectomy in an animal model.	[13,52-57]
Interventional radiology and angiography (N=7)	Arterial dilatation with balloon and umbrella devices, hepatic arterial chemoembolization and selective internal radiation therapy, abdominal computed tomography, and interventional neuroradiology.	[58-64]
Neurosurgery (N=7)	Biopsies, resection of brain gliomas, resection of a meningioma, ventriculostomy, and intraoperative image control.	[26,65-70]
Plastic surgery (N=3)	Measurement of breast implant volumes and measurement of distances on the breast surface.	[71-73]
Orthopedics (N=3)	Intraoperative image control.	[55,74,75]
Ear, nose, and throat (N=1)	Laryngoplasty.	[76]
Urology (N=2)	Enucleation of renal tumors and intraoperative image control.	[30,54]

Table 5. Use of gesture-based commercial off-the-shelf devices in surgery.

Use	Studies
Manipulation of images in interventional radiology environments or in the operating room (N=42)	
Image manipulation	[5,13,14,17,19,26,27,30,52,54,56,58-64,67-69,74,76-95]
Education and training	
Virtual or augmented reality for educational or interventional purposes (N=16)	[75,94,96-109]
Training in endoscopy (bronchoscopy and colonoscopy; N=3)	[110-112]
Robotic surgery (N=7)	
Robotics in surgery and in surgical instrumentation	[113-119]
Tracking of hand or instrument movements during open or minimally invasive surgery	
Instrument tracking in MIS[a] (N=7)	[108,120-125]
Tracking of hand movements during MIS (N=2)	[109,126]
Tracking of hand movements during open surgical knot tying (N=1)	[127]
Simulation for skills learning in MIS (N=4)	
Simulation for motor skills learning in MIS	[66,108,120]
Using patient-specific 3-dimensional images during MIS in real patients or simulators, and presurgical warm-up	[52,66,70,108]
Other uses	
Ethnographic studies (N=5)	[59,65,78,83,114]
Measurement of breast implant volumes and measurement of distances on the breast surface (N=3)	[71-73]
Manipulation of the operating table and lights (N=4)	[128-130]

[a] MIS: minimally invasive surgery.

Aims, Types of Study, Metrics, Samples, Results and Conclusions

In 78% (67/86) of the articles, the aim was to develop, create, present, describe, propose, examine, or explore a COTS-based system for gesture recognition in surgery. Most of the articles [65] identified in this systematic review were proof-of-concept or prototype user testing and observational and feasibility testing studies (Tables 1-3, see Multimedia Appendices 1-3 for the full Tables 1-3). In the 5 ethnographic studies included, the aim was to identify interactions between the staff and gesture-based COTS systems in interventional radiology departments or in the operating room [19,59,65,78,114]. In 4 studies, the aim was to compare the performance of MK with that of a mouse [5,79,80,96]; in 1 study, it was to compare the performance of the LMC with that of a mouse [81]; and in 4 studies, it was to compare different COTS devices [52,58,77,113]. In 10 studies, the aim was to evaluate face validity [97,120], content validity [97], construct validity [66,110,111,120,121,126,127,132], or concurrent validity of the devices [66,71,121,126]. A total of 7 studies involved experiments [19,26,113,115,122,123,131] and there was 1 patent application for an LMC-based application [124] and 1 interrater reliability study [72]. In addition, 1 study was a quasi-experimental prospective, blinded study with test-retest reliability [121]. Only 2 randomized controlled trials were identified [80,98], and when a tool for assessing risk of bias in randomized trials [133] was applied to them, it was found to be low in both.

In total, 25 out of 86 (29%) articles failed to describe the metric used, whereas 23 out of 86 (27%) used time as the main one. Given the varied nature of the design of the studies, the remaining 38 articles described multiple metrics such as performance rates, percentage of gesture recognition, accuracy of gesture recognition and/or speed of transmission thereof, measures of volume or distance, and questionnaires or interviews. Similarly, the sample types and numbers were very dissimilar: 17.4% of the articles did not describe the sample type, and the remainder stated that the samples comprised medical or veterinary students or specialists in several radiological or surgical specialties (Table 4).

Interventions

The most common intervention (42 studies) was image manipulation in general radiology, ultrasound imaging, interventional radiology, angiography, computed tomography, magnetic resonance imaging, and real-time elastography (in the operating room, in the operative dentistry setting, or in the interventional radiology suites; Tables 1-3; see Multimedia Appendices 1-3 for the full Tables 1-3). Table 5 shows other uses identified for gesture-based COTS devices in surgical environments.

Use of Commercial Off-The-Shelf Devices as Simulation Tools for Motor Skills Teaching in Minimally Invasive Surgery

In the field of skills learning in MIS, in 2013, Pérez et al first described the tracking of laparoscopic instruments using webcams, with encouraging results [122]. From 2016, several authors proposed the interesting possibility of using COTS devices for tracking laparoscopic instruments. Such devices include both the LMC [108,121,123,124] and MK [125]. In 2017, a portable low-cost simulator using the LMC [120] for basic motor skills learning in MIS was described, and so too were a simulator for endoscopic third ventriculostomy learning [66] and a head-mounted display system using Oculus Rift and the LMC to guide neuroendoscopic surgery by manipulating 3D images [70]. Others used the approach of tracking hand movements during MIS training [109,126]. Only 1 study explored the use of the LMC to assess surgical dexterity in tying surgical knots in open surgery [127].

Furthermore, 1 study compared 3 natural user interfaces (MK, the LMC, and the Myo armband) in combination with voice control to perform 2 hepatectomies and 2 partial nephrectomies on an experimental porcine model [52]; similar to the studies by Wright [66] and Xu [70], this study used 3D reconstructions of preoperative images of the patient, which were manipulated by gestures during surgery. However, the application of gesture control technology in these cases is not for training purposes but for surgical assistance and planification.

Discussion

Principal Findings

Using commercial devices to detect manual gestures in surgery is a very topical issue, given the need to manipulate medical images and for real-time 3D reconstructions during procedures without breaking asepsis and antisepsis protocols. Early studies published on this possibility used COTS systems with webcams, Complementary Metal-Oxide-Semiconductor-sensor cameras, and commercial digital cameras [26,27,53,82]. These pioneering studies showed that contactless interaction with images and medical information in environments such as operating rooms was possible using low-cost devices.

In this systematic review, MK and the LMC were identified as the most widely used COTS systems. MK was rated as a useful tool for the manipulation of medical data in sterile environments, with a positive rate of acceptance in 85% (39/46) of the studies on it. The LMC had a positive rate of acceptance in 83% (29/35) of the studies on it. The Myo armband was used to manipulate interventional neuroradiology images [58]. In addition, in a comparative study of the Myo armband, MK, and the LMC, they were used to manipulate images while hepatectomies and partial nephrectomies were being performed on an animal model [52]. In both cases, the device was rated highly. The main positive characteristics identified for the devices were the following: there was no need for contact; they were low-cost and portable; there was no need for calibration at the time of use; the gesture learning curve was easy; and the gesture recognition rates were high.

Performance of Individual Devices

MK [30] and the LMC [14,81,87,134,135] both use infrared cameras. The MK system is based on the time-of-flight principle [61], whereas the LMC is based on a sensor for infrared optical tracking with stereo vision accuracy. The MK depth sensor works at a distance between 0.8 m and 3.5 m, and the interface tracks the skeleton of the system operator. The wide range of distances at which the device recognizes gestures presents problems when using it in close interaction. The LMC detects the positions of fine objects such as finger tips or pen tips in a Cartesian plane. Its interaction zone is an inverted cone of approximately 0.23 m^3 and the motion detection range fluctuates between 20 mm and 600 mm [91,129]. The manufacturer reports an accuracy of 0.01 mm for fingertip detection, although 1 study showed an accuracy of 0.7 mm, which is considered superior to that achieved using MK [134,136]. The dimensions of the MK device are 280 mm (width), 71 mm (depth), and 66 mm (height) and its weight is 556 g, whereas those of the LMC are 76 mm (width), 30 mm (depth), and 13 mm (height) and its weight is 45 g.

Only 5 of the 46 (11%) studies that evaluated MK identified disadvantages relating to a longer latency time, difficulty in recreating an image when compared with a keyboard or mouse [5], limited gesture recognition, interference between the movements of different people in small environments

[85,89,130], and the users' preference for a mouse in a comparative study [96]. Various studies have highlighted the inaccuracy of MK in detecting finger movements [5,17,85,137], and the system also requires the use of large format screens [14,24,54,85,90]. The system was taken off the market in October 2017.

With regard to the LMC, once the 6 studies on robotics had been discarded, 4 articles were identified that presented limitations derived from using the device (18%). These studies noted alterations in performance when there was dirt on the surface of the device, as well as the limited number of gestures recognized owing to the occlusion phenomenon [87], alterations caused by ambient lighting [129], fatigue in some users [90], and a lack of studies validating the device for medical use [77].

The Myo armband was launched in 2013. This wearable wireless device is able to record electromyography via 8 stainless steel dry surface electrodes. It has a 9-axis inertial measurement unit sensor, haptic feedback, and Bluetooth communication capability. The main disadvantage is its limited sampling frequency of 200 Hz [138-140]. In total, 2 studies on the Myo armband were identified. The first concluded that the combination of the Myo armband and voice commands provided the most intuitive and accurate natural user interface [141]. The second compared the Myo armband and LMC with traditional image manipulation methods in surgery and concluded that the new input modalities had the potential to become more efficient [58].

Commercial Off-The-Shelf Devices in Robotic Surgery

Studies on the application of gesture-based COTS devices in robot-assisted surgery failed to demonstrate usefulness, owing to either the high cost of the robotic arm when using commercial cameras in surgical instrumentation [115] or, in the case of the LMC, the need for a more robust Application Programming Interface [116,117] and the lack of sufficient accuracy and robustness for manipulating a medical robot [113]. However, an ethnographic study found that MK was useful for workflow monitoring and for avoiding collisions between medical robots and operating room staff [114]. A simulation study of endonasal pituitary surgery comparing the LMC with the Phantom Omni showed that surgeons achieved a very similar percentage of tumor mass resection and procedure duration using the LMC to control the robot [118]. Another study found that the robotic tools could be controlled by gestures for training purposes but that the level of control had yet to reach that of a contact-based robotic controller [119].

Commercial Off-The-Shelf Devices in Training and Simulation

Studies on the use of COTS devices for gesture-based interfaces using the hand in the field of education in surgery refer to the use of virtual reality and augmented reality for teaching anatomy or for living the immersive experience within a virtual operating room. A total of 3 studies explored the possibility of using MK as a tool for skills learning in bronchoscopy and colonoscopy by means of simulation [110-112].

Various authors explored the possibility of hand tracking [109,126] or instrument tracking [108,121-125] using COTS devices to assess performance in MIS training. From these 2 approaches, Lahanas [120] eventually presented a portable low-cost model of a virtual reality simulator for basic motor skills learning in MIS, which was based on the LMC and capable of tracking instruments. The author also presented face and contrast validity studies. The original forceps tracking problems noted by the author were probably because of the fact that they were black. Problems caused by this color were also described in the study by Oropesa. This issue had already been raised by our group [108].

In the field of simulation for robotic surgery learning, the first studies published [113,115-117] found that the interfaces did not allow robots to be manipulated by gestures. However, the most recent publications [118,119] have suggested that the LMC could be a low-cost solution for creating control interfaces for surgical robots for the purposes of performing operations or training by means of simulation.

Ethnographic Studies

Ethnographic studies [59,65,78,83,114] deserve a separate mention as they transcend proofs-of-concept and user and prototype testing and approach gesture-based touchless interaction from a holistic viewpoint that includes the social practices of surgery, as well as the way in which medical images and manipulation devices are embedded and made meaningful within the collaborative practices of the surgery [10].

Requirements for the Future

There was found to be a shortage of objective validation studies (face validity: 1 study; concurrent validity: 3 studies; construct validity: 3 studies; discriminant validity: none; and predictive validity: none) of the different applications developed and presented as prototypes or proofs-of-concept for use in the clinical or teaching field. In teaching, the field of hand gesture–based interfaces should prioritize the following research objectives: first, to transcend studies on technical feasibility and individual hand gesture–based interaction with medical images so as to tackle the issue systematically within a framework of collaborative discussion, as happens in real surgical environments; and second, to conduct experimental studies in simulated surgical environments that allow hand gestures to be validated as a useful tool for touchless interaction in real operating rooms. To that end, the language of hand gestures for medical use would have to be standardized, so that the surgeons' cognitive load can be reduced. In turn, algorithms should be developed to allow differentiation between intentional and unintentional gestures (spotting) in the small spaces of the operating room. Finally, the problem of temporal segmentation ambiguity (how to define the gesture start and end points) and that of spatial-temporal variability (gestures can vary significantly from one individual to another) must be resolved.

From the range of evidence found, it is possible to infer that, with regard to the use of COTS devices, there is a very interesting field of study for the development and objective validation (contrast, concurrent, discriminant, and predictive validities) of portable low-cost virtual reality simulators for motor skills learning in MIS and robotic surgery. Such simulators will enable surgeons to do presurgical warm-ups anywhere at any time based on 3D reconstructions of specific patients’ images [52,66,70,108]. Thus, surgeons will be able to practice the surgery the night before they are due to perform it from the comfort of their own homes.

Despite the fact that MK was taken off the market in 2017 and that the LMC software only allows tool tracking up to V2 Tracking, the use of interaction with gesture-based virtual environments in the field of simulation identified in this review will enable new COTS devices (ie, the Myo armband) to be explored for skills learning in MIS and robotic surgery.

Limitations

A number of potential methodological limitations in our systematic review should be discussed. First, our inclusion criteria were limited to English-language publications. Second, although we used the most commonly used search engines in the health field (PubMed, EMBASE, ScienceDirect, Espacenet, OpenGrey, and IEEE) and complemented that by using the snowballing technique to identify relevant articles in the results generated by our search, we may have missed a few articles related to our research question. Finally, there may have been some potential for subjectivity in analyzing the findings, although 2 authors carefully reviewed each study independently and then discussed the results while double-checking each process and subsequently resolved any discrepancies through discussions with the third author whenever necessary.

Conclusions

As most of the articles identified in this systematic review are proof-of-concept or prototype user testing and feasibility testing studies, we can conclude that the field is still in the exploratory phase in areas requiring touchless manipulation within environments and settings that must adhere to asepsis and antisepsis protocols, such as angiography suites and operating rooms.

Without doubt, COTS devices applied to hand and instrument gesture–based interfaces in the field of simulation for skills learning and training in MIS could open up a promising field to achieve ubiquitous training and presurgical warm-up.

The withdrawal of MK from the market and suspension of the instrument tracking function in the latest LMC software versions constitute threats to the new developments identified in this review. Nevertheless, gesture-based interaction devices are clearly useful for manipulating images in interventional radiology environments or the operating room and for the development of virtual reality simulators for skills training in MIS and robotic surgery.

Authors' Contributions

All the authors contributed substantially to the study conception and design, data analysis and interpretation of the findings, and manuscript drafting. FAL participated in the collection and assembly of data. FSR is the guarantor of the paper. All the authors have read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of included studies evaluating Microsoft Kinect.

[PDF File (Adobe PDF File), 176KB - jmir_v21i5e11925_app1.pdf]

Multimedia Appendix 2

Summary of included studies evaluating the Leap Motion Controller.

[PDF File (Adobe PDF File), 132KB - jmir_v21i5e11925_app2.pdf]

Multimedia Appendix 3

Summary of included studies evaluating other devices.

[PDF File (Adobe PDF File), 66KB - jmir_v21i5e11925_app3.pdf]

References

1. Bures S, Fishbain JT, Uyehara CF, Parker JM, Berg BW. Computer keyboards and faucet handles as reservoirs of nosocomial pathogens in the intensive care unit. Am J Infect Control 2000 Dec;28(6):465-471. [doi: 10.1067/mic.2000.107267] [Medline: 11114617]
2. Schultz M, Gill J, Zubairi S, Huber R, Gordin F. Bacterial contamination of computer keyboards in a teaching hospital. Infect Control Hosp Epidemiol 2003 Apr;24(4):302-303. [doi: 10.1086/502200] [Medline: 12725363]
3. Hartmann B, Benson M, Junger A, Quinzio L, Röhrig R, Fengler B, et al. Computer keyboard and mouse as a reservoir of pathogens in an intensive care unit. J Clin Monit Comput 2004 Feb;18(1):7-12. [doi: 10.1023/B:JOCM.0000025279.27084.39] [Medline: 15139578]

4. Lu P, Siu LK, Chen T, Ma L, Chiang W, Chen Y, et al. Methicillin-resistant Staphylococcus aureus and Acinetobacter baumannii on computer interface surfaces of hospital wards and association with clinical isolates. BMC Infect Dis 2009 Oct 1;9:164 [FREE Full text] [doi: 10.1186/1471-2334-9-164] [Medline: 19796381]
5. Ebert LC, Hatch G, Ampanozi G, Thali MJ, Ross S. You can't touch this: touch-free navigation through radiological images. Surg Innov 2012 Sep;19(3):301-307. [doi: 10.1177/1553350611425508] [Medline: 22064490]
6. D'Antonio NN, Rihs JD, Stout JE, Yu VL. Computer keyboard covers impregnated with a novel antimicrobial polymer significantly reduce microbial contamination. Am J Infect Control 2013 Apr;41(4):337-339. [doi: 10.1016/j.ajic.2012.03.030] [Medline: 23036480]
7. Ionescu AV. A mouse in the OR Ambidextrous. Stanford Univ Journal of Design 2006;30:2 [FREE Full text]
8. van Veelen MA, Snijders CJ, van Leeuwen E, Goossens RH, Kazemier G. Improvement of foot pedals used during surgery based on new ergonomic guidelines. Surg Endosc 2003 Jul;17(7):1086-1091. [doi: 10.1007/s00464-002-9185-z] [Medline: 12728372]
9. Grätzel C, Fong T, Grange S, Baur C. A non-contact mouse for surgeon-computer interaction. Technol Health Care 2004;12(3):245-257 [FREE Full text] [Medline: 15328453]
10. O'Hara K, Dastur N, Carrell T, Gonzalez G, Sellen A, Penney G, et al. Touchless interaction in surgery. Commun ACM 2014 Jan 1;57(1):70-77 [FREE Full text] [doi: 10.1145/2541883.2541899]
11. El-Shallaly GE, Mohammed B, Muhtaseb MS, Hamouda AH, Nassar AH. Voice recognition interfaces (VRI) optimize the utilization of theatre staff and time during laparoscopic cholecystectomy. Minim Invasive Ther Allied Technol 2005;14(6):369-371. [doi: 10.1080/13645700500381685] [Medline: 16754183]
12. Nathan CO, Chakradeo V, Malhotra K, D'Agostino H, Patwardhan R. The voice-controlled robotic assist scope holder AESOP for the endoscopic approach to the sella. Skull Base 2006 Aug;16(3):123-131 [FREE Full text] [doi: 10.1055/s-2006-939679] [Medline: 17268585]
13. Strickland M, Tremaine J, Brigley G, Law C. Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. Can J Surg 2013 Jun;56(3):E1-E6 [FREE Full text] [doi: 10.1503/cjs.035311] [Medline: 23706851]
14. Rosa GM, Elizondo ML. Use of a gesture user interface as a touchless image navigation system in dental surgery: case series report. Imaging Sci Dent 2014 Jun;44(2):155-160 [FREE Full text] [doi: 10.5624/isd.2014.44.2.155] [Medline: 24944966]
15. Wachs J. Purdue Equestrain Team. 2007. Optimal Hand-Gesture Vocabulary Design Methodology for Virtual Robotic Control URL: https://web.ics.purdue.edu/~jpwachs/papers/PHD_JUAN_JW.pdf [accessed 2019-04-01] [WebCite Cache ID 77JJfnGoX]
16. Yanagihara Y, Hiromitsu H. System for selecting and generating images controlled by eye movements applicable to CT image display. Med Imaging Technol 2000;18:725 [FREE Full text] [doi: 10.11409/mit.18.725]
17. Gallo L, Placitelli A, Ciampi M. Controller-free exploration of medical image data: experiencing the Kinect. In: Proceedings of the 2011 24th International Symposium on Computer-Based Medical Systems. 2011 Jun 27 Presented at: CMBS'11; June 27-30, 2011; Bristol, UK p. 1-6. [doi: 10.1109/CBMS.2011.5999138]
18. Coddington J, Xu J, Sridharan S, Rege M, Bailey R. Gaze-based image retrieval system using dual eye-trackers. 2012 Jan 12 Presented at: 2012 IEEE International Conference on Emerging Signal Processing Applications; January 12-14, 2012; Las Vegas, NV, USA p. 37. [doi: 10.1109/ESPA.2012.6152440]
19. Jacob MG, Wachs JP, Packer RA. Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images. J Am Med Inform Assoc 2013 Jun;20(e1):e183-e186 [FREE Full text] [doi: 10.1136/amiajnl-2012-001212] [Medline: 23250787]
20. Tani B, Maia R, von Wangenheim A. A Gesture Interface for Radiological Workstations. In: Twentieth IEEE International Symposium on Computer-Based Medical Systems. Maribor: IEEE; 2007 Jun 20 Presented at: CMBS'07; June 20-22, 2007; Maribor, Slovenia p. 07. [doi: 10.1109/CBMS.2007.6]
21. Zudilova-Seinstra E, de Koning P, Suinesiaputra A, van Schooten B, van der Geest R, Reiber J, et al. Evaluation of 2D and 3D glove input applied to medical image analysis. Int J Hum Comput Stud 2010 Jun;68(6):355-369 [FREE Full text] [doi: 10.1016/j.ijhcs.2009.08.001]
22. Kirmizibayrak C. Interactive Volume Visualization and Editing Methods for Surgical Applications. Washington, DC: George Washington University; 2001.
23. Bigdelou A, Schwarz A, Navab N. An adaptive solution for intra-operative gesture-based human-machine interaction. In: Proceedings of the 2012 ACM international conference on Intelligent User Interfaces. New York, NY, USA: ACM; 2012 Presented at: IUI'12; February 14-17, 2012; Lisbon, Portugal p. 75-84. [doi: 10.1145/2166966.2166981]
24. Ren G, O'Neill E. 3D selection with freehand gesture. Comput Graph 2013 May;37(3):101-120. [doi: 10.1016/j.cag.2012.12.006]
25. Nishikawa A, Hosoi T, Koara K, Negoro D, Hikita A, Asano S, et al. FAce MOUSe: a novel human-machine interface for controlling the position of a laparoscope. IEEE Trans Robot Autom 2003 Oct;19(5):825-841. [doi: 10.1109/TRA.2003.817093]
26. Wachs JP, Stern HI, Edan Y, Gillam M, Handler J, Feied C, et al. A gesture-based tool for sterile browsing of radiology images. J Am Med Inform Assoc 2008;15(3):321-323 [FREE Full text] [doi: 10.1197/jamia.M241] [Medline: 18451034]

27. Wachs J, Stern H, Edan Y, Gillam M, Feied C, Smithd M, et al. Real-time hand gesture interface for browsing medical images. Int J Intell Comput Med Sci Image Process 2008 Jan;2(1):15-25 [FREE Full text] [doi: 10.1080/1931308X.2008.10644149]
28. Soutschek S, Penne J, Hornegger J, Kornhuber J. 3-D gesture-based scene navigation in medical imaging applications using Time-of-Flight cameras. In: IEEE. Anchorage, AK: IEEE; 2008 Presented at: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; June 23-28, 2008; Anchorage, AK, USA p. 08. [doi: 10.1109/CVPRW.2008.4563162]
29. Kipshagen T, Tronnier V, Bonsanto M, Hofmann UG. Touch-marker-free interaction with medical software. Berlin, Heidelberg: Springer; 2009 Sep 07 Presented at: World Congress on Medical Physics and Biomedical Engineering; September 7-12, 2009; Munich, Germany p. 7-12 URL: https://doi.org/10.1007/978-3-642-03906-5_21 [doi: 10.1007/978-3-642-03906-5_21]
30. Ruppert GC, Reis LO, Amorim PH, de Moraes TF, da Silva JV. Touchless gesture user interface for interactive image visualization in urological surgery. World J Urol 2012 Oct;30(5):687-691. [doi: 10.1007/s00345-012-0879-0] [Medline: 22580994]
31. Gallo L, De Pietro G, Marra I. 3D interaction with volumetric medical data: experiencing the Wiimote. In: Proceedings of the 1st international conference on Ambient media and systems. Quebec, Canada: ICTS, editor; 2008 Presented at: Ambi-Sys'08; February 11-14, 2008; Brussels, Belgium. [doi: 10.1145/1363163.1363177]
32. Hansen C, Köhn A, Schlichting S, Weiler F, Zidowitz S, Kleemann M, et al. Intraoperative modification of resection plans for liver surgery. Int J CARS 2008 Jun 4;3(3-4):291-297 [FREE Full text] [doi: 10.1007/s11548-008-0161-5]
33. Gallo L, De Pietro G, Coronato A. Toward a natural interface to virtual medical imaging environments. In: Proceedings of the working conference on Advanced visual interfaces. New York: ACM; 2008 Presented at: AVI'08; May 28-30, 2008; Napoli, Italy URL: https://dl.acm.org/citation.cfm?id=1385651
34. Gallo L, Pietro G. Input devices and interaction techniques for VR-enhanced medicine. In: Jeong J, Damiani E, editors. Multimedia Techniques for Device and Ambient Intelligence. Boston, MA: Springer US; 2009:115.
35. Gallo L, Minutolo A, de Pietro G. A user interface for VR-ready 3D medical imaging by off-the-shelf input devices. Comput Biol Med 2010 Mar;40(3):350-358. [doi: 10.1016/j.compbiomed.2010.01.006] [Medline: 20149912]
36. Gallo L. A glove-based interface for 3D medical image visualization. In: Tsihrintzis G, Damiani E, Virvou M, Howlett R, Jain L, editors. Intelligent Interactive Multimedia Systems and Services. Berlin Heidelberg: Springer; 2010:221.
37. Chang Y, Chen S, Huang J. A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. Res Dev Disabil 2011;32(6):2566-2570. [doi: 10.1016/j.ridd.2011.07.002] [Medline: 21784612]
38. Leiker AM, Miller M, Brewer L, Nelson M, Siow M, Lohse K. The relationship between engagement and neurophysiological measures of attention in motion-controlled video games: a randomized controlled trial. JMIR Serious Games 2016 Apr 21;4(1):e4 [FREE Full text] [doi: 10.2196/games.5460] [Medline: 27103052]
39. Simor FW, Brum MR, Schmidt JD, Rieder R, de Marchi AC. Usability evaluation methods for gesture-based games: a systematic review. JMIR Serious Games 2016 Oct 4;4(2):e17 [FREE Full text] [doi: 10.2196/games.5860] [Medline: 27702737]
40. Dimaguila GL, Gray K, Merolli M. Person-generated health data in simulated rehabilitation using Kinect for stroke: literature review. JMIR Rehabil Assist Technol 2018 May 8;5(1):e11 [FREE Full text] [doi: 10.2196/rehab.9123] [Medline: 29739739]
41. Gallagher A, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surg Endosc 2002 Dec;16(12):1746-1752. [doi: 10.1007/s00464-001-8215-6] [Medline: 12140641]
42. Korndorffer J, Clayton J, Tesfay S, Brunner W, Sierra R, Dunne J, et al. Multicenter construct validity for southwestern laparoscopic videotrainer stations. J Surg Res 2005 Sep;128(1):114-119. [doi: 10.1016/j.jss.2005.03.014] [Medline: 15916767]
43. Ritter E, Kindelan T, Michael C, Pimentel EA, Bowyer MW. Concurrent validity of augmented reality metrics applied to the fundamentals of laparoscopic surgery (FLS). Surg Endosc 2007 Aug;21(8):1441-1445. [doi: 10.1007/s00464-007-9261-5] [Medline: 17593461]
44. Hennessey I, Hewett P. Construct, concurrent, and content validity of the eoSim laparoscopic simulator. J Laparoendosc Adv Surg Tech A 2013 Oct;23(10):855-860. [doi: 10.1089/lap.2013.0229] [Medline: 23968255]
45. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-63; discussion 463. [doi: 10.1097/01.SLA.0000028969.51489.B4] [Medline: 12368674]
46. Schijven M, Jakimowicz J, Broeders IA, Tseng L. The Eindhoven laparoscopic cholecystectomy training course--improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum. Surg Endosc 2005 Sep;19(9):1220-1226. [doi: 10.1007/s00464-004-2240-1] [Medline: 16132332]
47. Gurusamy K, Aggarwal R, Palanivelu L, Davidson B. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. Br J Surg 2008 Sep;95(9):1088-1097. [doi: 10.1002/bjs.6344] [Medline: 18690637]

48. Larsen C, Oestergaard J, Ottesen B, Soerensen J. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta Obstet Gynecol Scand 2012 Sep;91(9):1015-1028. [doi: 10.1111/j.1600-0412.2012.01482.x] [Medline: 22693954]
49. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. Br Med J 2005 Nov 5;331(7524):1064-1065 [FREE Full text] [doi: 10.1136/bmj.38636.593461.68] [Medline: 16230312]
50. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
51. 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 healthcare interventions: explanation and elaboration. Br Med J 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
52. Sánchez-Margallo FM, Sánchez-Margallo JA, Moyano-Cuevas J, Pérez EM, Maestre J. Use of natural user interfaces for image navigation during laparoscopic surgery: initial experience. Minim Invasive Ther Allied Technol 2017 Oct;26(5):253-261. [doi: 10.1080/13645706.2017.1304964] [Medline: 28349758]
53. Grange S, Terrence W, Fong T, Baur C. M/ORIS: A medical/operating room interaction system. In: Proceedings of the 6th international conference on Multimodal interfaces. 2004 Presented at: ICMI'04; October 13-15, 2004; State College, PA, USA p. 159-166. [doi: 10.1145/1027933.1027962]
54. Bizzotto N, Costanzo A, Bizzotto L, Regis D, Sandri A, Magnan B. Leap motion gesture control with OsiriX in the operating room to control imaging: first experiences during live surgery. Surg Innov 2014 Dec;21(6):655-656. [doi: 10.1177/1553350614528384] [Medline: 24742500]
55. Bizzotto N, Costanzo A, Maluta T, Dall?Oca C, Lavini F, Sandri A. Preliminary experience with the use of leap motion gesture control to manage imaging in the operating room. J Orthopaed Traumatol 2014 Nov;15(Suppl 1):19-20 [FREE Full text]
56. Streba C, Gheonea I, Streba L, Sandulescu L, Saftoiu A, Gheone D. Virtual Palpation Model -combining spiral CT and elastography data: a proof-of-concept study. Gastroenterology 2014;146(5):344-345 [FREE Full text] [doi: 10.1016/S0016-5085(13)63684-7]
57. Nouei M, Kamyad A, Soroush A, Ghazalbash S. A comprehensive operating room information system using the Kinect sensors and RFID. J Clin Monit Comput 2015 Apr;29(2):251-261. [doi: 10.1007/s10877-014-9591-5] [Medline: 25017016]
58. Hettig J, Saalfeld P, Luz M, Becker M, Skalej M, Hansen C. Comparison of gesture and conventional interaction techniques for interventional neuroradiology. Int J Comput Assist Radiol Surg 2017 Sep;12(9):1643-1653. [doi: 10.1007/s11548-017-1523-7] [Medline: 28120179]
59. Johnson R, O?Hara K, Sellen A, Cousins C, Criminisi C. Exploring the Potential for Touchless Interaction in Image-Guided Interventional Radiology. 2011 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC, Canada p. 3323-3332 URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2011/05/chi2011_paper188.pdf [doi: 10.1145/1978942.1979436]
60. Hötker AM, Pitton MB, Mildenberger P, Düber C. Speech and motion control for interventional radiology: requirements and feasibility. Int J Comput Assist Radiol Surg 2013 Nov;8(6):997-1002. [doi: 10.1007/s11548-013-0841-7] [Medline: 23580026]
61. Tan JH, Chao C, Zawaideh M, Roberts AC, Kinney TB. Informatics in Radiology: developing a touchless user interface for intraoperative image control during interventional radiology procedures. Radiographics 2013;33(2):E61-E70. [doi: 10.1148/rg.332125101] [Medline: 23264282]
62. Iannessi A, Marcy P, Clatz O, Fillard P, Ayache N. Touchless intra-operative display for interventional radiologist. Diagn Interv Imaging 2014 Mar;95(3):333-337 [FREE Full text] [doi: 10.1016/j.diii.2013.09.007] [Medline: 24176864]
63. Bercu Z, Patil V, Patel RS, Kim E, Nowakowski F, Lookstein R. Abstracts of the BSIR 2013 Annual Scientific Meeting, November 13-15, 2013, Manchester, England. Cardiovasc Intervent Radiol 2014 Jan;37(Suppl 1):1-82 [FREE Full text] [doi: 10.1007/s00270-013-0835-4] [Medline: 24425448]
64. Bercu Z, Patil VV, Patel R, Kim E, Nowakowski S, Lookstein R, et al. Use of hands free gesture-based imaging control for vessel identification during hepatic transarterial chemoembolization and selective internal radiotherapy procedures. J Vasc Interv Radiol 2015 Feb;26(2):S186-S187 [FREE Full text] [doi: 10.1016/j.jvir.2014.12.499]
65. Mentis H, O'Hara K, Sellen A, Rikin TR. Interaction proxemics and image use in neurosurgery. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Conference on Computer-Human Interaction; 2012 Presented at: CHI'12; May 5-10, 2012; New York, NY, USA p. 927-936. [doi: 10.1145/2207676.2208536]
66. Wright T, de Ribaupierre S, Eagleson R. Design and evaluation of an augmented reality simulator using leap motion. Healthc Technol Lett 2017 Oct;4(5):210-215 [FREE Full text] [doi: 10.1049/htl.2017.0070] [Medline: 29184667]
67. Yoshimitsu K, Muragaki Y, Maruyama T, Saito T, Suzuki T, Ikuta S. Clinical trials of the non-touch intraoperative image controllable interface system using KINECT(TM). Int J Comput Assist Radiol Surg 2012;7(Suppl 1):S209-S210.
68. Yoshimitsu K, Muragaki Y, Maruyama T, Yamato M, Iseki H. Development and initial clinical testing of "OPECT": an innovative device for fully intangible control of the intraoperative image-displaying monitor by the surgeon. Neurosurgery 2014 Mar;10 Suppl 1:46-50; discussion 50. [doi: 10.1227/NEU.0000000000000214] [Medline: 24141478]

69. di Tommaso L, Aubry S, Godard J, Katranji H, Pauchot J. A new human machine interface in neurosurgery: The Leap Motion(®). Technical note regarding a new touchless interface. Neurochirurgie 2016 Jun;62(3):178-181. [doi: 10.1016/j.neuchi.2016.01.006] [Medline: 27234915]
70. Xu X, Zheng Y, Yao S, Sun G, Xu B, Chen X. A low-cost multimodal head-mounted display system for neuroendoscopic surgery. Brain Behav 2018 Dec;8(1):e00891 [FREE Full text] [doi: 10.1002/brb3.891] [Medline: 29568688]
71. Henseler H, Kuznetsova A, Vogt P, Rosenhahn B. Validation of the Kinect device as a new portable imaging system for three-dimensional breast assessment. J Plast Reconstr Aesthet Surg 2014 Apr;67(4):483-488. [doi: 10.1016/j.bjps.2013.12.025] [Medline: 24513562]
72. Wheat JS, Choppin S, Goyal A. Development and assessment of a Microsoft Kinect based system for imaging the breast in three dimensions. Med Eng Phys 2014 Jun;36(6):732-738. [doi: 10.1016/j.medengphy.2013.12.018] [Medline: 24507690]
73. Pöhlmann ST, Harkness E, Taylor C, Gandhi A, Astley S. Preoperative implant selection for unilateral breast reconstruction using 3D imaging with the Microsoft Kinect sensor. J Plast Reconstr Aesthet Surg 2017 Aug;70(8):1059-1067. [doi: 10.1016/j.bjps.2017.04.005] [Medline: 28595842]
74. Klumb F, Dubois-Ferriere V, Roduit N, Barea C, Strgar T, Ahmed K. CARS 2017-Computer Assisted Radiology and Surgery Proceedings of the 31st International Congress and Exhibition Barcelona, Spain, June 20-24, 2017. Int J Comput Assist Radiol Surg 2017 Jun;12(Suppl 1):1-286 [FREE Full text] [doi: 10.1007/s11548-017-1588-3] [Medline: 28527024]
75. Pauly O, Diotte B, Fallavollita P, Weidert S, Euler E, Navab N. Machine learning-based augmented reality for improved surgical scene understanding. Comput Med Imaging Graph 2015 Apr;41:55-60. [doi: 10.1016/j.compmedimag.2014.06.007] [Medline: 24998759]
76. Jacob MG, Wachs JP. Context-based hand gesture recognition for the operating room. Pattern Recognit Lett 2014 Jan;36:196-203. [doi: 10.1016/j.patrec.2013.05.024]
77. Hughes P, Nestorov N, Healy N, Sheehy N, O'Hare N. Comparing the utility and usability of the Microsoft Kinect and Leap Motion sensor devices in the context of their application for gesture control of biomedical images. 2015 Presented at: ECR 2015; March 4–8, 2015; Vienna. [doi: 10.1594/ecr2015/B-1192]
78. O’Hara K, Gonzalez G, Penney G, Sellen A, Corish R, Mentis H, et al. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. Comput Supported Coop Work 2014 May 7;23(3):299-337 [FREE Full text] [doi: 10.1007/s10606-014-9203-4]
79. Kirmizibayrak C, Radeva N, Wakid M, Philbeck J, Sibert J, Hahn J. Evaluation of gesture based interfaces for medical volume visualization tasks. Int J Virtual Real 2012:1-13 [FREE Full text] [doi: 10.1145/2087756.2087764]
80. Wipfli R, Dubois-Ferrière V, Budry S, Hoffmeyer P, Lovis C. Gesture-controlled image management for operating room: a randomized crossover study to compare interaction using gestures, mouse, and third person relaying. PLoS One 2016;11(4):e0153596 [FREE Full text] [doi: 10.1371/journal.pone.0153596] [Medline: 27082758]
81. Ogura T, Sato M, Ishida Y, Hayashi N, Doi K. Development of a novel method for manipulation of angiographic images by use of a motion sensor in operating rooms. Radiol Phys Technol 2014 Jul;7(2):228-234. [doi: 10.1007/s12194-014-0259-0] [Medline: 24609904]
82. Wachs J, Stern H, Edan Y, Gillam M, Feied C, Smith M. A Real-Time Hand Gesture Interface for Medical Visualization Applications. In: Tiwari A, Roy R, Knowles J, Avineri E, Dahal K, editors. Applications of Soft Computing, Volume 36 of Advances in Intelligent and Soft Computing. Berlin Heidelberg: Springer; 2006:153.
83. Jacob M, Cange C, Packer R, Wachs J. Intention, context and gesture recognition for sterile MRI navigation in the operating room. In: Alvarez L, Mejail M, Gomez L, Jacobo J, editors. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Volume 7441 of Lecture Notes in Computer Science. Berlin Heidelberg: Springer; 2012:220-227.
84. Frame M. A novel system for hands free manipulation of digital X-rays in a sterile environment using consumer electronics and software. Int J Comput Assist Radiol Surg 2012;7(Supplement 1):S208 [FREE Full text]
85. Ebert L, Hatch G, Thali M, Ross S. Invisible touch—Control of a DICOM viewer with finger gestures using the Kinect depth camera. J Forensic Radiol Imaging 2013 Jan;1(1):10-14 [FREE Full text] [doi: 10.1016/j.jofri.2012.11.006]
86. Ogura T, Sato M, Ishida Y, Hayashi N, Doi K. Development of a novel method for manipulation of angiographic images by use of a motion sensor in operating rooms. Radiol Phys Technol 2014 Jul;7(2):228-234. [doi: 10.1007/s12194-014-0259-0] [Medline: 24609904]
87. Ebert L, Flach P, Thali M, Ross S. Out of touch – a plugin for controlling OsiriX with gestures using the leap controller. J Forensic Radiol Imaging 2014 Jul;2(3):126-128 [FREE Full text] [doi: 10.1016/j.jofri.2014.05.006]
88. Rossol N, Cheng I, Shen R, Basu A. Touchfree medical interfaces. Conf Proc IEEE Eng Med Biol Soc 2014;2014:6597-6600. [doi: 10.1109/EMBC.2014.6945140] [Medline: 25571508]
89. Iannessi A, Marcy PY, Clatz O, Ayache N, Fillard P. Touchless user interface for intraoperative image control: almost there. Radiographics 2014;34(4):1142-1144. [doi: 10.1148/rg.344135158] [Medline: 25019447]
90. Widmer A, Schaer R, Markonis D, Müller H. Gesture interaction for content--based medical image retrieval. ACM New York, NY, USA: ACM International Conference on Multimedia Retrieval. Glasgow; 2014 Presented at: ICMR'14; 2014; Glasgow, United Kingdom. [doi: 10.1145/2578726.2578804]

91. Ogura T, Sato M, Kadowaki Y, Yasumoto Y, Okajima M, Tsutsumi S. Development of a new method for manipulation of dental images using a motion sensor in dentistry. 2015 Presented at: ECR 2015; March 4-8, 2015; Vienna, Austria. [doi: 10.1594/ecr2015/C-0251]
92. Ogura T, Sato M, Ishida Y, Hayashi N, Doi K. Development of a novel method for manipulation of angiographic images by use of a motion sensor in operating rooms. Radiol Phys Technol 2014 Jul;7(2):228-234. [doi: 10.1007/s12194-014-0259-0] [Medline: 24609904]
93. Mewes A, Saalfeld P, Riabikin O, Skalej M, Hansen C. A gesture-controlled projection display for CT-guided interventions. Int J Comput Assist Radiol Surg 2016 Jan;11(1):157-164. [doi: 10.1007/s11548-015-1215-0] [Medline: 25958060]
94. Nainggolan F, Siregar B, Fahmi F. Anatomy learning system on human skeleton using Leap Motion Controller. : IEEE; 2016 Aug 15 Presented at: 2016 3rd International Conference on Computer and Information Sciences (ICCOINS); August 15-17, 2016; Kuala Lumpur, Malaysia p. 2016-2013. [doi: 10.1109/ICCOINS.2016.7783260]
95. Virag I, Stoicu-Tivadar L, Crisan-Vida M. Gesture-based interaction in medical interfaces. 2016 Jul 11 Presented at: 2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI); May 12-14, 2016; Timisoara, Romania. [doi: 10.1109/SACI.2016.7507339]
96. Juhnke B, Berron M, Philip A, Williams J, Holub J, Winer E. Comparing the microsoft kinect to a traditional mouse for adjusting the viewed tissue densities of three-dimensional anatomical structures. 2013 Presented at: Medical Imaging 2013: Image Perception, Observer Performance, and Technology Assessment; 2013; Baltimore, Maryland, USA p. 86731 URL: https://doi.org/10.1117/12.2006994 [doi: 10.1117/12.2006994]
97. Pulijala Y, Ma M, Pears M, Peebles D, Ayoub A. An innovative virtual reality training tool for orthognathic surgery. Int J Oral Maxillofac Surg 2018 Sep;47(9):1199-1205. [doi: 10.1016/j.ijom.2018.01.005] [Medline: 29398172]
98. Pulijala Y, Ma M, Pears M, Peebles D, Ayoub A. Effectiveness of immersive virtual reality in surgical training-a randomized control trial. J Oral Maxillofac Surg 2018 May;76(5):1065-1072. [doi: 10.1016/j.joms.2017.10.002] [Medline: 29104028]
99. Placitelli A, Gallo L. 3D point cloud sensors for low-cost medical in-situ visualization. USA: IEEE; 2011 Presented at: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW); November 12-15, 2011; Atlanta, GA, USA p. 596. [doi: 10.1109/BIBMW.2011.6112435]
100. Samosky JT, Wang B, Nelson DA, Bregman R, Hosmer A, Weaver RA. BodyWindows: enhancing a mannequin with projective augmented reality for exploring anatomy, physiology and medical procedures. Stud Health Technol Inform 2012;173:433-439. [Medline: 22357032]
101. Blum T, Kleeberger V, Bichlmeier C, Navab N. Mirracle: an augmented reality magic mirror system for anatomy education. 2012 Presented at: 2012 IEEE Virtual Reality Workshops (VRW); March 4-8, 2012; Costa Mesa, CA, USA p. 433-439. [doi: 10.1109/VR.2012.6180909]
102. Dargar S, Nunno A, Sankaranarayanan G, De S. Microsoft Kinect based head tracking for Life Size Collaborative Surgical Simulation Environments (LS-CollaSSLE). Stud Health Technol Inform 2013;184:109-113. [doi: 10.3233/978-1-61499-209-7-109] [Medline: 23400140]
103. Juhnke B. Iowa State University. 2013. Evaluating the Microsoft Kinect compared to the mouse as an effective interaction device for medical imaging manipulations URL: https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=4362&context=etd
104. Guo X, Lopez L, Yu Z, Steiner KV, Barner K, Bauer T, et al. A portable immersive surgery training system using RGB-D sensors. Stud Health Technol Inform 2013;184:161-167. [doi: 10.3233/978-1-61499-209-7-161] [Medline: 23400150]
105. Yang Y, Guo X, Yu Z, Steiner KV, Barner KE, Bauer TL, et al. An immersive surgery training system with live streaming capability. Stud Health Technol Inform 2014;196:479-485. [doi: 10.3233/978-1-61499-375-9-479] [Medline: 24732560]
106. Hochman JB, Unger B, Kraut J, Pisa J, Hombach-Klonisch S. Gesture-controlled interactive three dimensional anatomy: a novel teaching tool in head and neck surgery. J Otolaryngol Head Neck Surg 2014;43:38 [FREE Full text] [doi: 10.1186/s40463-014-0038-2] [Medline: 25286966]
107. Kocev B, Ritter F, Linsen L. Projector-based surgeon-computer interaction on deformable surfaces. Int J Comput Assist Radiol Surg 2014 Mar;9(2):301-312. [doi: 10.1007/s11548-013-0928-1] [Medline: 23888316]
108. Alvarez-Lopez F, Maina MF, Saigí-Rubió F. Natural user interfaces: is it a solution to accomplish ubiquitous training in minimally invasive surgery? Surg Innov 2016 Aug;23(4):429-430. [doi: 10.1177/1553350616639145] [Medline: 27009688]
109. Juanes JA, Gómez JJ, Peguero PD, Ruisoto P. Digital environment for movement control in surgical skill training. J Med Syst 2016 Jun;40(6):133. [doi: 10.1007/s10916-016-0495-4] [Medline: 27091754]
110. Svendsen MB, Preisler L, Hillingsoe JG, Svendsen LB, Konge L. Using motion capture to assess colonoscopy experience level. World J Gastrointest Endosc 2014 May 16;6(5):193-199 [FREE Full text] [doi: 10.4253/wjge.v6.i5.193] [Medline: 24891932]
111. Colella S, Svendsen MB, Konge L, Svendsen LB, Sivapalan P, Clementsen P. Assessment of competence in simulated flexible bronchoscopy using motion analysis. Respiration 2015;89(2):155-161 [FREE Full text] [doi: 10.1159/000369471] [Medline: 25591730]
112. Coles T, Cao C, Dumas C. SAGES. 2014. ETrack: An affordable Ergonomic assessment tool for surgical settings URL: http://www.sages.org/meetings/annual-meeting/abstracts-archive/etrack-an-affordable-ergonomic-assessment-tool-for-surgical-settings/ [accessed 2019-04-02] [WebCite Cache ID 77KXM9p9R]

113. Kim Y, Kim P, Selle R, Shademan A, Krieger A. Experimental evaluation of contact-less hand tracking systems for tele-operation of surgical tasks. 2014 May 31 Presented at: 2014 IEEE International Conference on Robotics and Automation (ICRA); May 31-June 7, 2014; Hong Kong, China p. 2014. [doi: 10.1109/ICRA.2014.6907364]
114. Beyl T, Schreiter L, Nicolai P, Raczkowsky J, Wörn H. 3D perception technologies for surgical operating theatres. Stud Health Technol Inform 2016;220:45-50. [Medline: 27046552]
115. Jacob M, Li Y, Akingba G, Wachs JP. Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room. J Robot Surg 2012 Mar;6(1):53-63. [doi: 10.1007/s11701-011-0325-0] [Medline: 27637980]
116. Despinoy F, Sánchez A, Zemiti N, Jannin P, Poignet P. Comparative assessment of a novel optical human-machine interface for laparoscopic telesurgery. In: Stoyanov D, editor. Information Processing in Computer-Assisted Interventions. Cham: Springer; 2014:21.
117. Vargas H, Vivas O. Gesture recognition system for surgical robot's manipulation. 2014 Presented at: 2014 XIX Symposium on Image, Signal Processing and Artificial Vision; September 17-19, 2014; Armenia, Colombia. [doi: 10.1109/STSIVA.2014.7010172]
118. Travaglini TA, Swaney PJ, Weaver KD, Webster RJ. Initial experiments with the leap motion as a user interface in robotic endonasal surgery. Robot Mechatron (2015) 2016;37:171-179 [FREE Full text] [doi: 10.1007/978-3-319-22368-1_17] [Medline: 26752501]
119. Despinoy F, Zemiti N, Forestier G, Sánchez A, Jannin P, Poignet P. Evaluation of contactless human-machine interface for robotic surgical training. Int J Comput Assist Radiol Surg 2018 Jan;13(1):13-24. [doi: 10.1007/s11548-017-1666-6] [Medline: 28914409]
120. Lahanas V, Loukas C, Georgiou K, Lababidi H, Al-Jaroudi D. Virtual reality-based assessment of basic laparoscopic skills using the Leap Motion controller. Surg Endosc 2017 Dec;31(12):5012-5023. [doi: 10.1007/s00464-017-5503-3] [Medline: 28466361]
121. Kowalewski K, Hendrie JD, Schmidt MW, Garrow CR, Bruckner T, Proctor T, et al. Development and validation of a sensor- and expert model-based training system for laparoscopic surgery: the iSurgeon. Surg Endosc 2017 Dec;31(5):2155-2165. [doi: 10.1007/s00464-016-5213-2] [Medline: 27604368]
122. Pérez F, Sossa H, Martínez R, Lorias D. Video-based tracking of laparoscopic instruments using an orthogonal webcams system. Acad Sci Eng Technol Int J 2013;7(8):440-443 [FREE Full text] [doi: 10.5281/zenodo.1086517]
123. Oropesa I, de Jong T, Sánchez-González P, Dankelman J, Gómez E. Feasibility of tracking laparoscopic instruments in a box trainer using a Leap Motion Controller. Measurement 2016 Feb;80:115 [FREE Full text] [doi: 10.1016/j.measurement.2015.11.018]
124. Beck P. Free Patents Online. 2016. Accurate Three-dimensional Instrument Positioning URL: http://www.freepatentsonline.com/20160354152.pdf [accessed 2019-04-02] [WebCite Cache ID 77KST95Tx]
125. Owlia M, Khabbazan M, Mirbagheri MM, Mirbagheri A. Real-time tracking of laparoscopic instruments using kinect for training in virtual reality. Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:3945-3948. [doi: 10.1109/EMBC.2016.7591590] [Medline: 28269148]
126. Partridge RW, Brown FS, Brennan PM, Hennessey IA, Hughes MA. The LEAPTM gesture interface device and take-home laparoscopic simulators: a study of construct and concurrent validity. Surg Innov 2016 Feb;23(1):70-77. [doi: 10.1177/1553350615594734] [Medline: 26178693]
127. Sun X, Byrns S, Cheng I, Zheng B, Basu A. Smart sensor-based motion detection system for hand movement training in open surgery. J Med Syst 2017 Feb;41(2):24. [doi: 10.1007/s10916-016-0665-4] [Medline: 28000118]
128. Hartmann F, Schlaefer A. Feasibility of touch-less control of operating room lights. Int J Comput Assist Radiol Surg 2013 Mar;8(2):259-268. [doi: 10.1007/s11548-012-0778-2] [Medline: 22806717]
129. Mauser S, Burgert O. Touch-free, gesture-based control of medical devices and software based on the leap motion controller. Stud Health Technol Inform 2014;196:265-270. [doi: 10.1371/journal.pone.0176123] [Medline: 24732520]
130. Schröder S, Loftfield N, Langmann B, Frank K, Reithmeier E. Contactless operating table control based on 3D image processing. Conf Proc IEEE Eng Med Biol Soc 2014;2014:388-392. [doi: 10.1109/EMBC.2014.6943610] [Medline: 25569978]
131. Jacob M, Wachs J. Context-based hand gesture recognition for the operating room. Pattern Recognit Lett 2014 Jan;36:196-203 [FREE Full text] [doi: 10.1016/j.patrec.2013.05.024]
132. Sweet R, Kowalewski T, Oppenheimer P, Weghorst S, Satava R. Face, content and construct validity of the University of Washington virtual reality transurethral prostate resection trainer. J Urol 2004 Nov;172(5 Pt 1):1953-1957. [doi: 10.1097/01.ju.0000141298.06350.4c] [Medline: 15540764]
133. 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. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
134. Weichert F, Bachmann D, Rudak B, Fisseler D. Analysis of the accuracy and robustness of the Leap Motion Controller. Sensors (Switzerland) 2013 Jan;13(5):6380-6393. [Medline: 23673687]
135. Bachmann D, Weichert F, Rinkenauer G. Evaluation of the leap motion controller as a new contact-free pointing device. Sensors (Basel) 2014 Dec 24;15(1):214-233 [FREE Full text] [doi: 10.3390/s150100214] [Medline: 25609043]

136. Guna J, Jakus G, Pogačnik M, Tomažič S, Sodnik J. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors (Basel) 2014 Feb 21;14(2):3702-3720 [FREE Full text] [doi: 10.3390/s140203702] [Medline: 24566635]
137. Khoshelham K, Elberink SO. Accuracy and resolution of Kinect depth data for indoor mapping applications. Sensors (Basel) 2012;12(2):1437-1454 [FREE Full text] [doi: 10.3390/s120201437] [Medline: 22438718]
138. Mendez I, Hansen B, Grabow C, Smedegaard E, Skogberg N, Uth X, et al. Evaluation of the Myo armband for the classification of hand motions. IEEE Int Conf Rehabil Robot 2017 Dec;2017:1211-1214. [doi: 10.1109/ICORR.2017.8009414] [Medline: 28813986]
139. Li C, Ren J, Huang H, Wang B, Zhu Y, Hu H. PCA and deep learning based myoelectric grasping control of a prosthetic hand. Biomed Eng Online 2018 Aug 6;17(1):107 [FREE Full text] [doi: 10.1186/s12938-018-0539-8] [Medline: 30081927]
140. Ur Rehman MZ, Waris A, Gilani S, Jochumsen M, Niazi IK, Jamil M, et al. Multiday EMG-based classification of hand motions with deep learning techniques. Sensors (Basel) 2018 Aug 1;18(8):1-16 [FREE Full text] [doi: 10.3390/s18082497] [Medline: 30071617]
141. Sánchez-Margallo JA, Sánchez-Margallo FM, Pagador Carrasco JB, Oropesa García I, Gómez Aguilera EJ, Moreno del Pozo J. Usefulness of an optical tracking system in laparoscopic surgery for motor skills assessment. Cir Esp 2014;92(6):421-428. [doi: 10.1016/j.ciresp.2013.01.006] [Medline: 23668944]

Abbreviations

3D: 3-dimensional
COTS: commercial off-the-shelf
EMBASE: Excerpta Medica dataBASE
IEEE: Institute of Electrical and Electronics Engineers
LMC: Leap Motion Controller
MeSH: Medical Subject Headings
MIS: minimally invasive surgery
MK: Microsoft Kinect

Edited by G Eysenbach; submitted 12.08.18; peer-reviewed by K Kowalewski, JA Sánchez Margallo, B Davies; comments to author 13.10.18; revised version received 04.01.19; accepted 25.01.19; published 03.05.19.

Please cite as:
Alvarez-Lopez F, Maina MF, Saigí-Rubió F
Use of Commercial Off-The-Shelf Devices for the Detection of Manual Gestures in Surgery: Systematic Literature Review
J Med Internet Res 2019;21(5):e11925
URL: https://www.jmir.org/2019/5/e11925/
doi:10.2196/11925
PMID:31066679

Viewpoint

Qualitative Interview Studies of Working Mechanisms in Electronic Health: Tools to Enhance Study Quality

Marianne TS Holter[1*], MA; Ayna B Johansen[1*], PhD; Ottar Ness[2*], PhD; Svend Brinkmann[3*], PhD; Mette T Høybye[4,5*], PhD; Håvar Brendryen[1*], PhD

[1]Norwegian Centre for Addiction Research (SERAF), Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[2]Department of Education and Lifelong Learning, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[3]Department of Communication and Psychology, Aalborg University, Aalborg, Denmark
[4]Elective Surgery Centre, Regional Hospital Silkeborg, Silkeborg, Denmark
[5]Department of Clinical Medicine, Interacting Minds Centre, Aarhus University, Aarhus, Denmark
[*]all authors contributed equally

Corresponding Author:
Marianne TS Holter, MA
Norwegian Centre for Addiction Research (SERAF)
Institute of Clinical Medicine
University of Oslo
Postboks 1039 Blindern
Oslo, 0315
Norway
Phone: 47 93623061
Email: marianne.holter@protonmail.com

Abstract

Future development of electronic health (eHealth) programs (automated Web-based health interventions) will be furthered if program design can be based on the knowledge of eHealth's working mechanisms. A promising and pragmatic method for exploring potential working mechanisms is qualitative interview studies, in which eHealth working mechanisms can be explored through the perspective of the program user. Qualitative interview studies are promising as they are suited for exploring what is yet unknown, building new knowledge, and constructing theory. They are also pragmatic, as the development of eHealth programs often entails user interviews for applied purposes (eg, getting feedback for program improvement or identifying barriers for implementation). By capitalizing on these existing (applied) user interviews to also pursue (basic) research questions of how such programs work, the knowledge base of eHealth's working mechanisms can grow quickly. To be useful, such interview studies need to be of sufficient quality, which entails that the interviews should generate enough data of sufficient quality relevant to the research question (ie, *rich data*). However, getting rich interview data on eHealth working mechanisms can be surprisingly challenging, as several of the authors have experienced. Moreover, when encountering difficulties as we did, there are few places to turn to, there are currently no guidelines for conducting such interview studies in a way that ensure their quality. In this paper, we build on our experience as well as the qualitative literature to address this need, by describing 5 challenges that may arise in such interviews and presenting methodological tools to counteract each challenge. We hope the ideas we offer will spark methodological reflections and provide some options for researchers interested in using qualitative interview studies to explore eHealth's working mechanisms.

(J Med Internet Res 2019;21(5):e10354) doi:10.2196/10354

KEYWORDS

telemedicine; eHealth; mobile health; telehealth; mHealth; interviews as topic; health care evaluation mechanisms; data collection

The Need to Identify the Working Mechanisms of Automated Electronic Health Programs

Building the next generation of automated electronic health (eHealth) programs will require a shift of attention from the performance of individual programs to a joint effort of understanding eHealth's working mechanisms [1,2]. The term *eHealth* is a broad term that may refer to many forms of technological health support; this paper concerns itself with automated Web-based interventions for mental and physical health purposes [3], or *eHealth programs*. The outcomes of individual eHealth programs are well established; a vast majority of the research published between 1996 and 2013 concerned outcome (74%). However, much fewer publications focused on processes (26%) [4]. Consequently, the working mechanisms that underlie the outcomes of individual interventions are substantially less documented [5-7]. This is not only theoretically unsatisfactory; it is also problematic when it comes to designing new eHealth programs, as there are few, if any, field-specific theories of models that can be used to inform program development.

Instead, the development of eHealth programs often relies on rather static traditional behavior change theories [2] or models from face-to-face therapy [8-11], under the assumption that the principles are transferrable to automated eHealth therapy. However, the interaction between a program user and an automated eHealth program is in many cases *not* static; many programs include different degrees of interactivity and tailoring [3,5], making traditional behavior change theories potentially unsuitable [1,2]. On the other hand, using models from face-to-face therapy may not be appropriate either, as automated eHealth therapy by definition does not involve human contact. As automated eHealth programs are neither static nor involve human contact, it is possible (or even likely) that the way they achieve their effects is not explained with established theories and models [1,2]. This suggests a need for research that can identify eHealth's working mechanisms, knowledge on which it is possible to build eHealth-specific theories and models.

Qualitative Interviews: A Promising and Pragmatic Method for Studying Electronic Health Programs' Working Mechanisms

eHealth's working mechanisms can be studied using various methods, but a promising and pragmatic venue of investigation is the qualitative interview, that is, "professional conversations (...) where knowledge is constructed in the inter-action between the interviewer and the interviewee (...) about a theme of mutual interest" [12]. The qualitative interview is a promising method for investigating eHealth working mechanisms as it grants unique access to participants' experiences and as it is especially suited to explore what is unknown [13]. Providing a means to explore the unknown makes qualitative interviews a potent method for generating new knowledge and theory [14,15], and some interview studies have already demonstrated their potential for uncovering important insights about the processes that may be involved in automated eHealth therapy [16-19].

The qualitative interview is also a pragmatic research method, as many researchers already conduct interviews with program users as part of an applied research goal (developing or implementing an intervention). In the process of conducting interviews with program users, a researcher may become intrigued by a more basic research question and may perhaps consider the pragmatic solution of pursuing both the applied and the basic research goal in the same interviews by simply adding questions to the existing interview guide. We believe that such studies mixing basic and applied research goals have the potential of becoming an important asset to the field, by accumulating knowledge on more general issues that may help us understand how eHealth therapy works.

However, to become such an asset, the interviews conducted in these studies should provide what in qualitative methodology is known as *thick descriptions* or *rich data* [20,21]. Rich data are usually considered a requirement for a valid qualitative analysis, and the concept signifies having enough data of sufficient quality relevant to the research question, including both variation (ie, data breadth) as well as details and nuances (ie, data depth) [22,23]. Data that are *not* rich—that lack in breadth or depth—might threaten the study's quality or the potential reach of its conclusions. Thus, getting rich data that inform the research question is an important aspect of a qualitative study. In the case of qualitative interviews, producing rich data means conducting interviews in a way that makes the participants spend a lot of time talking about the aspects that are central to the investigation, including both breadth and depth in their descriptions. This may seem straightforward, but it can be surprisingly difficult in practice.

Difficult in Practice: The Case of a Study on a Person-To-Program Alliance

The reflections that are presented in this viewpoint paper arose from some of the authors' experiences with a specific interview study [24] (in review), the aim of which was to explore a potential person-to-program alliance. The study in question included the development of an alliance-supporting program [25], and in an early study phase, the interviews had both an applied and a basic purpose: the applied purpose was getting feedback for program improvement, and the basic purpose was exploring how the participants related to the program. By exploring how participants related to the program, we hoped to achieve a better understanding of a person-to-program alliance [26-28] as a potential eHealth working mechanism. However, it was surprisingly difficult to conduct interviews that would yield rich data on how the participants related to the program, and the initial interviews resulted in scant data to answer the research question. This left us with 3 options: (1) answering the basic research question with scant data, which would limit the conclusions we could draw, (2) abandoning the basic research question as unanswerable, or (3) trying to generate richer data by changing the way the interviews were being conducted.

We opted for trying to improve the interview method; however, we found no guidelines within the field of eHealth for how to conduct high-quality qualitative interview studies on potential eHealth working mechanisms. Therefore, we started looking more closely at the interviews we had conducted, asking ourselves what had gone wrong. This process led to the identification of a handful of problems that we believed were likely to have contributed to the difficulties in getting rich data on how the participants related to the program. As we started defining these problems, we discovered that we had also encountered several of them in other eHealth studies we had been involved in [29-33], and we therefore believed they could be relevant beyond the specific study we were currently engaged in.

We wanted our experiences to be of benefit to other researchers with similar agendas and interests, and we therefore sought to describe the problems we had encountered in a way that would maximize their generalizability. Thus, through discussion among ourselves and with other researchers, we conceptualized 5 interview challenges: achieving a joint understanding of the interview topic, keeping participants from straying off the focus of enquiry, aiding recall of specific program experiences, avoiding negative influence of the social interview situation, and structuring the dual-aim interview. Having identified the challenges, we consulted the literature on qualitative methodology to identify methodological tools to counteract each challenge.

Returning to the study that had started this process [24], we changed the interview method to include some of the tools we had identified. This markedly enhanced the quality of subsequent interviews, producing rich data to answer how the participants related to the eHealth program as well as whether this way of relating influenced change. Thus, although the original interview method generated scant data on the basic research question, the revised interview methodology led to interviews that could answer the same research question with rich data.

In short, although it seemed a pragmatic solution to use already-planned interviews to pursue the answer to a basic eHealth research question, we experienced that getting rich data on the basic research question was challenging. In the absence of guidelines for conducting high-quality qualitative interviews specifically adapted to the field of eHealth, the process we entered into led to an enhanced methodological awareness and specific methodological tools for increasing study quality. The main focus of this paper is to share the identified challenges and tools with the research community. However, before doing so, we will offer what we consider to be a handy heuristic for understanding some of these methodological challenges: the *invisible interaction* between eHealth program and program user.

A Handy Heuristic: The Invisible Interaction

We suggest that a person's interaction with any health intervention can be visualized in terms of a triangle, which includes the individual help seeker, the intervention, and the behavior change processes (Figure 1, adapted from Moen and Middelthon's discussion of interviews) [34]. A health intervention's working mechanisms can be conceptualized as how the interaction between the person and the health intervention influences the person's internal change processes. The interaction, in turn, can be described as a combination of the interaction's content (the *what* of the interaction) and the interactional processes (the *how* of the interaction). For example, psychotherapy's working mechanisms can be described as the therapy sessions' influence on the client's internal change processes. The therapy sessions, in turn, can be described as comprising 2 main elements: their content (eg, the topic discussed) and the interactional processes, when and how often interaction is initiated, how the interaction unfolds, how the next interaction is initiated, and so on.

However, interventions may differ according to how much the 2 interacting parties—the person and the intervention—influence the interactional content and the interactional processes. In the case of psychotherapy, both the client and the therapist highly influence both components of the interaction. Taking another example, a person reading a self-help book is also interacting with a health intervention: things also *act*, and people interact with them, in that the properties of a thing influence how a course of action involving that thing unfolds [35]. The working mechanisms of a self-help book can therefore also be described in terms of the interactional content and the interactional processes. However, the relative influence of the 2 interacting agents (reader and book) differ from the case of psychotherapy. The book decides the interactional content, although the reader largely decides the interactional processes: when and how often interaction is initiated (when to read again), how the interaction unfolds (what to read in what sequence), how the next interaction is initiated (picking up the book), and so on.

Considering the working mechanisms of an eHealth program, many programs will influence both the interactional content and the interactional processes. As with a self-help book, the interactional content will usually to a large extent be decided by the program. Moreover, just as a self-help book, the program is a *thing*, and many people are likely to think of things such as computer programs as inanimate objects with content. Indeed, the most prominent feature of eHealth programs is their content [3,10], even though they also may substantially influence the interactions with the user [3]. For example, eHealth programs may influence when and how often the interaction takes place (eg, through reminders to log on), how the interaction unfolds (eg, by responding with tailoring to user input), how the next interaction is initiated (eg, through invitation), and so on [5,10]. Some of these interactional processes may not be experienced directly by the individual user, for example, in the case of tailoring, the program may be adapted specifically to the user's input, but she or he nevertheless only sees 1 version of the program, masking the actual interaction. In sum, although eHealth programs may have a substantial influence on the interactional processes, the average program users may primarily focus on their content and think of them as inanimate objects that do not interact. In other words, to the user, the interaction with the program can be largely invisible (Figure 2).

The invisible interaction is a useful heuristic when considering the challenges of interview studies for exploring eHealth's working mechanisms. We previously stated that an asset of qualitative interviews is their potential to explore eHealth working mechanisms from the program user's perspective. However, from this perspective, part of the program's working mechanisms—the interactional processes— are maybe invisible to the participant, unless she or he purposefully directs his or her attention toward them. In other words, being largely invisible, the interactional processes may not be part of the participant's conscious experience that she or he is ready to share in an interview. This may create or contribute to certain challenges with exploring eHealth working mechanisms through interviews. We will now present 5 such challenges and suggest methodological tools to counteract them.

Figure 1. Working mechanisms of a behavior change intervention.

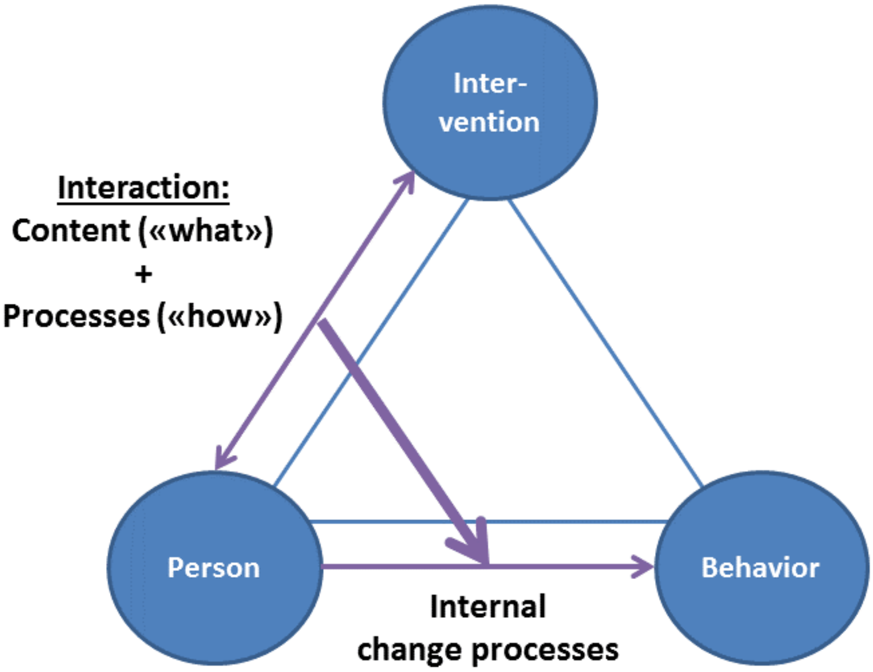


Figure 2. Working mechanisms of an automated electronic health intervention.

Interview Challenges and Tools

Achieving a Joint Understanding

When a researcher sets out to explore a potential eHealth working mechanism in an interview, it may be difficult to achieve a joint understanding of the interview topic together with the participant. For example, as mentioned previously, in the study that was the starting point for this paper, the researchers were interested in understanding how the participants related to the program [24]. The interview guide comprised mainly of descriptive interview questions—questions that ask the participant to describe a specific experience, which are usually recommended to get close to the participants' own experience [13]. Examples of descriptive interview questions were *could you tell me what you thought and felt the first time you used the program* and *could you describe the role the program has had in your quit attempt* (the program was for helping people quit smoking). However, the researchers struggled with superficial answers that did not seem to reveal anything about how the participants related to the program (such as *I thought the program was fine*) until 1 participant called the program *a secret friend*. The researchers were puzzled. Was this person's program experience unique? Why did other participants not talk about the program in this way at all? However, apart from a few statements similar to this one, the interviews were not generating data to answer the research question of how the participants related to the program.

Failure to get rich data on a research question may indicate a marginal phenomenon—or that the interviewer is failing to communicate the focus of enquiry in a way that facilitates joint understanding with the participant. We believe it is a truism that experience is multifaceted and that an experience can be described from many perspectives. For example, a client may describe a therapy session from a factual perspective of when and where it took place, from an experiential perspective of his or her emotions before, during, and after the session, from a historical perspective of the session as a stage in his or her spiritual development, and so on. If the interviewer's questions are mostly descriptive, there may be a scarcity of cues concerning which perspective to assume, leaving the decision up to the participant—and the participant's choice may not be the researcher's choice. This may be especially challenging in studies on eHealth working mechanisms, as the interactional processes may not be part of the participant's conscious experience. Therefore, descriptive questions asking for the participant's program experiences will perhaps not cause him or her to talk about the (invisible) person-program interaction but rather about the program as a thing with a content. On the other side of the conversation, the interviewer may fear that more direct questions onto the focus of enquiry will put words in the participant's mouth and disqualify any subsequent answer.

An interviewer can use several methodological tools to foster a joint understanding of the interview topic with the participant. One such tool is vignettes: vivid, exemplifying prose stories that guide the conversation toward a particular aspect of the participant's experience [36-38]. The interview vignette is constructed before the interviews and included in the interview guide. The vignette can be constructed on the basis of a participant account, on relevant literature, or on the researcher's current understanding of the processes under study. The interviewer might introduce the vignette by saying that she or he wants to share a story with the participant. After recounting the vignette, the interviewer can ask for the participant's reactions and ask follow-up questions (eg, if the participant has experienced anything similar or can provide a different perspective). Using several vignettes in the same interview can be a useful way of illustrating different perspectives on the research topic. This will implicitly communicate to the participant that all answers are acceptable, ensuring that the vignettes function to guide the conversation but not restrict the answers [36,38].

Another and more direct way of fostering a joint understanding of the interview topic is to involve the participants as coresearchers, or using epistemic interviewing [13,14,39,40]. In traditional qualitative interviewing, participants describe their subjective experiences and the data are analyzed and interpreted afterward by the researcher [13,14]. In contrast, a coresearcher design entails that the researcher shares his or her current understanding of the research topic and asks for the participant's views, and the research questions are investigated in collaboration. Involving participants as coresearchers also changes the roles of the interviewer, who becomes a sort of participant contributing with his or her perspective. This joint exploration entails that much of the analysis and validation is done in the interview [14,39].

A final tool to clarify and exhaust the interview topic is to ensure the possibility of conducting follow-up interviews [41]. A follow-up interview gives both the interviewer and the participant an opportunity to reflect on what was talked about in the first interview, allowing new insights or aspects to emerge [22]. It also gives the researcher an opportunity to clarify questions or test interpretations with the participant directly [39], giving more nuanced data and enhancing validity.

Coresearcher design and vignettes foster clearer communication, but they may also threaten the study's validity if the researcher holds on to his or her initial assumptions about the studied process, failing to acknowledge unexpected perspectives. To ensure that these tools strengthen and not hamper the quality of the study, the researcher should adopt what in psychotherapy is known as the *beginners mind*: remaining curious and receptive, open to all possibilities [42-44]. Furthermore, the interviewer must throughout the research process practice reflexivity, that is, considering how she or he may be affecting the study with "(…) thoughtful, conscious self-awareness" [44]. Reflexivity about, for example, preunderstandings, motivations, and the influence of previous experiences can lead to important insights [43,44]. Reflexive insights that could be of importance for analysis should be documented (eg, through memos or notes) [15,43,44]. Finally, documented reflections should be made part of the analysis and be made explicit to the reader [44].

Keeping Participants From Straying off the Focus of Enquiry

To allow time for joint exploration of the person-program interaction, it is necessary to limit the interview time spent on matters that are not at the core of the research question. Returning to Figure 2 and the triangle of program, user, and behavior change, the relative importance of each triangle endpoint will vary according to the research question: some parts of the triangle will be the focus of enquiry, whereas the other parts will be contextual. For example, in 1 study [33] (paper under preparation), the researchers interviewed patients who had gambling problems and had used a Web-based referral site to connect with problem gambling services. The focus of enquiry was their use of the website; the gambling problems were the context. However, the interviewer struggled with keeping the conversation focused on the website, as participants talked mostly about their personal history with gambling problems. When asked about their experiences with the website, they appeared to feel alienated and at a loss. Consequently, there was a lot of interview data on the participants' behavior change efforts—but little data on their use of the website.

If the participants continuously stray off the focus of enquiry by spending time on contextual aspects, it can threaten the data richness. Aspects that are contextual to the researcher may be aspects the participant wants to share or aspects she or he believes to be important to the investigation. The interviewer may try to lead the conversation back onto the focus of enquiry, but the participant may return to the contextual aspects, turning the interview into a battle over topic. Apart from being unpleasant for both, the result may be scant data on the focus of enquiry. When the focus of enquiry is potential eHealth working mechanisms, the invisible interaction may add to the challenge of straying off the topic. As the participant may be largely unaware of the interactional processes, she or he will instead talk about the aspects of which she or he is aware: the change processes (in isolation of the program) or the program (in isolation of the change processes). Information about the behavior change and about the program is certainly relevant contextual information, but talking about these aspects in isolation should not dominate the interview.

The interview conversation can be kept from straying off the research topic by using in-interview questionnaires to keep contextual answers short. The questionnaire can include questions addressing contextual issues (eg, *How long have you been worried about your gambling?* Or *Have you tried restricting how much you gamble before?*), together with any other questions that might serve as relevant analytic background (eg, demographics). The interviewer may fill out the questionnaire together with the participant at a suiting point during the interview. Using a piece of paper to fill out the answers will help keep the answers short, by providing limited space and communicating a wish for answers that the interviewer can write down. Short contextual answers will in turn leave more time for the focus of enquiry.

Aiding Recall of Specific Program Experiences

Sometimes participants may not recall program experiences in sufficient detail to answer the interviewer's questions. In the study that inspired this paper [24], the interviewer asked the participants to tell her about a program session they remembered especially well, thinking that she would use this session as a starting point for further descriptive interview questions [13]. To her surprise, several participants who were still active program users and had completed most sessions up until the time of the interview had difficulties remembering any particular program session at all.

Recalling specific program experiences may be challenging as although participants may be active program users at the time of the interview, they are not engaging with the program at that particular moment (unless you are combining the interview with a *think-aloud*-technique, discussed below) [45]. That means that to talk about program experiences, the participants must retrieve memories. However, program sessions may be short, and the participants are likely to use the program in between their other daily business. Consequently, program use may not be encoded as distinct episodic memories to begin with [46]; rather, these memories may be intertwined with other memories of everyday life. Thus, if the interviewer asks the participant to describe a program session, his or her question may not contain the right memory cues [47] to trigger memories of program use, and the participant may seemingly not recall any sessions at all. The invisible interaction may amplify this problem: if the participant is unaware of the program influencing the interaction, these program aspects will be even more difficult to retrieve on demand.

There are, however, methodological tools to amend the problem with recall in the interview situation: 1 such tool is to get *live* access to the person-program interaction through the think-aloud procedure [45]. In the think-aloud procedure, the participants go through (parts of) the program during the interview as the interviewer instructs the participant to *think aloud*, reporting all thoughts without censoring them. The interviewer should not interrupt the participant's flow of thoughts, and follow-up questions should be saved for after the think-aloud procedure is completed [45]. However, there are some limitations to this approach: unless the program comprises just 1 website or session, the researcher cannot use the think-aloud procedure to go through all program content, requiring him or her to select the most relevant sessions. Furthermore, when the focus of enquiry is working mechanisms within the invisible person-program interaction, the interviewer's presence may draw attention from the program's role in the interaction, adding to its invisibility. However, if these issues do not apply, the think-aloud procedure can enable a researcher to study possible eHealth working mechanisms as they happen, potentially removing the problem of recall.

Another tool for aiding recall is asking memory-facilitating interview questions. If program experiences have not been encoded as specific episodic memories, the interviewer's phrasing of questions becomes increasingly important, as the words she or he uses will influence the participant's memory-retrieval process by serving as memory cues [47]. The interviewer's choice of words can be guided by mapping the participant's program habits early in the interview. Knowledge of program habits can in turn be used to phrase questions in ways that contain memory cues; reflecting what the participant was doing before using the program, where she or he was, and his or her emotional state at the time of the experience [47]. Such memory-facilitating interview questions may help the participant disentangle the recall of program experiences from everyday life.

As a final note on program recall, it may not be necessary for the participant to remember any particular program session at all; the researcher must consider what level of detail is necessary to answer the research questions meaningfully. For some research questions, the sum of program experiences may be more important than any particular experience. If so, using the interview to discuss the participant's overall experience with the program can be more meaningful than facilitating recall of specific sessions [22].

Avoiding Negative Influence of the Social Interview Situation

All interviews are also social situations, and aspects of the social situation will influence the data [48]. In 1 of the interviews from the study that inspired this paper [24], a female interviewer interviewed a male participant, with the goal of understanding how he related to the eHealth program he had used. The interview was brief and disappointing; the participant's answers were short, and the topic was exhausted quickly. It was not until later that the interviewer became aware that she had been afraid of the participants judgment; that he would perceive her as a *typical woman*, valuing emotions (interactional processes; relating to the program) over facts (the program content). This subconscious fear had caused her to rush through the questions (which she during the interview had found awkward), partly answering some of them on behalf of the participant and ending the interview early.

Gender stereotypes are not the only potential social disturbances in an interview—other social roles may be prominent, and within eHealth research, the interviewer may be particularly prone to be perceived as an interviewer or clinician or interviewer or developer. Perceiving the interviewer as also a clinician may cause the participant to think of him or her as a therapeutic interactional partner and to be less attentive to the therapeutic agency of the eHealth program. Similarly, perceiving the interviewer as also a program developer may highlight the program as a thing made by someone else, making it more difficult to see the program's role as a therapeutic agent—or cause the participant to self-censor negative experiences, as 1 of the authors experienced in 2 different studies [29,31]. In both cases, the interviewer's presence may cause the participant to think of the *interviewer* as the interacting agent, pushing the experience of the *program* as an interacting agent to the background and adding to the interaction's invisibility. In sum, the social interview situation may cause the participant to talk differently about his or her program experiences than she or she would have otherwise. The consequence of this may be less rich data, or data that do not correctly represent the participant's experience.

The potentially negative influence of the social interview situation can be counteracted with methodological tools. The researcher acknowledging the potential negative influence of roles and stereotypes, both before and after the interviews, can minimize their negative effect. Before an interview, researchers should reflect on potentially salient social aspects and whether something should be done about them [43]. If circumstances can make the interviewer appear as a clinician or a program developer, the interviewer may try to change these circumstances beforehand, for example, by changing the interview location or considering how to dress or talk. Alternatively, these issues can be addressed explicitly in the beginning of the interview, clarifying the interviewer's role [19]. During the interview, the interviewer should try to monitor the social exchange [43], making notes of elements that may be impacting the conversation. After the interview, anything that might be of importance to the analysis should be documented [44]. These notes should be included somewhere easily accessible (eg, in the interview transcript or in a separate document) and analyzed as data that might inform, confirm, or qualify the analysis. Regarding the danger of additionally concealing the invisible interaction through the social exchange between the interviewer and the participant, the interviewer can try to arrange the interview situation so that it includes all 3 as potential agents: the participant, the program, and the interviewer [34].

Finally, it is important to acknowledge that although the social interview situation may sometimes be a negative influence on the data, it can also be an asset. Through the interviewer's reflexivity, the social situation may generate insights that would otherwise be missed. The interview in which the interviewer had rushed through the questions as she feared being labeled an *emotional woman* was considered as empirical material highlighting a possibly relevant aspect of how people *relate* to a program, namely, that relating to a program may go against social norms and produce feelings of embarrassment (in this case, as felt by the interviewer).

Structuring the Dual-Aim Interview

It was mentioned in the introduction that qualitative interviews are pragmatic for exploring potential eHealth working mechanisms as the development or implementation of eHealth programs often entail user interviews anyway. Therefore, researchers who are interested in exploring potential eHealth working mechanisms may do so through existing interviews with applied purposes. However, when applied and basic research goals are mixed like this in the same interview study, it may create an additional challenge in getting rich data on the basic research question. In the study that inspired this paper [24], early interviews had both an applied research goal (getting feedback for improving the program) and a basic research goal (understanding how the participants *related* to the program as a potential eHealth working mechanism). The interview guide started with questions addressing possible sources for program improvement (participants' likes or dislikes, specific program elements). Toward the end of the interview guide, questions on how the participants related to the program gradually increased in number (*Has the program ever made you happy*? *Have you ever been upset by the program?*). However, most participants answered interview questions on how they related to the program briefly and superficially, resulting in scant data.

Mixing applied and basic research aims can be problematic as different aims may require different interviewing modes. For the interviewer, changing from an *applied* interviewing mode to a *basic* interviewing mode will involve changing the point of his or her focal attention, that is, what to listen for and which follow-up questions to ask. For the participant, changing interviewing modes will involve changing how she or he is expected to answer, from talking more superficially about the breadth of his or her program experiences (*applied mode*) to talking in depth about a few aspects (*basic mode*). If the transition between the different modes is not explicit to the participant, she or he may answer interview questions with the wrong *mindset* —basic interview questions as if they were applied questions or applied interview questions as if they were basic. Unclear transitions may also cause the interviewer to miss important leads in the participant's answers because of the need to split his or her attention between the 2 research questions. The invisible interaction may exacerbate this challenge: in applied research, the program is treated as a thing, whereas in the search for basic working mechanisms, the program can be considered an interacting agent. An unclear transition between applied and basic research goals may make it more difficult for the participant to take the perspective of the program as an agent influencing the interaction.

Interviews with both applied and basic research aims may serve both aims through topical blocks and clear introductions. The transition can be facilitated by structuring the interview in topical blocks [49]: one covering the applied research question, another covering the basic research question. Topical blocks enable the interviewer to focus on 1 research question at a time, facilitating active listening and choosing following-up questions. The topical blocks should be kept separate; therefore, if the participant says something relevant for research question number 1 in the topical block of research question number 2, the interviewer's follow-up questions on this should be saved for the respective topical block. Furthermore, the transition between the different topical blocks should be made explicit through small introductions: first, a general introduction to the interview along with a presentation of the topical blocks, then separate introductions preceding each topical block. The introductions can even specify the interviewing modes and what the researcher expects of the participant in each section, for example, that the applied topical block involves factual questions and answers, whereas the basic topical block involves a coresearcher design with joint exploration. Providing the interview with structure and appropriate introductions helps both the participant and the interviewer into the right frame of mind, moving from 1 research question to another. In addition, structuring the dual-aim interview into topical blocks ensures that both research questions are being covered, instead of leaving this overview for analysis.

Concluding Thoughts

Conducting qualitative interviews is a promising and pragmatic approach for identifying the working mechanisms of automated eHealth programs. Existing user interviews for applied purposes can be used to also pursue basic research questions on eHealth working mechanisms. Researchers planning to conduct user interviews for applied purposes would be wise to ensure the possibility to pursue research questions concerning potential eHealth working mechanisms by including this purpose in the study information provided to ethics boards and prospective participants. However, getting rich data on eHealth working mechanisms through qualitative interviews may be challenging. In this paper, we suggest that challenges may arise partly due to what we have described as the *invisible interaction*: that eHealth programs affect the program users' change processes through their content *and* how they influence the person-program interaction, but that their influence on the interaction is largely invisible to the user. We have described 5 interview challenges and suggested tools from qualitative methodology to counteract each challenge. These tools may serve as a step toward a set of guidelines for conducting interview studies on eHealth working mechanisms, with the goal of generating rich data that will improve the quality and reach of the findings. Findings from high-quality interview studies can in turn be used to build more general, theoretical knowledge about the working mechanisms of automated eHealth programs. Through theorizing the general working mechanisms of eHealth interventions, we believe that the next generation of eHealth programs can be developed to fully take advantage of this medium's potential.

Acknowledgments

This research is funded by grant no. 228158/H10 from the Research Council of Norway. The authors wish to thank Filip Drozd, Caroline L Brandt, and Maja Wilhelmsen for contributing to this paper with their experiences with qualitative eHealth research. Their inputs were valuable corrections of the authors' initial presumptions, and they contributed to making this paper more applicable to a larger part of the field. The authors also wish to thank Professor Azy Barak for his encouragement and valuable insights regarding the need for theoretical work within the field of eHealth. The authors additionally wish to thank Anne-Lise Middelthon for inspiring insights and advice on qualitative methodology. The second author of this paper, Ayna B Johansen, tragically passed away before this paper was published. We are grateful for her enthusiasm for this paper and how she helped shape its content. She is remembered and missed.

Conflicts of Interest

None declared.

References

1. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
2. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
3. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
4. Barak A. University of Haifa. 2014. Emerging characteristics of internet-assisted psychological interventions: analysis of journal publications 1996-2013 URL: http://azy.edu.haifa.ac.il/recent-presentations [accessed 2019-01-16] [WebCite Cache ID 75TA11xyV]
5. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 1;2(1):91-101. [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
6. Michie S, Abraham C. Advancing the science of behaviour change: a plea for scientific reporting. Addiction 2008 Sep;103(9):1409-1410. [doi: 10.1111/j.1360-0443.2008.02291.x] [Medline: 18783495]

7. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 4;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
8. 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]
9. Drozd F, Haga SM, Brendryen H, Slinning K. An internet-based intervention (Mamma Mia) for postpartum depression: mapping the development from theory to practice. JMIR Res Protoc 2015 Oct 12;4(4):e120 [FREE Full text] [doi: 10.2196/resprot.4858] [Medline: 26476481]
10. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
11. Hales SA, Fossey J. Caring For Me and You: the co-production of a computerised cognitive behavioural therapy (cCBT) package for carers of people with dementia. Aging Ment Health 2017 Jul 10:1-8 [Epub ahead of print]. [doi: 10.1080/13607863.2017.1348475] [Medline: 28691520]
12. Brinkmann S, Kvale S. InterViews: Learning The Craft Of Qualitative Research Interviewing. Third edit. Thousand Oaks, California: Sage Publications, Inc; 2015. ISBN: 9781452275727.
13. Kvale S. InterViews: An Introduction To Qualitative Research Interviewing. Thousand Oaks, California: Sage Publications, Inc; 1994.
14. Brinkmann S. Could interviews be epistemic? An alternative to qualitative opinion polling. Qual Inq 2007 Nov 1;13(8):1116-1138. [doi: 10.1177/1077800407308222]
15. Charmaz K. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. Thousand Oaks, California: SAGE Publications, Inc; 2014.
16. Brandt CL, Dalum P, Thomsen TT. “I miss the care even though I know it’s just a machine”: an explorative study of the relationship between an Internet-based smoking cessation intervention and its participants. Health Informatics J 2013 Sep;19(3):233-243. [doi: 10.1177/1460458212470572] [Medline: 23981397]
17. Darvell MJ, Kavanagh DJ, Connolly JM. A qualitative exploration of internet-based treatment for comorbid depression and alcohol misuse. Internet Intervent 2015 May;2(2):174-182. [doi: 10.1016/j.invent.2015.03.003]
18. Kaplan B, Farzanfar R, Friedman RH. Personal relationships with an intelligent interactive telephone health behavior advisor system: a multimethod study using surveys and ethnographic interviews. Int J Med Inform 2003 Aug;71(1):33-41. [Medline: 12909156]
19. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen ML, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013 Nov 7;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
20. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ Commun Technol 1981;29(2):75-91 [FREE Full text] [doi: 10.1007/BF02766777]
21. Whittemore R, Chase SK, Mandle CL. Validity in qualitative research. Qual Health Res 2001 Jul;11(4):522-537. [doi: 10.1177/104973201129119299] [Medline: 11521609]
22. Polkinghorne DE. Language and meaning: data collection in qualitative research. J Couns Psychol 2005 Apr;52(2):137-145 [FREE Full text] [doi: 10.1037/0022-0167.52.2.137]
23. O'Reilly M, Parker N. 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. Qualitative Res 2013 May 17;13(2):190-197 [FREE Full text] [doi: 10.1177/1468794112446106]
24. Holter MT, Ness O, Johansen AB, Brendryen H. The emotional bond and the person-to-program alliance: A grounded theory study of how people relate to an automated eHealth program. In review 2019;-:1-33 (forthcoming).
25. Holter MT, Johansen A, Brendryen H. How a fully automated eHealth program simulates three therapeutic processes: a case study. J Med Internet Res 2016 Jun 28;18(6):e176 [FREE Full text] [doi: 10.2196/jmir.5415] [Medline: 27354373]
26. Peck DF. The therapist-client relationship, computerized self-help and active therapy ingredients. Clin Psychol Psychother 2010;17(2):147-153. [doi: 10.1002/cpp.669] [Medline: 19946947]
27. Barazzone N, Cavanagh K, Richards DA. Computerized cognitive behavioural therapy and the therapeutic alliance: a qualitative enquiry. Br J Clin Psychol 2012 Nov;51(4):396-417. [doi: 10.1111/j.2044-8260.2012.02035.x] [Medline: 23078210]
28. Kiluk BD, Serafini K, Frankforter T, Nich C, Carroll KM. Only connect: The working alliance in computer-based cognitive behavioral therapy. Behav Res Ther 2014 Dec;63:139-146 [FREE Full text] [doi: 10.1016/j.brat.2014.10.003] [Medline: 25461789]
29. Brendryen H. Happy Ending: The significance of age for treatment efficacy and intervention adherence. Res Pract Heal Promot 2009;14:79-88.
30. Brendryen H. Utprøving av Balance blant studenter - Balance er bra ... men ikke for meg. [Investigating reactions to Balance among students – Balance is excellent … but not for me.]. In: AKAN - Arbeidslivets kompetansesenter for rus- og

avhengighetsproblematikk [AKAN-Workplace competence center for drug and addiction issues]. Oslo: AKAN publikasjon; 2011.
31. Brendryen H, Johansen A, Duckert F, Nesvåg S. A pilot randomized controlled trial of an internet-based alcohol intervention in a workplace setting. Int J Behav Med 2017 Oct;24(5):768-777. [doi: 10.1007/s12529-017-9665-0] [Medline: 28755326]
32. Brendryen H, Kraft P. Happy ending: a randomized controlled trial of a digital multi-media smoking cessation intervention. Addiction 2008 Mar;103(3):478-484. [doi: 10.1111/j.1360-0443.2007.02119.x] [Medline: 18269367]
33. Johansen A, Helland P, Wennesland D, Henden E, Brendryen H. Exploring online problem gamblers’ motivation to change. Paper under preparation 2020:1-20 (forthcoming) [FREE Full text]
34. Moen K, Middelthon A. Qualitative research methods. In: Laake P, Benestad HB, Olsen BR, editors. Research in Medical and Biological Sciences. Amsterdam: Elsevier; 2015:321-328.
35. Latour B. Pandora's Hope: Essays on the Reality of Science Studies. Cambridge, Massachusetts: Harvard University Press; 1999. ISBN: 978-0674653368.
36. Finch J. The vignette technique in survey research. Sociology 1987 Feb 1;21(1):105-114. [doi: 10.1177/0038038587021001008]
37. Jenkins N, Bloor M, Fischer J, Berney L, Neale J. Putting it in context: the use of vignettes in qualitative interviewing. Qualitative Res 2010 Apr 12;10(2):175-198. [doi: 10.1177/1468794109356737]
38. Barter C, Renold E. The use of vignettes in qualitative research. Soc Res Updat 1999;-(25):1-7 [FREE Full text]
39. Smith JA. Towards reflexive practice: engaging participants as co-researchers or co-analysts in psychological inquiry. J Commun Appl Soc Psychol 1994 Oct;4(4):253-260. [doi: 10.1002/casp.2450040405]
40. Shotter J. “Getting It”: “Withness”-Thinking and the Dialogical... in Practice. London: Hampton Press Inc; 2007.
41. Warren CA. Qualitative interviewing. In: Gubrium JF, Holstein JA, editors. Handbook of Interview Research. Context and Method. Thousand Oaks, CA: Sage; 2001:83-101.
42. Safran J, Muran J. Negotiating The Therapeutic Alliance: A Relational Treatment Guide. New York: The Guilford Press; 2003.
43. Finlay L. Five lenses for the reflexive interviewer. In: Gubrium JF, Holstein JA, Marvasti AB, McKinney KD, editors. The Sage Handbook of Interview Research: The Complexity of the Craft. London, UK: Sage; 2012:317-332.
44. Finlay L. “Outing” the researcher: the provenance, process, and practice of reflexivity. Qual Health Res 2002 Apr;12(4):531-545. [doi: 10.1177/104973202129120052] [Medline: 11939252]
45. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 1993 Nov 01;3(4):430-441. [doi: 10.1177/104973239300300403]
46. Tulving E. Episodic and semantic memory. In: Organization of Mem 1972;1:381-403 doi:10.1017/S0140525X00047257.
47. Thomsen DK, Brinkmann S. An interviewer's guide to autobiographical memory: ways to elicit concrete experiences and to avoid pitfalls in interpreting them. Qual Res Psychol 2009 Nov 11;6(4):294-312. [doi: 10.1080/14780880802396806]
48. Brinkmann S. Methodological breaching experiments: steps toward theorizing the qualitative interview. Cult Psychol 2016 May 24;22(4):520-533. [doi: 10.1177/1354067X16650816]
49. Hermanowicz JC. The great interview: 25 strategies for studying people in bed. Qual Sociol 2002 Dec;25(4):479-499. [doi: 10.1023/A:1021062932081]

Abbreviations

eHealth: electronic health

Edited by G Eysenbach; submitted 09.03.18; peer-reviewed by A Lee, A Sheon; comments to author 18.08.18; revised version received 17.01.19; accepted 07.02.19; published 06.05.19.

Please cite as:
Holter MTS, Johansen AB, Ness O, Brinkmann S, Høybye MT, Brendryen H
Qualitative Interview Studies of Working Mechanisms in Electronic Health: Tools to Enhance Study Quality
J Med Internet Res 2019;21(5):e10354
URL: https://www.jmir.org/2019/5/e10354/
doi:10.2196/10354
PMID:31066683

Original Paper

Proactive Suicide Prevention Online (PSPO): Machine Identification and Crisis Management for Chinese Social Media Users With Suicidal Thoughts and Behaviors

Xingyun Liu[1,2,3], MA; Xiaoqian Liu[1], PhD; Jiumo Sun[1,2], BA; Nancy Xiaonan Yu[3], PhD; Bingli Sun[1], MA; Qing Li[4], PhD; Tingshao Zhu[1], PhD

[1]Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China
[3]Department of Social and Behavioural Sciences, City University of Hong Kong, Hong Kong, China (Hong Kong)
[4]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

Corresponding Author:
Tingshao Zhu, PhD
Institute of Psychology
Chinese Academy of Sciences
No.16, Lincui Road
Beijing,
China
Phone: 86 150 1096 5509
Email: tszhu@psych.ac.cn

Abstract

Background: Suicide is a great public health challenge. Two hundred million people attempt suicide in China annually. Existing suicide prevention programs require the help-seeking initiative of suicidal individuals, but many of them have a low motivation to seek the required help. We propose that a proactive and targeted suicide prevention strategy can prompt more people with suicidal thoughts and behaviors to seek help.

Objective: The goal of the research was to test the feasibility and acceptability of Proactive Suicide Prevention Online (PSPO), a new approach based on social media that combines proactive identification of suicide-prone individuals with specialized crisis management.

Methods: We first located a microblog group online. Their comments on a suicide note were analyzed by experts to provide a training set for the machine learning models for suicide identification. The best-performing model was used to automatically identify posts that suggested suicidal thoughts and behaviors. Next, a microblog direct message containing crisis management information, including measures that covered suicide-related issues, depression, help-seeking behavior and an acceptability test, was sent to users who had been identified by the model to be at risk of suicide. For those who replied to the message, trained counselors provided tailored crisis management. The Simplified Chinese Linguistic Inquiry and Word Count was also used to analyze the users' psycholinguistic texts in 1-month time slots prior to and postconsultation.

Results: A total of 27,007 comments made in April 2017 were analyzed. Among these, 2786 (10.32%) were classified as indicative of suicidal thoughts and behaviors. The performance of the detection model was good, with high precision (.86), recall (.78), F-measure (.86), and accuracy (.88). Between July 3, 2017, and July 3, 2018, we sent out a total of 24,727 direct messages to 12,486 social media users, and 5542 (44.39%) responded. Over one-third of the users who were contacted completed the questionnaires included in the direct message. Of the valid responses, 89.73% (1259/1403) reported suicidal ideation, but more than half (725/1403, 51.67%) reported that they had not sought help. The 9-Item Patient Health Questionnaire (PHQ-9) mean score was 17.40 (SD 5.98). More than two-thirds of the participants (968/1403, 69.00%) thought the PSPO approach was acceptable. Moreover, 2321 users replied to the direct message. In a comparison of the frequency of word usage in their microblog posts 1-month before and after the consultation, we found that the frequency of death-oriented words significantly declined while the frequency of future-oriented words significantly increased.

Conclusions: The PSPO model is suitable for identifying populations that are at risk of suicide. When followed up with proactive crisis management, it may be a useful supplement to existing prevention programs because it has the potential to increase the accessibility of antisuicide information to people with suicidal thoughts and behaviors but a low motivation to seek help.

(J Med Internet Res 2019;21(5):e11705) doi:10.2196/11705

KEYWORDS

suicide identification; crisis management; machine learning; microblog direct message; social network; Chinese young people

Introduction

Approximately one million people die by suicide globally every year. Aside from being a great challenge to public health, suicide also causes significant economic losses and aggravates labor shortages. It is estimated that by 2020, approximately 1.53 million people will die from suicide annually, and the number of people who attempt suicide will be 10 to 20 times greater [1]. Suicide is the leading cause of death among young people aged between 15 and 29 years [2]. In China, 200 million people attempt suicide annually, with two-thirds aged between 15 and 34 years [3]. Suicide prevention is thus crucial, particularly for young people.

Current suicide prevention methods practiced globally include school-based screening, screening by a primary care provider, and gatekeeper training, all of which are methods that are targeted at the general population and involve passively waiting for people to be in need [4]. However, many studies have found that because most people experiencing suicidal thoughts and behaviors tend not to participate in the aforementioned activities and have low motivation to seek help, existing methods have a rather weak effect on suicide prevention [5-8]. For example, only 17% of suicidal people in low-income countries, such as China, receive treatment in a timely manner [6]. The main reasons for not seeking help include the lack of a perceived need for services, high self-reliance, stigma, and structural factors such as time and cost [6-8]. However, the subjective judgment of at-risk individuals may not be good, and a high self-reliant tendency may lead to severe depressive symptoms and suicidal ideation among young people [9]. The passive suicide approach needs suicidal cases to actively seek help [10], such as the Columbia suicide screening program in which suicidal students filled in surveys in schools to get help [11]. In contrast, a proactive approach for suicide prevention, in which the program itself takes the initiative to identify suicidal people and invite them to use specific services, may increase the likelihood of service usage for the hidden people [10].

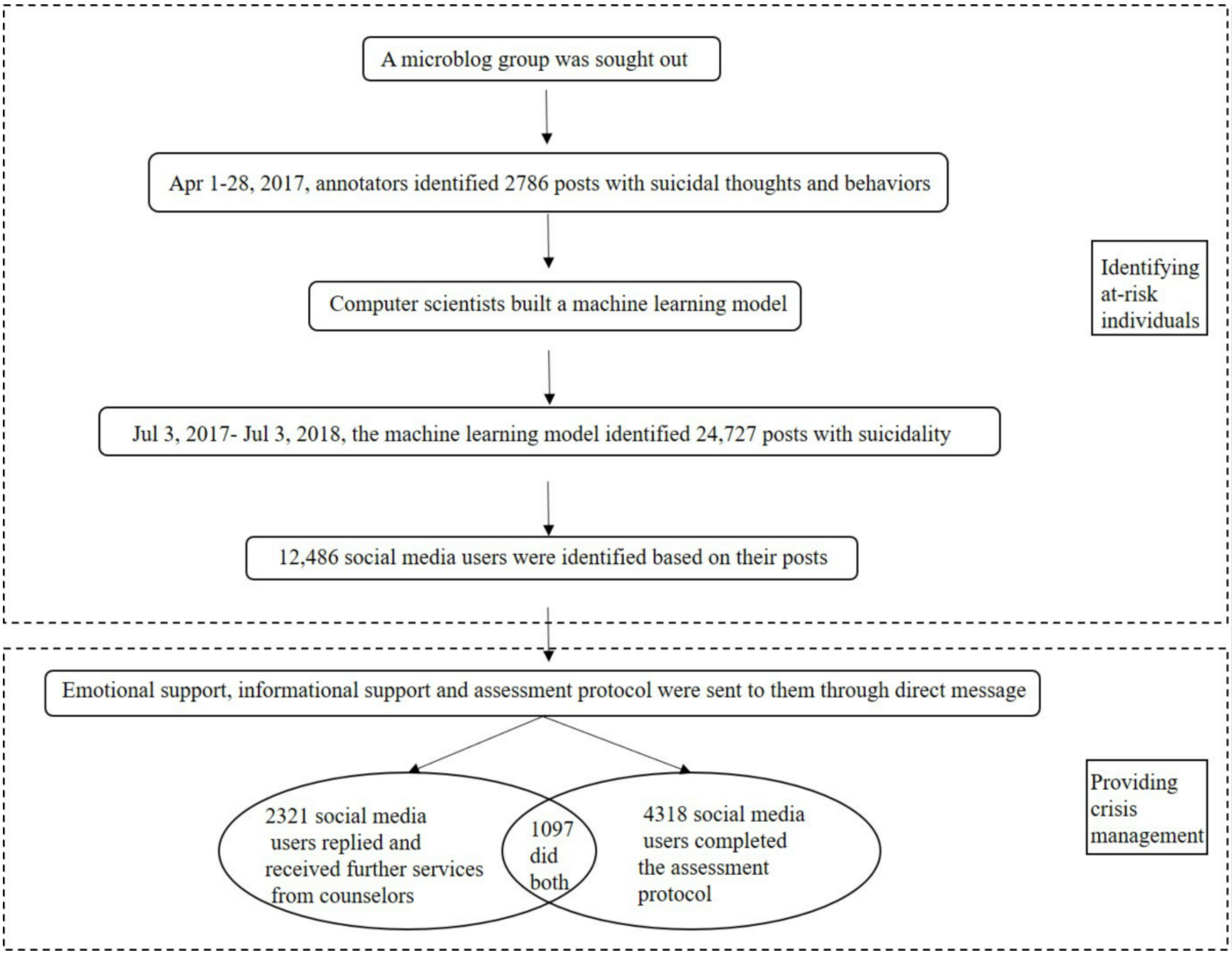
As in most developing countries, mental health care in China is at an early stage of development [12]. Moreover, because of China’s large population and uneven distribution of resources, it is hard to implement school-based screening or maintain primary care provider screenings nationally [13,14]. Gatekeeper training is still at an early stage [15], and thus new suicide prevention methods are urgently needed.

The internet has become an indispensable part of life for many people. As such, researchers have started to use people’s self-generated online messages to identify suicidal thoughts and behaviors either by manual [16] or machine learning analysis [17]. However, identification of suicidal thoughts and behaviors is just the first step in suicide prevention. Even though the internet has been used to manage suicide-prone crises [18,19], more effort is needed to prevent suicide. Previous studies have used the internet simply as a platform, and this means that they suffered from the same shortcomings as traditional prevention methods.

Half of the Chinese population uses the internet. Approximately two-fifths (40.9%) of China’s netizens use the Sina microblog, the Chinese version of Twitter [20]. Microbloggers (users of microblogs) can post microblog posts publicly, similar to Twitter. They can also send direct messages to other users that can be seen by the sender and receiver exclusively. Microbloggers can follow other users, along with replying to, commenting on, reposting, or liking others’ posts. An average of 139 million new posts are generated on the Sina microblog daily, with most (82%) microbloggers being under 30 years old [21]. These phenomena provide opportunities to prevent young people dying from suicide in China because existing findings suggest that young people feel they can freely discuss suicide-related topics on social networks [22,23]. The challenge is that the proportion of suicide posts is extremely low, making it nearly impossible to identify them manually, and that is why researchers have called for more efforts to build automated or semiautomated suicide ideation detectors to facilitate the provision of timely help and support to people at risk of suicide [24]. Suicidal people need to first be identified, and suicide identification and preventions can be carried out with the assistance of machine learning models.

To our knowledge, no research has previously been undertaken on the combination of proactive identification of suicidal individuals via social media and specialized crisis management. Moreover, a growing number of studies have demonstrated that (1) there is a shortage of online suicide prevention approaches [25], (2) unidirectional monologic suicide prevention information distributed by professional institutions is insufficient and more dialogic communication is needed between professionals and people at risk of suicide [26], (3) suicidal statuses can be detected from human language [27], and (4) a large improvement (such as enhancement of recall) can be achieved by applying machine learning algorithms to suicide identification [17]. Because language is an explicit behavior that indicates human mental status, people with suicide ideation are more likely to talk about suicide than people without suicide ideation [28]. Recently, researchers have started to use online longitudinal data to evaluate certain improvements after receiving psychosocial support services [29,30]. For example, people who used more future tense words online were found to have benefited more from online social support [31]. Previous studies have shown that higher future orientation was associated with less suicide ideation [32,33]. In our view, a reduction in death-oriented language and an increase in future-oriented language may serve as indicators for reduced suicide risk.

Figure 1. Study protocol for evaluating the Proactive Suicide Prevention Online (PSPO).



We proposed a new internet-based approach, Proactive Suicide Prevention Online (PSPO), for the identification and prevention of suicidal thoughts and behaviors (see Figure 1). We identified a microblog group online and manually annotated their comments on a suicide note to train a machine learning model. Next, the model was used to automatically identify posts that suggested suicidal thoughts and behaviors. We proactively provided crisis management in the form of emotional and informational support to microbloggers identified as at risk. Finally, we used the language changes in their posts as criteria to evaluate the efficacy of the PSPO approach. Based on the research gaps summarized above, we aimed to evaluate (1) the performance of PSPO in identifying high-risk social media users with suicidal thoughts and behaviors, (2) the acceptability of a proactive approach that offered help to social media users with suicidal thoughts and behaviors, (3) the improvement of PSPO in prompting suicidal social media users to seek help in comparison with traditional suicide prevention methods, and (4) the efficacy of PSPO for suicidal social media users in terms of changes in their language use (ie, reduced death-oriented words and increased future-oriented words).

Methods

Data Collection

Identifying At-Risk Individuals

A microblogger, Zoufan, died by suicide due to depression on March 17, 2012, with her suicide arousing wide attention online. Since her death, her blog has become a “secret garden” where suicidal people share their feelings and thoughts. By July 24, 2018, more than 1.3 million comments had been posted on her online suicide note, and many of them contained suicide information. This microblog group comprises people with and without suicide ideation. We analyzed the comments left on Zoufan’s online suicide note. The official Sina microblog application programming interface was used to obtain comments posted from April 1 to 28, 2017, which were manually annotated as the training set. Comments posted from July 3, 2017, to July 3, 2018, were obtained and automatically identified by the developed machine learning model. Further details regarding the procedures used for building the machine learning model (eg, the coding system for suicidal thoughts and behaviors in posts, feature selection for the machine learning models) are provided in Data Analysis.

Providing Crisis Management

All microbloggers who were identified by the machine learning model as expressing suicidal thoughts and behaviors were invited to join the study via direct message. There were no exclusion criteria because we aimed to reach out and provide support to as many suicidal social media users as possible.

The direct message, designed in our previous study [34], included (1) a brief introduction to the project; (2) URLs for assessment protocols on suicidal thoughts and behaviors, depressive symptoms, and help-seeking behaviors; (3) emotional support (empathy, recommendations such as having regular physical exercise and a healthy diet, and encouragement); (4) informational support (the URL for this study, along with referrals to hospitals and hotline services); and (5) details regarding the availability of one-to-one counseling by contacting counselors via direct messaging (see Multimedia Appendix 1).

If a user replied to the direct message, counselors provided support that was targeted to the user's specific problem. Twelve certified counselors (2 men and 10 women, mean age 23.08 [SD 1.08] years) with experience in handling suicidal cases were trained to provide counseling services through direct messaging.

Direct Message Assessment Protocol

Suicidal thoughts and behaviors were tested with the use of two items chosen based on previous research [35] and the 9-Item Patient Health Questionnaire (PHQ-9) [36]. The two items were "Do you have a plan to commit suicide?" and "Have you ever attempted suicide?" Participants responded with binary choices (yes/no), and if the answer was yes to the first item, they were required to indicate whether they had a specific or vague plan. A sample item on the PHQ-9 would be "having little interest or pleasure in doing things." Participants rated the frequency of the 9 symptoms over the past 2 weeks on a 4-point Likert scale (0 = not at all, 3 = nearly every day), and the total score of the PHQ-9 ranges from 0 to 27, with higher scores indicating a greater severity of depressive symptoms. The Chinese version has been shown to have good psychometric properties [37], and internal consistency was .84 in this study.

Help-seeking behavior was assessed by two items [35]: "What kind of psychological treatment have you received before?" and "Have you sought help when you had suicidal ideation?" and if the answer was positive to either of the questions, the effectiveness of the former help was rated on a 7-point Likert scale (1 = totally disagree, 7 = totally agree). If the participant's rating for the question was 3 or lower, we recorded that the former help was not useful.

Acceptability was measured with one item ("How acceptable do you find this proactive help?") and rated on a 7-point Likert scale (1 = totally disagree, 7 = totally agree). We considered a rating of 4 or higher to indicate that the program was acceptable.

One-to-One Counseling

The counselor training was based on problem-solving therapy [38], which begins with identifying a person's problem and then helping them to affirm feasible solutions to a specific issue by setting goals and comparing the pros and cons of every plausible solution. A concrete and feasible plan is then made to facilitate the client in overcoming the problem they face. The goal for counselors in crisis management was to persuade suicidal microbloggers to seek professional services and provide them with the appropriate referrals. The interaction between the microbloggers and our counselors also depended on the needs of the microbloggers. The training lasted 3 hours and included a theoretical explanation and practice of applying problem-solving therapy to this online situation. In addition to counselors being under monthly supervision by psychiatrists and senior counselors, we also formed an online chat group where the counselors could discuss the problems they encountered in consultations at any time.

Because the data used in this study were all publicly available, traditional informed consent was not appropriate. In the identification section, measures were taken to anonymize the data in the data analysis to minimize the inadvertent disclosure of personal information or information that may reveal clues with regard to an individual's online identity. In the crisis management part of the study, participants gave informed consent voluntarily when agreeing to take part in the program. The project received ethical approval from the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences, with the ethics approval number H16003.

Data Analysis

Building Machine Learning Models for Suicide Recognition

The first step for supervised learning was to obtain a training set. To achieve good performance of the machine learning model, this study decided not to use crowd-sourcing [39]. Rather, 5 psychology postgraduates with expertise in analyzing suicide annotated the microbloggers' comments. The annotation process was identical to the one used in our previous study [34]. Expressing a death wish or writing about suicide was coded as suicide ideation. Because a suicide plan is defined as suicide-related communication to account for its interpersonal nature, which is often expressed in verbal words regarding how one might advance from ideation to action [40], we operationalized a suicide plan as one that involved discussions regarding the act of dying (eg, a death kit, death place and time, making a will) after considering the nature of dialogue in social media. Attempted suicidal behavior within the preceding 2 weeks with current suicide ideation or the possibility of executing a suicide plan in the coming 1 to 2 weeks was coded as a suicide attempt. Posts were ranked as follows: 0 = no suicide risk; 1 = risk of suicide ideation but no detailed plan made; 2 = risk of suicide plan not requiring emergency aid; and 3 = significant risk of suicide attempt requiring emergency aid.

Comments identified as indicating suicidal thoughts and behaviors were labeled as positive training samples, and 10,000 posts without suicidal thoughts and behaviors were randomly selected to serve as negative training samples. Another strategy for the improvement of the performance of the machine learning model was the theoretical-based feature selection. Researchers in the field of computer science had a tendency to select predictive features by randomly using linguistic analyses, such as n-grams and sentiment analyses, without a theoretical or empirical basis [41,42]. In this study, we combined data-driven features (n-grams) that were derived from social media data, domain knowledge, and theoretical guidance [43,44] to select features. A knowledge-based generic suicide-related lexicon [45], which was manually developed by a panel of domain experts, was used. Theory-motivated features include personality and depression, which are the factors most commonly cited as being relevant to suicides [46-48]. Personal traits such as fearfulness, social inhibition, shyness, pessimism, immaturity, and lack of internal organization were associated with psychotic suicide attempts [43]. Moreover, there has been significant progress in predicting personality and depression from social media data [43,44,49,50], and depression has also been used to predict suicidal behaviors [51]. Thus, it is both theoretically sound and technically feasible to incorporate predictive features of personality traits and depression into a model for suicide ideation detection. To our knowledge, no prior studies have used the lexicon and predictive features (including personality traits and depression) that were included in the machine learning models.

A binary classification of suicidal thoughts and behaviors detection model was built to determine if the comment indicated suicidal thoughts and behaviors. We used support vector machine (SVM), decision tree, random forest, and logistic regression algorithms with 10-fold cross-validation to train the detection model because these machine learning algorithms are the most widely used methods of predicting psychological characteristics and emotions and detecting suicide ideation [45,52]. The performance of the detection model was evaluated through the use of four metrics: precision, recall, F-measure, and accuracy [53].

Using Language Frequency Changes as an Efficacy Indicator for Crisis Management

Because of the low response rate in the completion of the online survey, we were unable to collect 1-month post–PHQ-9 data. We used language frequency changes as an efficacy indicator for crisis management. To examine the microbloggers’ language changes between 1 month before the commencement of the program and 1 month after the completion of the program, the Simplified Chinese Linguistic Inquiry and Word Count (SCLIWC) was used. The SCLIWC is an amended version of the text analysis program LIWC designed to perform well in Simplified Chinese on a microblog [54]. It is composed of 7 main categories and 64 subcategories. Death-oriented and future-oriented words are 2 of the subcategories, and every participant’s posts were parsed into these 2 subcategories. Category scores were calculated by the ratio of words within the category to all the words in the posts. Furthermore, we used the change tendencies of the 2 subcategories as a measure of the efficacy of the program.

Results

Machine Learning Models for Suicide Recognition

From April 1 to 28, 2017, four weekly sessions of manual annotation of comments were conducted. Of the 27,007 comments, 10.32% (2786/27,007) were identified as indicating suicidal thoughts and behaviors. In those, 81.44% (2269/2786), 13.75% (383/2786), and 4.81% (134/2786) contained information coded as suicide ideation, suicide plan, and suicide attempt, respectively (Table 1).

Table 2 presents the means and standard deviations in the performances of the detection models with the whole feature set and two baseline feature sets developed by the selected classification algorithms. The best overall models were the SVM models. We compared the performances of SVM models, which were constructed with each feature set using a Tukey honestly significant difference post hoc test. The precision of the model of set C is lower than set A ($t=-6.32$, $P<.001$). However, the recall ($t=12.07$, $P<.001$), F-measure ($t=5.48$, $P<.001$), and accuracy ($t=3.32$, $P=.004$) of the model were all significantly higher in set C than with set A. Comparing the performance of SVM models using the feature sets C and B, and despite the precision of the SVM model with set C also being lower ($t=-5.80$, $P<.001$), the recall and F-measure of the model using set C were significantly higher ($t=12.23$, $P<.001$ and $t=3.87$, $P=.001$, respectively) than those of the model using set B, while their accuracies were equivalent ($t=1.34$, $P=.20$).

Of the 387,823 comments made between July 3, 2017, and July 3, 2018, 24,727 (6.38%) were identified as being indicative of suicidal thoughts and behaviors by the machine learning model.

Table 1. Manual identification of suicidal comments posted in April 2017.

Date	Comments (n)	Suicidal comments			
		Suicidal thoughts and behaviors, n (%)	Suicide ideation, n (%)	Suicide plan, n (%)	Suicide attempt, n (%)
4/01-4/07	6975	849 (12.17)	702 (82.69)	107 (12.60)	40 (4.71)
4/08-4/14	6201	682 (11.00)	561 (82.26)	90 (13.20)	31 (4.55)
4/15-4/21	6467	563 (8.71)	457 (81.17)	82 (14.56)	24 (4.26)
4/22-4/28	7364	692 (9.40)	549 (79.33)	104 (15.03)	39 (5.64)
Total	27,007	2786 (10.32)	2269 (81.44)	383 (13.75)	134 (4.81)

Table 2. Performance of the machine learning models.

Model performance and feature set	SVM[a], mean (SD)	DT[b], mean (SD)	RF[c], mean (SD)	LR[d], mean (SD)
Precision				
A[e]	.88 (.01)	.84 (.02)	.87 (.01)	.87 (.01)
B[f]	.88 (.01)	.76 (.01)	.85 (.01)	.87 (.01)
C[g]	.85 (.01)	.76 (.01)	.85 (.01)	.88 (.01)
Recall				
A	.78 (.02)	.68 (.05)	.75 (.02)	.79 (.02)
B	.80 (.01)	.75 (.01)	.74 (.01)	.79 (.01)
C	.85 (.01)	.75 (.01)	.73 (.01)	.80 (.01)
F-measure				
A	.83 (.01)	.74 (.03)	.80 (.01)	.83 (.01)
B	.84 (.01)	.76 (.01)	.79 (.01)	.83 (.01)
C	.85 (.01)	.76 (.01)	.78 (.01)	.84 (.01)
Accuracy				
A	.85 (.01)	.79 (.02)	.83 (.01)	.85 (.01)
B	.86 (.01)	.78 (.01)	.82 (.01)	.85 (.01)
C	.86 (.01)	.78 (.01)	.82 (.01)	.86 (.01)

[a] SVM: support vector machine.

[b] DT: decision tree.

[c] RF: random forest.

[d] LR: logistic regression.

[e] A: n-gram features.

[f] B: n-gram features + domain knowledge features.

[g] C: n-gram features + domain knowledge features + theory-motivated features.

Crisis Management

We sent direct messages to the 12,486 microbloggers (some microbloggers had multiple comments) who were identified as having expressed suicidal thoughts and behaviors in the 24,727 comments by the machine learning model. A total of 34.58% (4318/24,727) of individuals completed the assessment protocol, and there were 1403 valid samples (mean age 21.66 [SD 3.26] years). Females significantly outnumbered males (χ^2_1=647.33, *P*<.001), and most of the participants were students or employed, single, and had graduated from college (see Table 3).

In terms of suicide risk, most of the respondents (1259/1403, 89.73%) thought they would be better off dead or by hurting themselves in some way. Nearly half of these (699/1403, 49.82%) had a suicide plan, with 6.34% (89/1403) indicating a specific plan to commit suicide and 43.48% (610/1403) indicating a vague plan to do so. Of the 1403 participants, 545 (38.85%) had previously attempted suicide. The mean score for the PHQ-9 was 17.40 (SD 5.98), and this represents moderately severe depressive symptoms. Two-thirds of the participants (924/1403, 65.86%) had never received any kind of psychological treatment. Just over half (725/1403, 51.67%) had not sought help from anyone, 12.19% (171/1403) had sought help from professionals (such as a psychiatrist, therapist, or general practitioner), and 36.14% (507/1403) had sought help from people around them (such as family and friends). Of the participants who had sought help before, 48.33% (678/1403) rated the efficacy of the former help as 2.60 (SD 1.43), and 77.00% (522/678) thought that the former help rendered was of no use to them.

On a 7-point Likert scale, nearly 70% (968/1403, 69.00%) of participants rated proactive help through the use of direct messaging acceptable (4 or more on the scale). The average score for all participants was 4.35 (SD 1.81).

Between July 3, 2017, and July 3, 2018, microbloggers logged into the study website to view prevention information 12,300 times. A total of 2321 users replied to a direct message at least once. Figure 2 shows the total number of microbloggers who interacted with our counselors monthly from July 3, 2017, to July 3, 2018. On average, there were 234.08 (SD 88.70) microbloggers who interacted with our counselors every month. Table 4 shows the interaction between microbloggers and counselors, and this included the number of microblogger replies and interaction days. Approximately 90% (2043/2321, 90.12%) of the microbloggers replied fewer than 10 times. Nearly 97% (2246/2321, 96.77%) of the microbloggers interacted with our counselors for less than 5 days. A total of 1097 users completed an assessment protocol and consultation. Of the 12,486 microbloggers who were contacted, 5542 (44.39%) responded to our direct message either by completing the assessment protocol or by interacting with the counselors. Earlier studies found that the percentages of college students seeking help from professionals were 5.1% for all college students, 14.4% for college students with mental health problems, and 4.5% for college students without mental health problems [55]. Compared with the traditional methods used, we were able to prompt a larger number of people identified via the machine learning model as having posted suicidal content to seek help for their distress or suicide ideation.

Table 3. Demographic characteristics of participants.

Characteristic	Value, n (%)
Gender	
Male	225 (16.04)
Female	1178 (83.96)
Education level	
Junior middle school or below	82 (5.84)
Senior high school	272 (19.39)
College	1006 (71.71)
Graduate or above	43 (3.06)
Employment status	
Employed	472 (33.64)
Unemployed	229 (16.32)
Student	603 (42.98)
Marital status	
Single/divorced	1342 (95.65)
Married	61 (4.35)

Figure 2. Number of microbloggers who interacted with our counselors from July 3, 2017, to July 3, 2018.

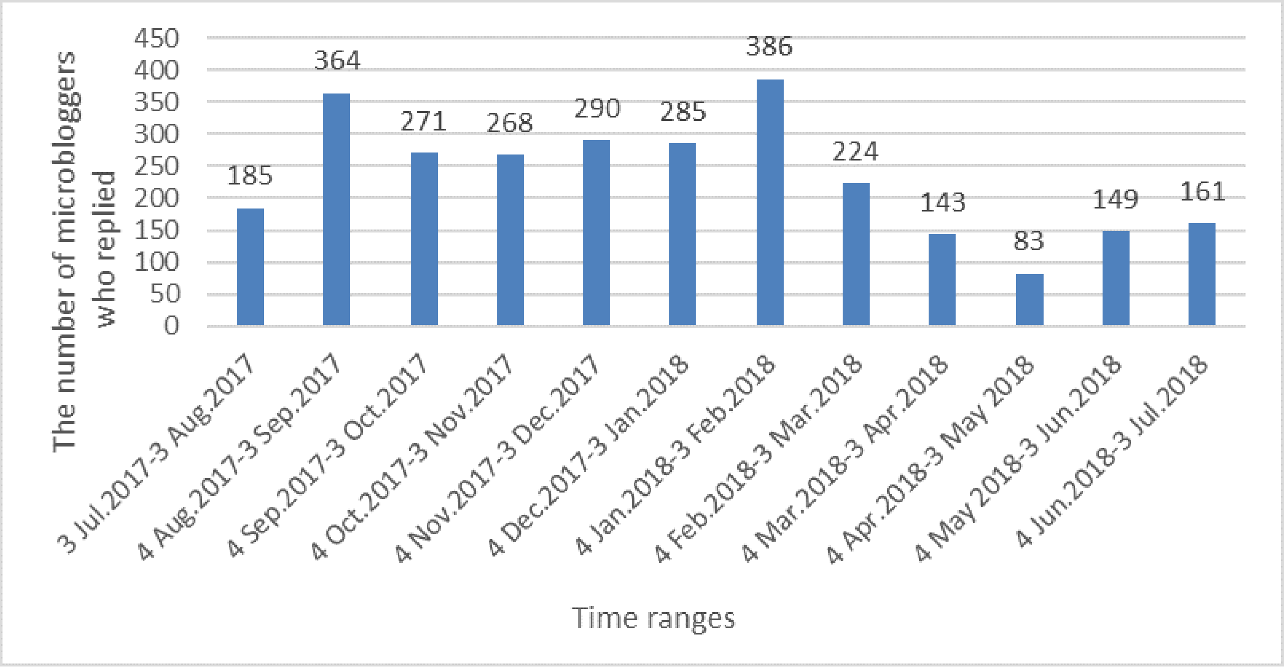


Table 4. Interactions between microbloggers and counselors.

Interactions	Value, n (%)
Number of microblogger replies	
≤10	2043 (90.12)
11-30	118 (5.21)
31-50	46 (2.03)
51-100	60 (2.65)
>100	54 (2.38)
Days of interactions with counselors	
≤5	2246 (96.77)
6-10	48 (2.07)
>10	27 (1.16)

Table 5. Changes in frequency of language use pre- and postprogram.

Category	Examples	Preprogram, % (SD)	Postprogram, % (SD)	*t* value	*P* value
Death-oriented words	Die/suicide/will	0.37 (0.01)	0.31 (0.01)	2.21	.03
Future-oriented words	After/soon/future	0.34 (0.01)	0.34 (0.01)	–2.29	.02

Finally, we used SCLIWC to detect language changes in the 2321 social media users who replied to the direct message. By tracing the accounts of those microbloggers, we compared their microblog posts from a month before and after receiving services from our counselors. After deleting users who did not complete the interaction with the counselors for one month, there were 2031 microbloggers. As shown in Table 5, the frequency of death words significantly declined (*P*=.03), and the frequency of future-oriented words significantly increased (*P*=.02). For the month before the program, the number of posts by a single user ranged from 1 to 1013. The mean was 30.59 (SD 84.36). For the month after the program, the number of posts by a single user ranged from 1 to 1279, and the mean was 27.41 (SD 74.04). The paired sample *t* test result showed that the difference in the total number of posts by a single user before and after the program was thus not significant (*t*=1.92, *P*=.06).

Discussion

Principal Findings

In our study, we first identified a microblog group formed around the Sina microblog account of a microblogger who committed suicide, which was an efficient way to identify a high-risk population. Then, we proactively pushed direct messages to invite all of the microbloggers identified by the machine learning model as people who had exhibited suicidal thoughts and behaviors to participate in our study. Our results provided some preliminary evidence that automatic identification of suicidal thoughts and behaviors along with proactive suicide prevention are acceptable and helpful.

For suicide ideation detection, recall is arguably more important than precision. The results of machine learning models demonstrate that in general, incorporating theory-related features and features based on domain knowledge can improve the recall, F-measure, and accuracy of a model for detecting suicide ideation. The best results were .88, .85, .85, and .86 for precision, recall, F-measures, and accuracy, respectively, which demonstrates the utility of the model in identifying suicide posts. Apart from multiresources and theoretical-based feature selection [39,41] being able to ensure that relevant features are included and redundant ones excluded, our results outperformed earlier suicide machine learning models to identify posts with suicide content [52,56] mainly because we did not rely on crowd-sourcing [39] and instead opted for postgraduate annotators specializing in studying suicide. In addition, the dataset size was almost twice the size of those used in similar studies done earlier [56,57].

In this study, most individuals who completed the questionnaires were single females (employed or students) with a college degree. This is consistent with a previous study showing that females are more likely to talk about suicide ideation to health professionals and use health services than males [6]. Those with a higher education and those who are never married also had significantly higher odds of receiving mental health treatment [6].

For the period of July 3, 2017, to July 3, 2018, 6.38% of the comments were identified as expressive of suicidal thoughts and behaviors. The self-reports of the 1403 participants identified as individuals who displayed suicidal proclivities testified to the utility of the machine learning model. The self-reported results in the survey showed that percentages for suicide ideation, suicide plan, and past suicide attempt were 89.73%, 49.82%, and 38.85%, respectively. These are much higher than the results in a meta-analysis study that showed percentages of suicide ideation and past suicide attempt for the general Chinese population were 3.9% and 0.8%, respectively [58]. Moreover, 6.34% of our participants self-reported a specific suicide plan. Their mean score for the PHQ-9 was 17.40 (SD 5.98), indicating moderately severe depressive symptoms.

If suicide ideation can be identified as early as possible, then at-risk individuals can be prevented from deteriorating to a point where they make a specific suicide plan [59]. PSPO enables timely crisis management only for those in need, without disturbing others.

Even with a high suicide ideation rate, 65.86% of participants who completed the questionnaires had never received any kind of psychological treatment. Moreover, 51.67% of them did not seek help from anyone regarding their suicide problem. This is consistent with earlier studies [60,61] and may explain why 69.00% of the participants accepted our PSPO—it provided a new way of accessing help for those who had experienced barriers to seeking help for suicidal thoughts and behaviors previously. Another possible reason for the acceptability of PSPO was its anonymous nature.

This study has demonstrated some primary evidence for the efficacy of PSPO. Eliminating suicide ideation can be a long process, but suicide crisis management can serve as “emotional cardiopulmonary resuscitation” for people at risk. We sent out 24,727 direct messages to 12,486 different social media users, and 5542 (44.39%) of them responded. Of these, 4318 individuals finished the assessment protocol and 2321 users replied to the direct message. On average, 234.08 (SD 88.70) microbloggers interacted with our counselors monthly, approximately 90% of the microbloggers replied fewer than 10 times, and nearly 97% of the microbloggers interacted with our counselors for less than 5 days. Those results indicate that PSPO might largely extend the potential for suicide prevention for those who have never sought help before when compared with traditional passive methods. Moreover, the prevention information on our website was viewed 12,300 times. Finally, after interacting with the counselors, the microbloggers with suicidal thoughts seemed to change the language they used on social media significantly. In particular, the frequency of death-oriented words was found to have been significantly reduced a month after the crisis management as compared with the frequency of those words one month prior to receiving crisis management. One possible reason could be that the microbloggers felt the concern, social support, and empathy of the counselors. Another possible reason is that these users started to seek help after the consultation. At the same time, the frequency of future-oriented words increased significantly, although it was a slight change. This may be due to the relatively small number of future words in the overall vocabulary used. Nevertheless, it may also signal that the users had less suicide ideation and became more willing to accept support than before.

Limitations and Future Work

In this study, we only focused on a microblog group. Future studies would be needed to establish whether our machine learning model can be applied to other similar suicide groups and other social media platforms such as school bulletin boards, suicide groups online, or online self-help groups for suicidal thoughts and behaviors. Because our suicidal thoughts and behaviors detection model is at an early stage of development, only a binary classification model was built, and it was mainly focused on finding suicidal candidates for the primary crisis management. Multiclass classification can be adopted and adapted in the future to facilitate customized suicide prevention for different social media users. Moreover, to detect suicide ideation, we concentrated mainly on text features extracted from posts, although other behaviors on social media, such as interactions with other users and posting frequencies and times, could also be effective predictors. Examining these potential factors may provide additional insights and guidance for building more effective models based on social media to detect suicide ideation.

Slightly more than one-third (34.58%) of the users completed the questionnaires provided in the direct message. Given the sensitivity of the subject, the relatively low response rate is understandable, although it is higher than those found in earlier studies (close to 10%) [62]. For future research, we plan to investigate the differences between those who participated in the program and those who did not to acquire more first-hand data to develop a deeper understanding of their psychology and behavior. Our aim is to involve more human resources in lifesaving, thereby increasing the prevention rate for suicide.

We only offered the primary crisis management information to microbloggers with suicidal thoughts and behaviors. More standardized and systematic emergency intervention protocols, mental health resources, and professional referrals are needed to guarantee reasonable retention in a future study. While PSPO provides an opportunity for longitudinal study, the effectiveness of various Web-based suicide prevention and intervention approaches including PSPO should be examined because a follow-up is crucial in suicide intervention [63]. We will thus try our best to provide follow-up measures and actions for the identified users.

Finally, because this was a preliminary study, we mainly used the changes in future-oriented word frequency to demonstrate the efficacy of PSPO. Our efficacy evidence should be interpreted with caution because the relationship between future-oriented words and reduced suicide risk would still need further verification. Future studies should consider using direct indicators of reduction of suicidal thoughts and behaviors to demonstrate the improvement. Moreover, there is a possibility that the results reflect regression to the mean. A few strategies in the study design stage (eg, using a randomized controlled trial, having multiple tests at different time points for actual behavior instead of just intention or attitude) and analysis stage (eg, using analysis of covariance) are desirable to reduce the regression to the mean [64,65].

Conclusion

This paper presents PSPO as a proactive suicide prevention method for identifying and preventing suicide incidents by social media users, especially young people. The results indicate that PSPO is feasible for identifying populations at risk of suicide and providing effective crisis management, and the identification of at-risk individuals is automatic and timely. The crisis management is also proactive, acceptable, and low cost. Our study may be a useful supplement to existing prevention programs, and suicide crisis management may increase public awareness of help-seeking related to suicide risk, thereby improving the well-being of the population. This approach could alleviate the problems associated with a huge population with weak psychological services and help with an imperfect suicide prevention system in large developing countries such as China.

Acknowledgments

The authors gratefully acknowledge the generous support of the National Basic Research Program of China (2014CB744600), China Social Science Fund (Y8JJ183010), National Social Science Fund of China (16AZD058), and the Research Grants Council of the Hong Kong Special Administrative Region, China (Collaborative Research Fund, project number C1031-18G). The sponsors had no role in the study design; the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The direct message sent to users with suicidal thoughts and behaviors.

[PDF File (Adobe PDF File), 67KB - jmir_v21i5e11705_app1.pdf]

References

1. Bertolote JM, Fleischmann A. A global perspective in the epidemiology of suicide. Suicidologi 2015;7(2):1. [doi: 10.5617/suicidologi.2330]
2. World Health Organization. Preventing suicide: a global imperative URL: http://www.who.int/mental_health/suicide-prevention/world_report_2014/en/ [accessed 2019-04-22] [WebCite Cache ID 71Aa9w67Y]
3. Statistical Information Center of the Ministry of Health of the People's Republic of China. China Health Statistics Yearbook 2013 URL: http://www.nhfpc.gov.cn/htmlfiles/zwgkzt/ptjnj/year2013/index2013.html [accessed 2019-04-22] [WebCite Cache ID 71AZ7kaVZ]
4. Schaffer A, Sinyor M, Kurdyak P, Vigod S, Sareen J, Reis C, et al. Population-based analysis of health care contacts among suicide decedents: identifying opportunities for more targeted suicide prevention strategies. World Psychiatry 2016 Jun;15(2):135-145 [FREE Full text] [doi: 10.1002/wps.20321] [Medline: 27265704]
5. Fountoulakis KN, Gonda X, Rihmer Z. Suicide prevention programs through community intervention. J Affect Disord 2011 Apr;130(1-2):10-16. [doi: 10.1016/j.jad.2010.06.009] [Medline: 20599277]
6. Bruffaerts R, Demyttenaere K, Hwang I, Chiu W, Sampson N, Kessler RC, et al. Treatment of suicidal people around the world. Br J Psychiatry 2011 Jul;199(1):64-70 [FREE Full text] [doi: 10.1192/bjp.bp.110.084129] [Medline: 21263012]
7. Hom MA, Stanley I, Joiner JT. Evaluating factors and interventions that influence help-seeking and mental health service utilization among suicidal individuals: a review of the literature. Clin Psychol Rev 2015 Aug;40:28-39. [doi: 10.1016/j.cpr.2015.05.006] [Medline: 26048165]
8. Han J, Batterham PJ, Calear AL, Randall R. Factors influencing professional help-seeking for suicidality. Crisis 2018 May;39(3):175-196. [doi: 10.1027/0227-5910/a000485] [Medline: 29052431]
9. Labouliere CD, Kleinman M, Gould MS. When self-reliance is not safe: associations between reduced help-seeking and subsequent mental health symptoms in suicidal adolescents. Int J Environ Res Public Health 2015 Apr 01;12(4):3741-3755 [FREE Full text] [doi: 10.3390/ijerph120403741] [Medline: 25837350]
10. World Health Organization. Towards evidence-based suicide prevention programmes URL: https://apps.who.int/iris/bitstream/handle/10665/207049/9789290614623_eng.pdf?sequence=1&isAllowed=y [accessed 2019-04-22] [WebCite Cache ID 75gj8a1Zn]
11. Scott MA, Wilcox HC, Schonfeld IS, Davies M, Hicks RC, Turner JB, et al. School-based screening to identify at-risk students not already known to school professionals: the Columbia suicide screen. Am J Public Health 2009 Feb;99(2):334-339. [doi: 10.2105/AJPH.2007.127928] [Medline: 19059865]
12. Gao X, Jackson T, Chen H, Liu Y, Wang R, Qian M, et al. There is a long way to go: a nationwide survey of professional training for mental health practitioners in China. Health Policy 2010 Apr;95(1):74-81. [doi: 10.1016/j.healthpol.2009.11.004] [Medline: 19962778]
13. Kong Y, Zhang J. Access to farming pesticides and risk for suicide in Chinese rural young people. Psychiatry Res 2010 Sep 30;179(2):217-221 [FREE Full text] [doi: 10.1016/j.psychres.2009.12.005] [Medline: 20483175]
14. Tang S, Meng Q, Chen L, Bekedam H, Evans T, Whitehead M. Tackling the challenges to health equity in China. Lancet 2008 Oct 25;372(9648):1493-1501. [doi: 10.1016/S0140-6736(08)61364-1] [Medline: 18930531]
15. Liang T, Zhang X, Wang Z. A review on the suicide gatekeeper training. Adv Psychol Sci 2012 Aug 20;20(8):1287-1295. [doi: 10.3724/sp.j.1042.2012.01287]

16. Cash SJ, Thelwall M, Peck SN, Ferrell JZ, Bridge JA. Adolescent suicide statements on MySpace. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):166-174. [doi: 10.1089/cyber.2012.0098] [Medline: 23374167]
17. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. J Med Internet Res 2017 Jul 10;19(7):e243 [FREE Full text] [doi: 10.2196/jmir.7276] [Medline: 28694239]
18. Barak A. Emotional support and suicide prevention through the Internet: a field project report. Comput Hum Behav 2007 Mar;23(2):971-984. [doi: 10.1016/j.chb.2005.08.001]
19. Robinson J, Hetrick S, Cox G, Bendall S, Yuen HP, Yung A, et al. Can an Internet-based intervention reduce suicidal ideation, depression and hopelessness among secondary school students: results from a pilot study. Early Interv Psychiatry 2016 Feb;10(1):28-35. [doi: 10.1111/eip.12137] [Medline: 24684946]
20. China Internet Network Information Center. 41st Statistical Report on Internet Development in China URL: http://www.cac.gov.cn/2018-01/31/c_1122347026.htm [accessed 2019-04-22] [WebCite Cache ID 71Bix5Mxc]
21. Microblog Sina. Sina microblog user development report URL: http://blog.sina.com.cn/s/blog_e9475ade0102x1wv.html [accessed 2019-04-22] [WebCite Cache ID 71Bj4HljO]
22. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
23. Barak A, Miron O. Writing characteristics of suicidal people on the Internet: a psychological investigation of emerging social environments. Suicide Life Threat Behav 2005 Oct;35(5):507-524. [doi: 10.1521/suli.2005.35.5.507] [Medline: 16268768]
24. Kumar M, Dredze M, Coppersmith G, De Choudhury M. Detecting changes in suicide content manifested in social media following celebrity suicides. Proc ACM Conf Hypertext Soc Media 2015 Sep;2015:85-94 [FREE Full text] [doi: 10.1145/2700171.2791026] [Medline: 28713876]
25. Jacob N, Scourfield J, Evans R. Suicide prevention via the Internet: a descriptive review. Crisis 2014 Jan 01;35(4):261-267. [doi: 10.1027/0227-5910/a000254] [Medline: 24984890]
26. Westerlund M, Hadlaczky G, Wasserman D. The representation of suicide on the Internet: implications for clinicians. J Med Internet Res 2012 Sep 26;14(5):e122 [FREE Full text] [doi: 10.2196/jmir.1979] [Medline: 23010086]
27. Zhang L, Huang X, Liu T, Li A, Chen Z, Zhu T. Using linguistic features to estimate suicide probability of Chinese microblog users. Lecture Notes in Computer Science 2015:549-559. [doi: 10.1007/978-3-319-15554-8_45]
28. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related twitter posts. Crisis 2017 Feb 23;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]
29. Saha K, Weber I, De Choudhury M. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. Proc Int AAAI Conf Weblogs Soc Media 2018 Jun:320-329 [FREE Full text] [Medline: 30505628]
30. Zhang Y, Ramesh A, Golbeck J, Sridhar D, Getoor L. A structured approach to understanding recovery and relapse in AA. 2018 Presented at: Proceedings of the 2018 World Wide Web Conference; 2018; Lyon. [doi: 10.1145/3178876.3186019]
31. De Choudhury M, Kıcıman E. The language of social support in social media and its effect on suicidal ideation risk. Proc Int AAAI Conf Weblogs Soc Media 2017 May;2017:32-41 [FREE Full text] [Medline: 28840079]
32. Hirsch JK, Duberstein PR, Conner KR, Heisel MJ, Beckman A, Franus N, et al. Future orientation and suicide ideation and attempts in depressed adults ages 50 and over. Am J Geriatr Psychiatry 2006 Sep;14(9):752-757 [FREE Full text] [doi: 10.1097/01.JGP.0000209219.06017.62] [Medline: 16943172]
33. Chang EC, Yu EA, Lee JY, Hirsch JK, Kupfermann Y, Kahle ER. An examination of optimism/pessimism and suicide risk in primary care patients: does belief in a changeable future make a difference? Cogn Ther Res 2012 Dec 16;37(4):796-804 [FREE Full text] [doi: 10.1007/s10608-012-9505-0]
34. Tan Z, Liu X, Liu X, Cheng Q, Zhu T. Designing microblog direct messages to engage social media users with suicide ideation: interview and survey study on Weibo. J Med Internet Res 2017 Dec 12;19(12):e381 [FREE Full text] [doi: 10.2196/jmir.8729] [Medline: 29233805]
35. Li X, Zhang Y, Wang Z, Yang S, Fei L. Characteristics of non-psychiatric patients with major depressive episode from general hospitals. Chin Mental Health J 2006;20(3):171-175. [doi: 10.3969/j.issn.1004-6763.2007.16.004]
36. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
37. Yu X, Tam WWS, Wong PTK, Lam TH, Stewart SM. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. Compr Psychiatry 2012 Jan;53(1):95-102. [doi: 10.1016/j.comppsych.2010.11.002] [Medline: 21193179]
38. D'Zurilla TJ, Goldfried MR. Problem solving and behavior modification. J Abnorm Psychol 1971 Aug;78(1):107-126. [Medline: 4938262]
39. Burnap P, Colombo W, Scourfield J. Machine classification and analysis of suicide-related communication on Twitter. 2015 Presented at: Proceedings of the 26th ACM Conference on Hypertext & Social Media; 2015; Guzelyurt. [doi: 10.1145/2700171.2791023]

40. Silverman MM, Berman AL, Sanddal ND, O'carroll PW, Joiner TE. Rebuilding the tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors. Part 2: Suicide-related ideations, communications, and behaviors. Suicide Life Threat Behav 2007 Jun;37(3):264-277 [FREE Full text] [doi: 10.1521/suli.2007.37.3.264] [Medline: 17579539]
41. Coppersmith G, Leary R, Whyne E, Wood T. Quantifying suicidal ideation via language usage on social media. URL: https://qntfy.com/static/papers/jsm2015.pdf [accessed 2018-07-26] [WebCite Cache ID 71BqRAN5y]
42. Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A. Suicide note classification using natural language processing: a content analysis. Biomed Inform Insights 2010 Aug 04;2010(3):19-28 [FREE Full text] [Medline: 21643548]
43. Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, et al. Automatic personality assessment through social media language. J Pers Soc Psychol 2015 Jun;108(6):934-952. [doi: 10.1037/pspp0000020] [Medline: 25365036]
44. Wang X, Zhang C, Ji Y, Sun L, Wu L, Bao Z. A depression detection model based on sentiment analysis in micro-blog social network. Revised Selected Papers of PAKDD 2013 International Workshops on Trends and Applications in Knowledge Discovery and Data Mining 2013;7867:201-213. [doi: 10.1007/978-3-642-40319-4_18]
45. Huang X, Zhang L, Chiu D, Liu T, Li X, Zhu T. Detecting suicidal ideation in Chinese microblogs with psychological lexicons. 2014 Presented at: UIC-ATC-SCALCOM '14 Proceedings of the 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom); 2014; Washington p. 182-187. [doi: 10.1109/uic-atc-scalcom.2014.48]
46. Brezo J, Paris J, Turecki G. Personality traits as correlates of suicidal ideation, suicide attempts, and suicide completions: a systematic review. Acta Psychiatr Scand 2006 Mar;113(3):180-206. [doi: 10.1111/j.1600-0447.2005.00702.x] [Medline: 16466403]
47. Harrington R. Depression, suicide and deliberate self-harm in adolescence. Brit Med Bull 2001;57(1):47-60. [doi: 10.1093/bmb/57.1.47] [Medline: 11719923]
48. Vandivort DS, Locke BZ. Suicide ideation: its relation to depression, suicide and suicide attempt. Suicide Life-Threat Behav 1979;9(4):205-218. [doi: 10.1111/j.1943-278x.1979.tb00439.x]
49. Golbeck J, Robles C, Edmondson M, Turner K. Predicting personality from Twitter. 2011 Presented at: Privacy, Security, Risk and Trust (PASSAT) and IEEE Third International Conference on Social Computing (SocialCom); 2011; Boston. [doi: 10.1109/passat/socialcom.2011.33]
50. Schwartz HA, Eichstaedt JC, Dziurzynski L, Kern ML, Blanco E, Kosinski M, et al. Toward personality insights from language exploration in social media. 2013 Presented at: AAAI Spring Symposium: Analyzing Microtext; 2013; Stanford.
51. Passos IC, Mwangi B, Cao B, Hamilton JE, Wu M, Zhang XY, et al. Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach. J Affect Disord 2016 Mar 15;193:109-116 [FREE Full text] [doi: 10.1016/j.jad.2015.12.066] [Medline: 26773901]
52. Guan L, Hao B, Cheng Q, Yip PS, Zhu T. Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. JMIR Ment Health 2015;2(2):e17 [FREE Full text] [doi: 10.2196/mental.4227] [Medline: 26543921]
53. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. 2055 Presented at: European Conference on Information Retrieval; 2005; Santiago de Compostela p. 345-359. [doi: 10.1007/978-3-540-31865-1_25]
54. Gao R, Hao B, Li H, Gao Y, Zhu T. Developing simplified Chinese psychological linguistic analysis dictionary for microblog. 2013 Presented at: International Conference on Brain and Health Informatics; 2013; Maebashi p. 359-368. [doi: 10.1007/978-3-319-02753-1_36]
55. Liu F, Zhou N, Cao H, Fang X, Deng L, Chen W, et al. Chinese college freshmen's mental health problems and their subsequent help-seeking behaviors: a cohort design (2005-2011). PLoS One 2017;12(10):e0185531 [FREE Full text] [doi: 10.1371/journal.pone.0185531] [Medline: 29040266]
56. Burnap P, Rana OF, Avis N, Williams M, Housley W, Edwards A, et al. Detecting tension in online communities with computational Twitter analysis. Technol Forecasting Soc Change 2015 Jun;95:96-108. [doi: 10.1016/j.techfore.2013.04.013]
57. Huang X, Li X, Zhang L, Liu T, Chiu D, Zhu T. Topic model for identifying suicidal ideation in Chinese. URL: http://www.aclweb.org/anthology/Y15-1064 [accessed 2018-07-26] [WebCite Cache ID 71BtBH4JP]
58. Cao X, Zhong B, Xiang Y, Ungvari GS, Lai KYC, Chiu HFK, et al. Prevalence of suicidal ideation and suicide attempts in the general population of China: a meta-analysis. Int J Psychiatry Med 2015;49(4):296-308 [FREE Full text] [doi: 10.1177/0091217415589306] [Medline: 26060259]
59. Bryan CJ, Butner JE, Sinclair S, Bryan ABO, Hesse CM, Rose AE. Predictors of emerging suicide death among military personnel on social media networks. Suicide Life Threat Behav 2017 Jul 28:1. [doi: 10.1111/sltb.12370]
60. Chen J. Seeking help for psychological distress in urban China. J Community Psychol 2012 Apr 04;40(3):319-341. [doi: 10.1002/jcop.20513]
61. Ma X, Xiang Y, Cai Z, Li S, Xiang Y, Guo H, et al. Lifetime prevalence of suicidal ideation, suicide plans and attempts in rural and urban regions of Beijing, China. Aust N Z J Psychiatry 2009 Feb;43(2):158-166. [doi: 10.1080/00048670802607170] [Medline: 19153924]

62. Sueki H, Yonemoto N, Takeshima T, Inagaki M. The impact of suicidality-related internet use: a prospective large cohort study with young and middle-aged internet users. PLoS One 2014;9(4):e94841 [FREE Full text] [doi: 10.1371/journal.pone.0094841] [Medline: 24740115]
63. Suominen K, Isometsä E, Suokas J, Haukka J, Achte K, Lönnqvist J. Completed suicide after a suicide attempt: a 37-year follow-up study. Am J Psychiatry 2004 Mar;161(3):562-563. [doi: 10.1176/appi.ajp.161.3.562] [Medline: 14992984]
64. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005 Feb;34(1):215-220 [FREE Full text] [doi: 10.1093/ije/dyh299] [Medline: 15333621]
65. Clifton L, Clifton DA. The correlation between baseline score and post-intervention score, and its implications for statistical analysis. Trials 2019 Jan 11;20(1):43-48 [FREE Full text] [doi: 10.1186/s13063-018-3108-3] [Medline: 30635021]

Abbreviations

DT: decision tree
PSPO: Proactive Suicide Prevention Online
SCLIWC: Simplified Chinese Linguistic Inquiry and Word Count
PHQ-9: 9-Item Patient Health Questionnaire
RF: random forest
SVM: support vector machine

Edited by G Eysenbach; submitted 26.07.18; peer-reviewed by J Han, K Bentley; comments to author 08.10.18; revised version received 02.12.18; accepted 30.03.19; published 08.05.19.

Please cite as:
Liu X, Liu X, Sun J, Yu NX, Sun B, Li Q, Zhu T
Proactive Suicide Prevention Online (PSPO): Machine Identification and Crisis Management for Chinese Social Media Users With Suicidal Thoughts and Behaviors
J Med Internet Res 2019;21(5):e11705
URL: https://www.jmir.org/2019/5/e11705/
doi:10.2196/11705
PMID:31344675

Original Paper

Self-Management Maintenance Inhalation Therapy With eHealth (SELFIE): Observational Study on the Use of an Electronic Monitoring Device in Respiratory Patient Care and Research

Esther Kuipers[1,2], PharmD; Charlotte C Poot[3], MSc; Michel Wensing[1,4], PhD; Niels H Chavannes[3], PhD; Peter AGM de Smet[1,5], PhD; Martina Teichert[6], PhD

[1]Department of IQ Healthcare, Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, Netherlands
[2]BENU Apotheek Zeist West, Zeist, Netherlands
[3]Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands
[4]Department of General Practice and Health Services Research, University Hospital Heidelberg, Heidelberg, Germany
[5]Department of Clinical Pharmacy, Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, Netherlands
[6]Department of Clinical Pharmacy & Toxicology, Leiden University Medical Centre, Leiden, Netherlands

Corresponding Author:
Esther Kuipers, PharmD
Department of IQ Healthcare
Radboud Institute for Health Sciences
Radboud University Medical Centre
PO Box 9101
Nijmegen, 6500HB
Netherlands
Phone: 31 243610591
Fax: 31 243540166
Email: esther.kuipers@radboudumc.nl

Abstract

Background: Electronic inhalation monitoring devices (EIMDs) are available to remind patients with respiratory diseases to take their medication and register inhalations for feedback to patients and health care providers as well as for data collection in research settings.

Objective: This study aimed to assess the validity as well as the patient-reported usability and acceptability of an EIMD.

Methods: This observational study planned to include 21 community pharmacies in the Netherlands. Patient-reported inhalations were collected and compared to EIMD registrations to evaluate the positive predictive value of these registrations as actual patient inhalations. Patients received questionnaires on their experiences and acceptance.

Results: A convenience sample of 32 patients was included from across 18 pharmacies, and 932 medication doses were validated. Of these, 796 registrations matched with patient-reported use (true-positive, 85.4%), and 33 inhalation registrations did not match with patient-reported use (false-positive, 3.5%). The positive predictive value was 96.0%, and 103 patient-reported inhalations were not recorded in the database (false-negative, 11.1%). Overall, patients considered the EIMD to be acceptable and easy to use, but many hesitated to continue its use. Reminders and motivational messages were not appreciated by all users, and more user-tailored features in the app were desired.

Conclusions: Patients' interaction with the device in real-world settings is critical for objective measurement of medication adherence. The positive predictive value of this EIMD was found to be acceptable. However, patients reported false-negative registrations and a desire to include more user-tailored features to increase the usability and acceptability of the EIMD.

(J Med Internet Res 2019;21(5):e13551) doi:10.2196/13551

KEYWORDS

eHealth; pharmacy; inhalation therapy; asthma; COPD; pharmacy practice research

Introduction

Electronic monitoring devices are increasingly important in the self-management of chronic diseases such as chronic obstructive pulmonary disease and asthma. These two chronic respiratory diseases affect an estimated 384 and 235 million people worldwide, respectively [1,2]. According to the prevailing clinical guidelines, daily intake of inhaled corticosteroids is the cornerstone for optimal asthma treatment, and inhaled corticosteroids are also used in chronic obstructive pulmonary disease, together with bronchodilators [2,3].

An electronic inhalation monitoring device (EIMD), which measures inhalation actuations, provides detailed data on patient adherence to treatment for both patient and health care providers [4-7]. EIMDs, in combination with audiovisual reminders and feedback, have the potential to improve medication adherence and health care outcomes by facilitating self-management and aiding in clinical (shared) decision making [8-15]. Previous studies have shown that involving patients in monitoring their own symptoms can lead to improved awareness and competence in disease management [16].

Besides clinical practice, EIMD data could be used as objective outcome measures for medication adherence in research [17-19]. The integrated time stamp enables comprehensive data collection in research populations on the timing and pattern of inhaler actuation, including dose dumping [20]. This way of data collection was stated to be more accurate and objective for the evaluation of medication use and is considered to be superior to self-report, canister weighing, or pharmacy dispensing data [17,21].

However, the widespread use of EIMDs for measuring patients' medication adherence in clinical practice and research setting depends on the acceptability by patients as well as health care providers and researchers, respectively. From the perspectives of both patient care and research, it is important that the EIMD is accurate and valid [20,22-24]. Earlier in-vitro studies (product validation studies) evaluating the validity of an EIMD, following a fixed protocol and simulating inhalations, found 99.2% overall accuracy of detection [25]. On the contrary, small-scale studies reported malfunctioning devices and potential loss of data as well as overrecording of doses that did not reflect actual inhalations [22-24,26]. Use of EIMDs as objective measures for medication adherence for research purposes in a real-world setting depends not only on technical capabilities, but also on how the user interacts with the system in real life. In other words, results in laboratory settings might not apply to EIMD validity and reliability in the broad use by community-dwelling patients. As EIMDs could be used for measuring real-time medication adherence in clinical trials, evidence on their potential to measure patient inhalations in community-dwelling patients is relevant.

For the implementation of EIMD in clinical practice, it is important to acknowledge the complexity of the implementation process, influenced by patients' acceptance and ability to use information technology tools [27,28]. Hence, before the EIMD is implemented and used as an accurate and reliable measure for patients' actual medication adherence in a real-world setting (in both clinical decision making and research), a rigorous evaluation of the technical performance, usability, and acceptability in clinical practice is required.

This study aimed to assess the validity and patient-reported usability and acceptability of an electronic adherence monitor and reminder device for patient care and research.

Methods

Study Design

This prospective observational study on agreement between EIMD measurements and patient-reported inhalations was conducted between April 18 and May 25, 2018.

Setting

This study was conducted by 21 students in the second year of their master's program in pharmacy from the University of Leiden, the Netherlands, during their internships in community pharmacies. The students were asked to validate EIMD registrations for two patients each, for 3 weeks, and they received additional training on the use of the EIMD program. The individual monitoring of patients was time consuming because of the protocol-specific information requirements, and consequently, dedicated persons with enough time were needed in the pharmacy. Additionally, the acquaintance with an innovative device to be implemented into daily practice and individual patient coaching were relevant learning objectives for the students.

Dutch pharmacists have a professional and legal responsibility to provide drug treatment for their patients and, as most patients in the Netherlands attend one community pharmacy, pharmacists usually possess the complete medication histories of their patients [29].

Ethics Approval

The study protocol was approved by the Ethical Committee of the Radboudumc Nijmegen (approval number, 2018-4153). Written informed consent was obtained from all individual participants included in this study, prior to the study.

Electronic Monitoring

The EIMD to be assessed was the Turbu+ V2.1 (AstraZeneca UK Limited), consisting of three components: (1) the electronic device that could be attached to the inhaler of the corresponding product "Symbicort Turbuhaler," (2) an app to be installed on the patient's mobile phone, and (3) an online portal allowing the health care professional access to the same actuation data. During patient inhalation, the device was actuated and the date and time of the actuations were recorded. The data were synchronized with the app on the mobile phone by Bluetooth, and the app visualized the timeline of these data up to the previous month. All EIMD data for measured actuations were automatically and electronically linked to the online portal of the health care provider for his/her patients and an additional research database containing the same data but anonymized. The research database has been setup to study patterns of medication adherence and to evaluate the effectiveness on interventions for medication adherence across multiple studies.

Actuation data from all health care professionals included in this study were aggregated into one central research database containing data on all included patients from the participating pharmacies.

Patients could only use the EIMD program after they were enrolled and trained by their health care provider. After entering the name, birth of date, sex, dose regimen settings, and email address, the patient automatically received an email with the link to download the app. The time(s) for a pop-up reminder in the app could be set individually by the patient at the time of scheduled inhalation(s). If patients did not take their medication, 30 minutes after the scheduled inhalation, a “missed-medication” motivational message was sent automatically. Patients received a weekly motivational push notification in the app (eg, “Great week. You’ve been following your prescription this week! Keep it up!”).

Patient Inclusion

Patients were eligible to participate if they (1) were current users of budesonide/formoterol Turbuhaler (Anatomic Therapeutic Chemical code R03AK07) [30] with at least two prescriptions in the previous 12 months; (2) were of age ≥ 18 years; (3) were regular patients in the pharmacy (registered in the pharmacy system and receiving dispensings from this pharmacy only); (4) had access to a smartphone, and (5) were able to use the internet. We aimed to include current users who were familiar with the inhaler. At random, eligible patients were invited during regular pharmacy visits or by phone. If interested, patients received an information leaflet on the study and an informed consent form. Patients interested in participating were asked to provide informed consent, allowing the student to collect general data about gender, age, use of short corticosteroid courses (indicating exacerbations), and refills from the pharmacy database (to assess adherence rates). Eligible patients were invited to the pharmacy for an intake visit.

Intake Visit

During the intake visit, the students provided the patient with an EIMD and (oral and written) instructions. The patients attached the device to the inhaler and installed the app during the visit or at home following the instructions. Additionally, the students verified the pharmacy’s information collected on drug use, patient’s refill adherence to inhaled corticosteroids, and the number of oral corticosteroid courses in the previous year. Subsequently, the patients completed the Beliefs About Medicines Questionnaire (BMQ), consisting of two parts: the BMQ Specific list, which measures perceptions of specific medicines, and the BMQ General list, which measures general beliefs about medicines. The BMQ Specific list comprises two scales—one assessing patients’ beliefs about the necessity of preventer medication for maintaining present and future health (Necessity scale) and the other for assessing their concerns about the potential adverse consequences of using such medication (Concerns scale) [31,32]. All items were rated on a 5-point Likert scale, with a range of 5-25 possible scores for both scales.

The students explained that they would call the patients 6 times within 14-21 days to check on their actual drug intake on the previous days and agreed on the most suitable days and time to call. If patients wished to, they were allowed to use a paper diary.

Accuracy and Patient-Reported Use of the Electronic Inhalation Monitoring Device: Procedures and Measures

During the follow-up period of 14-21 days, the students phoned the patients 6 times to check on their actual drug intake and EIMD performance from the previous day, and if possible, up to the day of the earlier phone call. The student collected and reported information on the number and time of daily inhalations, and all particularities or other circumstances that the patient reported were registered by the student in an Excel file. These were subsequently centralized into one patient self-reported database.

To avoid socially desirable answers and to collect reliable data, patients were instructed that this study focused on the accuracy of the device (instead of their medication use), and they were asked to use the medication at their usual dose and frequency and to use only the EIMD prepared device during the study period and no other inhalers of this medication. Semistructured questions were used during the phone call; for example, “When did you use your medication for the last time?” and “Do you see any registrations in your app that do not represent real medication use?”

During the phone call, the student had immediate access to the EIMD data and compared the data from the research database with the information reported by the patient (and registered in the patient self-reported database). Any discrepancies were directly discussed with the patient and registered.

Usability and Acceptability of the Patient Electronic Inhalation Monitoring Device: Procedures and Measures

To evaluate patients’ usability, patients received two questionnaires at the end of the study, including the System Usability Scale (SUS), which is a validated instrument for evaluating the usability of a wide range of products and services [33,34]. The SUS score can range from 0 to 100. For products with a score less than 70, improvement options should be considered [34].

The second questionnaire addressed the experiences and acceptability of the program and the expectations on the pharmacists’ role in counselling. This part consisted of 12 statements, formulated positively or negatively, measured on a 5-point Likert Scale (1=strongly disagree, 5=strongly agree), followed by 3 multiple-choice questions regarding the frequency of using the app and whether the patient would like to continue the program and by 4 open questions on advantages, disadvantages, and targets for improvements.

Data Analysis

The main outcome measure was the positive predictive value (PPV), calculated as (the number of correctly registered doses according to the patient self-reported database and registered database)/(the total number of registered doses [true and false positives])*100. Results of the questionnaires were analyzed using descriptive statistics. Responses to the open questions were coded and summarized for patients' opinions, wishes, and barriers encountered using the EIMD program to identify key topics via the grounded theory approach [35].

Patient-specific characteristics were assessed with the scores for the BMQ Specific and BMQ General questionnaires, represented on a continuous scale. The BMQ necessity and concerns scores were split at the scale midpoints to distinguish between four subgroups: accepting (high necessity, low concerns), ambivalent (high necessity, high concerns), skeptical (low necessity, high concerns), and indifferent (low necessity, low concerns) [36]. Refill adherence was calculated as the proportion of days covered in the prior 12 months (due to prior dispensings up to 15 months).

Results

Patient Demographics and Questionnaire Completion

A total of 32 patients were included by 18 students (Table 1). In 3 pharmacies, no patients eligible and willing to participate were available. The mean age of the included patients was 48.1 years (range, 20-69 years), and 17 patients were female (53.1%). Medication adherence in the previous year calculated from dispensing data was 81.3%. Most patients used their medication twice a day (n=26). Three patients (9%) used an additional diary to note their daily inhalations. Twenty-five patients (78%) completed the questionnaires and provided individual comments on the EIMD (Table 2 and Textbox 1).

Table 1. Patient characteristics (N=32).

Characteristic	Value
Age (years), mean (range)	48.1 (20-69)
Female, n (%)	17 (53.1)
Number of oral corticosteroid courses in the last year, mean (range)	0.31 (0-4)
Medication adherence in the previous year calculated from dispensing data (n=28), mean % (range)	81.32 (17.53-194.80)
BMQ[a] General Harm score (n=29; possible range: 4-20), mean (range)	10.21 (5-16)
BMQ General Overuse score (n=29; possible range: 4-20), mean (range)	8.04 (4-13)
BMQ Specific Necessity score (n=30; possible range: 5-25), mean (range)	17.70 (10-24)
BMQ Specific Concerns score (n=30; possible range: 5-25), mean (range)	10.97 (6-17)

[a] BMQ: Beliefs About Medicines Questionnaire.

Table 2. Patients' acceptability of the Electronic Inhalation Monitoring Device (EIMD) program. Values in italics indicate the highest percentage(s) per statement.

Statement	Disagree (score 1-2), n (%)	Neutral (score 3), n (%)	Agree (score 4-5), n (%)
The EIMD programme does provide useful insights in my medication use[a]	*11 (44)*	3 (12)	*11 (44)*
The EIMD programme is useful to me	6 (24)	8 (32)	*11 (44)*
The EIMD programme is time-consuming	*19 (76)*	2 (8)	4 (16)
The motivational messages in the EIMD app feel positive for me	*11 (44)*	10 (40)	4 (16)
The reminders of the EIMD programme help me to take my medication in time	*10 (40)*	6 (24)	9 (36)
The EIMD programme contributes to dealing with my illness[a]	*10 (40)*	5 (20)	*10 (40)*
The EIMD programme contributes to the interaction with my pharmacist	*10 (40)*	8 (32)	7 (28)
The EIMD programme helps establishing a habit to use my medication	*11 (44)*	9 (36)	5 (20)
The EIMD programme gives me confidence to help manage my disease	*10 (40)*	9 (36)	6 (24)
The pharmacist's monitoring of my medication use with this programme helps me to use my medication better[a]	*13 (52)*	6 (24)	6 (24)
When someone can monitor my medication intake, I take my medication as usual[a]	4 (16)	3 (12)	*18 (72)*
I see added value of the EIMD programme to manage my medication intake[a]	*11 (44)*	5 (20)	9 (36)
I would recommend the EIMD programme to other patients	7 (28)	*12 (48)*	6 (24)
The pharmacist can play an important role in counselling	2 (8)	3 (12)	*20 (80)*

[a] Item was phrased as a negative statement.

Textbox 1. Selection of patients' free-text comments about the Electronic Inhalation Monitoring Device program. Values in brackets represent gender and age of the participants (eg, F24=female, 24 years).

Technical functionalities:

> *If the app synchronises the data, this may take a long time (maybe this can be improved). In the end, the registration of my puffs did not go well.* [F24]
>
> *Until now, the app missed three inhalation registrations in the morning. I clicked twice within a minute, but only one [inhalation] was registered.* [M52]

Patient-technology interaction:

> *Not all inhalations were registered [in the app on my mobile phone], so I inhaled again and used too much medication.* [F45]
>
> *The app is not always working. If medication is taken, this is not always measured. Even if the medication has already been taken (before the time set), still a reminder was sent. The app does not provide insight into whether the medication has been taken in the right way.* [F20]
>
> *The alarm you can set will never go off. So you have no reminder.* [F45]
>
> *The Turbu+ did not register when the app was [physically] not around, and it was therefore not possible to register 2 inhalations in 1 day.* [M55]
>
> *The notifications did not work with my Galaxy s7 Edge.* [M57]

Patients' opinions and acceptability of the program:

Regarding the device/app:

> *The device itself feels rather rickety. The two parts did not really fit together.* [M51]
>
> *The idea is good, I also see the commercial need, but think about the return for the patient. Information in the app for how many days medication is still available in the inhaler, would be such convenient return for the patient. For example, before he goes on holiday, he knows if there are extra medicines needed.* [M55]
>
> *I feel that tracking medication use can be useful for many patients. However, it remains very difficult, because the connection of the device and the app is not clear. As a result, it [medication use] is registered at some time, but not on other times, for unclear reasons. With this, possible useful options of the app will immediately disappear.* [M43]

Regarding reminders:

> *If you are not adherent, the Turbu+ can be convenient as a reminder for using the medication.* [F39]
>
> *[Useful] when you forget [medication intake] (but I never forget this).* [M42]
>
> *I personally liked the reminders, and it was also very pleasant that I could monitor myself whether I had taken it or not!* [F24]
>
> *I would like to modify the app myself for what I agreed on with my doctor (about minimum and maximum use per day).* [F47]
>
> *[I would prefer] a clearer reminder sound signal.* [V56]

Regarding motivational messages:

> *The tone of the written messages deserves carefulness: the patient is the boss, the app only gives advice.* [F47]

Regarding attitude to electronic monitoring:

> *I would only use the app if only I can see it [the data]. Watching by pharmacist should not be the default for each patient.* [F47]
>
> *I am intrinsically motivated to use my medication and do not really need an app for that.* [M55]
>
> *Once the routine is there to take an inhalation twice a day, it has little added value for the users. It gives the impression that it is only useful for the pharmacist and the manufacturer to collect Big Data.* [M55]
>
> *I always take my medication when brushing my teeth. So there was no need for help with the intake. In addition, I am able to feel when I have to use my medication more often, according to my needs. I think that the app can help people who have difficulties with this.* [M28]
>
> *[An advantage is] to check when I am not sure about forgotten medicines. Insights into patterns in periods when I have more symptoms help me to anticipate.* [F48]

Thirty patients (94%) completed the BMQ General and Specific questionnaires at the start of the study. The majority of them could be classified as acceptant (n=19) regarding their inhalation medication: 4 patients were ambivalent and 7 were indifferent. No patients were classified as skeptical.

Accuracy and Patient-Reported Use of the Electronic Inhalation Monitoring Device

Of the 32 patients, 28 completed all 6 phone calls. Overall, the 18 students verified 932 medication doses (mean=29.1 doses verified per patient; range=3-88). A total of 796 doses registered in the research database matched patient-reported inhalations captured in the patient self-reported database (true-positive, 85.41%). In addition, 33 inhalations were registered in the research database that did not match the actual drug intake, as reported by the patients in the patient self-reported database (false-positive, 3.5%). Further, 22 patients reported an average of 4.7 inhalations (range, 1-20) that were not recorded in the research database, accounting for a total of 103 inhalations (false-negative, 11.1%).

Information on registration of 6 of the 32 devices was in full agreement with the patient-reported inhalations. The PPV of all registrations in the research database was 96.0%.Some patients reported technical problems with EIMD data recording and synchronization of the EIMD data with their mobile phone. Delayed data synchronization resulted in two patients taking more medication than prescribed under the assumption that they had forgotten their dose.

Usability and Acceptability of the Electronic Inhalation Monitoring Device

The majority of patients indicated that the app was easy to use and not unnecessarily complex; the mean SUS score was 68.9 (SD 11.34; range, 52.5-90).

In the acceptability questionnaire, patients rated the EIMD generally as useful (76% neutral or agree) and 84% rated the EIMD program as not time-consuming (Table 2). With regard to recommending the EIMD to other patients, the majority rated this item as neutral (n=12, 48%). In addition, 80% of the patients indicated that they were not willing (n=15) or uncertain (“maybe,” n=5) about continuing the program themselves, and 70% (n=14) of these patients reported that they had only participated in this study because they wanted to help the students fulfil their assignment and facilitate the research. The five patients who were positive about continuation reported that they participated mainly to gain personal insight into their inhalation patterns. The patients generally felt positive about the pharmacists’ role in counselling.

Written feedback on the advantages, disadvantages, and targets for improvements was provided by 25 patients. This feedback was clustered into four themes: EIMD functionality, reminders, motivational messages, and attitude toward electronic monitoring (Textbox 1).

Some of the previously mentioned technical issues reported by some patients resulted in a level of frustration or confusion because of missed or unnecessary reminders due to missing data, which impacted their acceptance of the app (Textbox 1). The reminders and motivational messages were appreciated by 9 and 4 patients, respectively. Several patients suggested more individualization of the settings in the app, such as inclusion of a personal choice to share data with a specific health care provider and management of their dose regimen settings.

Discussion

Principal Findings

In this study, we found an acceptable PPV: 96.0% of the registered doses represented patient-reported drug intake. However, we found a high number of false-negative registrations: 11.1% of patient-reported inhalations were not recorded by the EIMD.

These unrecorded patient-reported inhalations could have been the result of a number of factors, either technical or user-related issues, that should be discussed. Although in earlier studies, loss of data or missing data were associated with technical issues such as battery drain, this was not likely to have occurred during the short duration of our study [7,24,26]. The high number of false-negative registrations could possibly be the result of overreporting by the patients or suboptimal use of the device (eg, Bluetooth not activated, EIMD not paired to the phone, or EIMD not within a 5-meter distance from the phone during inhaler actuation). Assessment of user experience revealed that the written instruction did not contain detailed information about the data synchronization protocol, which, combined with instructions at enrolment, may have led to some of the reported observations. In earlier research with the same EIMD, the researchers had presumed the possibility of false-negative as well as false-positive registrations, but they were unable to verify the registrations by patients’ actual inhalations or the data on user interaction with the system to interpret their findings [15].

This study demonstrated the importance of validating medication adherence data in real-world settings. Patients’ interaction with the device is critical for objective measurement of medication adherence in research and clinical settings. We emphasized on the importance of evaluating technical performance to identify technical/user issues and stressed the need of evaluating usability and acceptability across multiple components of the EIMD.

Previous studies on EIMD performance focused on accuracy and reliability in laboratory settings and lacked data collected in a real-life setting where patients interact with the EIMD. Furthermore, previous studies did not access patient acceptability and user experience [7,20,23,24,26,37,38], both of which are essential for successful implementation and sustained use in daily practice [39]. This is the first study to demonstrate the importance of evaluating and validating EIMDs in a real-world setting. The accuracy of inhalation measurement is essential from not only a research point of view, but also a clinical perspective, as false-negative registrations lead to unnecessary signals and reminder messages. Furthermore, it falsely reports patients as nonadherent, and this could lead to underestimation of adherence, incorrect clinical decision making, and overuse of medication when patients assume to have forgotten their dose.

Patient Experiences and Acceptability of the Electronic Inhalation Monitoring Device

Our findings on acceptability and user experiences further underline the importance of evaluating EIMDs on acceptability, preferably early in the implementation process. There is a growing body of evidence on electronic health apps, in general, that do not perform as expected in clinical practice, because the app turns out to be unacceptable or does not fit the users' needs [40-43]. Although patients indicated that the app was easy to use and not complex, the majority did not intend to use the app in the future, apparently because there was no clear personal need to use the EIMD. They were primarily motivated to use the EIMD to facilitate the research rather than having an intrinsic motivation to gain personal insights into their inhalation patterns, and their medication adherence in the previous year calculated from dispensing data was already high. Patients with intrinsic motivation to improve their disease management (eg, based on low adherence or impaired disease control) would possibly benefit more from the EIMD. Therefore, understanding different types of patient segments is important to succeed in the implementation; the EIMD needs to match with the patient profile (eg, adherence and asthma control), needs, and preferences. In addition, the fact that the majority of patients did not intend to continue with the EIMD may have been the result of a suboptimal technical performance, for example, loss or troubles with the Bluetooth connection, which was regularly experienced; delay in the synchronization of data from the EIMD to the mobile app; and inhalations taken just before midnight not visible on the intended day. These technicalities can probably easily be improved and thereby increase the chance of acceptance and successful implementation. Moreover, health care providers and patients would benefit from further development of the EIMD, so that it can not only detect inhaler actuation, but also check the inhalation quality, breath force, and inhalation technique.

Our findings emphasized the need for clear patient selection and a more individualized, tailored device. In this study, half of the people found the reminders helpful in taking the medication on time, while the other half found them useless. We found similar results for the motivational messages. Hence, when designing a self-management intervention containing multiple self-management strategies (ie, motivational messaging, reminders, and audio-visual behavioral feedback), it is important to critically review each component on usability separately and preferably tailor the intervention to the needs of the patient.

Strengths and Limitations

This study has some limitations. First, the actual inhalations were self-reported by the patients, which could have led to bias by overreporting actual use (and thus also an overestimation of unrecorded actuations), as patients might tend to provide socially desirable answers. However, as a 24-hour patient observation was not feasible, this was the best way to collect data on patients' actual drug intake within this setting. To prevent socially desirable answers, patients were instructed at the intake that this study was on verifying the registrations of the EIMD rather than their medication use or adherence, and during the phone calls, the students asked open questions. Additionally, patients were questioned about their actual inhalation mainly at the same and previous days, with a maximum of 5 days prior. Thus, the phone calls for data evaluation focused on the most recent moments of drug intake in order to reduce possible recall bias. In further analysis regarding the possible impact of recall bias, we did not observe any differences between measurements on day 5 and those on the earlier days, or between different age groups (data not shown). As earlier studies have shown that there is considerable variation in the accuracy of diaries to note medication intake [44], the use of a diary was not mandatory. As patients frequently reported technical issues as a possible explanation for both missing and extra registered inhalations, we do not expect much bias from this setting.

Second, all pharmacies were related to the master's education program of pharmacy from the University of Leiden. Data were collected in different pharmacies and by different students, to prevent bias from specific settings. However, such bias could not be fully excluded. Although the students were not yet registered health care providers, they were quite motivated and technically skillful. The students selected a convenience sample of patients, although it seemed difficult to find patients with the original inhalation medication that was fit for the EIMD; in some pharmacies, no patients could be selected at all. Third, the short inclusion time of this study, due to the internship period, could have led to selection of patients who were more willing to help the students with their task rather than being interested in their own medication performance or adherence. This was reflected in patients' individual comments: They regularly indicated that the program could be especially useful for other patients, but that it was of little value to them. Some patients reported that they did not need "help with the inhalation" or "an app for taking the medication." With a mean medication adherence of >80%, this seems to be a group with relatively high adherence and could indicate selection bias of the more adherent patients, for which the device may be less useful. Further research is needed on how health care providers should preselect patients for an EIMD on the basis of their experiences.

A strength of this study was the intensive follow-up from students, with more than 900 validated measurements. Although the number of included patients was limited, the number of drug intake comparisons was sufficient to detect omissions in the recorded actuations. However, the small population may have affected the representativeness of the results of the acceptability questionnaire. The statements in this questionnaire were formulated positively or negatively to reduce the risk of positively biased answers. This questionnaire was adapted to this specific EIMD and not validated beforehand. As a consequence, some patients might have experienced difficulties in comprehending the language used or the variety in both positively and negatively formulated questions.

Implications for the Future

It is recommended that the discussed technical issues should be further elucidated and solved before using EIMD data as an objective adherence measurement, and medication overuse may also be of interest. In order to fully benefit from the EIMD and guarantee reliability and validity, an EIMD should be validated in a setting where the users interact with the system and can encounter technical or user issues. Furthermore, EIMDs and the accompanying self-management program should be evaluated on usability in daily practice. This study provided an example of how to do this. Validation of EIMDs in real-world settings is likely to improve usability in daily practice; the EIMD should be easy to use and measure all actuations correctly, even when the patient is not technically skilled. Future research should pay sufficient attention to different types of patient segments, as the EIMD needs to match with the patient profile (eg, adherence and asthma control), needs, and preferences.

Conclusions

Comparison of EIMD data with patient-reported inhalations showed that EIMD registrations represented patient inhalations to an acceptable degree, with a PPV of 96%, but these registrations were likely to underreport actual drug intake by 11%. Technical improvements should address the Bluetooth connection and data synchronization. Additionally, patient characteristics contribute to the validity of EIMD measurements, and larger sample sizes are needed to explore their influence. For the acceptance of a self-management program with an EIMD, patients who benefit from self-monitoring and reminders should be targeted by tailoring the possibilities to the needs of the individual user.

Acknowledgments

We thank all the patients, students, and community pharmacists who participated in this study. This work was supported by unconditional research grants from the Royal Dutch Association for the advancement of pharmacy (grant number PR16_1410) and AstraZeneca (grant number 2603949255).

Conflicts of Interest

None declared.

References

1. World Health Organization. Asthma URL: https://www.who.int/en/news-room/fact-sheets/detail/asthma [accessed 2019-01-30] [WebCite Cache ID 75oPywQIx]
2. The Global Initiative for Chronic Obstructive Lung Disease. 2018 Global Strategy for Diagnosis, Management, and Prevention of COPD URL: https://goldcopd.org/gold-reports-2017/ [accessed 2019-01-30] [WebCite Cache ID 75oQX0EUM]
3. 2018 Global Strategy for Asthma Management and Prevention. Global Initiative for Asthma URL: https://ginasthma.org/gina-reports/ [accessed 2019-01-30] [WebCite Cache ID 75oQTDICX]
4. Melvin E, Cushing A, Tam A, Kitada R, Manice M. Assessing the use of BreatheSmart® mobile technology in adult patients with asthma: a remote observational study. BMJ Open Respir Res 2017;4(1):e000204 [FREE Full text] [doi: 10.1136/bmjresp-2017-000204] [Medline: 29071077]
5. Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. BMC Med 2015;13:58 [FREE Full text] [doi: 10.1186/s12916-015-0303-x] [Medline: 25857569]
6. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013;11:CD010013. [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
7. Foster JM, Smith L, Usherwood T, Sawyer SM, Rand CS, Reddel HK. The reliability and patient acceptability of the SmartTrack device: a new electronic monitor and reminder device for metered dose inhalers. J Asthma 2012 Aug;49(6):657-662. [doi: 10.3109/02770903.2012.684253] [Medline: 22741746]
8. Silow-Carroll S, Smith B. Clinical management apps: creating partnerships between providers and patients. Issue Brief (Commonw Fund) 2013 Nov;30:1-10. [Medline: 24312989]
9. Roberts CA, Geryk LL, Sage AJ, Sleath BL, Tate DF, Carpenter DM. Adolescent, caregiver, and friend preferences for integrating social support and communication features into an asthma self-management app. J Asthma 2016 Nov;53(9):948-954. [doi: 10.3109/02770903.2016.1171339] [Medline: 27116540]
10. Chan AHY, Stewart AW, Harrison J, Camargo CA, Black PN, Mitchell EA. The effect of an electronic monitoring device with audiovisual reminder function on adherence to inhaled corticosteroids and school attendance in children with asthma: a randomised controlled trial. Lancet Respir Med 2015 Mar;3(3):210-219. [doi: 10.1016/S2213-2600(15)00008-9] [Medline: 25617215]
11. Chan AHY, Stewart AW, Harrison J, Black PN, Mitchell EA, Foster JM. Electronic adherence monitoring device performance and patient acceptability: a randomized control trial. Expert Rev Med Devices 2017 May;14(5):401-411. [doi: 10.1080/17434440.2017.1322505] [Medline: 28434269]
12. Charles T, Quinn D, Weatherall M, Aldington S, Beasley R, Holt S. An audiovisual reminder function improves adherence with inhaled corticosteroid therapy in asthma. J Allergy Clin Immunol 2007 Apr;119(4):811-816. [doi: 10.1016/j.jaci.2006.11.700] [Medline: 17320942]
13. Foster JM, Usherwood T, Smith L, Sawyer SM, Xuan W, Rand CS, et al. Inhaler reminders improve adherence with controller treatment in primary care patients with asthma. J Allergy Clin Immunol 2014 Dec;134(6):1260-1268.e3. [doi: 10.1016/j.jaci.2014.05.041] [Medline: 25062783]
14. Burgess SW, Sly PD, Devadason SG. Providing feedback on adherence increases use of preventive medication by asthmatic children. J Asthma 2010 Mar;47(2):198-201 [FREE Full text] [doi: 10.3109/02770900903483840] [Medline: 20170329]

15. Kuipers E, Wensing M, de Smet P, Teichert M. Self-management research of asthma and good drug use (SMARAGD study): a pilot trial. Int J Clin Pharm 2017 Aug;39(4):888-896 [FREE Full text] [doi: 10.1007/s11096-017-0495-6] [Medline: 28597176]
16. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [Medline: 12435261]
17. Riekert KA, Rand CS. Electronic Monitoring of Medication Adherence: When Is High-Tech Best? Journal of Clinical Psychology in Medical Settings 2002 Mar;9(1):25-34 [FREE Full text] [doi: 10.1023/A:1014131928789]
18. Chan AHY, Harrison J, Black PN, Mitchell EA, Foster JM. Using electronic monitoring devices to measure inhaler adherence: a practical guide for clinicians. J Allergy Clin Immunol Pract 2015;3(3):335-49.e1. [doi: 10.1016/j.jaip.2015.01.024] [Medline: 25840665]
19. Chan AHY, Reddel HK, Apter A, Eakin M, Riekert K, Foster JM. Adherence monitoring and e-health: how clinicians and researchers can use technology to promote inhaler adherence for asthma. J Allergy Clin Immunol Pract 2013;1(5):446-454. [doi: 10.1016/j.jaip.2013.06.015] [Medline: 24565615]
20. Simmons MS, Nides MA, Kleerup EC, Chapman KR, Milgrom H, Rand CS, et al. Validation of the Doser, a new device for monitoring metered-dose inhaler use. J Allergy Clin Immunol 1998 Sep;102(3):409-413. [Medline: 9768581]
21. Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. Clin Ther 1999 Jun;21(6):1074-90; discussion 1073. [doi: 10.1016/S0149-2918(99)80026-5] [Medline: 10440628]
22. Tashkin DP. Microelectronic adherence monitors for metered-dose inhalers: who monitors the monitors? J Allergy Clin Immunol 1999 Jul;104(1):22-24. [Medline: 10400834]
23. Julius SM, Sherman JM, Hendeles L. Accuracy of three electronic monitors for metered-dose inhalers. Chest 2002 Mar;121(3):871-876. [Medline: 11888975]
24. Wamboldt FS, Bender BG, O'Connor SL, Gavin LA, Wamboldt MZ, Milgrom H, et al. Reliability of the model MC-311 MDI chronolog. J Allergy Clin Immunol 1999 Jul;104(1):53-57. [Medline: 10400839]
25. Gradinarsky L, Lööf T. Inhalation Adherence Monitoring Using Smart Electronic Add-on Device - Accuracy evaluation using simulated real-life test program. 2014 Presented at: 4th International Conference on Wireless Mobile Communication and Healthcare; Nov 3-5, 2014; Athens, Greece URL: https://ieeexplore.ieee.org/document/7015930
26. Brueckner JW, Marshik P, Sherman J, Hendeles L. Reliability of the Medtrac MDI Chronolog. J Allergy Clin Immunol 1997 Oct;100(4):488-491. [Medline: 9338542]
27. Hardiker NR, Grant MJ. Factors that influence public engagement with eHealth: A literature review. Int J Med Inform 2011 Jan;80(1):1-12. [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
28. Grol R, Wensing M. Implementatie. Effectieve verbetering van de patiëntenzorg. Maarssen: Elsevier Gezondheidszorg; 2011.
29. Buurma H, Bouvy ML, De Smet PA, Floor-Schreudering A, Leufkens HG, Egberts AC. Prevalence and determinants of pharmacy shopping behaviour. J Clin Pharm Ther 2008;33(1):17-23. [Medline: 18211612]
30. World Health Organization Collaborating Centre for Drug Statistics Methodology. ATC/DDD Index 2019 URL: https://www.whocc.no/atc_ddd_index/ [accessed 2019-01-30] [WebCite Cache ID 75oQplXpI8]
31. Horne R, Weinman J. Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. J Psychosom Res 1999 Dec;47(6):555-567. [Medline: 10661603]
32. Horne R, Weinman J, Hankins M. The beliefs about medicines questionnaire: The development and evaluation of a new method for assessing the cognitive representation of medication. Psychology & Health 1999 Jan;14(1):1-24. [doi: 10.1080/08870449908407311]
33. Brooke J. SUS - A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London: Taylor and Francis; 1996:189-194.
34. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
35. Green J, Thorogood N. Qualitative Methods for Health Research. London: Sage Publications; 2014.
36. Menckeberg TT, Bouvy ML, Bracke M, Kaptein AA, Leufkens HG, Raaijmakers JA, et al. Beliefs about medicines predict refill adherence to inhaled corticosteroids. J Psychosom Res 2008 Jan;64(1):47-54. [doi: 10.1016/j.jpsychores.2007.07.016] [Medline: 18157999]
37. Patel M, Pilcher J, Travers J, Perrin K, Shaw D, Black P, et al. Use of metered-dose inhaler electronic monitoring in a real-world asthma randomized controlled trial. J Allergy Clin Immunol Pract 2013 Jan;1(1):83-91. [doi: 10.1016/j.jaip.2012.08.004] [Medline: 24229826]
38. Pilcher J, Shirtcliffe P, Patel M, McKinstry S, Cripps T, Weatherall M, et al. Three-month validation of a turbuhaler electronic monitoring device: implications for asthma clinical trial use. BMJ Open Respir Res 2015;2(1):e000097. [doi: 10.1136/bmjresp-2015-000097] [Medline: 26629345]
39. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]

40. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
41. Kelders SM, van Gemert-Pijnen JE, Werkman A, Seydel ER. Evaluation of a web-based lifestyle coach designed to maintain a healthy bodyweight. J Telemed Telecare 2010;16(1):3-7. [doi: 10.1258/jtt.2009.001003] [Medline: 20086259]
42. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May 01;4(3):666-684 [FREE Full text] [doi: 10.1177/193229681000400323] [Medline: 20513335]
43. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
44. Jentzsch NS, Camargos PAM. Methods of assessing adherence to inhaled corticosteroid therapy in children and adolescents: adherence rates and their implications for clinical practice. J Bras Pneumol 2008 Aug;34(8):614-621 [FREE Full text] [Medline: 18797747]

Abbreviations

BMQ: Beliefs About Medicines Questionnaire
EIMD: electronic inhalation monitoring device
PPV: positive predictive value
SUS: System Usability Scale

Edited by G Eysenbach; submitted 30.01.19; peer-reviewed by J Boven, C Li; comments to author 28.02.19; revised version received 20.03.19; accepted 23.03.19; published 30.05.19.

Please cite as:
Kuipers E, Poot CC, Wensing M, Chavannes NH, de Smet PAGM, Teichert M
Self-Management Maintenance Inhalation Therapy With eHealth (SELFIE): Observational Study on the Use of an Electronic Monitoring Device in Respiratory Patient Care and Research
J Med Internet Res 2019;21(5):e13551
URL: http://www.jmir.org/2019/5/e13551/
doi:10.2196/13551
PMID:31148542

Original Paper

OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes

Jan Walker[1,2*], RN, MBA; Suzanne Leveille[1,3*], RN, PhD; Sigall Bell[1,2], MD; Hannah Chimowitz[1], BA; Zhiyong Dong[1], MSc; Joann G Elmore[4], MD, MPH; Leonor Fernandez[1,2], MD; Alan Fossa[1], MPH; Macda Gerard[1], BA; Patricia Fitzgerald[1], MSc; Kendall Harcourt[1], MPA; Sara Jackson[5], MD, MPH; Thomas H Payne[5,6], MD; Jocelyn Perez[1], BSc; Hannah Shucard[5], MS; Rebecca Stametz[7], MPH, DEd; Catherine DesRoches[1,2], DrPH; Tom Delbanco[1,2], MD

[1]Division of General Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States
[2]Harvard Medical School, Boston, MA, United States
[3]College of Nursing and Health Sciences, University of Massachusetts, Boston, MA, United States
[4]David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States
[5]Division of General Internal Medicine, School of Medicine, University of Washington, Seattle, WA, United States
[6]Department of Medicine, Medicine Information Technology Services, University of Washington, Seattle, WA, United States
[7]Steele Institute for Health Innovation, Geisinger, Danville, PA, United States
[*]these authors contributed equally

Corresponding Author:
Jan Walker, RN, MBA
Division of General Medicine
Beth Israel Deaconess Medical Center
133 Brookline Avenue HV-2200
Boston, MA, 02215
United States
Phone: 1 517 975 7616
Email: jwalker1@bidmc.harvard.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/4/e18639/

Abstract

Background: Following a 2010-2011 pilot intervention in which a limited sample of primary care doctors offered their patients secure Web-based portal access to their office visit notes, the participating sites expanded OpenNotes to nearly all clinicians in primary care, medical, and surgical specialty practices.

Objective: The aim of this study was to examine the ongoing experiences and perceptions of patients who read ambulatory visit notes written by a broad range of doctors, nurses, and other clinicians.

Methods: A total of 3 large US health systems in Boston, Seattle, and rural Pennsylvania conducted a Web-based survey of adult patients who used portal accounts and had at least 1 visit note available in a recent 12-month period. The main outcome measures included patient-reported behaviors and their perceptions concerning benefits versus risks.

Results: Among 136,815 patients who received invitations, 21.68% (29,656/136,815) responded. Of the 28,782 patient respondents, 62.82% (18,081/28,782) were female, 72.90% (20,982/28,782) were aged 45 years or older, 76.94% (22,146/28,782) were white, and 14.30% (4115/28,782) reported fair or poor health. Among the 22,947 who reported reading 1 or more notes, 3 out of 4 reported reading them for 1 year or longer, half reported reading at least 4 notes, and 37.74% (8588/22,753) shared a note with someone else. Patients rated note reading as very important for helping take care of their health (16,354/22,520, 72.62%), feeling in control of their care (15,726/22,515, 69.85%), and remembering the plan of care (14,821/22,516, 65.82%). Few were very confused (737/22,304, 3.3%) or more worried (1078/22,303, 4.83%) after reading notes. About a third reported being encouraged by their clinicians to read notes and a third told their clinicians they had read them. Less educated, nonwhite, older, and Hispanic patients, and individuals who usually did not speak English at home, were those most likely to report major benefits from note reading. Nearly all respondents (22,593/22,947, 98.46%) thought Web-based access to visit notes a good idea, and 62.38% (13,427/21,525) rated this practice as very important for choosing a future provider.

Conclusions: In this first large-scale survey of patient experiences with a broad range of clinicians working in practices in which shared notes are well established, patients find note reading very important for their health management and share their notes frequently with others. Patients are rarely troubled by what they read, and those traditionally underserved in the United States report particular benefit. However, fewer than half of clinicians and patients actively address their shared notes during visits. As the practice continues to spread rapidly in the United States and internationally, our findings indicate that OpenNotes brings benefits to patients that largely outweigh the risks.

(J Med Internet Res 2019;21(5):e13876) doi:10.2196/13876

KEYWORDS

patient portal; physician-patient relations; electronic health record; health care survey; patient participation

Introduction

Secure Patient Portals

Patients who engage and participate actively in their health care appear to achieve better health outcomes and incur lower health care costs [1]. In part, to stimulate such engagement, the US federal government passed a legislation in 2009 incentivizing doctors and health systems to adopt new technologies offering patients electronic access to their health data via secure electronic patient portals [2]. Patients were invited to review their test results and problem and medication lists and, in many cases, to send secure messages to their clinicians. However, very few offered patients access to the notes written by their clinicians during or following face-to-face encounters.

The OpenNotes Initiative

In 2010, the Robert Wood Johnson Foundation funded the OpenNotes initiative, designed initially to examine the feasibility and effects of having primary care physicians (PCPs) share their notes routinely with patients. For 12 months, PCPs invited patients registered on portals to read these notes and 105 doctors and 5500 of their patients subsequently completed surveys. Representing 3 geographically dispersed and very different health care settings, the respondents were highly positive, with patients reporting a wide range of clinically important benefits and doctors noting little impact on their workflow [3]. Subsequent studies extending beyond primary care to medication adherence, inpatient care, oncology, mental illness, and other specialties suggested similar benefits, and in recent years, the practice of inviting patients to review their clinicians' notes has spread [4-13]. Presently, US clinicians offer more than 38 million patients electronic access to their notes through patient portals, policy makers are considering mandating such practice, and fully transparent records are spreading in several other nations [14-16]. Many prominent American institutions now offer open notes in the vast majority of their ambulatory practices, including medical and surgical offices and those focusing on mental illness. However, at this point, we know little about patient experiences with open notes over time throughout the broad spectrum of ambulatory care.

Following their initial limited experiment in primary care, the 3 institutions that originally piloted OpenNotes adopted the practice throughout, and many of their patients have grown accustomed to reading notes following visits to nearly all clinicians in their associated primary care, medical, surgical, and mental health practices. On the basis of the reports and observations over the past few years, we developed 4 primary hypotheses: (1) over time, patients would continue to report important benefits from reading visit notes; (2) patients would often share or discuss notes with others; (3) those traditionally at risk for experiencing substandard care would report the greatest benefit; and (4) patients and clinicians would communicate about these notes actively during visits. We report findings from a large survey of patients conducted in the institutions that participated in the pilot OpenNotes inquiry.

Methods

Setting

We conducted a Web-based survey of patients who had been seen in hospital offices and community practices at 3 health systems: Beth Israel Deaconess Medical Center (BIDMC), an urban academic health system in and around Boston; Geisinger, a large rural integrated health system in Pennsylvania; and University of Washington Medicine (UW) in Seattle, which includes both private and community-funded safety net practices affiliated with the University of Washington. All 3 systems participated in the original 2010-2011 OpenNotes pilot involving PCPs [3] and, by 2014, all 3 had expanded open notes to virtually all outpatient offices and clinicians, thereby providing access to visit notes in specialty as well as primary care settings. Open notes became the standard for virtually all types of outpatient clinicians who sign notes in the patient's medical record, including doctors, physical, occupational, speech, and other types of therapists, dieticians, nurses, nurse practitioners, and physician assistants. At the same time, the systems developed policies allowing some individual clinicians to opt out of participation and enabled participating clinicians to manually block the release of individual notes to the patient portal [17]. BIDMC and UW sent automated email messages informing patients when new notes were made available, but Geisinger did not [18]. At the time of the survey, somewhat fewer than half of all ambulatory patients were registered for the patient portal at each site (personal communications from Rebecca Stametz and Thomas Payne, July 24, 2018, and Amy Goldman, July 25, 2018).

Participants

The survey included patients seen in primary care and specialty offices: at the hospital and 6 affiliated sites at BIDMC, at 3 hospitals and 9 freestanding offices at UW, and at 7 hospitals and 53 outlying practices at Geisinger. Eligible patients were aged 18 years or older, had logged into the portal at least once in the previous 12 months, and had at least 1 ambulatory visit note available in the previous 12 months. We excluded patients who had been invited to participate in focus groups or other surveys related to OpenNotes within the preceding 12 months. Using portal tracking data, we identified patients who had, and had not, accessed available visit notes in the previous 12 months and described them as readers and nonreaders, respectively. We did not exclude nonreaders because we wanted to gain some understanding of why they had not read the notes. BIDMC and UW included all eligible readers and a random sample of eligible nonreaders in the survey sample. For administrative reasons, Geisinger drew random samples from both groups, resulting in smaller samples than the other 2 sites. Across the 3 sites, 109,904 readers and 27,959 nonreaders were sent invitations for the survey. The Institutional Review Boards at BIDMC, Geisinger, and UW approved the survey and study protocol at their respective sites.

Constructing the Questionnaire

This survey draws heavily on questions used in the original demonstration project in primary care; its development has been previously described [19]. We updated the questionnaire based on comments from outside reviewers, focus groups with diverse patients from a community health center, assessment of the distribution of responses in the original study, and an evaluation of psychometric properties of different versions of some of the items, including those related to benefits and risks of reading notes (manuscript under review). Except for a few site-specific modifications, the questionnaire was the same for all patients, and with the exception of skip patterns, free text, and demographic questions, all items required a response. Both the original and updated versions are available on request from the authors.

Conducting the Survey

We surveyed patients between June and October 2017, using a Web-based survey platform, Survey Gizmo. Patients were sent invitations by email either to their portal accounts or to the personal email address associated with the accounts. Each patient's invitation contained his or her study identification embedded in a unique link to the survey, and each study identification could be used only once. Following the original invitation, patients received 2 reminders 1 week apart if they had not completed the survey. Knowing that patients sometimes confuse notes with other parts of the medical record, or with the portal itself, we described clinical notes in both the survey invitation and in a survey question and we also showed a screenshot of the location of visit notes on their institutions' portals to increase the likelihood that patients would report on visit notes. Knowing also that invitations might be opened by care partners rather than patients, each respondent had an option to complete the survey as a patient or as a care partner, and care partners were automatically linked to a different questionnaire.

We offered respondents an incentive for completing the survey: a raffle of 50 prizes of US $25 or US $50 at each site.

Statistical Analysis

To maximize the chances that we were including responses about clinical notes rather than another part of the record, as a final step, we excluded responses from patients whose self-report of note reading in the past 12 months did not match portal data; for example, patients reported they had read notes, but the portal tracking data showed they had not. We also excluded respondents who reported reading notes for a week or less, or did not answer the question about length of time reading notes, since our objective was to assess patients' experiences over the prior 12 months. Except as noted in the tables, all questions included in this analysis had < 4% missing responses, and denominators include all nonmissing responses to each item. Items related to medication management are addressed in another paper [20].

Most items addressing potential benefits and risks asked for ratings on an 11-point scale, ranging from 0 (not at all important, confusing, or concerned) to 10 (extremely important, confusing, or concerned). Responses to these items were collapsed into 4 categories a priori for analysis: 0-1, 2-4, 5-7, and 8-10, and we reported the 8-10 category as *very important, confused, or concerned*. Four-level agreement responses were dichotomized as agree or somewhat agree, and disagree or somewhat disagree.

Using percentages and chi-square tests, we compared respondents with nonrespondents using administrative data available at each institution: age and sex at all 3 sites, Hispanic ethnicity at Geisinger and UW, and insurance type at BIDMC and Geisinger.

We used descriptive statistics to examine respondents' sociodemographic and health characteristics and experiences with note reading, both overall and stratified by study site. We used the chi-square test for independence (degrees of freedom: row-1 × column-1) to test for differences according to demographic characteristics in patients' experiences in note reading. We performed a multivariable analysis using log Poisson regression models to estimate overall relative risks and 95% CIs for reporting benefits as very important according to patient characteristics, adjusted for other sociodemographic and health characteristics and number of notes read. Owing to the large sample size, we expected even slight differences between groups to be statistically significant. Therefore, we interpreted intercategory differences in proportions or relative risks of 10% or more as meaningful differences. All analyses were completed using SAS software version 9.4 (SAS Institute Inc).

Results

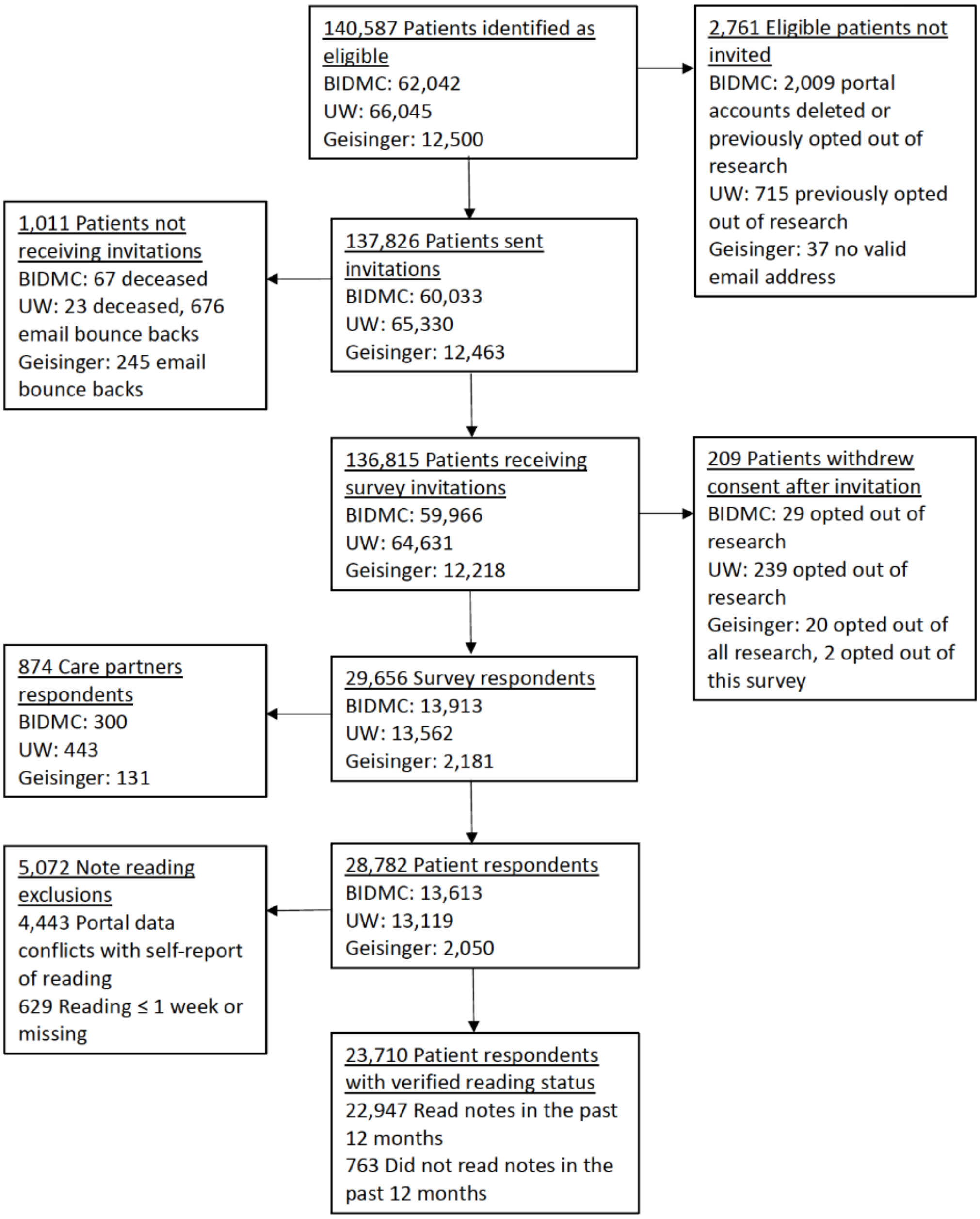
Participants

Of 136,815 patients who received survey invitations, 21.68% (29,656/136,815) responded; 28,782 were patients, and 874 were care partners (Figure 1). Compared with nonrespondents, respondents were older at all 3 sites. At BIDMC and UW, white patients were more likely to respond than nonwhite patients. No differences were noted between responders and nonresponders regarding sex at any of the 3 sites, regarding

Hispanic ethnicity at UW and Geisinger, and regarding insurance type at BIDMC and Geisinger (data not shown).

Of the 28,782 patient respondents, 62.82% (18,081/28,782) were female; 72.90% (20,982/28,782) were aged 45 years or older; 76.94% (22,146/28,782) were white; 53.06% (15,271/28,782) were employed; and 14.30% (4115/28,782) reported fair or poor general health. Overall, 64.92% (18,685/28,782) had completed college, but the sites differed substantially: the proportion of those with a high school education or less was 4.8% (655/13,613) at BIDMC and 5.4% (708/13,119) at UW, but 25.9% (530/2050) at Geisinger.

Figure 1. Study flow diagram. BIDMC: Beth Israel Deaconess Medical Center; UW: University of Washington Medicine.



Accessing and Reading Notes

After note reading exclusions, 23,710 responses were included in the analysis: 22,947 note readers and 763 nonreaders. Among note readers, three-quarters reported reading notes for a year or more and half reported reading 4 or more notes. In general, they reported locating notes easily on the portals and considered email notifications about new notes useful. About one-third of readers at BIDMC (4065/11,899, 34.16%) and UW (2998/9719, 30.85%), but only 20.1% (237/1178) at Geisinger, reported mentioning to their clinicians that they had read a note, and about the same proportion (7324/22,798, 32.13%) said they were encouraged by their clinicians to read notes. Among those who were encouraged, 60% read 4 or more notes compared with 46% of those who were not encouraged to read notes. In total, 37.74% (8588/22,753) reported sharing or discussing a note with a family member or someone else. Overall, 98.46% (22,593/22,947) of readers thought making notes available to patients a good idea, and 62.38% (13,427/21,525) said access to visit notes would be very important in choosing a future provider.

Benefits and Risks of Note Reading

As shown in Figure 2, between 50.43% (11,346/22,514) and 72.62% (16,354/22,520) of patients in all sites combined rated note reading as very important for helping them achieve 6 different benefits. In contrast, 3.3% (737/22,304) said they were very confused by their notes, 4.83% (1078/22,303) reported being more worried after reading notes, and 11.46% (2529/22,067) reported being very concerned about privacy. The positive perceptions of benefits were evident across the 3 sites, with patients in rural Pennsylvania reporting the highest ratings of importance of note reading.

We found meaningful differences (≥ 10%) in patients' perceptions of the benefits of reading notes according to sociodemographic characteristics (Table 1). Black patients were more likely than white patients to rate note reading as very important for 5 of the 6 benefits and patients aged 45 years or older rated it very important for 4 of the benefits compared with those aged 18 to 24 years. Those who usually spoke a language other than English at home were more likely than English speakers to use notes to make the most of visits, remember the plan of care, and prepare for visits. Patients with the fewest years of education and Hispanic patients were more likely than others to cite note reading as very important for remembering the care plan and preparing for visits. Patients who read a greater number of notes were more likely to cite reading notes as very important for all 6 benefits.

Few patients reported being very confused or more worried from reading notes, with only minor differences according to sociodemographic characteristics (Table 2). Though differences did not reach the 10% threshold, greater proportions of black patients, Asians, and other minorities reported being very concerned about privacy related to open notes, 16.2% (174/1076) to 19.2% (104/543), compared with 10.15% (1828/18,012) of white patients.

After adjusting for other factors, education level continued to be inversely associated with patient ratings of the importance of note reading for all 6 benefits (Table 3). Similarly, after adjusting for other characteristics, black patients were more likely than white patients to rate note reading as very important for achieving benefits. Older patients—especially those aged > 45 years—were more likely to rate notes as important for 4 of the benefits. Hispanic patients and those who spoke a language other than English at home found notes very important for 3 benefits. Notably, the importance of note reading in preparing for office visits was the benefit most endorsed among more vulnerable demographic groups.

When we asked the 763 nonreaders about the main reason they had not looked at visit notes, about half selected, *I forgot or did not know my visit notes were available*. A total of 10.4% (79/761) indicated they did not know they had a right to look at notes, 8.8% (67/761) were too busy, 7.2% (55/761) did not think reading would be useful, and 6.3% (48/761) were not able to find the notes. Among the 12.0% (95/761) reporting another reason, the majority wrote that they had no need to read notes because they trusted their clinicians, had received printed copies, or had no health issues or recent visits. Only 9.1% (69/760) of nonreaders reported being encouraged by a clinician to read their notes. We found no material differences in demographic characteristics in the nonreaders compared with the readers (data not shown). Even though they had not read notes, 89.0% (679/763) of nonreaders agreed that making notes available to patients on the Web is a good idea.

Figure 2. Benefits and risks of reading notes. BIDMC: Beth Israel Deaconess Medical Center; UW: University of Washington Medicine.

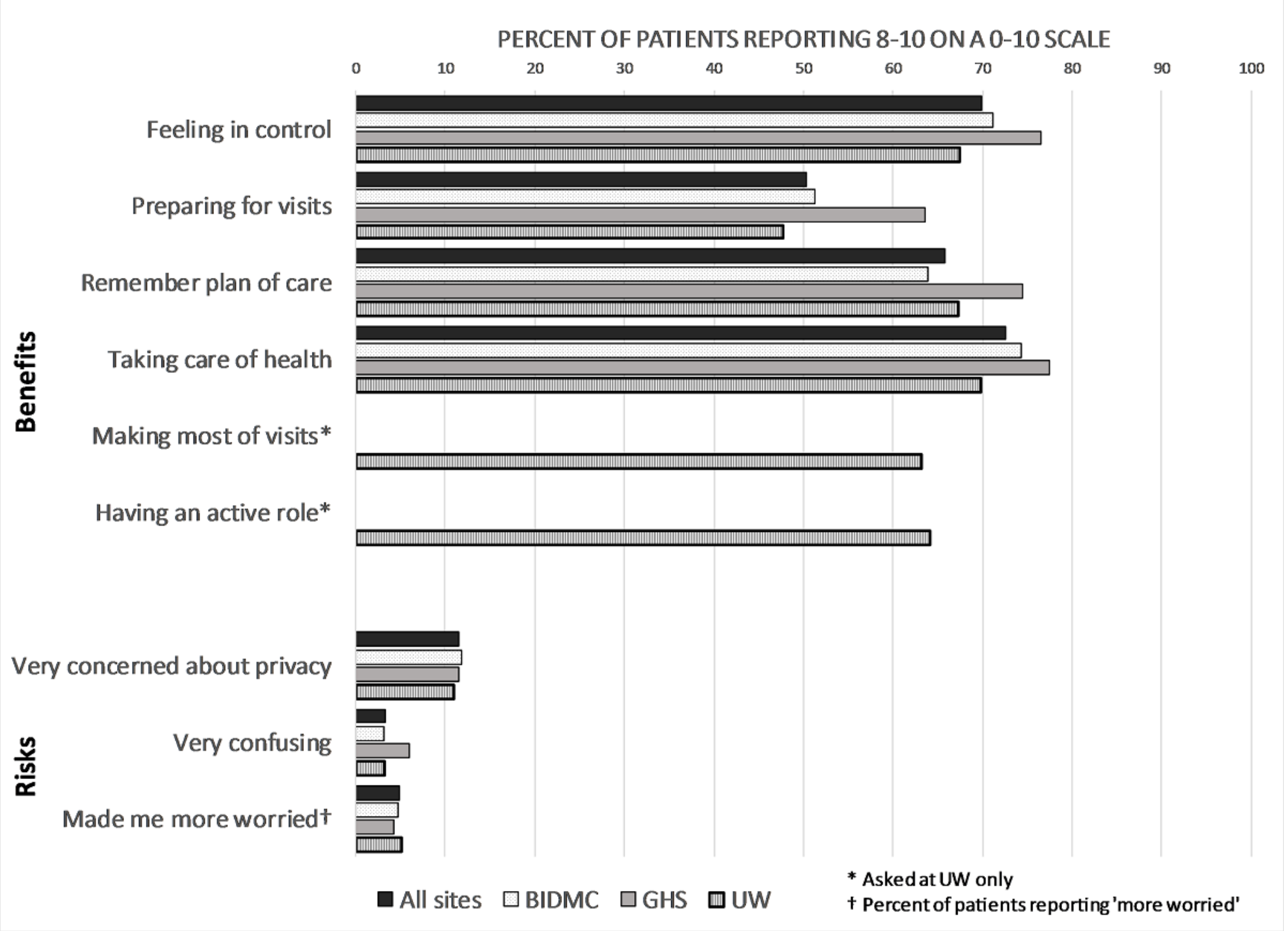


Table 1. Proportion of patients identifying notes as *extremely important* in achieving benefits.

Demographics	Taking care of health[a]		Active role in care[a,b]		Feeling in control of care[a]		Making the most of visits[a,b]		Remembering the plan of care[a]		Preparing for visits[a]	
	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]
Age (years)		**<.001**		**<.001**		**.007**		**<.001**		**.004**		**<.001**
18-24	436 (58.80)		233 (53.60)		498 (67.10)		227 (52.30)		473 (63.80)		303 (40.80)	
25-44	3358 (67.63)		1471 (62.38)		3452 (69.54)		1392 (59.06)		3263 (65.75)		2249 (45.32)	
45-64	7073 (75.47)		2453 (66.07)		6652 (70.99)		2411 (64.86)		6285 (67.07)		4893 (52.22)	
65+	5487 (73.74)		1977 (64.88)		5124 (68.88)		2003 (65.76)		4800 (64.52)		3901 (52.44)	
Sex		**.14**		**<.001**		**<.001**		**.06**		**<.001**		**.15**
Female	10,376 (72.96)		3954 (65.99)		10,212 (71.82)		3829 (63.87)		9604 (67.54)		7217 (50.76)	
Male	5978 (72.04)		2180 (61.22)		5514 (66.46)		2204 (61.93)		5217 (62.89)		4129 (49.76)	
Race		**<.001**		**.01**		**<.001**		**<.001**		**<.001**		**<.001**
White	13,012 (72.24)		4757 (64.03)		12,523 (69.53)		4653 (62.63)		11,726 (65.1)		8918 (49.51)	
Black	455 (83.80)		93 (73.00)		436 (80.30)		93 (73.00)		423 (77.90)		342 (63.00)	
Asian	824 (76.60)		433 (67.30)		797 (74.10)		443 (68.90)		777 (72.20)		597 (55.50)	
Other	535 (76.00)		236 (68.80)		522 (74.20)		233 (67.90)		511 (72.60)		424 (60.20)	
Multiple races	578 (75.70)		322 (69.00)		560 (73.30)		317 (67.90)		535 (70.00)		420 (55.00)	
Ethnicity		**<.001**		**.31**		**<.001**		**.54**		**<.001**		**<.001**
Hispanic or Latino	626 (80.50)		217 (67.40)		612 (78.70)		210 (65.20)		599 (77.00)		472 (60.70)	
Non-Hispanic	14,845 (72.69)		5641 (64.66)		14,282 (69.93)		5545 (63.56)		13,426 (65.74)		10,271 (50.29)	
Education		**<.001**		**<.001**		**<.001**		**<.001**		**<.001**		**<.001**
Masters or doctoral degree	5713 (71.33)		1958 (62.12)		5410 (67.55)		1923 (61.01)		4873 (60.84)		3700 (46.20)	
4-year college degree or some graduate school	5210 (71.33)		2052 (64.33)		5101 (69.84)		2018 (63.26)		4824 (66.05)		3526 (48.27)	
Some college or technical school	3591 (76.40)		1563 (68.07)		3450 (73.40)		1533 (66.77)		3413 (72.62)		2719 (57.85)	
High school or less	1113 (79.73)		335 (68.90)		106 (76.50)		336 (69.10)		1053 (75.43)		902 (64.60)	
Language		**<.001**		**.03**		**<.001**		**.002**		**<.001**		**<.001**
Other	372 (80.50)		142 (72.10)		361 (78.10)		146 (74.10)		357 (77.30)		315 (68.20)	
English	15,136 (72.84)		5728 (64.58)		14,567 (70.10)		5627 (63.44)		13,700 (65.93)		10,466 (50.36)	
General health		**.08**		**.19**		**.07**		**.97**		**<.001**		**<.001**

Demographics	Taking care of health[a]		Active role in care[a,b]		Feeling in control of care[a]		Making the most of visits[a,b]		Remembering the plan of care[a]		Preparing for visits[a]	
	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]
Excellent, very good, or good	13,134 (72.76)		4853 (65.09)		12,717 (70.45)		4752 (63.73)		11,853 (65.66)		9050 (50.14)	
Fair or poor	2452 (74.21)		1047 (63.38)		2275 (68.86)		1052 (63.68)		2283 (69.10)		1778 (53.81)	
Number of notes read		**<.001**		**<.001**		**<.001**		**<.001**		**<.001**		**<.001**
1	869 (54.80)		429 (51.40)		892 (56.20)		421 (50.40)		830 (52.30)		552 (34.80)	
2 or 3	5815 (66.86)		2244 (59.81)		5619 (64.64)		2176 (57.96)		5251 (60.40)		3780 (43.48)	
4 or more	9098 (80.56)		3223 (71.34)		867 6 (76.83)		3202 (70.89)		8194 (72.55)		6615 (58.59)	
Do not know or not sure	572 (60.70)		238 (53.00)		539 (57.20)		234 (52.20)		546 (57.90)		399 (42.30)	

[a] Answering 8 to 10 on a 0-10 scale.

[b] Only asked at University of Washington Medicine.

[c] *P* values from the chi-square test for independence between categorical measures.

Table 2. Proportions of patients reporting risks from reading notes.

Demographics	Notes were very confusing[a]		Very concerned about privacy[a]		More worried after reading notes[b]	
	n (%)	P value[c]	n (%)	P value[c]	n (%)	P value[c]
Age (years)		**<.001**		**<.001**		**<.001**
18-24	8 (1.00)		48 (7.00)		46 (6.00)	
25-44	112 (2.30)		433 (9.00)		256 (5.20)	
45-64	285 (3.10)		1284 (13.96)		471 (5.10)	
65+	332 (4.50)		764 (10.40)		305 (4.10)	
Sex		**.23**		**<.001**		**<.001**
Female	481 (3.40)		1488 (10.68)		721 (5.10)	
Male	256 (3.10)		1041 (12.81)		357 (4.40)	
Race		**.72**		**<.001**		**<.001**
White	570 (3.20)		1828 (10.10)		807 (4.50)	
Black	17 (3.10)		104 (19.20)		32 (6.00)	
Asian	42 (4.00)		174 (16.20)		65 (6.00)	
Other	25 (4.00)		129 (18.30)		50 (7.00)	
Multiple races	25 (3.00)		105 (13.70)		45 (6.00)	
Ethnicity		**.67**		**.11**		**.20**
Hispanic or Latino	23 (3.00)		101 (13.00)		43 (6.00)	
Non-Hispanic	659 (3.20)		2280 (11.16)		962 (4.70)	
Education		**<.001**		**.04**		**.03**
Master’s or doctoral degree	218 (2.70)		891 (11.10)		330 (4.10)	
4-year college degree or some graduate school	208 (2.90)		796 (10.90)		359 (4.90)	
Some college or technical school	183 (3.90)		585 (12.50)		245 (5.20)	
High school or less	78 (6.00)		149 (10.70)		82 (6.00)	
Language		**.10**		**.03**		**.001**
Other	21 (5.00)		67 (15.00)		29 (6.00)	
English	664 (3.20)		2327 (11.20)		978 (4.70)	
General health		**<.001**		**.001**		**<.001**
Excellent, very good, or good	547 (3.00)		1996 (11.06)		710 (3.90)	
Fair or poor	138 (4.20)		429 (13.00)		301 (9.10)	
Number of notes read		**<.001**		**<.001**		**<.001**
1	33 (2.00)		165 (10.70)		51 (3.00)	
2 or 3	239 (2.80)		904 (10.60)		369 (4.30)	
4 or more	420 (3.80)		1331 (12.02)		614 (5.50)	
Do not know or not sure	45 (5.00)		129 (14.50)		44 (5.00)	

[a] Proportion answering 8 to 10 on a 0-10 scale.

[b] Proportion answering *more worried*.

[c] *P* values from chi-square test of independence between categorical measures.

Table 3. Adjusted relative risk and 95% CI of identifying notes as very important for potential benefits.

Demographics	Taking care of health[a] , risk ratio (95% CI)	Having an active role in care[a], risk ratio (95% CI)	Feeling in control of care[a], risk ratio (95% CI)	Making the most of visits[a], risk ratio (95% CI)	Remembering the plan of care[a], risk ratio (95% CI)	Preparing for office visits[a], risk ratio (95% CI)
Age (years)[b,c]						
18-24	1.00 (ref[d])	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
25-44	1.14 (1.07-1.22)	1.18 (1.07-1.29)	1.04 (0.99-1.10)	1.16 (1.05-1.28)	1.05 (0.99-1.12)	1.14 (1.03-1.25)
45-64	1.27 (1.20-1.36)	1.24 (1.13-1.36)	1.06 (1.01-1.12)	1.28 (1.16-1.40)	1.07 (1.01-1.14)	1.30 (1.18-1.42)
65+	1.25 (1.18-1.33)	1.25 (1.14-1.38)	1.05 (0.99-1.11)	1.32 (1.2-1.46)	1.05 (0.99-1.11)	1.33 (1.21-1.45)
Sex[b,c]						
Male	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Female	1.02 (1.00-1.04)	1.08 (1.04-1.11)	1.07 (1.05-1.09)	1.04 (1.01-1.08)	1.06 (1.04-1.08)	1.02 (1.00-1.05)
Race[b,c]						
White	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (rref)	1.00 (ref)
Black	1.13 (1.09-1.18)	1.11 (1.00-1.24)	1.11 (1.07-1.16)	1.14 (1.03-1.27)	1.14 (1.08-1.19)	1.22 (1.14-1.30)
Asian	1.11 (1.07-1.15)	1.08 (1.02-1.15)	1.07 (1.03-1.11)	1.15 (1.08-1.22)	1.12 (1.08-1.17)	1.18 (1.11-1.25)
Multiple races	1.05 (1.01-1.10)	1.07 (1.01-1.14)	1.04 (0.99-1.08)	1.10 (1.03-1.17)	1.04 (0.99-1.09)	1.10 (1.03-1.18)
Other	1.02 (0.98-1.07)	1.05 (0.98-1.14)	1.04 (0.99-1.09)	1.06 (0.98-1.15)	1.05 (1.00-1.10)	1.14 (1.07-1.22)
Education[b,c]						
Master's or doctoral degree	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
4-year degree or some grad school	1.01 (0.99-1.03)	1.05 (1.01-1.09)	1.03 (1.01-1.06)	1.05 (1.01-1.09)	1.08 (1.06-1.11)	1.06 (1.02-1.09)
Some college or technical school	1.06 (1.04-1.08)	1.09 (1.05-1.14)	1.08 (1.06-1.11)	1.09 (1.05-1.14)	1.17 (1.14-1.20)	1.23 (1.18-1.27)
High school or less	1.11 (1.08-1.14)	1.12 (1.05-1.20)	1.15 (1.11-1.18)	1.14 (1.07-1.22)	1.23 (1.19-1.27)	1.37 (1.31-1.44)
Ethnicity[b,c]						
Non-Hispanic	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Hispanic or Latino	1.11 (1.07-1.16)	1.02 (0.94-1.11)	1.08 (1.04-1.13)	1.02 (0.94-1.11)	1.11 (1.07-1.16)	1.14 (1.07-1.22)
General health[b,c]						
Excellent or very good	1.0 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Good	0.95 (0.93-0.98)	0.91 (0.87-0.95)	0.91 (0.88-0.93)	0.92 (0.88-0.96)	0.99 (0.96-1.02)	0.95 (0.92-0.99)
Fair or poor	0.98 (0.96-1.00)	0.96 (0.93-1.00)	0.95 (0.93-0.96)	0.95 (0.92-0.98)	1.01 (0.98-1.03)	0.99 (0.96-1.02)
Primary language[b,c]						
English	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Other	1.07 (1.02-1.13)	1.09 (0.99-1.20)	1.09 (1.04-1.15)	1.12 (1.03-1.23)	1.11 (1.06-1.18)	1.28 (1.19-1.37)
Number of notes read[b,c]						
1	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
2-3	1.20 (1.14-1.26)	1.16 (1.08-1.25)	1.15 (1.09-1.20)	1.15 (1.07-1.24)	1.14 (1.09-1.20)	1.23 (1.14-1.32)
4+	1.44 (1.37-1.51)	1.38 (1.29-1.48)	1.37 (1.31-1.43)	1.40 (1.30-1.51)	1.36 (1.30-1.43)	1.62 (1.51-1.74)
Do not know	1.08 (1.01-1.16)	1.03 (0.92-1.15)	1.02 (0.95-1.10)	1.04 (0.93-1.17)	1.07 (0.99-1.15)	1.15 (1.03-1.28)

[a] Answering 8 to 10 on a 0-10 scale.

[b] Obtained from log Poisson regression adjusted for age, sex, race, education, ethnicity, language, self-reported health, and number of notes read.

[c] 22,947 patients included in the model with no missing data on the dependent or independent variable.

[d] ref: reference group.

Discussion

Principal Findings

In this study, the largest assessment to date of patients who read a wide variety of clinicians' notes over time, patients report that reading clinical notes brings them substantial benefit. The respondents represent urban and rural settings, varying education levels, and broad age and racial distributions, and they had access to notes composed by most clinicians providing primary care and ambulatory care in medical, surgical, and mental health specialties. About two-thirds of patients describe notes as extremely important in increasing their sense of control, improving recall and understanding of their plans for care, and better preparing them for visits. Few reported confusion or increased worries, and those patients from medically underserved groups reported the most benefit. Whether or not they chose to review their notes, patients overwhelmingly approved the practice, with a majority reporting that access to notes would be extremely important in determining their future choice of clinicians. On the contrary, only a third of patients recalled discussing their notes during visits or having their clinicians recommend that they read them.

Comparison With Other Studies

These reports from almost 23,000 patients who read their notes amplify and reinforce the findings from smaller, more targeted, and shorter studies within primary care practices and discrete clinical specialties [5-7,9,12,13,21]. Many respondents had read notes over several years, and these results may foretell a *steady state* of patient experiences over the longer term. The high proportion of patients reporting benefits from reading notes did not diminish compared with our first survey in primary care patients several years ago, and more patients reported mentioning their note reading to their providers (32%, up from 19%) [3].

Implications of the Findings

Focusing on clinically important process measures, these results strongly suggest that transparency helps patients feel more engaged in their care. This is an important finding, given that a growing body of evidence indicates that engaged patients are more likely to adhere to treatment plans and medications, follow through on screening and prevention protocols, detect and prevent errors, and adopt more effective management strategies for chronic illnesses [22-26].

Although the time-honored principle of patient-clinician confidentiality is not in itself affected by open notes, it is up to patients to decide whether or not to disclose their medical information to others. More than a third of respondents reported sharing notes with someone else, almost twice the rate reported in 2012 in the study of primary care patients in the same institutions [3]. Shared notes may be particularly helpful to informal care partners [27,28] and to those in search of informal second opinions from both lay and professional associates.

In this survey, patients who are potentially the most vulnerable—those who are older, less educated, non-white, Hispanic, or not English speakers at home—reported the most benefit. By virtue, both of taking the initiative to sign onto a patient portal and reading their notes, they may not be representative of other Americans with similar demographics, as is true for everyone in our sample. But this finding extends and amplifies findings in earlier inquiries [5-7,9,10,12,13,21]. Vulnerable patients may come to a visit with a lower baseline sense of control and knowledge than others; they may also have more difficulty understanding or retaining what practitioners say or emphasize. Assisted at times by family members or other acquaintances who can help interpret and research points made in notes, the possibility of review in their homes may contribute to their particular enthusiasm for this new opportunity. Moreover, in several studies, patients who are disadvantaged may trust their health professionals less than those with social and educational backgrounds similar to their clinicians [29]. Shared notes can increase trust, as studies and anecdotes from both patients and clinicians suggest [24,26]. That particularly vulnerable patients may gain the most from open notes is worthy of further inquiry. Similarly, the fact that minority patients are more concerned about privacy than white patients also deserves further study.

As some of our findings suggest, embedding open notes into clinical practice faces many challenges. Even in these 3 *mature* institutions, the majority of their patients have not registered on their patient portals, and half of the responding nonreaders did not know that notes were available—even though they were using their portals for other purposes. Moreover, contrary to our predictions, interchange in the clinician's office about past notes was infrequent. Here, there were also substantial differences across the sites. Patients from rural Pennsylvania were less likely than those in Boston and Seattle to report speaking with their providers about note reading. However, they were also more likely to report benefits from reading, including greater feelings of control of their care and better preparation for office visits. Access to clinicians' notes may offer particular benefits to rural populations, a possibility that warrants further study.

Strengths and Limitations

This is a cross-sectional study examining patients' self-reported experiences from only 3 regions of the United States, and results may not be generalizable to other regions or practices. Moreover, the response rate was modest, although it was similar to the response rate in a recent Consumer Assessment of Healthcare Provider and Systems survey [30]. It is possible that the survey respondents were those most enthusiastic about open notes. Although there were some demographic differences between responders and nonresponders, these differences were small, and the size of the sample analyzed gives further weight to the findings. It should be noted also that the majority of respondents were white, in good health, and highly educated. It is difficult to draw firm conclusions about the benefits to non-English speakers because the survey was only offered in English; subsequent surveys could be administered in other languages and also explore issues of health literacy. Finally, the literature on the impact of portals on patients is often confused by lack of specificity about different functionalities. A strength of our study is that we took several steps to make sure patients had experience reading notes and were reporting on reading visit notes, rather than on other information available on their portals.

Unanswered Questions and Future Directions

Some argue today that fully open medical records are simply *...the right thing to do*, and this large survey of patients furnishes further evidence that their benefits outweigh their risks, certainly from the point of view of patients. Although the findings confirm that the benefits of note reading extend beyond primary care practices to virtually all specialties and types of clinicians, more needs to be learned about using open notes as a tool for communication and promoting interaction between patients and clinicians across health care venues and populations. The Department of Health and Human Services has proposed new rules that would increase patient access and control of their medical information, and easy electronic access to notes could become the law of the land [31]. To what degree can active educational interventions help patients learn optimally from what their records document? And in the future, might patients also contribute to their records by providing interval histories and articulating their goals for a visit in their own words, thereby enriching narratives, promoting focused interactions, and hopefully off-loading work from beleaguered clinicians? [24].

At a time when medical practice in the United States is moving toward shared notes as a new standard, the patients participating in this study offer both affirmative and provocative reports, but building shared notes into the fabric of care remains a work in progress.

Acknowledgments

The authors thank the Robert Wood Johnson Foundation, Gordon and Betty Moore Foundation, Peterson Center on Healthcare, and Cambia Health Foundation for their financial support, as well as James Ralston, MD, MPH, Lee Hargraves, PhD, and Julie A Wright, PhD, for insightful comments on drafts of the questionnaire. The authors are also grateful to the patients for their input and responses and to all the colleagues on the OpenNotes team.

Conflicts of Interest

All authors had financial support for the submitted work from the Robert Wood Johnson Foundation, Gordon and Betty Moore Foundation, Peterson Center on Healthcare, and Cambia Health Foundation. The funders had no role in designing or conducting the study, analyzing the data, preparing the manuscript, or deciding to submit this manuscript for publication.

References

1. Hibbard J. Patient activation and the use of information to support informed health decisions. Patient Educ Couns 2017 Dec;100(1):5-7. [doi: 10.1016/j.pec.2016.07.006] [Medline: 27432014]
2. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
3. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
4. Dobscha S, Denneson L, Jacobson L, Williams H, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
5. Kayastha N, Pollak KI, LeBlanc TW. Open notes: a qualitative study of oncology patients' experiences reading their cancer care notes. J Clin Oncol 2017 Nov;35(31_suppl):33-33. [doi: 10.1200/JCO.2017.35.31_suppl.33]
6. Peck P, Torous J, Shanahan M, Fossa A, Greenberg W. Patient access to electronic psychiatric records: a pilot study. Health Pol Technol 2017 Sep;6(3):309-315. [doi: 10.1016/j.hlpt.2017.06.003]
7. Weinert C. Giving doctors' daily progress notes to hospitalized patients and families to improve patient experience. Am J Med Qual 2017 Jul;32(1):58-65. [doi: 10.1177/1062860615610424] [Medline: 26453323]
8. Lee CI, Langlotz CP, Elmore JG. Implications of direct patient online access to radiology reports through patient web portals. J Am Coll Radiol 2016 Dec;13(12 Pt B):1608-1614. [doi: 10.1016/j.jacr.2016.09.007] [Medline: 27888949]
9. Oster NV, Jackson SL, Dhanireddy S, Mejilla R, Ralston JD, Leveille S, et al. Patient access to online visit notes: perceptions of doctors and patients at an urban HIV/AIDS clinic. J Int Assoc Provid AIDS Care 2015 Jul;14(4):306-312 [FREE Full text] [doi: 10.1177/2325957414526783] [Medline: 24729072]
10. Wright E, Darer J, Tang X, Thompson J, Tusing L, Fossa A, et al. Sharing physician notes through an electronic portal is associated with improved medication adherence: quasi-experimental study. J Med Internet Res 2015 Oct 8;17(10):e226 [FREE Full text] [doi: 10.2196/jmir.4872] [Medline: 26449757]

11. Lee BS, Oster NV, Chen GY, Ding LL, Walker JD, Elmore JG. Ophthalmology patients' interest in online access to clinic notes at three US clinics. Ophthalmic Physiol Opt 2017 Dec;37(4):420-427. [doi: 10.1111/opo.12367] [Medline: 28417473]
12. Mishra V, Hoyt R, Wolver S, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with OpenNotes. Appl Clin Inform 2019 Jan;10(1):10-18. [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
13. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Informatics Assoc 2014 Oct 28:380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
14. Walker J, Meltsner M, Delbanco T. US experience with doctors and patients sharing clinical notes. Br Med J 2015 Feb 10;350:g7785. [doi: 10.1136/bmj.g7785] [Medline: 25670182]
15. Moll J, Rexhepi H, Cajander Å, Grünloh C, Lic T, Huvila I, et al. Patients’ Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden. J Med Internet Res [Internet]; 2018;20. [doi: 10.2196/jmir.949]
16. OpenNotes. 2018. URL: https://www.opennotes.org/ [accessed 2019-04-08] [WebCite Cache ID 77U0ukbna]
17. Payne TH, Beahan S, Fellner J, Martin D, Elmore JG. Health records all access pass. Patient Portals that allow viewing of clinical notes and hospital discharge summaries: the University of Washington OpenNotes implementation experience. J AHIMA 2016 Dec;87(8):36-39. [Medline: 29425012]
18. Mafi JN, Mejilla R, Feldman H, Ngo L, Delbanco T, Darer J, et al. Patients learning to read their doctors' notes: the importance of reminders. J Am Med Inform Assoc 2016 Sep;23(5):951-955. [doi: 10.1093/jamia/ocv167] [Medline: 26911830]
19. Leveille S, Walker J, Ralston J, Ross S, Elmore J, Delbanco T. Evaluating the impact of patients' online access to doctors' visit notes: designing and executing the OpenNotes project. BMC Med Inform Decis Mak 2012 Apr 13;12:32 [FREE Full text] [doi: 10.1186/1472-6947-12-32] [Medline: 22500560]
20. DesRoches C, Bell S, Dong Z, Elmore J, Fernandez L, Fitzgerald P, et al. Patients managing medications and reading their visit notes: A survey of OpenNotes participants. Ann Intern Med 2019:1-1 (forthcoming).
21. Cromer R, Denneson L, Pisciotta M, Williams H, Woods S, Dobscha S. Trust in mental health clinicians among patients who access clinical notes online. Psychiatr Serv 2017;68(5):520-523. [doi: 10.1176/appi.ps.2016001685]
22. Náfrádi L, Nakamoto K, Schulz PJ. Is patient empowerment the key to promote adherence? A systematic review of the relationship between self-efficacy, health locus of control and medication adherence. PLoS One 2017 Oct;12(10):e0186458 [FREE Full text] [doi: 10.1371/journal.pone.0186458] [Medline: 29040335]
23. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [doi: 10.2307/3350391] [Medline: 8941260]
24. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2016 May 18:511-544. [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
25. Bell SK, Folcarelli PH, Anselmo MK, Crotty BH, Flier LA, Walker J. Connecting patients and clinicians: the anticipated effects of open notes on patient safety and quality of care. Jt Comm J Qual Patient Saf 2015 Aug;41(8):378-384. [doi: 10.1016/S1553-7250(15)41049-9] [Medline: 26215527]
26. Bell SK, Folcarelli P, Fossa A, Gerard M, Harper M, Leveille S, et al. Tackling ambulatory safety risks through patient engagement: what 10,000 patients and families say about safety-related knowledge, behaviors, and attitudes after reading visit notes. J Patient Saf 2018 Apr 27:-. [doi: 10.1097/PTS.0000000000000494] [Medline: 29781979]
27. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2016 Aug 7;24(e1):e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
28. Jackson S, Mejilla R, Darer J, Oster NV, Ralston JD, Leveille SG, et al. Patients who share transparent visit notes with others: characteristics, risks, and benefits. J Med Internet Res 2014 Nov 12;16(11):1-10 [FREE Full text] [doi: 10.2196/jmir.3363] [Medline: 25405911]
29. Street RL, O'Malley KJ, Cooper LA, Haidet P. Understanding concordance in patient-physician relationships: personal and ethnic dimensions of shared identity. Ann Fam Med 2008 May;6(3):198-205 [FREE Full text] [doi: 10.1370/afm.821] [Medline: 18474881]
30. Medicaid. 2015. Methodology Report Nationwide CAHPS Surveys of Adults Enrolled in Medicaid between October and December 2013 URL: https://www.medicaid.gov/medicaid/quality-of-care/downloads/performance-measurement/methodology-report.pdf [accessed 2019-02-05] [WebCite Cache ID 75xwSsaeb]
31. Department of Health and Human Services. HealthIT.gov. 2019. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certifaction Program URL: https://www.healthit.gov/sites/default/files/page/2019-02/HITACNPRMPresentation.pdf [accessed 2019-04-12] [WebCite Cache ID 77ZYSA5h8]

Abbreviations

BIDMC: Beth Israel Deaconess Medical Center

PCP: primary care physician
UW: University of Washington Medicine

Edited by G Eysenbach; submitted 01.03.19; peer-reviewed by S Wolver, D Klein; comments to author 20.03.19; revised version received 01.04.19; accepted 02.04.19; published 06.05.19.

Please cite as:
Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, Fernandez L, Fossa A, Gerard M, Fitzgerald P, Harcourt K, Jackson S, Payne TH, Perez J, Shucard H, Stametz R, DesRoches C, Delbanco T
OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes
J Med Internet Res 2019;21(5):e13876
URL: https://www.jmir.org/2019/5/e13876/
doi:10.2196/13876
PMID:31066717

Original Paper

A Novel Instrument for Measuring Older People's Attitudes Toward Technology (TechPH): Development and Validation

Peter Anderberg[1], PhD; Shahryar Eivazzadeh[1], MSc; Johan Sanmartin Berglund[1], MD, PhD

Department of Health, Blekinge Institute of Technology, Karlskrona, Sweden

Corresponding Author:
Peter Anderberg, PhD
Department of Health
Blekinge Institute of Technology
Valhallavägen 1
Karlskrona, 37139
Sweden
Phone: 46 734223736
Email: peter.anderberg@bth.se

Abstract

Background: The use of health technology by older people is coming increasingly in focus with the demographic changes. Health information technology is generally perceived as an important factor in enabling increased quality of life and reducing the cost of care for this group. Age-appropriate design and facilitation of technology adoption are important to ensure functionality and removal of various barriers to usage. Development of assessment tools and instruments for evaluating older persons' technology adoption and usage as well as measuring the effects of the interventions are of high priority. Both usability and acceptance of a specific technology or service are important factors in evaluating the impact of a health information technology intervention. Psychometric measures are seldom included in evaluations of health technology. However, basic attitudes and sentiments toward technology (eg, technophilia) could be argued to influence both the level of satisfaction with the technology itself as well as the perception of the health intervention outcome.

Objective: The purpose of this study is to develop a reduced and refined instrument for measuring older people's attitudes and enthusiasm for technology based on relevant existing instruments for measuring technophilia. A requirement of the new instrument is that it should be short and simple to make it usable for evaluation of health technology for older people.

Methods: Initial items for the TechPH questionnaire were drawn from a content analysis of relevant existing technophilia measure instruments. An exploratory factor analysis was conducted in a random selection of persons aged 65 years or older (N=374) on eight initial items. The scale was reduced to six items, and the internal consistency and reliability of the scale were examined. Further validation was made by a confirmatory factor analysis (CFA).

Results: The exploratory factor analysis resulted in two factors. These factors were analyzed and labeled techEnthusiasm and techAnxiety. They demonstrated relatively good internal consistency (Cronbach alpha=.72 and .68, respectively). The factors were confirmed in the CFA and showed good model fit (χ^2_8=21.2, χ^2/df=2.65, comparative fit index=0.97, adjusted goodness-of-fit index=0.95, root mean square error of approximation=0.067, standardized root mean square residual=0.036).

Conclusions: The construed TechPH score showed expected relations to external real-world criteria, and the two factors showed interesting internal relations. Different technophilia personality traits distinguish clusters with different behaviors of adaptation as well as usage of new technology. Whether there is an independent association with the TechPH score against outcomes in health technology projects needs to be shown in further studies. The instrument must also be validated in different contexts, such as other countries.

(J Med Internet Res 2019;21(5):e13951) doi:10.2196/13951

KEYWORDS

technophilia; aging; internet; health technology; eHealth

Introduction

Background

Older people's use of technology is increasingly coming in focus with the demographic changes. Gerontechnology (technology for the aging population) is a growing field in transdisciplinary research as well as in the development of new products [1,2]. Previous research into older people's technology use has identified that the design [3,4] and technology adoption [5,6] perspectives are important to ensure appropriate functionality and remove various barriers for use. Personal factors, such as self-efficacy and proficiency [5,7,8], as well as subjective technology adaptivity [9,10] have also been identified as significant predictors of technology use in old age.

One area that has attracted interest is the use of technology by older adults in various health settings, both in formal health care [11,12] and from salutogenic perspectives [13,14] targeting social isolation and participation. However, there is a strong need to find evidence of the effectiveness and efficiency of health technology interventions, which is becoming increasingly important as the number of available health technology solutions grows [15].

Several instruments for evaluating interactions with health technology exist today. Common theoretical concepts addressed in those instruments are effectiveness, efficiency, hardware/software, perceived ease of use, and satisfaction, although validated psychometric instruments measuring personality traits are sparse [16].

One of the most widely used instruments is the technology acceptance model (TAM) [17] and its subsequent developments [18] for evaluating attitudes predicting intentions to use and how users come to accept and use a particular information technology. In this model, two specific factors determine the user's acceptance: perceived usefulness and perceived ease of use. Although TAM was not developed specifically for health information technology (HIT), it has found its way into this area as a measurement of end users' reactions to HIT [19]. Expansions to adapt TAM to a more specific HIT context have been made: the health information technology acceptance model (HITAM) [20]. Application and problematization of TAM toward older persons have also been made: the senior technology acceptance model [21] and HITAM of older persons [22]. The importance of contextual factors [23,24], as well as the usability and acceptability for older adults with mild cognitive impairment and dementia [25] have recently been in focus.

Another widely used model is the Delone and McLean information systems (IS) success model, which considers constructs of intention to use, user satisfaction, and net benefits as the outcomes of three sets of indicators that are information quality, service quality, and system quality [26]. The Delone and McLean IS success model and its extended variants, such as human, organization, and technology-fit [27], have been widely used in health technology [28]. How traits affected by or closely related to technophobia, such as technostress [29], can impact the satisfaction construct in the Delone and McLean IS model has been discussed [30].

Usability is another essential aspect of HIT evaluation. Both the design and evaluation of artifact, efficiency, effectiveness, and satisfaction are important. The Health Information Technology Usability Evaluation Scale [31] and the System Usability Scale [32] are instruments for usability evaluation that have been used in a HIT context [33,34]. In what is an extension of the TAM perspective, Kamin and Lang [10] explored the motivational resources for older persons' technology use by the concept of subjective personal adaptivity. They found that positive beliefs about the benefits of technology, the time and effort invested to learn how to use technology, and a sense of trustworthiness and safety while using technology is connected to both perceived technology competence and technology usage.

Usability, acceptance, motivation, and adoption of a specific technology or service are important factors in evaluating the impact of a HIT intervention. However, basic attitudes and sentiments toward technology (eg, technophilia) could be argued to influence both the level of satisfaction with the technology itself as well as the perception of the health intervention outcome. This would constitute a personality trait, an underlying factor that would create a preintervention entry level of acceptance and interest, positive or negative. Edison and Geissler [35] argue that this "affinity" toward technology is a more general attitude that precedes the more specific attitudes resulting from the rational (reasoned or planned) process that is measured in TAM. Plociennik et al [36] found that technophilia has a direct influence on perceived usefulness in TAM. Ronit [37] sees technophilia as the enthusiasm toward technology with its rewarded and knowledgeable adoption correlated with both perceived usefulness and perceived ease of use. Kamin and Lang [9,10] showed a correlation between usability and utility and a subjectively perceived interest and competence.

Technophilia

Technophilia has no universally established definition but generally refers to a strong enthusiasm and love for modern technology. Seebauer et al [38] define it as "an attitude toward ICT [internet and communication technology], representing a subaspect of technology-related values, just as ICT are a subcategory of modern technology." Osiceanu [39] defines technophilia as an "attraction, enthusiasm of the human individual determined by the activities which involve the use of advanced technologies. It is expressed by easy adaptation to the social changes brought by technological innovations." Martínez-Córcoles et al [40] suggest that merely enthusiasm and desire is not enough for technophilia, but also an acquired need for (dependency) and joy of having and displaying the latest products/versions (technoreputation). In a working paper, Li and Fuller [41] suggest a definition of technophilia as "positive affective states that arise momentarily in response to an individual's ICT context that is appraised to exceed his or her expectations and goals."

In their review of previous related research on technophilia and similar concepts, Seebauer et al [38] found three hierarchically nested perspectives on technophilia. At the top level, they found values connected to general technology-related values reflecting global beliefs in societal progress through technology. At the intermediate level were attitudes referring more specifically to ICT as a part of modern technologies. The lowest level was constituted by a keen interest in and use of a specific technology or service as a subcategory of ICT. Openness toward technology and innovation influences personal dedication to certain technological artifacts and services, whereas feelings of low enthusiasm may work in the opposite direction [38]. Donat et al [42] refer to this lower (negative) end of technology enthusiasm as *technophobia*, and their construct is made up from two opposite ends where technophilia is at the positive end. Nimrod [43] includes this positive end of the spectrum in her construct of a technophobia scale while investigating ICT use among older adults. Osiceanu [39] further views technophobia as the negative feelings about using technology, but also includes perceptions on the adverse effects that technology may have on society in the construct. This is a common way to conceptualize technophobia construct with both factors concerning personal feelings toward technology use together with overall perceptions of technology in society.

In this study, we choose to avoid the complexity of bringing the two concepts of technophilia and technophobia together and simply let technophilia refer to a person’s enthusiasm for and positive feelings toward their technology use and absence of the fears and doubts some older people could have about their ability to manage using new technology. It constitutes a personality trait, an underlying psychological construct [38] that would create a pretechnology and preintervention entry level of acceptance and interest. This could also be connected to other psychological characteristics (ie, the personality traits of openness and neuroticism) [44].

Steinerman et al [45] report that technophilia is a consistent predictor of openness to research participation for older adults and that research with older adults that incorporates technology should consider technophilia to be more successful in recruiting participants for the study.

Older people compared to younger persons are sometimes reported to score lower on technophilia scales [38]. Still, older people are not a homogeneous group, and there may be differences between “younger” old and “older” old individuals. Therefore, it is of importance to create an instrument with the ability to discriminate between high and low technophilia individuals within various groups of older people.

Objectives

The purpose of this study is to develop a reduced and refined instrument for measuring older people’s attitudes and enthusiasm for technology based on relevant existing instruments for measuring technophilia (named TechPH, short for technophilia).

Technophilia is a general quality for any individual’s relationship to technology that could potentially influence a wide range of aspects of technology use, such as adoption, continuity, and perceived outcome.

In this text, we contextualize this general quality with a focus on older people for use in health technology intervention research as a complement to existing instruments. The new instrument requires that it should be short and simple to make it usable for older people.

Methods

Data Collection and Sample

Data were obtained from a sample of participants in the Swedish National Study of Aging and Care (SNAC). SNAC is a longitudinal cohort study of a representative sample of the aging Swedish population that began data collection in 2001. It is a comprehensive, interdisciplinary study that investigates the health and living conditions of the Swedish population aged 60 years and older. A detailed outline of the SNAC study is available by Lagergren et al [46]. Our study sample was based on participants from one of the four regions in the SNAC study, the SNAC Blekinge (SNAC-B) cohort with individuals living in the municipality of Karlskrona.

Data were collected through a questionnaire that was sent out in October 2017 to all participants in the SNAC-B study who were alive in January 2017 (N=878). Of these, 18 had deceased before answering the questionnaire. A total of 659 persons responded, corresponding to a response rate of 77% (659/860). Among nonresponders, 28% (57/201) were unable to respond due to their health conditions (eg, severe dementia or other diseases) and were considered nonusers of ICT. In this study, only individuals who responded that they were ICT users were included (N=374). Demographic data for the study population (ie, age, gender, and educational level) are presented in Table 1.

Measures and Scale Development

Development of the new short instrument, TechPH, was made in three steps. First, a search was made in Web of Science and Google Scholar for technophilia measurement instruments to use as a background for building the new instrument. To be included, the articles had to contain a psychometric instrument for use on the individual level. Eight relevant instruments were found to match the criteria (Multimedia Appendix 1), of which seven reported a full instrument and were included in this study. Instrument 8 [40] did not include the full instrument in their article and could not be used in the content analysis. Overarching themes (emotional, personal gain, openness/curiosity, competence, general attitudes) were identified through content analysis and items corresponding to these themes were constructed with a five-point Likert scale questionnaire, ranging from 1 (fully disagree) to 5 (fully agree). Following the instruments in the analysis, both questions with a positive and a negative direction were constructed to make sure both the lower and the higher end of the spectrum of technophilia were sufficiently covered. The instruments used in the analysis also gave reason to assume that more than one factor could be present. A total of six to eight questions was considered ideal both for the length of the instrument and for the case that more than one factor would emerge from the factor analysis.

Table 1. Demographic data for the study population, including age, gender, and educational level (N=374).

Category	Participants
Sex, n (%)	
Male	196 (52.4)
Female	178 (47.6)
Age (years), mean (SD)	**72.6 (7.1)**
65-75, n (%)	232 (62.0)
76-96, n (%)	142 (38.0)
Education level,[a] n (%)	
Low	94 (26.0)
Medium	135 (38.0)
High	131 (36.0)

[a] N=360 for education level. Education was categorized in three groups according to the previous Swedish education system, relevant for the age groups in this study (low: those who did not finish secondary school; medium: those who finished secondary school but no further education; high: those with some form of higher education).

The new instrument resulting from the content analysis was able to build on the previous instruments by including all themes in the same instrument. The analysis showed that none of the seven existing instruments investigated covered all the themes resulting from the analysis, making the new developed short instrument (TechPH) more comprehensive than its predecessors.

Expert knowledge in gerontology from both the medical side (gerontologist) and the technical side (design expert) was used to phrase and formulate the questions for the target group.

Secondly, the questionnaire was pretested and cognitive interviews [47] were made with the target group (eight individuals of both sexes varying in age between 60 and 82 years). The interview persons were given the 8-item questionnaire and were encouraged to think-aloud when they read them. The interviewer would also follow-up with verbal probing (ie, questions about how the interview persons understood the question) based on item wording, terminology, and if the structure was clear and easy to understand. Specifically, the questions "Can you repeat the question I just asked in your own words? Was there anything confusing about this question? What does the word <term> mean to you as it is used in the question? Tell me what you were thinking when I asked about <topic of question>."

The item questions were then revised according to the feedback from the interviews with respect to the verbal probing. Especially important was to make sure that the questionnaire was using terminology relevant to older people using technology to ensure face validity. The resulting questionnaire items (Table 2) was also pretested on four persons from the target group.

Finally, factor analyses were made. First exploratory factor analysis to see the factor structure and decide on a one or multiple factor solution. Before the exploratory factor analysis, Bartlett test of sphericity was used to ensure significant correlation and Kaiser-Meyer-Olkin test of sampling adequacy for sufficient variance among the items. Maximum likelihood factoring and Promax with Kaiser normalization were used, and only factors with eigenvalues greater than 1 were included.

The reliability of the questionnaire was calculated with the Cronbach alpha coefficient to ensure sufficient internal consistency.

Table 2. Descriptive statistics of suggested instrument items (N=374).

Questionnaire item	Mean (SD)	Median
1. I think it's fun with new technological gadgets	3.40 (1.19)	3.00
2. Using technology makes life easier for me	3.78 (1.27)	4.00
3. I like to acquire the latest models or updates	2.53 (1.34)	2.00
4. I am sometimes afraid of not being able to use the new technical things	2.90 (1.35)	3.00
5. Today, the technological progress is so fast that it's hard to keep up	3.73 (1.22)	4.00
6. I would have dared to try new technical gadgets to a greater extent if I had had more support and help than I have today	3.10 (1.41)	3.00
7. People who do not have access to the internet have a real disadvantage because of all that they are missing out on	4.13 (1.16)	5.00
8. Too much technology makes society vulnerable	4.10 (1.07)	4.00

Confirmatory factor analysis was then used to verify the factor structure. Measures of fit are reported, such as chi-square statistic and its significance, the adjusted goodness-of-fit Index (AGFI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR).

Finally, an index based on sum scores was calculated [48], and sociodemographic attributes and self-assessed technical competence were assessed for group comparisons and as indicators of criterion validity.

Results

Exploratory Factor Analysis

The Kaiser-Meyer-Olkin test of sampling adequacy was in the adequate range of 0.76 [49] and the Bartlett test of sphericity (χ^2_{28}=554.1 was significant (P<.001), indicating that the items were appropriate for a factor analysis [50].

A two-factor solution for technophilia emerged with an eigenvalue greater than 1 and examination of the scree plot. A one-factor solution was also tested for but showed low internal consistency and was discarded. The two factors were distinctly different with respective clear loadings. The factor with questions regarding positive feelings toward technology (ie, items 1-3) contained items with various aspects of enthusiasm toward technology was accordingly named *techEnthusiasm*. The factor with more negative feelings (items 4-6), contained items with different aspects of anxiety toward technology with respect to use and competency and was named *techAnxiety*.

The questions regarding general attitudes (items 7 and 8) gave low loadings and cross-loaded on both factors above the recommended maximum threshold of 0.32 [51]; therefore, they were removed.

The final six-item solution (Table 3) gave satisfactory loadings of above 0.5 [50], and a total variance explained of 63.5% for the two factors together.

Convergent validity with average variance extracted above 0.5 [50] and discriminant validity shown with only small cross-loadings, together with a factor correlation of −.48. The reliability of the questionnaire, in terms of internal consistency, was calculated by Cronbach alpha (Table 3) with satisfactory results for the small item number (techEnthusiasm Cronbach alpha=.72 and techAnxiety Cronbach alpha=.68).

Confirmatory Factor Analysis

The final confirmatory factor analysis conducted showed relatively good [52] fit indexes for the two-factor model (χ^2_8=21.2, χ^2/df=2.65, CFI=0.97, AGFI=0.95, RMSEA=0.067, SRMR=0.036). The model showed satisfactory (>0.5) standardized factor loadings given the sample size [50] confirming construct validity. Table 4 shows the standardized parameter estimates.

Table 3. Exploratory factor analysis loadings and Cronbach alphas.

Item	techEnthusiasm	techAnxiety
1. I think it’s fun with new technological gadgets	0.86	0.01
2. Using technology makes life easier for me	0.62	0.04
3. I like to acquire the latest models or updates	0.60	−0.02
4. I am sometimes afraid of not being able to use the new technical things	−0.07	0.68
5. Today, the technological progress is so fast that it’s hard for me to keep up	0.00	0.75
6. I would have dared to try new technical gadgets to a greater extent if I had had more support and help than I have today	0.09	0.53
Cronbach alpha	.72	.68

Table 4. Confirmatory factor analysis standardized factor loadings for TechPH.

Item	techEnthusiasm	techAnxiety
1. I think it’s fun with new technological gadgets	0.88	—[a]
2. Using technology makes life easier for me	0.63	—
3. I like to acquire the latest models or updates	0.61	—
4. I am sometimes afraid of not being able to use the new technical things	—	0.74
5. Today, the technological progress is so fast that it’s hard to keep up	—	0.72
6. I would have dared to try new technical gadgets to a greater extent if I had had more support and help than I have today	—	0.53

[a] Not applicable.

TechPH Index

A composite score [50] (see Table 5) was created from the six items in the two factors, techEnthusiasm and techAnxiety (the latter reversely coded due to the negative correlation). Each item was weighted with its loading before sum scores were created and averaged [48] and standardized back to a 1 to 5 scale so that the TechPH index could be interpreted on a five-point response scale, ranging from 1 (fully disagree) to 5 (fully agree), where the higher the index indicates a higher level of technophilia.

Distribution of all individuals' scores in the two factors, techEnthusiasm and techAnxiety, is presented in a scatterplot (see Figure 1). A low negative correlation between the two factors was observed (R^2=.12).

Group Comparison

The TechPH index (Table 5) was used for group comparison for gender, age group, level of education, self-assessed technical skills, and internet use frequency. This comparison yielded significant (P<.05) results. Cronbach alpha was .71, which signifies moderately good reliability of the index and internal consistency.

The TechPH index scores reflect the same finding or assumptions in technology acceptance by demographic variables; the index decreases in the older old (≥75 years) group and is slightly higher for men, which confirms other findings with regard to age groups and gender [53].

Table 5. TechPH index: descriptives and group test statistics.

Group	N	TechPH index, mean (SD)	*t* test (*df*)	*F* test (*df*1,*df*2)	*P* value
All	374	3.01 (0.86)	—[a]	—	—
Gender			1.71 (372)	—	.046
Men	196	3.08 (0.88)			
Women	178	2.93 (0.82)			
Age (years)			2.89 (372)	—	.004
<75	232	3.11 (0.81)			
≥75	142	2.85 (0.90)			
Education			—	0.65 (2,357)	.52
Low	94	2.96 (0.92)			
Medium	135	2.96 (0.81)			
High	131	3.07 (0.87)			
Self-assessed technical skill[b]			—	86.40 (2,337)	<.001[c]
Low	105	2.38 (0.64)			
Medium	200	3.18 (0.70)			
High	35	4.02 (0.82)			
Internet use frequency[d]			—	29.26 (2,338)	<.001[c]
Low	77	2.59 (0.82)			
Medium	104	2.87 (0.67)			
High	160	3.38 (0.84)			

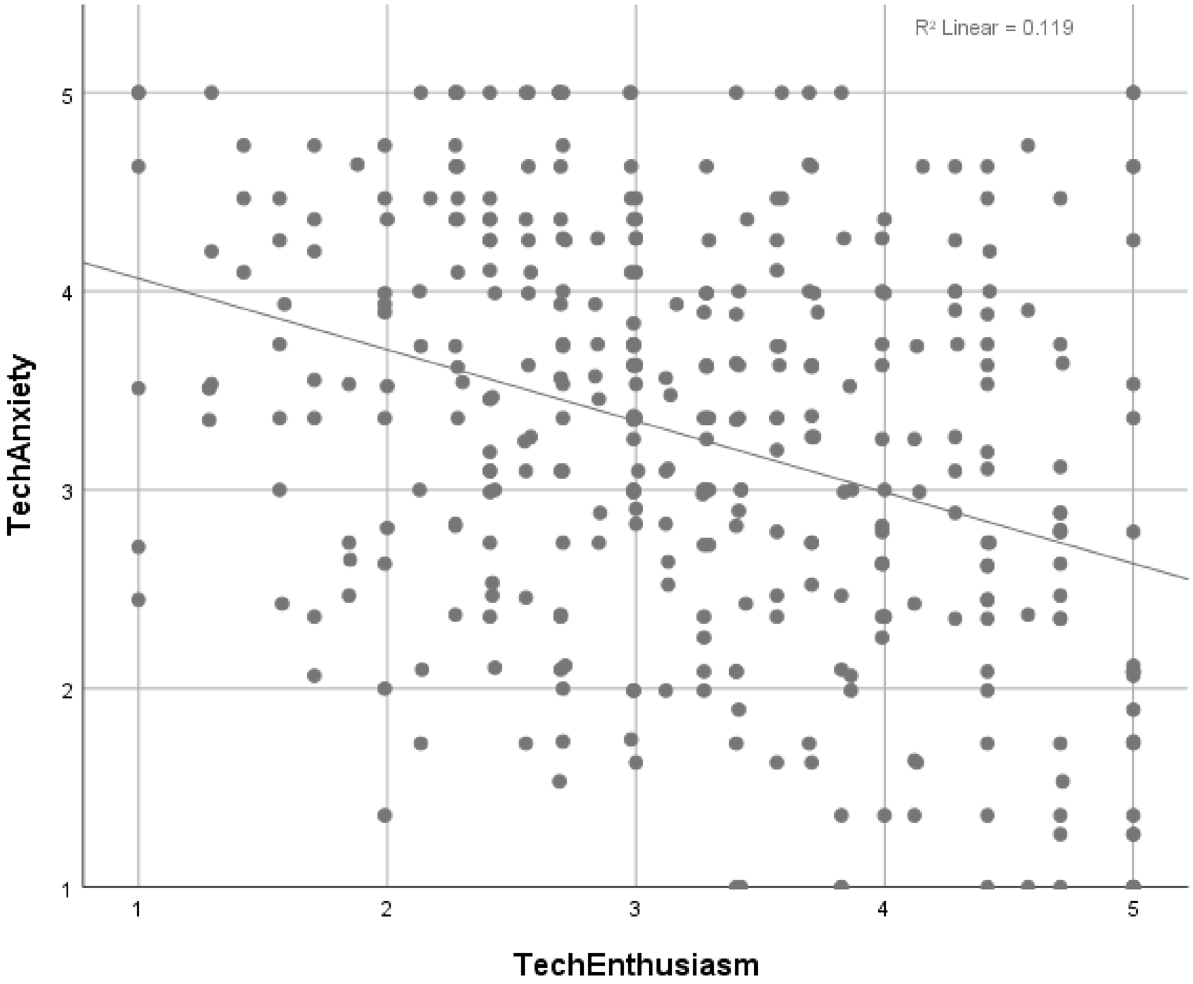
[a] Not applicable.

[b] "How skilled do you consider yourself when it comes to using a smartphone or a tablet?" Low=not at all skilled, medium=average skilled, high=very skilled.

[c] All post hoc (Tukey) group mean differences were significant at the .05 level.

[d] The participants were categorized as high=daily, medium=at least once a week but not daily, low=less than once a week.

Figure 1. Scatterplot of techEnthusiasm (y-axis) and techAnxiety (x-axis). Individuals showing high technophilia (TechPH) are found in the second quadrant; low TechPH are found in the fourth quadrant.



Discussion

This project set out to create a short instrument to enable measuring of technophilia for use among older persons participating in health technology research projects. The requirements were that it should be based on existing validated instruments and that it should be short and simple to make it usable for older people.

The resulting instrument consists of six items in two factors measuring techEnthusiasm and techAnxiety as factors of technophilia. Factor analysis of the instrument showed the feasibility of a two-factor model. Better fit for a two-factor model, compared to a one-factor model, shows that techEnthusiasm and techAnxiety are not just inverse ends of the same continuum, but two independent factors that could influence various aspects of the use of technology. There is a slight reverse correlation between them (Figure 1).

The results in Table 5 complies with general findings, and assumptions about gender, age, technical skills, and internet use frequency correlates with technology adoption. The correlation between technical skills and TechPH confirms other findings regarding reduced fear of technology by gaining technical skills [54,55]. Seifert and Schelling [56] showed that affinity for technology has a positive impact on internet use. This is also consistent with the findings of Nimrod [43] with similar relations to education, gender, and age, as well as use of technology (in this case, the internet) as in our study. Nimrod investigated technophobia with Sinkovics et al [57] with a 13-item and 3-factor instrument (personal failure, human versus machine ambiguity, perceived convenience). This is a similar setup to TechPH, with different factors pointing both to a lower and higher end of the technophilia/technophobia spectrum. Nimrod measured technophobia specifically toward ICT, such as “computers, internet, and mobile phones,” thus making some of the questions border a measure of a utility perspective rather than a personality trait with feelings toward technology in general.

Concerning techEnthusiasm, item 1 (“I think it’s fun with new technological gadgets”) loaded strongly on the techEnthusiasm factor, which was expected based on the theoretically assumed relationship with the latent variable, confirming that an item (observed variable) that closely reflects the latent variable should be highly correlated with that for a valid model. Item 2 (“Using technology makes life easier for me”) and item 3 (“I like to acquire the latest models or updates”) loaded somewhat weaker but are still seen as conceptually valid for the construct as a whole.

On the techAnxiety side, item 4 (“I am sometimes afraid of not being able to use the new technical things”) closely reflects techAnxiety both in articulation and relatively high correlation. It should be noted that this item considers an internal cause, that is an inability to use technology. Item 5 (“Today, the technological progress is so fast that it’s hard to keep up”) reflects the anxiety over a perceived inability to internalize and relate to the fast, technological progress and connects to a “technostress” [58]. Item 6 (“I would have dared to try new technical gadgets to a greater extent if I had had more support and help than I have today”) refer to the anxiety older people can feel about their lacking ability to handle technology on their own and fear of social isolation and lack of support of the aging population [59].

The two questions regarding general attitudes, items 7 and 8 (ie, “People who do not have access to the internet have a real disadvantage because of all that they are missing out on” and “Too much technology makes society vulnerable”), did not load sufficiently on any of the factors and did not make up a factor of their own. Both these items showed high means and medians and had poor discriminant value, signifying that these attitudes are shared between persons both with high and low techPH. This result is similar to that of Seifert and Schelling [56], in which both onliners and offliners with a high affinity for technology attributed a high value to the internet (in this case, for staying independent longer in old age).

It might be assumed that a high techEnthusiasm score is associated with a low techAnxiety score and vice versa, but it is still possible that both scores could be high or low together. This could have implications for a medium score and needs to be investigated further when TechPH is tested in health technology projects.

An interpretation of a set of high scores in both techEnthusiasm and techAnxiety factors could be that the individual has a basic positive attitude or enthusiasm to technology, but also feels limitations. A lack of interest in technology could be the reason why a person might show low degrees of technology enthusiasm and anxiety simultaneously. However, this lack of interest in technology does not necessarily indicate ignoring technology benefits in general, but it might be the personal attitude about their necessity for the current situation of the respondent [23].

From this perspective, TechPH could be hypothesized to have an effect on the outcome that is separate from the planned intention to use or the perceived usefulness of the application itself. Another assumption is that this could have an impact on how a person perceives problems with use and nonuse friendliness and make a person more error tolerant. This could possibly skew the usability measurements and constitute a confounder to the measured health effect outcome. In smaller studies, especially in randomized controlled studies, this would be a variable of interest to study. This is similar to the effect that Kamin and Lang [10] suggest while exploring the motivational resources for older persons’ technology use by the concept of subjective personal adaptivity. They argue that usability testing might be misleading if motivational factors moderating task performance in person technology transactions are not considered.

Whether this is a personal trait influencing attitudes toward technology that is related to age or physical or cognitive problems will be tested in further studies. We can also assume that the impact on the factors in TechPH is affected differently depending on the type of health technology being evaluated. It could be assumed that the techAnxiety factor has a greater impact on technology that influences items such as personal privacy.

A strength of the instrument introduced in this study is that it is based on previously validated, relevant instruments. It is shortened as much as possible, to three variables per factor [60], and articulated by expert analysis to be suitable for older people. The factor analysis is based on a satisfactory sample size of the general population of older adults from a midsized community in Sweden. Overall, the instrument performed as expected and will now be tested for its prediction ability of the outcome for a health technology project with older people.

In conclusion, we suggest that different technophilia traits distinguish clusters with different behaviors of adaptation as well as usage of new technology and hypothesize that this can be measured with the TechPH score. Whether there is an independent association with the TechPH score or either of the two factors contributing to the score, techEnthusiasm and techAnxiety, against outcomes in health technology projects needs to be shown in further studies. The instrument must also be validated in different contexts, such as other countries.

Acknowledgments

The authors would like to thank Professor Stefan Renvert, Professor Tobias Larsson, and Professor Mikael Rennemark for valuable comments during the work with the text. We are also grateful to the study participants and the participating counties and municipalities. The Swedish National Study on Aging and Care is financially supported by the Swedish Ministry of Health and Social Affairs and the participating county councils, municipalities, and university departments.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Content analysis of existing Technophilia instruments.

[XLSX File (Microsoft Excel File), 79KB - jmir_v21i5e13951_app1.xlsx]

References

1. Schulz R, Wahl H, Matthews JT, De Vito DA, Beach SR, Czaja SJ. Advancing the aging and technology agenda in gerontology. Gerontologist 2015 Oct;55(5):724-734. [doi: 10.1093/geront/gnu071] [Medline: 25165042]
2. Kwon S. In: Kwon S, editor. Gerontechnology: Research, Practice, and Principles in the Field of Technology and Aging. New York: Springer Publishing Company; 2016.
3. Peine A. Technology and ageing-theoretical propositions from science and technology studies (STS). In: Neves BB, Vetere F, editors. Ageing and Digital Technology: Designing and Evaluating Emerging Technologies for Older Adults. Singapore: Springer Singapore; 2019:51-64.
4. Fisk A, Czaja S, Rogers W, Charness N, Sharit J. Designing for Older Adults-Principles and Creative Human Factors Approaches. Boca Raton, FL: CRC Press; Nov 14, 2018.
5. Mitzner TL, Savla J, Boot WR, Sharit J, Charness N, Czaja SJ, et al. Technology adoption by older adults: findings from the PRISM trial. Gerontologist 2019 Dec 09;59(1):34-44. [doi: 10.1093/geront/gny113] [Medline: 30265294]
6. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
7. Roque NA, Boot WR. A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire. J Appl Gerontol 2018 Feb;37(2):131-156. [doi: 10.1177/0733464816642582] [Medline: 27255686]
8. Boot WR, Charness N, Czaja SJ, Sharit J, Rogers WA, Fisk AD, et al. Computer proficiency questionnaire: assessing low and high computer proficient seniors. Gerontologist 2015 Jun;55(3):404-411 [FREE Full text] [doi: 10.1093/geront/gnt117] [Medline: 24107443]
9. Kamin ST, Lang FR, Beyer A. Subjective technology adaptivity predicts technology use in old age. Gerontology 2017;63(4):385-392. [doi: 10.1159/000471802] [Medline: 28423371]
10. Kamin S, Lang F. The Subjective Technology Adaptivity Inventory (STAI): a motivational measure of technology usage in old age. Gerontechnology 2013 Oct 18;12(1):16-25. [doi: 10.4017/gt.2013.12.1.008.00]
11. Ramprasad C, Tamariz L, Garcia-Barcena J, Nemeth Z, Palacio A. The use of tablet technology by older adults in health care settings-is it effective and satisfying? A systematic review and meta analysis. Clin Gerontol 2019;42(1):17-26. [doi: 10.1080/07317115.2017.1322162] [Medline: 28665775]
12. Jakobsson E, Nygård L, Kottorp A, Malinowsky C. Experiences from using eHealth in contact with health care among older adults with cognitive impairment. Scand J Caring Sci 2019 Jan 10:1-10. [doi: 10.1111/scs.12634] [Medline: 30628736]
13. Barbosa Neves B, Franz R, Judges R, Beermann C, Baecker R. Can digital technology enhance social connectedness among older adults? A feasibility study. J Appl Gerontol 2019 Jan;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]
14. Baker S, Warburton J, Waycott J, Batchelor F, Hoang T, Dow B, et al. Combatting social isolation and increasing social participation of older adults through the use of technology: A systematic review of existing evidence. Australas J Ageing 2018 Sep;37(3):184-193. [doi: 10.1111/ajag.12572] [Medline: 30022583]
15. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of eHealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
16. Wakefield B, Turvey C, Nazi K, Holman J, Hogan T, Shimada S, et al. Psychometric properties of patient-facing eHealth evaluation measures: systematic review and analysis. J Med Internet Res 2017 Dec 11;19(10):e346 [FREE Full text] [doi: 10.2196/jmir.7638] [Medline: 29021128]
17. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
18. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
19. 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]
20. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
21. Chen K, Chan AH. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
22. Middlemass JB, Vos J, Siriwardena AN. Perceptions on use of home telemonitoring in patients with long term conditions-concordance with the Health Information Technology Acceptance Model: a qualitative collective case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):89 [FREE Full text] [doi: 10.1186/s12911-017-0486-5] [Medline: 28651588]

23. Peek STM, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
24. Merkel S, Enste P, Hilbert J, Chen K, Chan A, Kwon S. Technology acceptance and aging. In: Kwon S, editor. Gerontechnology Research, Practice, and Principles in the Field of Technology and Aging. New York: Springer Publishing Company; 2017:335-350.
25. Holthe T, Halvorsrud L, Karterud D, Hoel K, Lund A. Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: a systematic literature review. Clin Interv Aging 2018;13:863-886 [FREE Full text] [doi: 10.2147/CIA.S154717] [Medline: 29765211]
26. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. Inform Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]
27. Yusof MM, Paul RJ, Stergioulas LK. Towards a framework for health information systems evaluation. 2006 Presented at: HICSS '06 39th Annual Hawaii International Conference on System Sciences; Jan 4-7, 2006; Kauia, HI. [doi: 10.1109/hicss.2006.491]
28. Yu P, Li H, Gagnon M. Health IT acceptance factors in long-term care facilities: a cross-sectional survey. Int J Med Inform 2009 Apr;78(4):219-229. [doi: 10.1016/j.ijmedinf.2008.07.006] [Medline: 18768345]
29. Self R, Aquilina C. TechnoStress in the 21st century; does it still exist and how does it affect knowledge management and other information systems initiatives. In: Advances in Intelligent Systems and Computing. Berlin: Springer; 2013:117-127.
30. Tarafdar M, Tu Q, Ragu-Nathan TS. Impact of technostress on end-user satisfaction and performance. J Manage Inform Syst 2014 Dec 09;27(3):303-334. [doi: 10.2753/mis0742-1222270311]
31. Yen P, Wantland D, Bakken S. Development of a customizable health IT usability evaluation scale. AMIA Annu Symp Proc 2010 Nov 13;2010:17-21. [Medline: 21347112]
32. Brooke J. SUS: A quick and dirty usability scale. In: Usability Evaluation in Industry. London: Taylor and Francis; 1996:189-194.
33. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11. [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
34. Schnall R, Cho H, Liu J. Health Information Technology Usability Evaluation Scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
35. Edison SW, Geissler GL. Measuring attitudes towards general technology: Antecedents, hypotheses and scale development. J Target Meas Anal Mark 2003 Nov 1;12(2):137-156. [doi: 10.1057/palgrave.jt.5740104]
36. Plociennik C, Wandke H, Kirste T. What influences user acceptance of ad-hoc assistance systems? A quantitative study. In: Proceedings of MMS European Mathematical Information Service. 2010 Presented at: MMS European Mathematical Information Service; Feb 23-25, 2010; Bonn p. 57-70.
37. Ronit P. Technophilia: A new model for technology adoption. In: Proceedings of the UK Academy for Information Systems Conference. 2011 Presented at: UK Academy for Information Systems Conference; Apr 11, 2011; Oxford, UK p. 41.
38. Seebauer S, Stolz R, Berger M. Technophilia as a driver for using advanced traveler information systems. Transport Res C-Emer 2015 Nov;60:498-510 [FREE Full text] [doi: 10.1016/j.trc.2015.10.009]
39. Osiceanu M. Psychological implications of modern technologies: “technofobia” versus “technophilia”. Procd Soc Behv 2015 May;180:1137-1144 [FREE Full text] [doi: 10.1016/j.sbspro.2015.02.229]
40. Martínez-Córcoles M, Teichmann M, Murdvee M. Assessing technophobia and technophilia: development and validation of a questionnaire. Technol Soc 2017 Nov;51:183-188 [FREE Full text] [doi: 10.1016/j.techsoc.2017.09.007]
41. Li Y, Fuller B. I'm lovin' IT: toward a technophilia model of user adaptation to ICT. In: Proceedings Association for Information Systems (AIS). 2017 Presented at: AMCIS 2017; Aug 10, 2017; Boston, MA.
42. Donat E, Brandtweiner R, Kerschbaum J. Attitudes and the digital divide: attitude measurement as instrument to predict internet usage. Inform Sci 2009;12(1):37-56. [doi: 10.28945/427]
43. Nimrod G. Technophobia among older Internet users. Educ Gerontol 2018 Jan 24;44(2-3):148-162 [FREE Full text] [doi: 10.1080/03601277.2018.1428145]
44. Devaraj S, Easley RF, Crant JM. How does personality matter? Relating the five-factor model to technology acceptance and use. Inform Syst Res 2008 Mar;19(1):93-105. [doi: 10.1287/isre.1070.0153]
45. Steinerman J, Lipton R, Rapkin B, Quaranto B, Schwartz C. Factors associated with openness to research participation in an aging community: the importance of technophilia and social cohesion. Gerontechnology 2013 Mar 21;11(4):504-512. [doi: 10.4017/gt.2013.11.4.012.00]
46. Lagergren M, Fratiglioni L, Hallberg I, Berglund J, Elmståhl S, Hagberg B, et al. A longitudinal study integrating population, care and social services data. The Swedish National study on Aging and Care (SNAC). Aging Clin Exp Res 2004 Apr;16(2):158-168. [Medline: 15195992]
47. Peterson CH, Peterson NA, Powell KG. Cognitive interviewing for item development: validity evidence based on content and response processes. Meas Eval Couns Dev 2017 Oct 04;50(4):217-223. [doi: 10.1080/07481756.2017.1339564]

48. DiStefano C, Zhu M, Mindrila D. Understanding and Using Factor Scores: Considerations for the Applied Researcher. Prac Assess Res Eval 2009 Oct 01;14(20):1-10 [FREE Full text]
49. DiIorio CK. Measurement In Health Behavior: Methods For Research And Evaluation. San Francisco, CA: John Wiley & Sons; Mar 06, 2006:1-336.
50. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis (seventh edition). Upper Saddle River, NJ: Prentice Hall; 2019.
51. Osborne J, Costello A. Sample size and subject to item ratio in principal components analysis. Prac Assess Res Eval 2004;9(11):1-10 [FREE Full text]
52. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
53. Or CK, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560. [doi: 10.1197/jamia.M2888] [Medline: 19390112]
54. Wild K, Mattek N, Maxwell S, Dodge H, Jimison H, Kaye J. Computer-related self-efficacy and anxiety in older adults with and without mild cognitive impairment. Alzheimers Dement 2012 Nov;8(6):544-552. [doi: 10.1016/j.jalz.2011.12.008] [Medline: 23102124]
55. Shedletsky L. Internet training for older adult learners: an intergenerational mentoring approach. LLI Rev 2006;1:43 [FREE Full text]
56. Seifert A, Schelling H. Seniors online: attitudes toward the internet and coping with everyday life. J Appl Gerontol 2018 Jan;37(1):99-109. [doi: 10.1177/0733464816669805] [Medline: 27655237]
57. Sinkovics RR, Stöttinger B, Schlegelmilch BB, Ram S. Reluctance to use technology-related products: development of a technophobia scale. Thunderbird Int Bus Rev 2002 Jun 27;44(4):477-494 [FREE Full text] [doi: 10.1002/tie.10033]
58. Nimrod G. Technostress: measuring a new threat to well-being in later life. Aging Ment Health 2018 Aug;22(8):1080-1087. [doi: 10.1080/13607863.2017.1334037] [Medline: 28562064]
59. Berner J, Rennemark M, Jogréus C, Anderberg P, Sköldunger A, Wahlberg M, et al. Factors influencing Internet usage in older adults (65 years and above) living in rural and urban Sweden. Health Informatics J 2015 Sep;21(3):237-249. [doi: 10.1177/1460458214521226] [Medline: 24567416]
60. Bollen KA. Structural Equations With Latent Variables. New York: John Wiley & Sons; Apr 28, 1989.

Abbreviations

AGFI: adjusted goodness-of-fit index
CFI: comparative fit index
IS: information system
HIT: health information technology
HITAM: health information technology acceptance model
ICT: information and communication technology
RMSEA: root mean square error of approximation
SNAC: Swedish National Study of Aging and Care
SNAC-B: SNAC Blekinge
SRMR: standardized root mean square residual
TAM: technology acceptance model

Edited by G Eysenbach; submitted 07.03.19; peer-reviewed by N Roque, A Seifert; comments to author 28.03.19; revised version received 17.04.19; accepted 24.04.19; published 23.05.19.

Please cite as:
Anderberg P, Eivazzadeh S, Berglund JS
A Novel Instrument for Measuring Older People's Attitudes Toward Technology (TechPH): Development and Validation
J Med Internet Res 2019;21(5):e13951
URL: http://www.jmir.org/2019/5/e13951/
doi:10.2196/13951
PMID:31124467

Original Paper

Finding the Right Blend of Technologically Enhanced Learning Environments: Randomized Controlled Study of the Effect of Instructional Sequences on Interprofessional Learning

Sok Ying Liaw[1], BHSN, MHS, PhD; Khoon Kiat Tan[2], BSc (Hons), MEd, PhD; Ling Ting Wu[1], BSc, MSc; Seng Chee Tan[3], BSc, MEd, PhD; Hyekyung Choo[4], BA, MA, MSW, PhD; John Yap[5], BA, MA; Sok Mui Lim[6], BSc (Hons), PhD; Lilian Wong[7], BSc, MBA, PharmD; Jeanette Ignacio[1], BSc, MD, PhD

[1]Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore
[2]School of Health Sciences, Nanyang Polytechnic, Singapore, Singapore
[3]National Institute of Education, Learning Sciences & Technologies, Nanyang Technological University, Singapore, Singapore
[4]Department of Social Work, National University of Singapore, Singapore, Singapore
[5]National University of Singapore Information Technology, National University of Singapore, Singapore, Singapore
[6]Centre for Learning Environment and Assessment Development, Singapore Institute of Technology, Singapore, Singapore
[7]Department of Pharmacy, National University of Singapore, Singapore, Singapore

Corresponding Author:
Sok Ying Liaw, BHSN, MHS, PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
Clinical Research Centre
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 65167451
Email: nurliaw@nus.edu.sg

Abstract

Background: With the availability and capabilities of varied technologically enhanced learning activities, the blended learning approach has become increasingly popular in interprofessional education. The combined use of different technologically enhanced learning activities has not been fully examined, particularly to determine the effects of instructional sequences for effective learning outcomes.

Objective: The objective of this study was to investigate whether the instructional sequences of a blended learning approach can improve students' learning outcomes on interprofessional competencies.

Methods: A randomized controlled study was conducted with 40 interprofessional health care teams. These teams undertook three technologically enhanced learning activities—Web-based instruction (WI), virtual reality (VR), and simulation exercise (SE)—after random assignment to three groups based on three different instructional sequences (WI-VR-SE, WI-SE-VR, and SE-WI-VR). Pretests and posttests were conducted to evaluate the students' learning outcomes on interprofessional competencies.

Results: A total of 198 participants from the three groups completed the questionnaires. All three groups reported significant improvement in their levels of self-efficacy ($P<.05$) and attitudes ($P<.001$) toward interprofessional team care about 1 month after the interprofessional learning activity. Although no significant difference was found ($P=.06$) between the WI-VR-SE and WI-SE-VR groups in the self-efficacy posttests, participants in the SE-WI-VR group reported significantly lower ($P<.05$) posttest scores than those in the WI-SE-VR group. The majority of the participants (137/198, 69.1%) selected the instructional sequence "WI-VR-SE" as their top preference.

Conclusions: This study shows that the instructional sequence of a blended learning approach can have a significant impact on students' learning outcomes. The learning of concepts from WI followed by problem-solving activity in the SE was found to be a more effective learning sequence than the reverse sequence. We recommend that future studies focus on scaffolding students' learning when planning instructional sequences for technologically enhanced learning activities within blended learning environments.

(J Med Internet Res 2019;21(5):e12537) doi:10.2196/12537

KEYWORDS

blended learning; constructivism; instructional sequence; interprofessional learning; simulation; technologically-enhanced learning; virtual reality; web-based instruction

Introduction

Technologically enhanced learning in health professional education has evolved rapidly from basic text-based learning to inclusion of more multimedia features (eg, video and animation), also known as Web-based multimedia instruction, to provide self-directed learning opportunities for learners on didactic material. Yavner et al [1] described Web-based multimedia instruction as a screen-based set of learning material in a personal computer or other multimedia devices that learners can read, listen to, and watch. Although Web-based multimedia instruction can be used to replace didactic learning methods (eg, lecture), integrating it with group learning activities and experiential learning (eg, patient encounter) can significantly enhance learning [2].

With the development of communication tools, there is a growing trend of electronic learning (e-learning) use in interprofessional learning to support interaction and promote information sharing among different health care professionals [3]. E-learning has been identified as a practical and accessible learning tool to overcome logistical challenges often associated with scheduling interprofessional learning activities across different health care courses [4]. Although the evidence of e-learning in improving interprofessional collaboration is significant, feelings of isolation among learners are commonly reported. As a result of diminishing face-to-face interactions, the use of e-learning may also affect interprofessional interactions/dynamics [5]. Therefore, it will be beneficial to use a blended learning approach that combines e-learning and face-to-face interactions to develop interprofessional competencies. A study by Riesen et al [6] found that a blended learning environment that included online, virtual face-to-face, and traditional face-to-face interactions improved health care students’ interprofessional competencies.

Advanced interactive technologies including game-based learning, virtual patient, and virtual reality have recently gained attention in health care education. A virtual reality environment generated by computers to create 3D realms allows every user to don the role of a virtual avatar and enables him/her to interact with the avatar in a real-time environment [7]. The use of this environment for collaborative learning is gaining popularity in health care education, as more health care professionals are searching for ways to develop their interprofessional collaborative practice competencies, as encouraged by the government and professional bodies. Many were also spurred by the success of the use of physical simulation for team training [8]. Implementing such training in a virtual environment offers several advantages, including overcoming logistical challenges (eg, facilities and scheduling) associated with physical simulation [9]. Studies have shown that the use of virtual simulations was found to be as effective as physical simulation in improving performance in acute care [10,11]. Blending virtual and physical simulations may optimize learning effectiveness [12].

Blended learning refers to combining computer-mediated learning with face-to-face interactions. This can involve a mix of Web-based technologies or various pedagogical approaches to support learning [13]. A systematic review on the effectiveness of blended learning in health professionals showed a consistent positive effect when compared with no intervention, which is better than or at least comparable to nonblended instruction for the acquisition of knowledge [14]. Another systematic review, which focused on the role of blended learning in health care clinical education, reported the potential of learning in improving clinical competencies and suggested that future research should go beyond a mere comparison with traditional approaches [2]. Rather, research into blended learning should pay attention to the ways of implementing a blended course effectively [14], including different blends of effective approaches, tools, and technologies [15].

Although the combined use of different learning modalities within blended learning environments has become increasingly popular in the delivery of interprofessional education [6], it remains unclear whether the instructional sequences of these learning modalities affect learning outcomes. In this study, a technologically enabled blended learning approach was designed to deliver an interprofessional learning activity on patient-centered team care for health care students. Based on the learning processes of concept building, experiential learning, and problem solving, three technologically enabled learning modalities—Web-based instruction (WI), virtual reality (VR) environment, and face-to-face simulation exercise (SE)—were implemented. The conventional instruction sequence often involves scaffolding of students’ learning that starts with the acquisition of content (concept building) using WI, engagement in experiential learning in VR to internalize the learned concepts, and the application of learning to problem solve in SE (ie, WI-VR-SE). Would other alternative sequences such as SE-WI-VR and WI-SE-VR be as effective? An initial exploratory problem-solving SE followed by concept building or experiential learning can unfold benefits by activating learners’ prior knowledge, enhancing their awareness of knowledge gaps, and helping them to relate new knowledge delivered by instructional and experiential learning [16].

This study aimed to investigate whether the instructional sequences of a blended learning approach can improve students’ learning outcomes on interprofessional competencies. The study also aimed to explore students’ evaluations of the different technologically enhanced learning modalities and their instructional sequences for the delivery of interprofessional education.

Methods

Study Designs, Setting, and Participants

After obtaining approval from institutional review boards of higher educational institutions, a prospective randomized controlled trial study with a pre-post test design was conducted on students undertaking health care courses (medicine, nursing, pharmacy, physiotherapy, occupational therapy, and medical social work) who were in their senior year at three tertiary educational institutions. Participants were recruited via convenience sampling using email and Facebook. The results of a previous study were used to estimate the sample size [17]. For a moderate effect size, a power analysis suggested at least 52 samples in each group to achieve 80% power at 5% level of significance for one-way analysis of variance testing the differences among the three groups. With an estimate of 20% overall dropout rate, a minimum of 195 students (65 per group) were targeted for recruitment. The recruited participants were assigned to interprofessional teams, each consisting five to six health care students, with one member from each health care course. These interprofessional teams were randomly assigned to one of the three groups: WI-VR-SE (14 teams), WI-SE-VR (13 teams), and SE-WI-VR (13 teams). The grouping allocations were made known to the researcher but concealed from the participants.

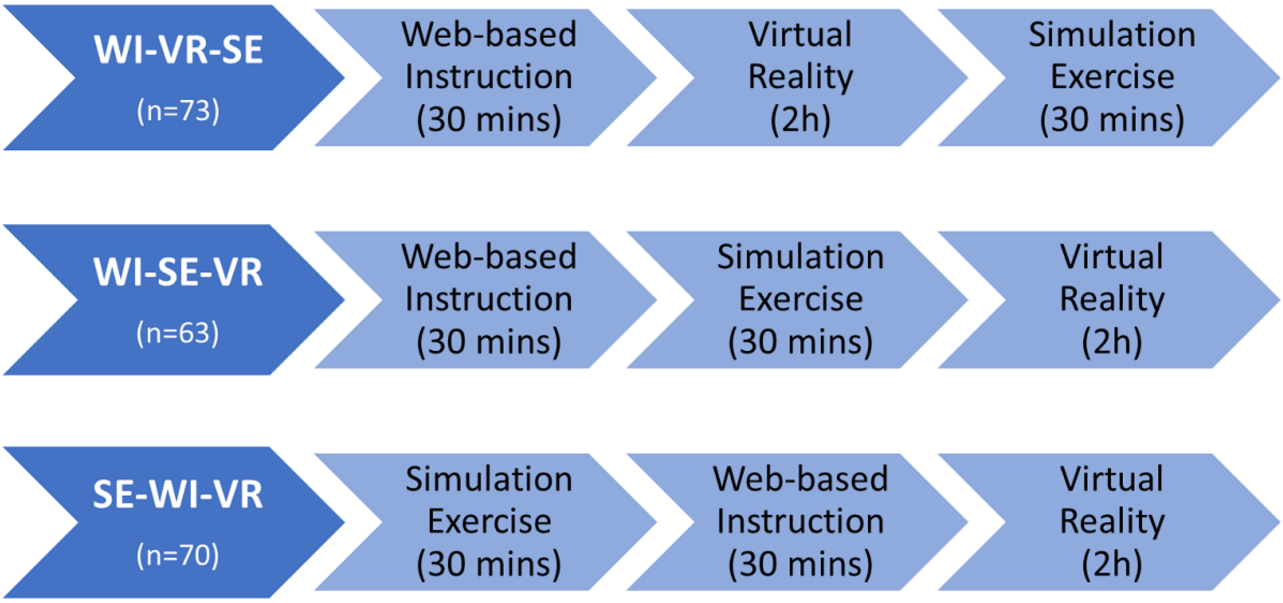
Implementation of Blended Learning Strategies

The interprofessional training on multidisciplinary rounds for patient-centered team care was selected as the focus of the learning content. As shown in Figure 1, the interprofessional teams were assigned to undertake interprofessional learning using three different instructional sequences (WI-VR-SE, WI-SE-VR, and SE-WI-VR) at a university simulation center. At the WI station, the participants were brought into an individual room with a computer set up. At their own paces, they acquired the concept of team mental models for multidisciplinary rounds by watching a video of a patient case presentation. At the VR and SE stations, the participants formed an interprofessional team of five to six participants to engage in the learning activities in the multidisciplinary rounds. At the VR station, using their own health care professional avatars, the participants worked through virtual simulation scenarios for the delivery of team care on multidisciplinary round. This was followed by debriefing sessions led by a trained facilitator in the VR environment. At the SE station, after spending some time exploring a case study and discussing the plan of care, an interprofessional team would enter a simulated physical ward environment to perform a multidisciplinary round on a simulated patient with physical and psychosocial problems. Taken together, the entire learning process on multidisciplinary rounds for patient-centered team care involved concept building using WI, facilitator-led experiential learning in VR, and the application of learning to problem solve in SE.

Data Collection and Instruments

The participants were asked to complete questionnaires before and 1 month after the blended learning activity to evaluate their interprofessional competencies. A 5-item confidence scale with a 10-point scale, developed by Grundy [18] for measurement of confidence level related to a performance, was used to measure the participants' level of self-efficacy in contributing to patient-centered care in an interprofessional team. The Cronbach alpha for this study was 0.79. A 24-item Interprofessional Socialization and Valuing Scale (ISVS) with three subscales (self-perceived ability to work with others, value in working with others, and comfort in working with others) was used to measure the participants' beliefs, behaviors, and attitudes in interprofessional socialization [19]. Each item was rated on a 5-point Likert scale. A high internal consistency of Cronbach alpha of 0.80 was obtained in this study. A 14-item Attitudes Toward Interprofessional Health Care Teams (ATIHCT) was administered to measure the participants' attitudes toward health care teams [20]; this study obtained a high internal consistency with a Cronbach alpha of 0.80.

Figure 1. Three blended learning approaches. WI: Web-based instruction; VR: virtual reality; SE: simulation exercise.



An evaluation questionnaire was administered immediately after the blended learning activity to evaluate three learning modalities in relation to the levels of enthusiasm, help, and satisfaction on a 5-point Likert scale. A question on ranking the three instructional sequences in order of the learner's preference was included in the evaluation questionnaire. All questionnaires were completed electronically to ensure that participants answered all the questions.

Data Analysis

Descriptive statistics were used to represent the demographic characteristics of the study population and the evaluation of the learning strategies. Analysis of variance with a posthoc test was carried out to determine the differences between the groups on demographic characteristics. A paired *t* test was used to examine any significant changes between the baseline and posttest attitude scores measured by the ATIHCT and ISVS. Analysis of covariance was employed to evaluate the effect of the blended learning approach on attitudes and self-efficacy posttest scores by using pretest scores as a covariate.

Results

A total of 207 health care students participated in the blended learning activity. However, only 198 completed the questionnaires (WI-VR-SE: 73; WI-SE-VR: 60, SE-WI-VR: 65), with a response rate of 95.7% for the one-month posttest questionnaires (Figure 2).

As shown in Table 1, most of the participants were female (65.2%) and undertaking a degree course (69.7%). There were no significant differences in demographic characteristics among the three groups, including age (*P*=.62), gender (*P*=.81), type of qualification (*P*=.81), and type of health care course (*P*=.21). This suggested homogeneity of the participants between the three groups.

As shown in Figure 3, all three groups reported significantly higher levels (*P*<.05) of self-efficacy in performing interprofessional team care after interprofessional learning. However, the SE-WI-VR group had the lowest self-efficacy posttest mean scores. Between-group comparisons using analysis of covariance revealed a significant difference (*P*=.03) among the three groups in terms of self-efficacy posttest mean scores, with the SE-WI-VR group reporting significantly lower posttest scores than the WI-SE-VR groups after controlling the pretest scores (*P*=.04). There were no significant differences between the WI-VR-SE and WI-SE-VR groups in self-efficacy posttest scores after controlling the pretest scores (*P*=.06).

Table 2 shows that the posttest scores on interprofessional socialization using the ISVS increased significantly (*P*<.001) from the baseline scores for all three groups. However, no significant differences were found between the baseline and posttest scores on the ATIHCT for all three groups. Between-group comparisons also did not identify any significant differences in the attitude posttest scores for both the ISVS and ATIHCT after controlling the pretest scores.

As shown in Table 3, the mean score ratings on 5-point scales indicated that the participants were enthusiastic, satisfied, and able to perceive the helpfulness of the individual learning strategies. Among them, SE had the highest mean scores for level of help (mean 4.19, SD 0.79), level of satisfaction (mean 4.17, SD 0.79), and level of enthusiasm (mean 3.99, SD 0.68).

Figure 2. CONSORT diagram. WI: Web-based instruction; VR: virtual reality; SE: simulation exercise.

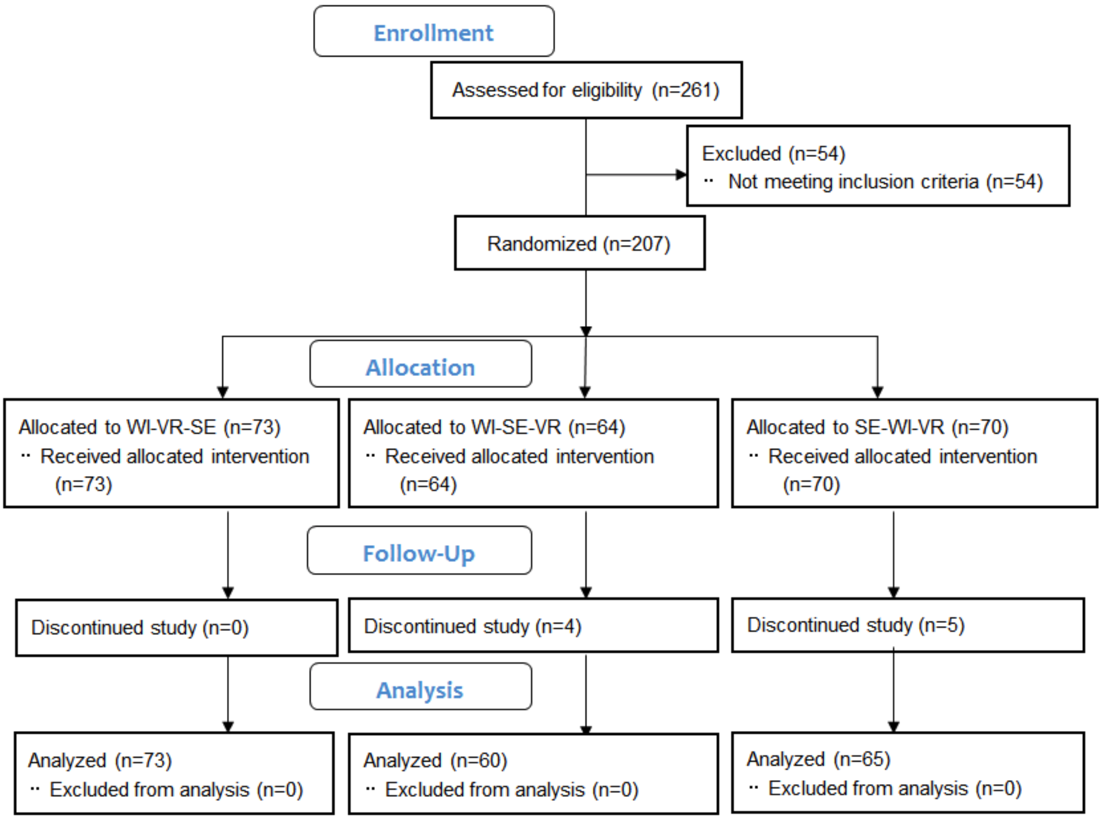


Table 1. Demographic characteristics.

Characteristic	Overall	WI[a]-VR[b]-SE[c] group (n=73)	WI-SE-VR group (n=60)	SE-WI-VR group (n=65)	*P* value
Age, mean (SD)	22.60 (1.73)	22.73 (1.83)	2263 (2.08)	22.44 (1.20)	0.62
Gender, n (%)					**0.81**
Male	69 (34.8)	27 (37.0)	19 (31.7)	23 (35.4)	
Female	129 (65.2)	46 (63.0)	41 (68.3)	42 (64.6)	
Type of qualification, n (%)					**0.81**
Degree	138 (69.7)	49 (67.1)	47 (78.3)	42 (64.6)	
Diploma	60 (30.3)	24 (32.9)	13 (21.7)	23 (35.4)	
Type of health care course, n (%)					**0.21**
Medicine	38 (19.2)	14 (19.2)	13 (21.7)	11 (16.9)	
Nursing	36 (18.2)	14 (19.2)	10 (16.7)	12 (18.5)	
Pharmacy	38 (19.2)	14 (19.2)	12 (20.0)	12 (18.5)	
Occupational therapy	29 (14.6)	11 (15.1)	9 (15.0)	9 (13.8)	
Physiotherapy	22 (11.1)	7 (9.6)	6 (10.0)	9 (13.8)	
Social work	35 (17.7)	13 (17.8)	10 (16.7)	12 (18.5)	

[a] WI: Web-based instruction.

[b] VR: virtual reality.

[c] SE: simulation exercise.

Figure 3. Mean (SD) self-efficacy scores at pretest and posttest. WI: Web-based instruction; VR: virtual reality; SE: simulation exercise.

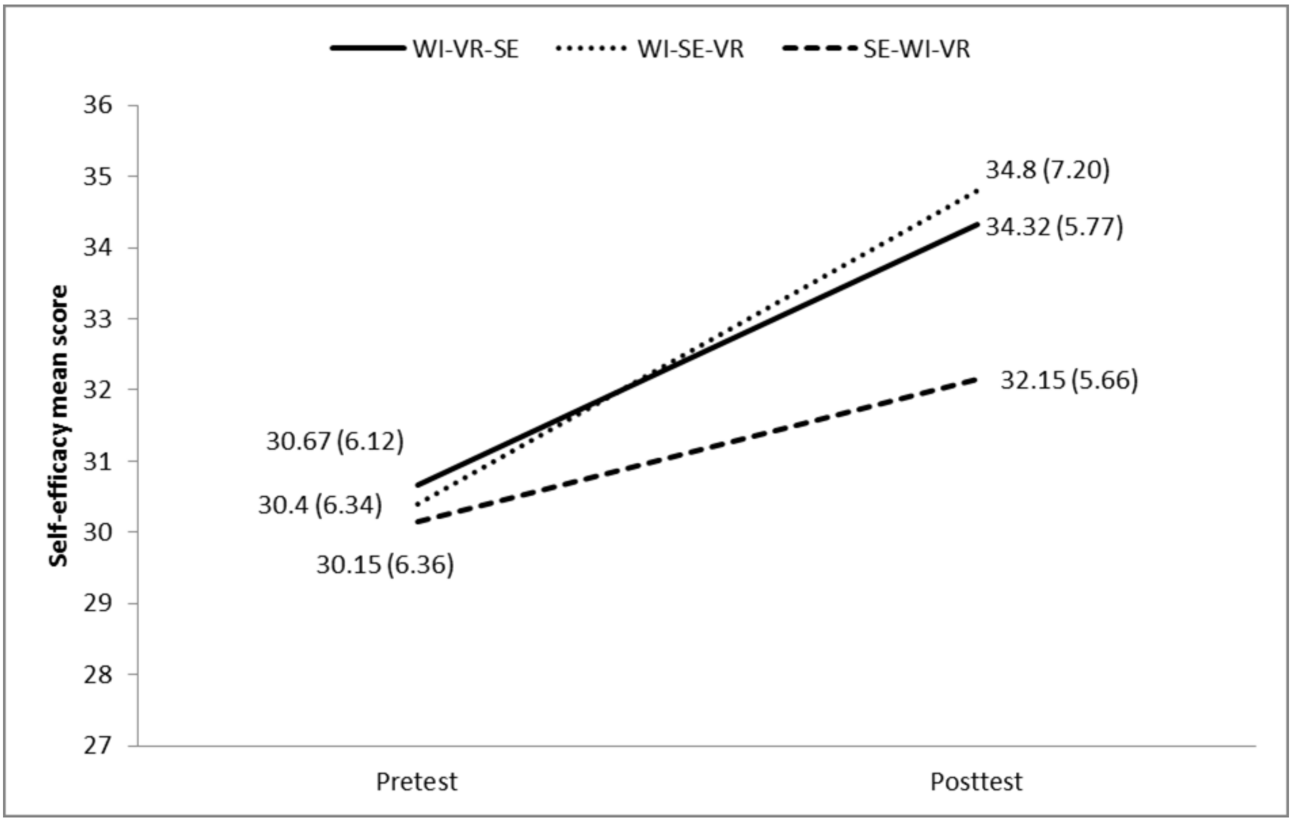


Table 2. Pretest and posttest scores on Attitudes Toward Interprofessional Health Care Teams and Interprofessional Socialization and Valuing Scale.

Instrument (range of possible scores) and group	Baseline, mean (SD)	Post score, mean (SD)	Change of scores (post-pre score), mean (SD)	Within group difference, *t* (*df*)	Between group, *F* (*df*)
Interprofessional Socialization and Valuing Scale (24-168)					**0.22 (2)**
WI-VR-SE	124.49 (14.10)	137.25 (16.87)	12.75 (14.49)	7.52 (72)[a]	
WI-SE-VR	123.92 (14.99)	135.77 (20.05)	11.85 (20.41)	4.49 (59)[a]	
SE-WI-VR	121.94 (13.57)	137.65 (12.41)	15.70 (16.55)	7.65 (64)[a]	
Attitudes Toward Interprofessional Health Care Teams (14-70)					**0.33 (2)**
WI-VR-SE	55.26 (5.38)	56.44 (6.94)	1.18 (6.43)	1.57 (72)	
WI-SE-VR	56.02 (5.12)	56.45 (6.31)	0.43 (5.59)	0.60 (59)	
SE-WI-VR	55.42 (5.66)	55.60 (7.14)	0.18 (6.23)	0.24 (64)	

[a] $P<.001$

Table 3. Participants' evaluation of the learning strategies.

Evaluation items and learning strategies	Mean score (SD)
Level of enthusiasm	
Simulation exercise	3.99 (0.68)
Virtual reality	3.92 (0.63)
Web-based instruction	3.61 (0.68)
Level of help	
Simulation exercise	4.19 (0.79)
Virtual reality	4.07 (0.73)
Web-based instruction	3.75 (0.84)
Level of satisfaction	
Simulation exercise	4.17 (0.78)
Virtual reality	4.03 (0.73)
Web-based instruction	3.90 (0.73)

In terms of the participants' top preferences for interprofessional learning through the three different instructional sequences of a blended learning approach, the majority of the participants (69.1%) selected the instructional sequence "WI-VR-SE" as their top preference; in addition, less than 10% of the participants chose the sequence "SE-WI-VR" as their top preferences and about 20% chose the sequence "WI-SE-VR."

Discussion

The evaluation of the participants' levels of self-efficacy in contributing to interprofessional patient-centered care, which was conducted 1 month following the interprofessional learning activity, indicated significant improvements from the baseline scores for all three groups. Drawing from the constructivist learning theory, case-based and problem-based learning approaches were incorporated into the blended learning strategies to help learners develop competencies in performing interprofessional patient-centered care. These approaches are known to enable participants to actively construct their knowledge based on their interpretations of experiences [21]. Furthermore, according to the situated learning theory, which is another aspect of constructivism [22], situating learning in authentic contexts that reflect the way knowledge and skills can be applied in actual life provides learners with meaningful learning experiences and thus deepens their learning [23]. As SE provided the most authentic learning environment, participants reported that it contributed to the highest levels of help and satisfaction toward their learning.

Among the three groups, participants who undertook the SE-WI-VR sequence were found to have the lowest improvement of self-efficacy levels, which was significantly lower than those who undertook WI-SE-VR. These findings suggest that the instructional sequences of a blended learning approach can significantly influence learners' perceptions of learning outcomes. Our findings show that learning starting with instructional prompts on concepts followed by problem solving in SE was a more effective learning sequence than the reverse sequence. Although the problem-solving activity in SE may prepare learners to benefit from WI learning, the introduction of concepts seemed to be necessary, particularly for novice health care students who may not have sufficient domain knowledge to engage in problem-solving tasks in SE

[16]. This study therefore supports the introduction of concepts on team mental models, as they serve as cognitive tools that scaffold students' abilities to perform tasks on interprofessional team care delivery [24].

Although no significant differences in self-efficacy scores were found between the instructional sequences WI-SE-VR and WI-VR-SE, the instructional sequence WI-VR-SE was chosen by the majority of the students as their top preferences. In this WI-VR-SE instructional sequence, scaffolding support was provided initially in WI through a video demonstration of desired performances using cognitive tools. This was followed by facilitator-led experiential learning in VR that fostered reflection and metacognition. The support was gradually decreased in SE, where students were given the opportunity to collaborate among themselves to apply their learning to problem solve a case scenario. According to Kim and Hannafin [25], in the context of technologically enhanced learning, scaffolding is defined as “cognitive and social supports designed to augment student problem-solving inquiry.” Our findings therefore support a previous study on the use of scaffolding to facilitate the alignment of technologically enhanced learning activities in blended learning [25].

Unlike self-efficacy in performing interprofessional team care, no significant differences were reported among the three groups in their attitudes toward interprofessional socialization and the health care team. Given the opportunities to engage in collaborative learning in VR and SE, all groups reported a significant improvement in attitudes toward interprofessional socialization. According to the social constructivist views of learning, the exchange and discussion of ideas through social interaction are critical for learners to construct meaningful knowledge [26]. Although SE provided a more authentic social learning environment, most students in our study preferred to undertake interprofessional learning in VR before engaging in SE. A possible reason could be that, unlike face-to-face social interactions, the anonymity embedded in VR may provide a less stressful and less threatening learning environment for different health care students to communicate and collaborate with one another. The presence of stress in simulation training was commonly reported in previous studies [27]. The study therefore supports the use of VR to prepare students for physical simulation.

Although a robust study using a randomized controlled trial and a large sample size was employed in this study, an important limitation is the lack of performance measurement as the outcome measure. The use of self-reported surveys that measure changes in attitudes and self-perceived efficacy may not predict actual performances and may be subjected to social desirability. Future studies can measure the impact on team performance. Another limitation is that we did not incorporate a debriefing or feedback session in the SE after the students' role-play experiences, which could have served as expert feedback and engage learners in reflection on actions.

In conclusion, this study provides evidence of the effectiveness of a blended learning approach using a randomized controlled study. Technologically enhanced learning strategies based on the constructivist learning theory improves health care students' interprofessional competencies. The study suggests that the instructional sequence of a blended learning approach can have a significant impact on students' learning outcomes. The more effective learning sequence allows learners to grasp concepts from WI before presenting them with a problem-solving activity in SE. From the learners' perspective, they preferred to start with concept building using WI, followed by experiential learning in VR and subsequent application through SE. This study provides recommendation for future practice to scaffold students' learning when planning the instructional sequence of technologically enhanced learning activities within blended learning environments. Future studies can undertake a more robust outcome measurement by evaluating the effects of the instructional sequence on team performance.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2MB - jmir_v21i5e12537_app1.pdf]

References

1. Yavner SD, Pusic MV, Kalet AL, Song HS, Hopkins MA, Nick MW, et al. Twelve tips for improving the effectiveness of web-based multimedia instruction for clinical learners. Med Teach 2015 Mar;37(3):239-244. [doi: 10.3109/0142159X.2014.933202] [Medline: 25109353]
2. Rowe M, Frantz J, Bozalek V. The role of blended learning in the clinical education of healthcare students: a systematic review. Med Teach 2012 Mar;34(4):e216-e221. [doi: 10.3109/0142159X.2012.642831] [Medline: 22455712]
3. Curran V, Reid A, Reis P, Doucet S, Price S, Alcock L, et al. The use of information and communications technologies in the delivery of interprofessional education: A review of evaluation outcome levels. J Interprof Care 2015;29(6):541-550. [doi: 10.3109/13561820.2015.1021002] [Medline: 25955607]
4. Cain J, Chretien K. Exploring Social Media's Potential in Interprofessional Education. JRIPE 2013 Aug 14;3(2):541-550. [doi: 10.22230/jripe.2013v3n2a110] [Medline: 25955607]

5. Reeves S, Fletcher S, McLoughlin C, Yim A, Patel KD. Interprofessional online learning for primary healthcare: findings from a scoping review. BMJ Open 2017 Aug 04;7(8):e016872 [FREE Full text] [doi: 10.1136/bmjopen-2017-016872] [Medline: 28780560]
6. Riesen E, Morley M, Clendinneng D, Ogilvie S, Ann MM. Improving interprofessional competence in undergraduate students using a novel blended learning approach. J Interprof Care 2012 Jul;26(4):312-318. [doi: 10.3109/13561820.2012.660286] [Medline: 22559163]
7. Veltman M, Connor K, Honey M, Diener S, Bodily D. Collaborative practice through simulations in a multiuser virtual environment. Comput Inform Nurs 2012 Feb;30(2):63-67. [doi: 10.1097/NXN.0b013e31824a8b89] [Medline: 22367490]
8. Liaw SY, Carpio GAC, Lau Y, Tan SC, Lim WS, Goh PS. Multiuser virtual worlds in healthcare education: A systematic review. Nurse Educ Today 2018 Jun;65:136-149. [doi: 10.1016/j.nedt.2018.01.006] [Medline: 29571002]
9. Liaw SY, Chan SW, Scherpbier A, Rethans J, Pua GG. Recognizing, responding to and reporting patient deterioration: transferring simulation learning to patient care settings. Resuscitation 2012 Mar;83(3):395-398. [doi: 10.1016/j.resuscitation.2011.08.021] [Medline: 21946262]
10. Luigi Ingrassia P, Ragazzoni L, Carenzo L, Colombo D, Ripoll Gallardo A, Della Corte F. Virtual reality and live simulation: a comparison between two simulation tools for assessing mass casualty triage skills. Eur J Emerg Med 2015 Apr;22(2):121-127. [doi: 10.1097/MEJ.0000000000000132] [Medline: 24841770]
11. Liaw SY, Chan SW, Chen F, Hooi SC, Siau C. Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: randomized controlled trial. J Med Internet Res 2014 Sep 17;16(9):e214 [FREE Full text] [doi: 10.2196/jmir.3322] [Medline: 25230684]
12. Liaw SY, Wong LF, Lim EYP, Ang SBL, Mujumdar S, Ho JTY, et al. Effectiveness of a Web-Based Simulation in Improving Nurses' Workplace Practice With Deteriorating Ward Patients: A Pre- and Postintervention Study. J Med Internet Res 2016 Feb 19;18(2):e37 [FREE Full text] [doi: 10.2196/jmir.5294] [Medline: 26895723]
13. Driscoll M, Carliner S. Advanced Web-Based Training Strategies: Unlocking instructionally sound online learning. San Francisco, CA: Pfeiffer; 2005.
14. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The Effectiveness of Blended Learning in Health Professions: Systematic Review and Meta-Analysis. J Med Internet Res 2016 Jan 04;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]
15. Ayala JS. Blended learning as a new approach to social work education. Journal of Social Work Education 2013 Mar 16;45(2):277-288. [doi: 10.5175/JSWE.2009.200700112]
16. Schalk L, Schumacher R, Barth A, Stern E. When problem-solving followed by instruction is superior to the traditional tell-and-practice sequence. Journal of Educational Psychology 2018 May;110(4):596-610. [doi: 10.1037/edu0000234]
17. Khanal P, Vankipuram A, Ashby A, Vankipuram M, Gupta A, Drumm-Gurnee D, et al. Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles. J Biomed Inform 2014 Oct;51:49-59 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.005] [Medline: 24732098]
18. Grundy SE. The confidence scale: development and psychometric characteristics. Nurse Educ 1993;18(1):6-9. [Medline: 8487987]
19. King G, Shaw L, Orchard CA, Miller S. The interprofessional socialization and valuing scale: a tool for evaluating the shift toward collaborative care approaches in health care settings. Work 2010;35(1):77-85. [doi: 10.3233/WOR-2010-0959] [Medline: 20164627]
20. Kim K, Ko J. Attitudes toward interprofessional health care teams scale: a confirmatory factor analysis. J Interprof Care 2014 Mar;28(2):149-154. [doi: 10.3109/13561820.2013.857645] [Medline: 24329713]
21. Kala S, Isaramalai S, Pohthong A. Electronic learning and constructivism: a model for nursing education. Nurse Educ Today 2010 Jan;30(1):61-66. [doi: 10.1016/j.nedt.2009.06.002] [Medline: 19573956]
22. Pritchard A. Ways of learning: Learning theories and learning styles in the classroom. Abingdon, United Kingdom: Routledge; 2019.
23. Chen G, Nurkhamid, Wang C, Yang S, Lu W, Chang C. Digital Learning Playground: supporting authentic learning experiences in the classroom. Interactive Learning Environments 2013 Apr;21(2):172-183. [doi: 10.1080/10494820.2012.705856]
24. Jonassen DH, Rohrer-Murphy L. Activity theory as a framework for designing constructivist learning environments. ETR&D 1999 Mar;47(1):61-79. [doi: 10.1007/BF02299477]
25. Kim MC, Hannafin MJ. Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. Computers & Education 2011 Feb;56(2):403-417. [doi: 10.1016/j.compedu.2010.08.024]
26. Vygotsky LS, Cole M. Mind in Society: The Development of Higher Psychological Processes. Cambridge: Harvard University Press; 1978.
27. Ignacio J, Dolmans D, Scherpbier A, Rethans J, Chan S, Liaw S. Comparison of standardized patients with high-fidelity simulators for managing stress and improving performance in clinical deterioration: A mixed methods study. Nurse Educ Today 2015 Dec;35(12):1161-1168. [doi: 10.1016/j.nedt.2015.05.009] [Medline: 26047602]

Abbreviations

ATIHCT: Attitudes Toward Interprofessional Health Care Teams
ISVS: Interprofessional Socialization and Valuing Scale
SE: simulation exercise
VR: virtual reality
WI: Web-based instruction

Edited by G Eysenbach; submitted 22.10.18; peer-reviewed by Y Lau, S Chan; comments to author 21.03.19; revised version received 26.03.19; accepted 30.03.19; published 28.05.19.

Please cite as:
Liaw SY, Tan KK, Wu LT, Tan SC, Choo H, Yap J, Lim SM, Wong L, Ignacio J
Finding the Right Blend of Technologically Enhanced Learning Environments: Randomized Controlled Study of the Effect of Instructional Sequences on Interprofessional Learning
J Med Internet Res 2019;21(5):e12537
URL: http://www.jmir.org/2019/5/e12537/
doi:10.2196/12537
PMID:31140432

Original Paper

Health Information Technologies in the Support Systems of Pregnant Women and Their Caregivers: Mixed-Methods Study

Marian Taylor Dorst[1], BA; Shilo H Anders[2], PhD; Sai Chennupati[1]; Qingxia Chen[2], PhD; Gretchen Purcell Jackson[2], MD, PhD

[1]Vanderbilt University, Nashville, TN, United States

[2]Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Marian Taylor Dorst, BA
Vanderbilt University
2201 West End Ave
Nashville, TN, 37235
United States
Phone: 1 8582005906
Email: DorstM@calmedu.org

Abstract

Background: The quality and quantity of families' support systems during pregnancy can affect maternal and fetal outcomes. The support systems of expecting families can include many elements, such as family members, friends, and work or community groups. Emerging health information technologies (eg, social media, internet websites, and mobile apps) provide new resources for pregnant families to augment their support systems and to fill information gaps.

Objective: This study sought to determine the number and nature of the components of the support systems of pregnant women and their caregivers (eg, family members) and the role of health information technologies in these support systems. We examined the differences between pregnant women's support systems and those of their caregivers and the associations between support system composition and stress levels.

Methods: We enrolled pregnant women and caregivers from advanced maternal-fetal and group prenatal care clinics. Participants completed surveys assessing sociodemographic characteristics, health literacy, numeracy, and stress levels and were asked to draw a picture of their support system. Support system elements were extracted from drawings, categorized by type (ie, individual persons, groups, technologies, and other) and summarized for pregnant women and caregivers. Participant characteristics and support system elements were compared using the Pearson chi-square test for categorical variables and Wilcoxon ranked sum test for continuous variables. Associations between support system characteristics and stress levels were measured with Spearman correlation coefficient.

Results: The study enrolled 100 participants: 71 pregnant women and 29 caregivers. The support systems of pregnant women were significantly larger than those of caregivers—an average of 7.4 components for pregnant women and 5.4 components for caregivers (P=.003). For all participants, the most commonly reported support system elements were individual persons (408/680, 60.0%), followed by people groups (132/680, 19.4%), technologies (112/680, 16.5%), and other resources (28/680, 4.1%). Pregnant women's and caregivers' technology preferences within their support systems differed—pregnant women more often identified informational websites, apps, and social media as parts of their support systems, whereas caregivers more frequently reported general internet search engines. The size and components of these support systems were not associated with levels of stress.

Conclusions: This study is one of the first demonstrating that technologies comprise a substantial portion of the support systems of pregnant women and their caregivers. Pregnant women more frequently reported specific medical information websites as part of their support system, whereas caregivers more often reported general internet search engines. Although social support is important for maternal and fetal health outcomes, no associations among stress, support system size, and support system components were found in this study. As health information technologies continue to evolve and their adoption increases, their role in patient and caregiver support systems and their effects should be further explored.

(J Med Internet Res 2019;21(5):e10865) doi:10.2196/10865

KEYWORDS

pregnancy; social networks; social media; health information technology; caregivers; life stress

Introduction

Background

Significant and meaningful social support has been generally accepted as a means of positively affecting a pregnant mother's well-being. Specifically, it mitigates her stress, improves her eventual birth outcomes, and buffers against prenatal and postpartum depression [1-5]. Social support also has functional value as members of the support system can provide material and physical assistance as well as emotional and informational feedback [6,7]. Although some literature has suggested that the amount and quality of social support does not improve a pregnant mother's stress coping mechanisms or birth outcomes, a positive association between social support and mother's and child's well-being has prevailed in many studies [8,9]. There are many types of social support that contribute to a pregnant mother's well-being. Social relationships describe an individual's more immediate social ties, including family members, marriage partners, and close friends [10]; social resources are those relationships that are identified by an individual as helpful [11]. This important distinction differentiates between a pregnant mother's social connections and the subset of these personal connections whose support she finds most valuable and the least stress inducing. The combination of a pregnant woman's social relationships and social resources creates her unique social support system that helps her navigate pregnancy [12].

Not only is the physical presence of a social support system important but also is the perceived support of this system. Perceived social support has been shown to have a positive effect on stress levels and quality of life [13,14]. Specifically for pregnant women, perceived social support has been shown to be increased when using health information technologies such as online blogs and Facebook [13,15]. This greater perceived social support leads to higher perceptions of individual empowerment and improved maternal well-being, especially when looking at the variables of stress, depression, and marital satisfaction and conflict [16,17].

To build a strong and well-perceived support system, pregnant women may employ the help of caregivers to assist with their emotional and physical needs during this life stage [14,18]. In the context of this study, a caregiver is any informal, that is, any nonprofessional assistant who provides care for the pregnant woman [19,20]. These caregivers could include individuals such as spouses, parents, or friends, and they are an invaluable source of support [18,20]. However, it is important to recognize that providing this support can cause stress for the caregivers themselves. Previous research has shown that caregivers can experience negative health outcomes and psychological distress while fulfilling the caregiver role [20]. Therefore, it is necessary to investigate the support systems of both pregnant women and their caregivers as both these groups may experience stress during the pregnancy.

As consumer health information technologies have proliferated and evolved, they have also become popular resources for pregnancy support. According to a recent study of 613 women from throughout the world, 97% of pregnant mothers with internet access used general search engines to seek pregnancy-related information for a variety of reasons, including unsatisfactory experiences with health care providers, social networking, and general pregnancy information [21]. Women's confidence in making pregnancy-related decisions can increase significantly after consulting internet sources. Researchers have observed that smartphone apps and social media platforms combine the "expert patient" ideal, the desire to feel up to date with all available health information on the internet, with the ideologies of responsible motherhood [22]. Many of these apps are considered useful by mothers because they *push* content. Instead of mothers having to seek out relevant information, pregnancy apps often send regular emails and updates to their users about temporal fetal development and the stages of pregnancy. There have also been recent albeit limited efforts to create health-related technologies for expectant fathers and other caregivers of the pregnant mothers. For example, an Australian app, mPregnancy, designed for men, includes more masculine-oriented descriptions of the phases of fetal development and stages of pregnancy [22].

Goal of This Study

Little has been published about the presence and role of health information technologies in the support networks of pregnant women and their caregivers. As part of a comprehensive study of health information needs and information management practices in pregnant families, we aimed to elicit the components of the social networks of pregnant women and their caregivers, the types of technologies considered part of this support system, and any relationships between social network characteristics and stress. With these data, this study sought to identify the prevalence of technology in the social networks of pregnant women and their caregivers to expose a potential need for increased and improved technological resources for this population.

Methods

Study Population

We enrolled 100 pregnant women and their caregivers from 2 diverse prenatal care settings at the Vanderbilt University Medical Center (VUMC): the Junior League Fetal Center at Vanderbilt (FCV) and the Expect with Me (EWM) group prenatal care program. The FCV is an advanced maternal-fetal care setting that incorporates a clinical program in fetal diagnosis and therapy. Multidisciplinary teams bring a group of expert medical providers from different specialties to deliver care at 1 location. Most patients seen at the FCV experience a pregnancy with a fetal anomaly or other complications of pregnancy. EWM is an innovative group prenatal care program that combines the components of traditional prenatal care with health education and support delivered in a group setting. Small groups of pregnant women (ie, 8-12 individuals) with similar gestational ages and their caregivers meet for 10 group sessions during pregnancy. This model is aligned with the Institute of Medicine's six domains of quality and has been demonstrated to reduce preterm births and health care costs during and after birth [23]. Inclusion of these 2 settings ensured that this study included both normal pregnancies and pregnancies with complications.

Adult pregnant women and then their caregivers were approached at both sites for participation in the study; if interested, they contacted us to set up an interview time. Pregnant women were approached for participation first. If she agreed to participate, she could then choose whether or not to include up to 3 caregivers in the study as well. Not all pregnant women chose to include caregivers in the study; nor did all pregnant women have caregivers available for participation. The eligibility criteria for all participants included age greater than or equal to 18 years, pregnancy with gestational age less than 36 weeks, home within 100 miles of VUMC, and the ability to speak English or Spanish. All participants provided written informed consent. This study was approved by the VUMC institutional review board.

Measures

This research project contains analyses of data from a comprehensive study of information needs and information management practices in pregnant women and their caregivers. All participants completed an individual research visit to VUMC during which they completed surveys and participated in an interview, all of which typically lasted between 1 and 2 hours. For Spanish-speaking participants, all research materials were translated using a forward and backward translation process into Spanish, and a Spanish interpreter was present for the entire research visit. Participants were compensated for their time with a US $25 gift card and reimbursement for their travel to and from VUMC. The measures relevant to this specific research project are described below.

The participants completed a demographic questionnaire assessing age, race, ethnicity, home location, marital status, parity (number of children previously born), employment status, individual and household income, education, and access to various technologies. We recorded the gestational age of the pregnancy at the date of the interview and the relationship between the pregnant woman and any participating caregivers. Additional surveys were used to assess health literacy, numeracy, and levels of stress. For English-speaking participants, health literacy was measured using the Rapid Estimate of Adult Literacy in Medicine (REALM), a well-validated and widely used measure of health literacy [24,25]. For Spanish-speaking participants, health literacy was assessed with the Short Assessment of Health Literacy for Spanish-speaking Adults, which is based on and highly correlated with REALM [26]. Numeracy was measured by the General Health Numeracy Test, a 6-item validated questionnaire for assessing general health numeracy [27]. Stress was determined using the Cohen Perceived Stress Scale, a widely used psychological instrument for measuring the perception of stress that has demonstrated high reliability in pregnant populations [28,29] and has been translated into Spanish with adequate fit [30,31]. This instrument is a questionnaire in which various feelings are assessed on a 5-point scale, ranging from never (0) to almost always (4). Positively worded items are reverse scored, and the ratings are averaged, such that a higher score indicates more perceived stress. We examined stress levels because in the pregnant population, high stress levels have been associated with decreased birth weight of babies as well as preterm delivery [32] and because of reported associations between social networks and stress during pregnancy. Studies involving other health-related diagnoses have found correlations between meeting information needs and reductions in stress [33].

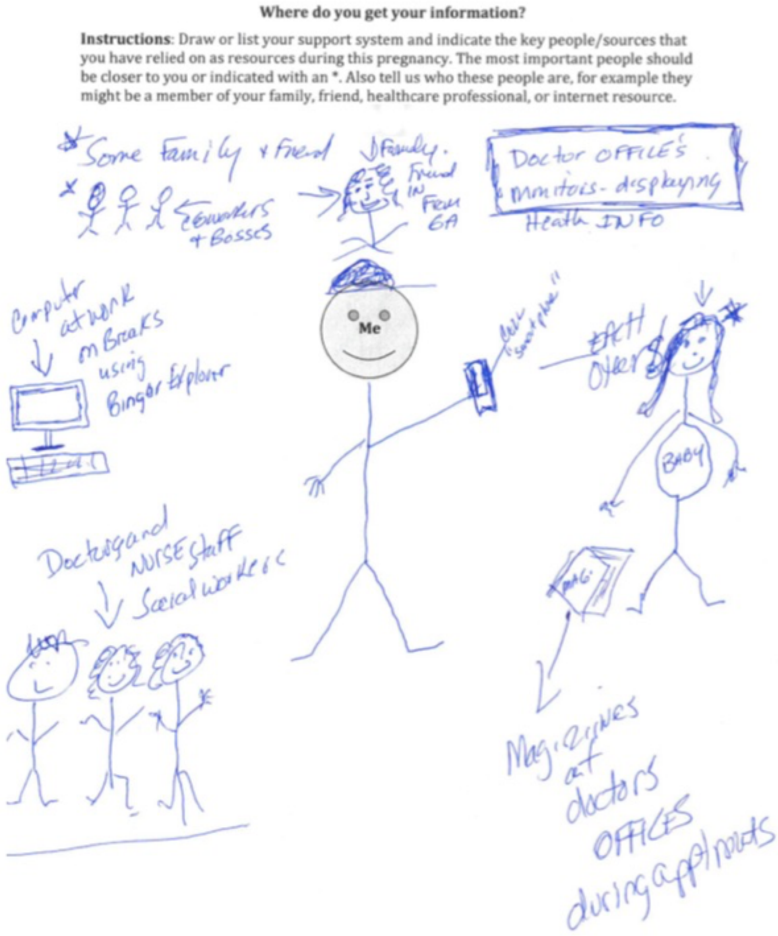
Support resources were assessed by giving participants a sheet of paper with a face in the center, representing the participant, and the following instructions:

> *Draw or list your support system and indicate the key people/sources that you have relied on as resources during this pregnancy. The most important people should be closer to you or indicated with an *. Also, tell us who these people are. For example, they might be a member of your family, friend, healthcare professional, or internet resource (e.g., contacts from blogs, social networking sites, or other web or mobile application resources).*

This method for gathering this information is in line with participatory design methods and has been used successfully in other studies [34]. The interviewer would read the instructions and answer any questions that the participant had about this measure.

An example of a caregivers' support network is shown in Figure 1.

Figure 1. Example of caregiver support network.



Where do you get your information?

Instructions: Draw or list your support system and indicate the key people/sources that you have relied on as resources during this pregnancy. The most important people should be closer to you or indicated with an *. Also tell us who these people are, for example they might be a member of your family, friend, healthcare professional, or internet resource.

Analyses

From each social resource drawing, research team members extracted the total number of elements and determined the number of individual people, groups of people, technologies, and other elements. The independent variable we noted was the role of each person (eg, husband, sister, or mother), and the dependent variables were the name or type of each group and the types of technologies. Subtypes of technologies were categorized as general internet search engines (eg, Google), educational or informational resources (eg, WebMD), apps (eg, BabyBump), or social media (eg, Facebook).

For continuous characteristics, we summarized the data using mean and SD, and we compared these variables between pregnant women and caregivers with the use of the Wilcoxon rank sum test. For categorical characteristics, we reported the percentages of each category and compared between roles using the Pearson chi-square test. Stress level was evaluated by 10 items, with each given a value ranging from 0 to 4. The internal consistency of items was measured by the Cronbach coefficient. With good to excellent internal consistency (Cronbach alpha >.8), the stress score would be calculated as the average of nonmissing 10-item values. Associations among support system size, components, and stress were examined with the Spearman correlation for continuous variables and Pearson chi-square test for categorical variables. The 95% CI was constructed for Spearman correlation coefficient based on 5000 bootstrap samples. The 2-sided *P* values less than .05 were considered significant. All analyses were performed using R3.4.4 (the R Project for Statistical Computing). These analyses were conducted using the lens of grounded theory: empirical data are coded, categorized, and interpreted to recognize patterns and theories [35].

Results

Population

This study enrolled a total of 100 participants, including 71 pregnant women and 29 caregivers. For the 29 caregivers, the relationships with the mother consisted of 11 spouses, 7 significant others, 7 parents, 2 siblings, 1 child older than 18 years, and 1 ex-husband. The characteristics of the study population are shown in Table 1. Overall, 83 participants were recruited from the FCV and 17 from EWM. In addition, 79.0% (79/100) of our sample was female and 21.0% (21/100) male. Pregnant participants had an average age of 27.6 years (SD 6.3 years), whereas caregivers were significantly older, with an average age of 33.9 years (SD 12.0 years; *P*=.04). Furthermore, 20.0% (20/100) of participants identified as black or African American, 71.0% (71/100) as white, 1.0% (1/100) as Native Hawaiian or other Pacific Islander, and 8.0% (8/100) as other. In addition, 5 participants (4 pregnant women and 1 caregiver) were Spanish speaking.

Table 1. Demographic characteristics of the study population.

Demographic characteristic	Pregnant women (n=71)	Caregivers (n=29)	Total (N=100)
Age, mean (SD)	27.6 (6.3)	33.9 (12.0)	29.4 (8.8)
Estimated gestational age, mean (SD)	30.8 (4.0)	31.1 (3.9)	30.9 (4.0)
Race, n (%)			
American Indian or Alaska Native	0 (0)	0 (0)	0 (0)
Asian	0 (0)	0 (0)	0 (0)
Black or African American	14 (20)	6 (21)	20 (20)
Native Hawaiian or other Pacific Islander	1 (1)	0 (0)	1 (1)
White	50 (70)	21 (72)	71 (71)
Other	6 (8)	2 (7)	8 (8)
Gender, n (%)			
Female	71 (100)	8 (28)	79 (79)
Male	0 (0)	21 (72)	21 (21)
Other children, n (%)			
Yes	36 (51)	17 (59)	53 (53)
No	35 (49)	12 (41)	47 (47)
Education level, n (%)			
Eighth grade	0 (0)	1 (3)	1 (1)
High school	22 (31)	11 (38)	34 (34)
Some college courses	17 (24)	8 (28)	25 (25)
2-year degree	6 (8)	2 (7)	8 (8)
4-year degree	14 (20)	5 (18)	19 (19)
Master’s degree	7 (10)	1 (3)	8 (8)
PhD or equivalent	2 (3)	0 (0)	2 (2)
Other	3 (4)	1 (3)	3 (3)
Individual income, n (%)			
Under US $15,000	34 (51)	7 (26)	41 (41)
US $15,000-US $29,999	12 (18)	9 (33)	21 (21)
US $30,000-US $44,999	8 (12)	6 (22)	14 (14)
US $45,000-US $59,999	4 (6)	2 (7)	6 (6)
US $60,000-US $79,999	5 (7)	1 (4)	6 (6)
US $80,000-US $99,999	2 (3)	1 (4)	3 (3)
US $100,000 and above	2 (3)	1 (4)	3 (3)
Household income, n (%)			
Under US $15,000	13 (19)	6 (22)	19 (19)
US $15,000-US $29,999	13 (19)	7 (26)	20 (20)
US $30,000-US $44,999	11 (16)	7 (26)	18 (18)
US $45,000-US $59,999	9 (13)	4 (15)	13 (13)
US $60,000-US $79,999	5 (8)	0 (0)	5 (5)
US $80,000-US $99,999	4 (6)	0 (0)	4 (4)
US $100,000-US $119,000	5 (8)	0 (0)	5 (5)
US $120,000-US $139,000	2 (3)	2 (7)	4 (4)
US $140,000 and above	5 (8)	1 (4)	6 (6)

Demographic characteristic	Pregnant women (n=71)	Caregivers (n=29)	Total (N=100)
Stress, mean (SD)	1.4 (0.6)	1.7 (0.7)	1.6 (0.7)
Health literacy, mean (SD)	62.6 (5.2)	61.1 (6.4)	62.2 (5.6)

The average gestational age of the fetus at the time participants were interviewed was 30.9 weeks, 30.8 weeks for pregnant participants, and the pregnant woman associated with the caregiver was 31.1 weeks' average gestational age at the time of their caregiver's interview. Overall, 53.0% (53/100) of participants had children before the pregnancy looked at in this study. Health literacy scores (calculated out of 66 maximum points) averaged 62.2 for all study participants, 62.6 for pregnant women, and 61.1 for caregivers. A score of 61 or greater correlates to a ninth-grade literacy level. Individuals at this literacy level will be able to read most health education materials [24,25]. Health numeracy scores (calculated out of 6 maximum points) averaged 2.2 for all participants, 2.1 for pregnant women, and 2.3 for caregivers. Stress levels (range 0-4) averaged 1.6 for all participants, 1.4 for pregnant women, and 1.7 for caregivers. Between pregnant women and their caregivers, there were no significant differences in any of the variables (age, race, gender, etc) shown in Table 1 (range of *P*=.06 for stress levels between the groups to *P*=.92 for difference in race between the 2 groups).

Technology Access

Table 2 presents the technologies that pregnant women and their caregivers reported having access to at the time of their participation. The most participants in both groups reported having smartphones, including 96% (68/71) of pregnant women and 93% (27/29) of caregivers. In both groups, fewer participants had home phones, specifically 37% (26/71) of pregnant women and 41% (12/29) of caregivers. Within the caregiver subset, there was a difference in access to technologies based on the caregiver's sex. Overall, female caregivers reported lower rates of technology access across all categories. Female caregivers had the least access to home phones and gaming consoles, with 25% (2/8) reporting access in both categories, whereas male caregivers had the least access to just home phones (10/21, 48%).

Support System Components

Tables 3 and 4 present the types and numbers of support system components for pregnant women and their caregivers. The 100 participants reported a total of 680 support system components. Individual persons were the most common support system elements, comprising 60.3% (408/680) of the support system for all participants, 58.4% (307/525) for pregnant women, and 57.7% (90/156) for all caregivers. Among the individuals cited, first-degree relatives such as spouses, parents, or siblings were the most common, although many individuals also included specific health care providers or named friends. The next most common support component was groups of people, making up 19.3% (132/680) of the support system for the entire sample, 21.0% (110/525) for pregnant women, and 21.2% (35/156) for caregivers. Friends, especially those with children, were one of the most common groups in support networks. Other commonly mentioned groups included colleagues, health care practices, and members of one's church.

Both groups identified technologies as parts of their support system: 16.2% (112/680) for all participants, 16.8% (88/525) for pregnant women, and 16.0% (25/156) for caregivers. Google (20/112, 17.9% technology components), internet searches (21/112, 18.8%), and BabyCenter (14/112, 12.5%) were the most common support network components, but a wide variety of health information technologies were identified, including blogs, online support groups, and online journals. Finally, other resources comprised the smallest part of the support system: 4.2% (28/680) of the system for all participants, 3.8% (20/525) for pregnant women, and 5.1% (8/156) for caregivers. In this category, magazines and books, most notably the book *What to Expect When You're Expecting*, were the most common elements.

Table 2. Pregnant women's and caregivers' current access to technologies.

Technology	All participants (N=100), n (%)	Pregnant Women (n=71), n (%)	Caregivers (n=29), n (%)	Female caregivers (n=8), n (%)	Male caregivers (n=21), n (%)
Computer at home	81 (81)	59 (83)	22 (76)	5 (63)	17 (81)
Computer at work	63 (63)	42 (59)	21 (72)	3 (38)	18 (86)
Smartphone	95 (95)	68 (93)	27 (93)	7 (89)	20 (95)
Cell phone	92 (92)	66 (93)	26 (90)	6 (75)	20 (95)
Home phone	38 (38)	26 (37)	12 (41)	2 (25)	18 (86)
Gaming console	66 (66)	46 (65)	20 (69)	2 (25)	10 (48)
Tablet	68 (68)	49 (69)	19 (66)	3 (38)	16 (76)

Table 3. Types of support system components.

Participants	Support system components									
	Persons		Groups		Technology		Others		Totals	
	Total, n (%)	Mean (SD)	Total, n (%)	Mean (SD)	Total, n (%)	Mean (SD)	Total, n (%)	Mean (SD)	Total, n	Mean (SD)
All participants (N=100)	408 (60.0)	3.97 (2.98)	132 (19.4)	1.45 (1.48)	112 (16.5)	1.13 (1.19)	28 (4.1)	0.28 (0.51)	680	6.8 (3.4)
Pregnant women (n=71)	307 (58.4)	4.32 (3.32)	110 (21.0)	1.55 (1.50)	88 (16.8)	1.24 (1.21)	20 (3.8)	0.28 (0.54)	525	7.4 (3.40)
Caregivers (n=29)	90 (57.7)	3.10 (1.80)	35 (22.4)	1.13 (1.38)	25 (16.0)	0.86 (1.13)	8 (5.1)	0.28 (0.45)	156	5.4 (3.0)
Female (n=8)	16 (57.1)	2.00 (0.93)	9 (32.1)	1.13 (0.32)	1 (3.6)	0.13 (0.35)	2 (7.1)	0.25 (0.46)	28	3.5 (2.0)
Male (n=21)	74 (57.8)	3.52 (1.89)	24 (18.8)	1.14 (1.20)	24 (18.8)	1.14 (1.20)	6 (4.7)	0.29 (0.46)	128	6.1 (3.0)

Table 4. Technology support system components.

Participants	Technology support system components				
	General internet search engines, n (%)	Informational or educational websites, n (%)	Apps, n (%)	Social media tools, n (%)	Total, n (%)
All participants (N=100)	40 (37.3)	45 (42.1)	10 (9.3)	12 (11.2)	107 (100)
Pregnant women (n=71)	29 (33.7)	36 (41.9)	10 (11.6)	11 (12.8)	86 (100)
Caregivers (n=29)	11 (50)	9 (40.9)	0 (0)	2 (9.1)	22 (100)
Female caregivers (n=8)	1 (100)	0 (0)	0 (0)	0 (0)	1 (100)
Male caregivers (n=21)	10 (47.6)	9 (42.9)	0 (0)	2 (9.5)	21 (100)

When the caregiver subset was divided into men and women, there were some apparent differences in the support system composition. As seen in Table 3, male and female caregivers showed a similar proportion of persons in their support system but differed in their proportions of groups, technology, and other elements. Female caregivers' support systems included a greater percentage of groups (9/28, 32.1%) than did the support systems of males (24/128, 18.8%), and men reported a greater proportion of technology in their support systems (24/128, 18.8%) than did women (1/28, 3.6%). Men also showed a greater mean number of support elements in every category. Pregnant women had slightly higher mean number of components, mainly from persons and technology. For caregivers, the most notable differences between the genders were within the categories of persons and technology. Male caregivers' support systems, on average, included 3.52 support persons and 1.14 technology elements, whereas female caregivers' support systems consisted of 2.00 persons and 0.13 technology elements.

Associations With Stress

We examined whether the participant characteristics, including the social network size and components, were associated with levels of stress. Stress was not found to significantly correlate with age (Spearman rho=−0.176; 95% CI −0.361 to 0.021; *P*=.08), race (*P*=.93), household income (Spearman rho=−0.196; 95% CI −0.395 to 0.008; *P*=.06), or education level (Spearman rho=−0.111; 95% CI −0.286 to 0.07; *P*=.27) within the entire sample. None of these variables within the pregnant women and caregiver data subsets significantly correlated with stress (all *P*>.05). In addition, none of the support system components described in Table 3 significantly correlated with stress; among all participant data, pregnant women subset data, and caregiver subset data, there was no significant correlation between stress and the number of persons, groups, technology, and other resources in an individual's support system. The composition of an individual's support system (ie, percentage of support system distributed across persons, groups, technology, and other support components) also did not correlate significantly with stress. Only 1 variable, individual income, correlated with stress level (Spearman rho=−0.229; 95% CI −0.425 to −0.026; *P*=.03).

Technology Support System Components

Table 4 presents the types of technologies pregnant women and their caregivers reported as part of their support systems. Among all the participants, informational websites were the most popular form of technology support resources (45/107, 42.1%), followed by general internet search engines (40/107, 37.3%), social media tools (12/107, 11.2%), and apps (10/107, 9.3%). Among pregnant women, informational websites are the most popular form of technological support (36/86, 41.9%), followed by general internet search engines (29/86, 33.7%), social media tools (11/86, 12.8%), and apps (10/86, 11.6%). Caregivers more frequently reported general internet search engines (11/22, 50%) as part of their support systems, followed by informational websites (9/22, 40.9%), social media tools (2/22, 9.1%), and then apps (0/22, 0%). Consistent with Table 2, only 1 female caregiver reported a technology support element (general internet search engine), whereas male caregivers reported support using general internet search engines (10/21, 47.6%), informational websites (9/21, 42.9%), and social media tools (2/21, 9.5%).

Discussion

Principal Findings

This study is one of the first to examine the presence of health information technologies as part of the support systems of pregnant women and their caregivers [15-17,22]. We observed that both pregnant women and a diverse set of caregivers reported technologies as a substantial component of the support systems for pregnancy. Pregnant women more frequently cited medical informational sites rather than broad internet search engines as components of their support system, with social media and apps identified less frequently. Caregivers more often included general internet search engines in their support networks, followed by informational websites, social media, and apps. These differences suggest that pregnant women may have more specific questions regarding their own pregnancy and the medical conditions of both themselves and the fetus. Therefore, they seek more specific advice from reputable informational websites such as hospital or government sites rather than search broadly using sites such as Google. Previous research has suggested that prenatal visits are ineffective in addressing the information needs of pregnant women, and therefore, they may turn to technologies and online resources to fill these gaps [36].

We also noted that male caregivers reported more individual and technology support system components than female caregivers did. This finding may expose a current need for pregnancy-related technology and resources specifically designed for male caregivers as most pregnancy resources are likely tailored for expectant mothers and a female audience. Previous research has found that during pregnancy, male caregivers are often portrayed as the “bumbling assistants,” and thus, the resources available to them are written with oversimplified information [22]. In actuality, male caregivers may wish to play a more serious and informed role, and current technology resources do not support this role. Our analysis of information needs reported by participants in this study revealed that male caregivers do indeed want information about the phases of pregnancy, normal versus abnormal symptoms of a pregnant woman, and other relevant information to assist the pregnant individual [37]. However, currently available information may be written primarily with a pregnant audience in mind, which could limit the understanding of a male caregiver and reinforce the “bumbling assistant” stereotype [22]. Fathers often feel excluded from the prenatal education process [38], and our findings might suggest that our caregivers sought additional information and support through a greater number of persons and technologies in their network. Male caregivers also reported greater access to all forms of technology than did female caregivers, so their greater inclusion of technology-based resources as support system components may have been observed simply because of this greater access.

Overall, pregnant women had larger support systems than their caregivers, suggesting that expectant mothers seek larger and broader support networks during the vulnerable life stage of pregnancy. Pregnant women identified more persons, groups, and technologies as parts of their support systems than did caregivers. This finding supports the results of previous research, which demonstrated that first-time mothers reported more numerous sources of support than their husbands [38]. This may insinuate that a pregnant woman is offered more help during her pregnancy through a variety of personal and informational resources. Alternatively, she may simply be more receptive of the aid provided during this time in her life.

In our study, women also valued information and support from experienced mothers and women who had recently given birth. One pregnant participant in our sample reported that she regularly sought information on Facebook from her high school friends who lived in another state; they were also pregnant and having children, and they would provide information tailored to her. This finding supports previous research that pregnant women who invested time into online blogs and Facebook groups reported feeling more supported in their transition to motherhood and more connected to extended family, friends, and the outside world [15-17]. Technologies such as social media may provide a way for both pregnant women and caregivers to obtain support from family and friends who are geographically remote.

Female caregivers, such as paternal and maternal grandmothers of the child, have been shown to play a central role in pregnancy decision making and in postnatal decisions such as breastfeeding by sharing their experience [39]. Female caregivers such as grandmothers are often seen as possessors of time-tested knowledge, and thus, they are able to significantly influence a pregnant woman, both positively and negatively [39]. One caregiver in our study, who was a mother to the pregnant participant, reported using social media to share her knowledge and experience with expectant mothers within and outside her family. As older age groups continue to adopt and use social media, this type of support may become more commonplace. However, in our study, we see that both pregnant women and female caregivers reported people-based resources such as individuals and groups in their support systems much more frequently than technologies. Female caregivers may both contribute toward and benefit from prenatal group-style programs such as EWM where they can share, compare, and gain knowledge. The predominance of persons in the support networks of pregnant women may have been affected by recruitment from such a group prenatal care setting, giving these participants and their caregivers greater access to persons with shared experiences to build their support networks.

Both pregnant women and their caregivers also indicated that groups comprised a substantial portion of their support systems. Many cited groups in support networks were large, for example, a church congregation or a physician group/midwifery team. Previous studies have shown that church attendance and involvement in clubs during pregnancy were not associated with increased perceived social support [40,41]. The high prevalence of religious and health care groups observed in the support networks in our study may have been a reflection of cultural norms in the southeastern United States and the intensive care provided in both of our recruitment sites.

Pregnant women more frequently noted social media tools as part of their support networks. The social media category includes interactive information searching, such as online forums, blogs, Facebook, and Pinterest, which may appeal more to the pregnant woman who seeks a more tailored, immediate, or relatable response to a question. The role of social media in meeting pregnancy information needs and augmenting a pregnant woman's support network has not been extensively explored, but the potential health benefits to extending social networks through technologies are numerous. It is well known that health behaviors and conditions such as smoking and obesity are spread through in-person social networks, but the exact nature of how such health states are propagated has yet to be elucidated fully [41,42]. Positive health-related behaviors, such as smoking cessation and happiness, have also been shown to spread through social networks [43,44]. Online social networks and social media tools, such as Facebook and Twitter, are communication tools by which social ties can be formed, strengthened, and maintained, which can increase perceived social support [13,14]. One could envision pregnant women and their caregivers obtaining social support for important perinatal activities such as exercise and breastfeeding through health information technologies. No caregivers acknowledged using any sort of app in their support system, and only 1 caregiver reported the use of 2 social media tools. This supports the notion that most pregnancy-related apps are more directly targeted toward pregnant women rather than toward their caregivers.

Our findings correspond with a fundamental tenet of social network theory: when an individual is exposed to components of her social network, she may adjust her behavior via mechanisms such as social pressure and information gathering [45]. Though our study was not designed to collect data within a social network theory framework, it is interesting to consider how the strength, quality, and variety of our participants' support systems may influence information-seeking behaviors.

Unlike previous studies, we did not see any relationship between the size or nature of the support system and stress [13]. The generally large breadth of our pregnant women's and caregivers' support systems may have offered the emotional and informational support that could counter stress. Our only variable that did significantly correlate with stress, individual income, is an external variable that is not easily stabilized or impacted by a support system. In addition, this research project was a cross-sectional study that examined support networks and stress before delivery, and both may change significantly after the birth of the child. The stress levels observed in our population were relatively low, and it is also possible that our sample of 100 participants may not be large enough to observe such a correlation between support system resources and stress. Our ongoing research includes a similar longitudinal study, which will examine changes in support systems and stress levels at various points during pregnancy and after delivery.

Limitations

Our study has several limitations. First, only English- and Spanish-speaking participants were included, and relatively few Spanish-speaking individuals participated. The demographics of our population reflected those of the state of Tennessee and nearby communities, and our findings may not generalize to other settings. The size and nature of support systems likely have cultural influences, and our study may have excluded other populations with different types of social networks. Our study instrument also provided participants with potential examples that may have impacted their responses when completing this sheet. In addition, our study was cross-sectional, and the varying gestational ages of the participants could have influenced the responses given. For example, a pregnant woman at a gestational age of 34 weeks with delivery in the near future would consider her midwife and doctor as valuable support resources, but a pregnant woman at a gestational age of only 18 weeks may be more concerned with day-to-day support provided by family and colleagues. Our research team is currently conducting a longitudinal study in which support systems, technology usage, and stress will be evaluated during the course of pregnancy and after delivery. Finally, this study did not look more deeply into the types of support provided by each component of the support network. Therefore, we could not assess the quality of any support resource or make direct conclusions about its usefulness or purpose during the pregnancy experience of each participant.

Conclusions

Social support is critical for optimal maternal and fetal outcomes during pregnancy. This study is one of the first to demonstrate that technologies formed a substantial component of the support networks for pregnant women and their caregivers. Pregnant women more frequently cited specific medical informational or educational resources, whereas caregivers more commonly reported general internet search engines in their support networks. Male caregivers reported more access to technologies and identified them in their support networks more frequently than did female caregivers. Female caregivers' support networks during pregnancy included more individuals and groups of persons. Additional research is needed to explore exactly how technologies provide support to families during pregnancy and the effects that use of health information technologies might have on pregnancy outcomes.

Acknowledgments

This research was funded by the Agency for Health Research and Quality, Grant R01 HS021496. Study data were collected and managed using Research Electronic Data Capture tools hosted at VUMC and funded by grant UL1 TR000445 from NCATS/NIH [46]. This research was also supported in part by the Surgical Outcomes Center for Kids at Monroe Carell Jr Children's Hospital at Vanderbilt and through the Section for Surgical Sciences at VUMC. The authors are indebted to the many clinical staff, research coordinators, and trainees who participated in recruitment, coordination, and data collection for this study, including Laurie Novak, Christopher Simpson, Jamie Tice, Amy Rudin, Patty Jones, Tracy Perry, Kelly Bennett, Mary Anne Carroll, Deborah

Wage, Anne Pruett, Martha Dudek, Caitlin Gabarits, Jill Simmons, Ebone Ingram, Mary Masterman, Hannah Huth, and Sharon Davis.

Conflicts of Interest

None declared.

References

1. Collins NL, Dunkel-Schetter C, Lobel M, Scrimshaw SC. Social support in pregnancy: psychosocial correlates of birth outcomes and postpartum depression. J Pers Soc Psychol 1993 Dec;65(6):1243-1258. [Medline: 8295121]
2. Kessler R, McLeod J. Social support and mental health in community samples. In: Social Support and Health. San Diego: Academic Press; 1985:219-240.
3. Pagel MD, Smilkstein G, Regen H, Montano D. Psychosocial influences on new born outcomes: a controlled prospective study. Soc Sci Med 1990;30(5):597-604. [Medline: 2309138]
4. Li Y, Long Z, Cao D, Cao F. Social support and depression across the perinatal period: a longitudinal study. J Clin Nurs 2017 Sep;26(17-18):2776-2783. [doi: 10.1111/jocn.13817] [Medline: 28334472]
5. Morikawa M, Okada T, Ando M, Aleksic B, Kunimoto S, Nakamura Y, et al. Relationship between social support during pregnancy and postpartum depressive state: a prospective cohort study. Sci Rep 2015 May 29;5:10520 [FREE Full text] [doi: 10.1038/srep10520] [Medline: 26022720]
6. Israel BA. Social networks and health status: linking theory, research, and practice. Patient Couns Health Educ 1982;4(2):65-79. [Medline: 10258419]
7. Norwood S. Social support and pregnancy outcomes: the effects and interactions of formal and informal sources. Gonzaga University 1992:1.
8. Feldman PJ, Dunkel-Schetter C, Sandman CA, Wadhwa PD. Maternal social support predicts birth weight and fetal growth in human pregnancy. Psychosom Med 2000;62(5):715-725. [Medline: 11020102]
9. Hester S. Psychosocial factors in low and high risk pregnancy: impact on birth outcome. Pasadena, CA: Fuller Theological Seminary; 1998.
10. Gottlieb BH. Social networks and social support in community mental health. In: Gottlieb BH, editor. Social Networks and Social Support. Beverly Hills, CA: Sage Publications; 1981.
11. Barrera M. Social support in the adjustment of pregnant adolescents: assessment issues. In: Social Networks and Social Support. Beverly Hills: Sage Publications Inc; 1981:69-96.
12. Oakley A. Is social support good for the health of mothers and babies? J Reprod Infant Psychol 1988 Jan;6(1):3-21. [doi: 10.1080/02646838808404014]
13. Panayiotou G, Karekla M. Perceived social support helps, but does not buffer the negative impact of anxiety disorders on quality of life and perceived stress. Soc Psychiatry Psychiatr Epidemiol 2013 Feb;48(2):283-294. [doi: 10.1007/s00127-012-0533-6] [Medline: 22711064]
14. Stapleton LR, Schetter CD, Westling E, Rini C, Glynn LM, Hobel CJ, et al. Perceived partner support in pregnancy predicts lower maternal and infant distress. J Fam Psychol 2012 Jun;26(3):453-463 [FREE Full text] [doi: 10.1037/a0028332] [Medline: 22662772]
15. Holtz B, Smock A, Reyes-Gastelum D. Connected motherhood: social support for moms and moms-to-be on Facebook. Telemed J E Health 2015 May;21(5):415-421. [doi: 10.1089/tmj.2014.0118] [Medline: 25665177]
16. McDaniel BT, Coyne SM, Holmes EK. New mothers and media use: associations between blogging, social networking, and maternal well-being. Matern Child Health J 2012 Oct;16(7):1509-1517. [doi: 10.1007/s10995-011-0918-2] [Medline: 22094592]
17. Evans M, Donelle L, Hume-Loveland L. Social support and online postpartum depression discussion groups: a content analysis. Patient Educ Couns 2012 Jun;87(3):405-410. [doi: 10.1016/j.pec.2011.09.011] [Medline: 22019021]
18. Alio AP, Lewis CA, Scarborough K, Harris K, Fiscella K. A community perspective on the role of fathers during pregnancy: a qualitative study. BMC Pregnancy Childbirth 2013 Mar 07;13:60 [FREE Full text] [doi: 10.1186/1471-2393-13-60] [Medline: 23497131]
19. Brown M, Brown S. Informal caregiving: a reappraisal of effects on caregivers. Social Issues and Policy Review 2014 Jan 13;8(1):74-102. [doi: 10.1111/sipr.12002]
20. Roth DL, Fredman L, Haley WE. Informal caregiving and its impact on health: a reappraisal from population-based studies. Gerontologist 2015 Apr;55(2):309-319. [doi: 10.1093/geront/gnu177] [Medline: 26035608]
21. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
22. Johnson S. “Maternal Devices”, social media and the self-management of pregnancy, mothering and child health. Societies 2014 Jun 13;4(2):330-350. [doi: 10.3390/soc4020330]
23. Mooney S, Russell MA, Prairie B, Savage C, Weeks WB. Group prenatal care: an analysis of cost. J Health Care Finance 2008;34(4):31-41. [Medline: 21110479]

24. Davis TC, Crouch MA, Long SW, Jackson RH, Bates P, George RB, et al. Rapid assessment of literacy levels of adult primary care patients. Fam Med 1991 Aug;23(6):433-435. [Medline: 1936717]
25. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]
26. Lee SY, Bender DE, Ruiz RE, Cho YI. Development of an easy-to-use Spanish Health Literacy test. Health Serv Res 2006 Aug;41(4 Pt 1):1392-1412 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00532.x] [Medline: 16899014]
27. Osborn CY, Wallston KA, Shpigel A, Cavanaugh K, Kripalani S, Rothman RL. Development and validation of the General Health Numeracy Test (GHNT). Patient Educ Couns 2013 Jun;91(3):350-356 [FREE Full text] [doi: 10.1016/j.pec.2013.01.001] [Medline: 23433635]
28. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
29. Lobel M, DeVincent CJ, Kaminer A, Meyer BA. The impact of prenatal maternal stress and optimistic disposition on birth outcomes in medically high-risk women. Health Psychol 2000 Nov;19(6):544-553. [Medline: 11129357]
30. Ramírez MT, Hernández RL. Factor structure of the Perceived Stress Scale (PSS) in a sample from Mexico. Span J Psychol 2007 May;10(1):199-206. [Medline: 17549893]
31. Remor E. Psychometric properties of a European Spanish version of the Perceived Stress Scale (PSS). Span J Psychol 2006 May;9(1):86-93. [Medline: 16673626]
32. Sable MR, Wilkinson DS. Impact of perceived stress, major life events and pregnancy attitudes on low birth weight. Fam Plann Perspect 2000;32(6):288-294 [FREE Full text] [Medline: 11138865]
33. Czaja R, Manfredi C, Price J. The determinants and consequences of information seeking among cancer patients. J Health Commun 2003;8(6):529-562. [doi: 10.1080/716100418] [Medline: 14690888]
34. Utrankar A, Mayo-Gamble TL, Allen W, Novak L, Kassim AA, Bonnet K, et al. Technology use and preferences to support clinical practice guideline awareness and adherence in individuals with sickle cell disease. J Am Med Inform Assoc 2018 May 07;25(8):976-988. [doi: 10.1093/jamia/ocy036] [Medline: 29741695]
35. Charmaz K, Belgrave L. Grounded Theory. In: The Blackwell Encyclopedia of Sociology. Hoboken, NJ: John Wiley & Sons; 2015:1-6.
36. Kraschnewski JL, Chuang CH, Poole ES, Peyton T, Blubaugh I, Pauli J, et al. Paging "Dr. Google": does technology fill the gap created by the prenatal care visit structure? Qualitative focus group study with pregnant women. J Med Internet Res 2014 Jun 03;16(6):e147 [FREE Full text] [doi: 10.2196/jmir.3385] [Medline: 24892583]
37. Anders S, Aaron H, Jackson G, Novak L. Support caregivers in pregnancy: a qualitative study of their activities and roles. J Patient Exp 2018 Jul 16:- [FREE Full text] [doi: 10.1177/2374373518785570]
38. Deave T, Johnson D, Ingram J. Transition to parenthood: the needs of parents in pregnancy and early parenthood. BMC Pregnancy Childbirth 2008 Jul 29;8:30 [FREE Full text] [doi: 10.1186/1471-2393-8-30] [Medline: 18664251]
39. Negin J, Coffman J, Vizintin P, Raynes-Greenow C. The influence of grandmothers on breastfeeding rates: a systematic review. BMC Pregnancy Childbirth 2016 Apr 27;16:91 [FREE Full text] [doi: 10.1186/s12884-016-0880-5] [Medline: 27121708]
40. Spoozak L, Gotman N, Smith MV, Belanger K, Yonkers KA. Evaluation of a social support measure that may indicate risk of depression during pregnancy. J Affect Disord 2009 Apr;114(1-3):216-223 [FREE Full text] [doi: 10.1016/j.jad.2008.07.015] [Medline: 18765164]
41. Harley K, Eskenazi B. Time in the United States, social support and health behaviors during pregnancy among women of Mexican descent. Soc Sci Med 2006 Jun;62(12):3048-3061 [FREE Full text] [doi: 10.1016/j.socscimed.2005.11.036] [Medline: 16403596]
42. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
43. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258 [FREE Full text] [doi: 10.1056/NEJMsa0706154] [Medline: 18499567]
44. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. Br Med J 2008 Dec 04;337:a2338 [FREE Full text] [Medline: 19056788]
45. Valente TW, Pitts SR. An appraisal of social network theory and analysis as applied to public health: challenges and opportunities. Annu Rev Public Health 2017 Mar 20;38:103-118. [doi: 10.1146/annurev-publhealth-031816-044528] [Medline: 27992729]
46. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

Abbreviations

EWM: Expect with Me
FCV: Junior League Fetal Center at Vanderbilt

REALM: Rapid Estimate of Adult Literacy in Medicine
VUMC: Vanderbilt University Medical Center

Edited by G Eysenbach; submitted 24.04.18; peer-reviewed by B Holtz, A Efthymiou; comments to author 28.08.18; revised version received 31.01.19; accepted 30.03.19; published 09.05.19.

Please cite as:
Dorst MT, Anders SH, Chennupati S, Chen Q, Purcell Jackson G
Health Information Technologies in the Support Systems of Pregnant Women and Their Caregivers: Mixed-Methods Study
J Med Internet Res 2019;21(5):e10865
URL: https://www.jmir.org/2019/5/e10865/
doi:10.2196/10865
PMID:31094327

Original Paper

DiabetesSistersVoices: Virtual Patient Community to Identify Research Priorities for Women Living With Diabetes

Peijin Han[1], MD, MHS; Wanda Nicholson[2,3], MD, MPH, MBA; Anna Norton[4], MS; Karen Graffeo[4], BSc; Richard Singerman[5], PhD; Steven King[5], BS; Aditi Sundaresan[2], MPH; Wendy Bennett[1,6], MD, MPH

[1]The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Center for Women's Health Research, Department of Obstetrics and Gynecology, Chapel Hill, NC, United States

[3]Public Health Leadership Program, The University of North Carolina School of Medicine and, Gillings Global School of Public Health, Chapel Hill, NC, United States

[4]DiabetesSisters, #180, 1112 W Boughton Road, Bolingbrook, IL, United States

[5]TrustNetMD, Washington, DC, United States

[6]Johns Hopkins University School of Medicin, Division of General Internal Medicine, Baltimore, MD, United States

Corresponding Author:
Wendy Bennett, MD, MPH
Johns Hopkins University School of Medicin, Division of General Internal Medicine
2024 E Monument Street
Baltimore, MD, 21210
United States
Phone: 1 4105026081
Email: wendy.bennett@jhu.edu

Abstract

Background: Women with or at high risk of diabetes have unique health concerns across their life course. Effective methods are needed to engage women living with diabetes to develop and carry out a patient-centered research agenda.

Objective: This study aimed to (1) describe the creation of DiabetesSistersVoices, a virtual patient community for women living with and at risk for diabetes and (2) assess the feasibility and acceptability of DiabetesSistersVoices for engaging women in talking about their experiences, health care, and research priorities.

Methods: We partnered with a national advocacy organization to create DiabetesSistersVoices and to develop recruitment strategies, which included use of social media, Web-based newsletters, and weblinks through partnering organizations. Study inclusion criteria were as follows: Being a woman aged ≥18 years, residing in the United States, and self-reporting a diagnosis of diabetes or risk of diabetes. Eligible participants were given access to DiabetesSistersVoices and completed online surveys at enrollment and 6 months. We assessed trends in participants' activities, including posting questions, sharing experiences about living with diabetes, and searching for posted resources.

Results: We enrolled 332 women (white: 86.5%; type 1 diabetes: 76.2%; median age: 51 years [interquartile range: 31 to 59 years]) over 8 months. Most (41.6%, 138/332) were classified as being active users (ie, posting) of the virtual community, 36.1% (120/332) as observers (ie, logged in but no posts), and 22.3% (74/332) as never users (ie, completed baseline surveys but then never logged in). Online activities were constant during the study, although participants had the highest website usage during the first 10 weeks after their enrollment.

Conclusions: We demonstrated the feasibility and acceptability of an online patient community for women living with diabetes by showing durability of recruitment and online usage over 6 months of testing. Next steps are to address barriers to joining a virtual patient community for women of color and women with type 2 diabetes to enhance inclusiveness and gain diverse perspectives to inform diabetes research.

(J Med Internet Res 2019;21(5):e13312) doi:10.2196/13312

KEYWORDS

social media; online social networking; women's health; diabetes mellitus

Introduction

Background

Diabetes is increasingly common among men and women, affecting 9.4% of the US population [1,2]. Women are disproportionately affected by diabetes, with a greater risk of heart attack and stroke, compared with men with diabetes [3-6]. Compared with women without diabetes, women with diabetes have unique health needs and concerns across their life course, including preconception health, pregnancy, postpartum care, and the menopausal transition. In addition, women with diabetes experience higher rates of eating disorders [6], pregnancy complications [7], sexual dysfunction [8], and higher rates of diabetes-related urinary incontinence [9]. There is a need to understand the patient experience for women living with diabetes to inform and improve patient-centered research, advocacy, and health care delivery for this important, high-risk population.

Increasingly, patients with a variety of conditions, including diabetes, are going online to learn more about their disease, create social support networks, and communicate with other patients such as themselves [10]. In a recent Pew survey, 16% of adult internet users reported going online to find other patients with similar experiences and diagnoses [11]. In fact, there has been a proliferation of online patient communities, including PatientsLikeMe and patient bloggers, especially for people living with and seeking support for living with chronic diseases [12]. The term Web 2.0 was introduced in 2004 to describe the improved communication and collaboration tools available via social networking [13-15]. Growing evidence shows online support groups or "learning health communities" as having benefits for patients [12,16-18]. Along with peer-to-peer support, online patient communities are useful for patient engagement and as an innovative tool to connect patients with researchers so that they can be involved in all stages of the research process [19,20].

Study Objectives

Engaging patients about what is most important to them is crucial to direct researchers, consumer and advocacy organizations, policy makers, health systems leaders, and funders toward clinical research and health system improvements that are patient-centered and meaningful [21]. Our goal was to develop a virtual patient community for women with diabetes to communicate with each other about their experiences to help identify research and health care priorities.

In partnership with DiabetesSisters, a national advocacy organization for women living with any type of diabetes, we developed and launched DiabetesSistersVoices, a virtual patient community to engage women with diabetes to talk about patient-centered research priorities and create a platform for peer support. We aimed to engage women to join the community across racial/ethnic groups, geographical regions, and types of diabetes. The objectives of this study were to (1) describe the creation of DiabetesSistersVoices, a virtual patient community for women living with and at risk for diabetes and (2) assess the feasibility and acceptability of DiabetesSistersVoices for engaging women about their experiences, health care, and research priorities.

Methods

We began with the development phase and refinement of the website, followed by a recruitment and enrollment phase to engage women with diabetes. The institutional review boards at Johns Hopkins University and the University of North Carolina approved the study.

Phase 1: Development and Refinement of DiabetesSistersVoices

The DiabetesSistersVoices virtual patient community's platform began with refining a recently developed and tested platform. The platform, developed by Lehmann and colleagues [22], had been built for communication between community health workers. We chose this platform's software approach because it utilized open-source software with customized interfaces and modules, making it easily adaptable and scalable, and it already contained multiple features that enabled communication and information sharing. Several key features of the platform were the ability to post a question or topic, respond to other posts, search for or post resources for others, tag or relabel, or "like" the topics. "Liking" a post provided points for the post to show its popularity or helpfulness.

As an initial step to refining the platform for the needs of women with diabetes, the chief executive officer of DiabetesSisters (AN) invited 8 patients and advocates who were active in the organization and willing to donate time to participate in semistructured phone interviews with the platform developer and research team. We demonstrated the functions of the platform and inquired about their desired functionalities and features for the new site.

We convened a diverse stakeholder advisory board to gather early input about the design and launch of DiabetesSistersVoices [23]. Stakeholder advisory board members represented 5 stakeholder groups including patient partners from DiabetesSisters, a national nonprofit diabetes organization for women and our main partner in this study. We additionally included leaders from the Black Women's Health Imperative and PatientsLikeMe and a researcher conducting community-based work related to diabetes in racial minority communities. Members met every 3 months and they provided insights and feedback about the platform's clarity and usability, font size, white space, and images and instructions for participants during the meetings. To ensure broad applicability of the platform, we iteratively refined the Web pages by conducting one-on-one "hands-on" semistructured interviews with each stakeholder advisory board member and additional patient partners identified by the stakeholder advisory board's representative organizations (Multimedia Appendix 1 contains the interview guide). Comments from the semistructured interviews were presented back to the research team and used to refine the Web platform. From these interviews, we gathered valuable feedback on the ease of use of traversing the website and the ability to express ideas and exchange information with other participants.

Figure 1. Landing page of DiabetesSistersVoices. There were 4 basic functions of the website, including “share your experience,” “join a forum for conversation,” “find a resource,” and “explore a topic”. The size of popular topics listed below was based on their popularity, that is, the number of times they were clicked.

Share Your Experience
Share experiences or questions about living with diabetes

Join a Forum for Conversation
Let's discuss your health, diabetes and healthcare

Find a Resource
Search for helpful resources and information

Explore a Topic
Pick your favorite topic from the topic list

Popular Topics

type 2 type 1 and lada resource
healthy living complications newly diagnosed
diabetes and other conditions q&a emotional wellness treatment
gestational advanced duration diabetes diet food nutrition
glucose monitoring relationships and intimacy minority health
physical activity and exercise psychosocial support pregnancy
heart disease type 1 insulin

DiabetesSistersVoices is a national online community for women living with pre-diabetes or diabetes to ask questions, share experiences and search for resources. Our goal is to provide you with a forum to talk about your health and health care needs. Your views will help researchers better understand what health topics are most important to you and guide priorities for research on management, treatment, health services and prevention of diabetes in across all phases of life- young adulthood to pregnancy to menopause to the elder years.

The final version of the website used in Phase 2 (www.diabetessistersvoices.org) comprised multiple features to facilitate online conversation. Additional features beyond those in the original platform included (1) a video orientation guide for participants that reviewed the goals of the project and how to use the site and (2) an online moderator who monitored discussions (and could take down posts if not appropriate), welcomed participants, and sent weekly emails (see Multimedia Appendix 2 for the list of topics). Figure 1 shows the landing page of DiabetesSistersVoices that participants were able to view after completion of the sign-in process using their individually selected username and password.

Phase 2: Observational Cohort Study

In Phase 2, we recruited women with diabetes to use the DiabetesSistersVoices virtual patient community over 6 months.

Recruitment Strategies

In collaboration with the stakeholder advisory board, we developed and employed a wide range of recruitment strategies and activities to attract women to the site and engage participants across different age, racial, and socioeconomic strata. The recruitment strategies included (Multimedia Appendix 3):

1. Social media promotion through Facebook and Twitter and monthly electronic newsletters from DiabetesSisters. Facebook posts were “boosted” monthly to showcase the posts to more users.

2. Printed brochures, posters, and postcards distributed in person at churches in Baltimore, health clinics, health fairs, and conferences for patients with diabetes.
3. A toolkit aimed at partnering organizations, which included a template for the invitation from the partner, a description of DiabetesSistersVoices, and promotional materials including flyers and postcards.
4. The site moderator used online “Q&A” threads to keep promoting new discussion topics and emailed a weekly topic to all the registered users to get them to come back (see Multimedia Appendix 2 for the list of weekly topics).

Study Participants: Eligibility and Online Consent Process

Participants registered to participate on the DiabetesSistersVoices website. The website contained information about the project; and if they were interested, they clicked “Register to Participate” and were prompted to provide their name and email address. Our study team received a daily list of emails from the website. Email addresses were entered into the QUALTRICS survey platform to send an email message with further details about the study and a unique link to online screening questions. Participants were unable to access this unique link again, once consent was completed.

Study inclusion criteria were of the female gender, aged ≥18 years, and self-report of type 1, 2, or prediabetes (defined as “at risk for diabetes” or gestational diabetes mellitus history) and residing in the United States. Following successful completion of the screening questions, participants completed an electronic consent process and downloaded a copy of the consent form. We conducted a monthly raffle of items related to healthy living with diabetes (eg, hand weights, yoga mats, and books) to participants who logged in ≥2 times in the past month, starting December 2016.

Data Collection

We assessed with 2 time points of data collection using online surveys (at baseline enrollment and after 6 months since the launch of the website) and had continuous surveillance of website utilization using Google Analytics, a Web analytics service.

Online Surveys

Participants completed a baseline survey following consent. Participants who had been enrolled for ≥1 month completed an end of study online survey (administrated 6 months after the initiation of the study). Participants received surveys as an email with a unique survey link. Participants were able to review and change their answers through a back button and then click submit, and the survey completeness was further verified after it was submitted. Survey response data were captured and stored automatically using the Johns Hopkins QUALTRICS server, which was protected under the Johns Hopkins University firewall, and only the principal investigators, coinvestigator, and study coordinator had access to the data. Baseline survey questionnaires assessed sociodemographic characteristics, health status, internet use, social support, and health-related quality of life using standard measures (Multimedia Appendix 4) [24,25]. We used the Medical Outcomes Study Social Support Survey Instrument to assess emotional and informational social support [26] and the Patient-Reported Outcomes Measurement Information System (PROMIS) scale version 1.2 Global Health survey instrument to assess users’ self-reported health status [27]. The end of study surveys contained the Medical Outcomes Study (MOS) Social Support Survey instrument, PROMIS scale version 1.2 Global Health survey instrument, and satisfaction with the site (see Multimedia Appendix 5 for site satisfaction).

To measure the utilization of the DiabetesSistersVoices online community, we used Google Analytics, a Web analytics service, to track and report DiabetesSistersVoices’ website traffic and activities, including the length of time of individual sessions, number of downloads, page views, “clicks” on topic tags, and “likes” that were captured by this application. We were able to link website users with survey data using a deidentified unique user identification number, assigned by Google Analytics at the time of registration.

Data Analysis

We defined “never users” as participants who enrolled for the study but never logged in to the community, “observers” as participants who logged in to the site at least once but never posted on the site, and “active users” as participants who posted on the site at least once. We reported engagement activities as “posting” comments, “liking” comments or resources, clicking on a topic, and conducting a search. We calculated the proportion of participants who engaged in each of these activities at least once over the previous 2 weeks.

We used descriptive statistics to describe baseline characteristics of participants over time and by level of online engagement. We presented descriptive statistics for the satisfaction level of the participants who completed the end of study satisfaction survey. For participants who were eligible to complete the online survey (ie, enrolled in the study for at least 1 month), we compared the sociodemographic of those who completed the end of study social support and health-related quality of life survey with those who did not. Both completed questionnaires and questionnaires terminated early were included in the analysis.

To test for statistically significant differences between groups, we performed a nonparametric k-sample test on the equality of medians for continuous variables and chi-square test for categorical variables. A 2-sided P value ≤.05 was considered statistically significant. All the statistical analyses were performed using STATA (StataCorp LLC).

Results

Enrollment into DiabetesSistersVoices Virtual Patient Community

From November 2016 to June 2017 (30 weeks), 511 women registered at the DiabetesSistersVoices website. Among them, 395 women began the online screening and consent process and 332 women completed the consent form and enrolled in the study (Figure 2).

Figure 3 demonstrates the cumulative enrollment of participants over the 30 weeks, overall, and stratified by race and diabetes type (type 1 vs type 2 diabetes). At the beginning of the study, all participants were white with type 1 diabetes, when the majority of promotions were by DiabetesSisters' online and in-person (at conferences) advertising. Over time and with the expansion of recruitment methods, the diversity of study participants increased. The study was "boosted" on Facebook multiple times for women with interests in diabetes-related topics.

Figure 2. Enrollment flow of participants into the DiabetesSistersVoices study. MOS: Medical Outcomes Study; PROMIS: Patient-Reported Outcome Measurement Information System.

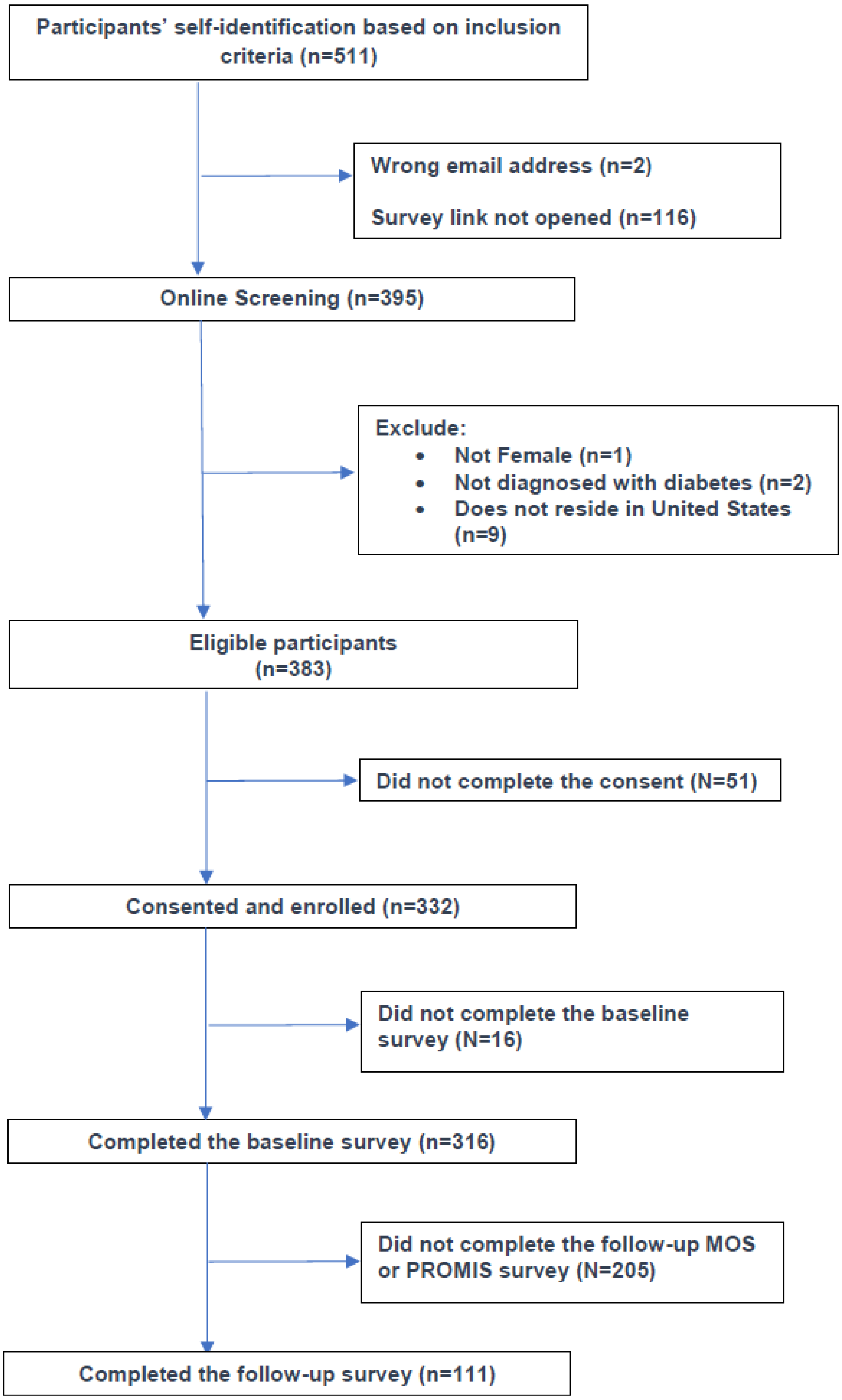
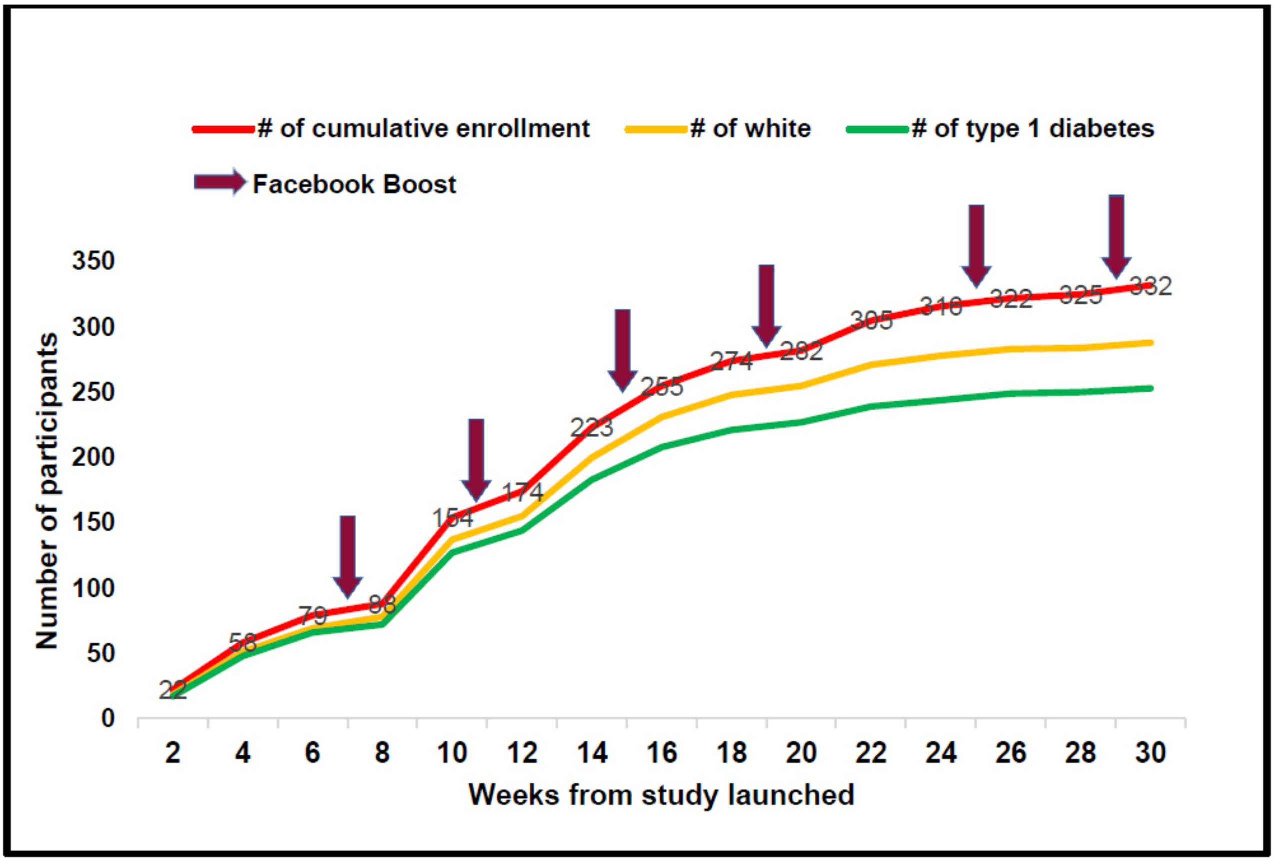


Figure 3. Participants' enrollment over time–overall, by race, and type of diabetes. Red arrows indicate the time of Facebook boost.



Participant Characteristics

Table 1 shows the baseline characteristics of study participants who enrolled in the DiabetesSistersVoices virtual patient community. Overall, participants were 86.5% (287/332) white, 6.3% (21/332) black, 3.3% (11/332) Hispanic, 0.9% (3/332) Asian, and 3.0% (10/332) elected not to report race. The median age was 51 years (interquartile range [IQR]: 37 to 59). Compared with the "never users" and "observers," "active users" were older (aged 54 years vs 49 and 50 years; *P*=.04). A total of 68.4% (225/332) of the participants had a college degree or above, followed by 30.4% (100/332) who had high school or some college, and 1.2% (4/332) had less than a high school diploma. The majority (76.2%, 253/332) of participants had type 1 diabetes (vs 22.0% [73/332] with type 2 diabetes). No participants reported having "prediabetes." On average, participants were in the study for 5 months (IQR: 4 to 6 months). Most women reported frequent email use of at least once daily (90.0%, 297/332) and use of social networks such as Facebook (89.9%, 295/332).

A total of 74 (22.3%) of participants were classified as "never users," 120 (36.1%) as "observers," and 138 (41.6%) as "active users." Similar proportions of white and nonwhite women and women with type 1 (vs type 2 diabetes) were active users on the site (Table 1). Compared with "never users," "observers" and "active users" were more likely to use email daily (93.3% and 91.3% vs 82.2%) and use social networking sites (93.2% and 90.0% vs 84.7%). "Active users" had higher baseline social support scores compared with "never users" and "observers," but there was no difference in the quality of life scores (Table 1).

Table 1. Characteristics of study participants who enrolled in the DiabetesSistersVoices virtual patient community.

Participants' characteristics	All users (n=332)	Never users (n=74)	Observers (n=120)	Active users (n=138)	*P* value[a]
Age (years), median (interquartile range)	51 (37-59)	49 (34-56)	50 (36-58)	54 (40-63)	.04
Education[b] **, n (%)**					
Less than high school or GED	4 (1.2)	2 (2.7)	0 (0.0)	2 (1.5)	.13
High school or some college	100 (30.4)	29 (39.2)	31 (26.1)	40 (29.4)	.13
College degree and above	225 (68.4)	43 (58.1)	88 (74.0)	94 (69.1)	.13
Race/ethnicity[b] **, n (%)**					
White	287 (86.5)	61 (82.4)	104 (86.7)	122 (88.4)	.27
Black	21 (6.3)	8 (10.8)	7 (5.8)	6 (4.4)	.27
Asian	3 (0.9)	2 (2.7)	0 (0.0)	1 (0.7)	.27
Hispanic	11 (3.3)	2 (2.7)	5 (4.1)	4 (2.9)	.27
Elected not to report	10 (3.0)	1 (1.4)	4 (3.3)	5 (3.6)	.27
Diabetes type[b] **, n (%)**					
Type 1	253 (76.2)	53 (71.6)	90 (75.0)	110 (79.7)	.45
Type 2	73 (22.0)	20 (27.0)	26 (21.7)	27 (19.6)	.45
Diabetes treatment duration (years), mean (SD)	22.0 (15.0)	21.7 (14.6)	20.5 (13.2)	23.6 (16.1)	.18
Email use[b] **, n (%)**					
Daily	297 (90.0)	60 (82.2)	111 (93.3)	126 (91.3%)	.08
Every few days	30 (9.1)	11 (15.1)	8 (6.7)	11 (8.0%)	.08
Less than weekly	3 (0.9)	2 (2.7)	0 (0.0)	1 (0.7%)	.08
Ever use social networking sites[b], n (%)	295 (89.9%)	61 (84.7)	110 (93.2)	124 (90.0)	.17
Website engagement					
Number of sessions, median (IQR)	3 (1.0-5.5)	—[c]	2 (1.0-4.0)	5.5 (4.0-12.0)	<.001
Total topic clicks	421	—	96	325	—
Total posts/responses	904	—	0	904	—
Total likes	530	—	42	488	—
Total searches	167	—	32	135	—
Total downloads	671	—	121	550	—
Total social support score, mean (SD)[d]	68.0 (23.6)	64.3 (25.6)	71.1 (22.5)	67.2 (23.8)	.22
PROMIS-physical health score, mean (SD)[d]	36.4 (4.7)	36.7 (4.5)	36.1 (4.0)	36.4 (5.2)	.68
PROMIS-mental health score, mean (SD)[d]	38.4 (8.7)	39.4 (7.8)	38.0 (8.6)	38.4 (9.1)	.28

[a] Nonparametric k-sample test on the equality of medians for continuous variables and chi-square test for categorical variables.

[b] Percentage may not add up to 100% because of missing data.

[c] —Not applicable.

[d] A total of 316 out of 332 completed the Patient-Reported Outcomes Measurement Information System (PROMIS) Quality of Life and Medical Outcomes Study (MOS) social support questionnaires. MOS scores range from 0 (lowest) to 100 (highest). PROMIS scores were presented as *t* score. The norm in population is mean 50 (SD 10). Scores 0.5 to 1.0 SD or worse than the mean=mild symptoms/impairment, scores 1.0 to 2.0 SD or worse than the mean=moderate symptoms/impairment, and scores 2.0 SD or worse than the mean indicate more severe symptoms/impairment.

Online Activities on DiabetesSistersVoices Virtual Patient Community

Over the 30-week study, study participants clicked on topics 421 times, posted or replied on the site 904 times, "liked" posts 530 times, searched for resources 167 times, and downloaded resources 671 times (Table 1). Figure 4 shows that online activities on the site were constant during the study, with about one-third of participants at any given time posting or clicking a topic on the site at least once every other week. Fewer participants "liked" posts on the site (approximately 20%) or performed searches (<10%). Figure 5 shows the proportion of participants engaged over time following their enrollment into the study. On average, participants' website usage was highest during their first 10 weeks of enrollment. A total of 10 participants were enrolled during the entire study duration of 30 weeks.

Figure 4. Participants' online activities over time. The proportion of participants who were engaged in each of the online activities was accumulatively calculated every 2 weeks since the initiation of the study.

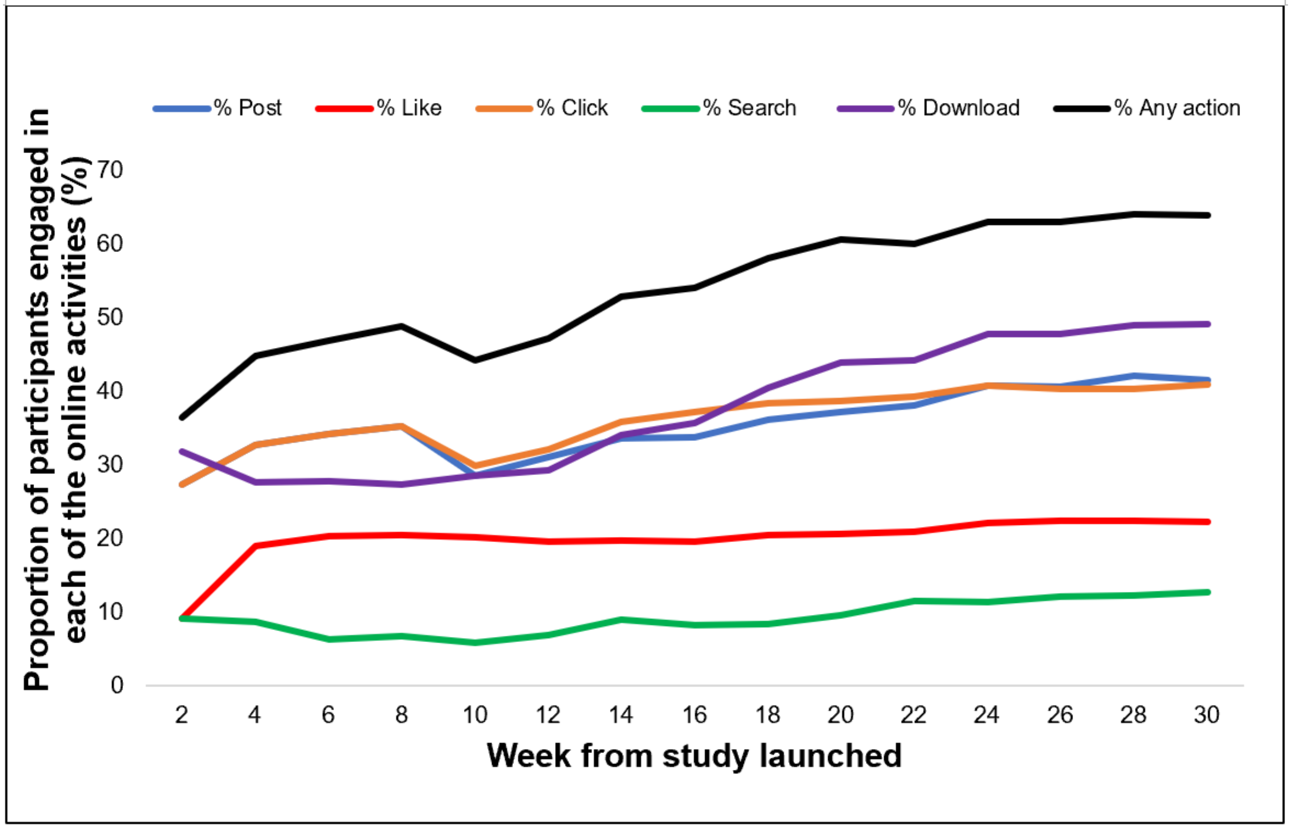
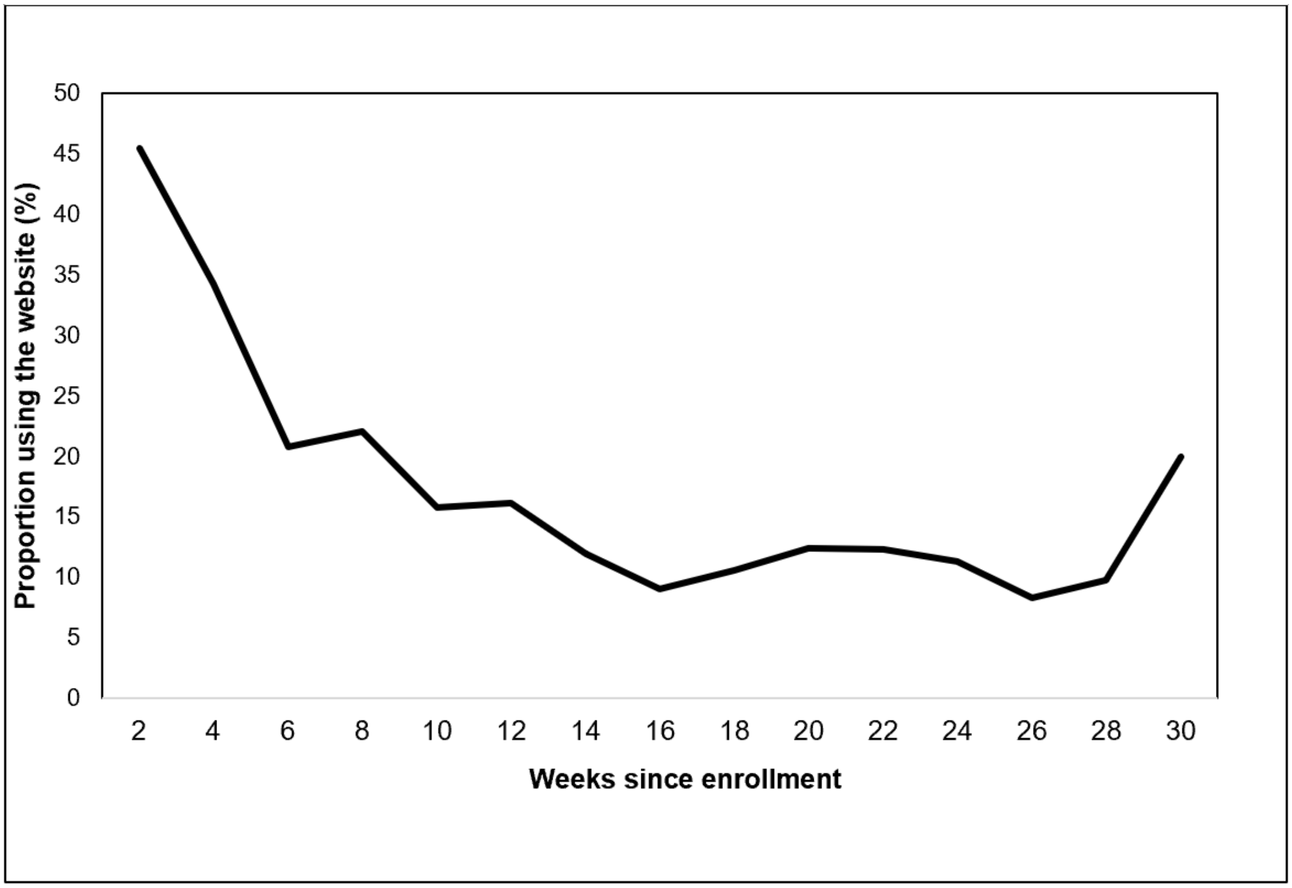


Figure 5. Proportion of participants engaged over time (in weeks) following their study enrollment. The numbers below the x-axis represent the total number of participants in the study.



Satisfaction With DiabetesSistersVoices Virtual Patient Community

Table 2 describes the responses for the end of study satisfaction survey, completed by 120 out of 323 eligible participants (37.2%). A total of 67.5% of these participants met the definition of “active users” (Table 2). The majority (94%) of participants were not satisfied with the electronic consent process. However, participants were generally satisfied with the major functions of the DiabetesSistersVoices platform, including logging in, posting questions, searching for comments, and learning topics from weekly email sent by the site moderator. Active website users tend to have a higher satisfaction level compared with the “never users” (data not shown).

Compared with participants who completed only the baseline social support and quality of life survey, participants who completed both baseline and end-of-study surveys were senior (aged 54 years vs 49 years; *P*<.001) and had a higher education level (Multimedia Appendix 6). The distributions of race and diabetes diagnosis had no statistically significant difference between those 2 groups (Multimedia Appendix 6).

Table 2. Participants' satisfaction with the study and DiabetesSistersVoices virtual patient community's platform features (N=120).

Aspects of the DiabetesSistersVoices virtual patient community	Statistics, n (%)
Consenting to participate in the study[a]	
Dissatisfied	21 (17.5)
Neutral	92 (76.7)
Satisfied	7 (5.8)
Online survey about you and your health at the start of the study	
Dissatisfied	1 (0.8)
Neutral	29 (24.2)
Satisfied	80 (66.7)
Logging on to the website	
Dissatisfied	16 (13.3)
Neutral	17 (14.1)
Satisfied	82 (68.3)
Posting questions or comments	
Dissatisfied	16 (13.3)
Neutral	24 (20.0)
Satisfied	66 (55.0)
Searching for resources	
Dissatisfied	14 (11.7)
Neutral	22 (18.3)
Satisfied	67 (55.8)
Communicating with other members of the community	
Dissatisfied	15 (12.5)
Neutral	25 (20.8)
Satisfied	62 (51.7)
Topic of weekly email	
Dissatisfied	11 (9.2)
Neutral	24 (20.0)
Satisfied	76 (63.3)

[a] Percentage may not add up to 100% because some participants did not use the feature.

Discussion

Principal Findings

In partnership with DiabetesSisters, a national organization serving women living with diabetes, and a diverse stakeholder advisory board, we created the DiabetesSistersVoices virtual patient community and enrolled 332 women with all types of diabetes to provide peer support and to identify high priority research areas for women living with diabetes. Our results demonstrate strong feasibility and acceptability of the online patient community, showing high levels of website engagement, even over a short time period. However, like other "online" only studies, the major limitation was the lack of broad representation of women living with diabetes (ie, older women, women of color, and women with type 2 diabetes). In addition, participants reported high levels of satisfaction with the DiabetesSistersVoices virtual patient community but were less satisfied with the electronic consent for research, which may have created a barrier for those who never enrolled.

Advantages of and Barriers to Online Patient Engagement

Virtual patient communities have multiple purposes, including peer-to-peer patient support (eg, PatientsLikeMe) [12,20,28], dissemination and sharing of resources either coming from health professionals or from patients to other patients, [12,29,30], as well as recruitment into research studies [31]. In addition, like DiabetesSistersVoices, virtual patient communities can provide an opportunity for researchers to engage with patients about what issues are most important to them for ongoing or to guide future research [12,19,32,33]. Patients value peer-to-peer communication as being "more real." In a Web-based smoking cessation trial, the "crowdsourced" messages by patients about how to quit smoking were associated with greater return visits to the smoking cessation clinic [34]. However, most patient engagement websites are not moderated, and we identified that patient-led moderation is an important component of patient engagement. To encourage engagement and re-engagement, the site moderator used online “Q&A” threads to keep promoting new discussion topics and emailed a weekly topic to all the registered users to encourage them to come back (Multimedia Appendix 2 lists the weekly topics). In addition, growing evidence supports virtual support groups or “learning health communities” as having some benefits for patients [12,16,17,35,36]. A 2004 systematic review identified 38 studies reporting computer-based peer-to-peer communities, and electronic self-support groups showed a need for rigorously designed studies to evaluate the effects of social media interventions on health behaviors and long-term health outcomes [36]. Given the strength of evidence supporting *in-person* peer-to-peer support for complex health behaviors and disease management in a wide variety of settings and disease [37,38]. In addition, few studies have used electronic modalities for patient engagement around research and to identify research priorities [21].

Despite our goal to engage diverse women with both type 1 and type 2 diabetes, the majority of DiabetesSistersVoices participants were of white race and had type 1 diabetes, likely because of our strong affiliation with the DiabetesSisters organization, which has a strong allegiance and trust within the community of women with type 1 diabetes. Given the disparities in the diabetes epidemic [39], there is a high need to develop Web-based engagement strategies specifically tailored for African American and Latina women [40,41] as well as women with type 2 diabetes and prediabetes [42]. African American women with diabetes are diverse with regard to age, region (urban and rural population), access to care, socioeconomic status, and types of health disparities they experience. For example, diabetes self-management interventions have been designed and specifically tailored for African American women to take into account women’s values, multiple caregiving roles, cultural implications around eating and weight, socioeconomic barriers to healthful eating and medication adherence, and spiritual and social support needs, as well as day-to-day barriers to self-management (eg, high stress, family conflicts, and coping styles) [43-45]. Web-based modalities also need to specifically address the needs and values of African American women to have greatest impact.

Online engagement of African American women using Web-based modalities is underutilized in diabetes research despite the fact that African American women are “connected” and use mobile technology at very high rates [46]. In 2014, African Americans trailed whites by only 7% in internet use (80% vs 87%, respectively) [47] and the gap continues to close [48]. However, among them, compared with the traditional recruiting strategies (eg, paper flyers, media, and face-to-face advertising), the online enrollment rate was proportionally smaller [49]. Other studies have identified barriers to online recruitment for African Americans to enroll into research studies, include a learning curve for computer usability [50], preference to face-to-face support [51], and mistrust of the researchers [51-53]. In addition, barriers specific to diabetes, such as “diabetes stigma” ie, diabetes viewed as a “lifestyle disease” with blaming of the individual), may also prevent women to self-identify (eg, discuss their diagnosis online), seek out support and resources, and join a virtual peer community with other patients [54-58].

Study Limitations

There are several limitations to this study. First, because this was a research study, the consent process posed a barrier to women who otherwise may have joined an online peer support group. Second, we had lower enrollment in the study for minority women and women with type 2 diabetes. As described above, the majority of our participants were white and with type 1 diabetes, in part because our main stakeholder partner, DiabetesSisters, has a community in which 65% are women living with type 1 diabetes. We also engaged fewer older women who may use the internet for information but not as a method for social interaction with peers. Third, this was a short-term study, totaling 6 months, but many women used the site for less than 6 months (range 1 to 6 months), limiting our ability to assess changes in social support or diabetes knowledge as a results of website participation. Finally, overall rates of “posting” new content on the site was low but consistent with typical Web behaviors described on other websites, where the majority of users are considered “observers” but not active posters or leaders in social networking sites [59-61].

Conclusions

In summary, our study findings suggest that a virtual patient community can be an engaging and efficient tool for women with diabetes to interact with each other and provide their perspectives about diabetes care to inform the next generation of research questions. We identified multiple approaches to engage women to share their perspectives on a range of topics and interest related to diabetes. Further study is needed in a larger cohort of women with type 2 and gestational diabetes as the majority of study participants had type 2 diabetes. Furthermore, future studies might consider targeting aspects of diabetes care (eg, fertility and menopause) that are specific to women and address conditions across the woman’s lifespan.

Acknowledgments

The DiabetesSistersVoices team would like to acknowledge and offer gratitude to their committed stakeholder advisory board members who provided valuable and meaningful feedback every step of the way: AC (UT Health San Antonio School of Nursing); AF (The Black Women’s Health Imperative); MF (PatientsLikeMe); AG AM (DiabetesSisters); and ES (DiabetesSisters).

This project was funded by a Eugene Washington PCORI Engagement Award Program Agreement # 2605-JHU.

We would also like to acknowledge Harold Lehmann, MD, PhD (The John's Hopkins University School of Medicine) for key advisory input throughout this research project.

Conflicts of Interest

RS and SK both have an ownership interest in TrustNetMD.

Multimedia Appendix 1

Instructions for DiabetesSistersVoices users: interview guide.

[DOCX File, 266KB - jmir_v21i5e13312_app1.docx]

Multimedia Appendix 2

Topic list in April newsletter.

[DOCX File, 13KB - jmir_v21i5e13312_app2.docx]

Multimedia Appendix 3

Examples of social media and recruitment materials. Exhibit from left to right: a Facebook post, a flyer used online and for in-person recruitment, and the title page for the Toolkit we distributed to stakeholders with information about the study and recruitment materials.

[PNG File, 1MB - jmir_v21i5e13312_app3.PNG]

Multimedia Appendix 4

DiabetesSistersVoices baseline survey questionnaires.

[DOCX File, 25KB - jmir_v21i5e13312_app4.docx]

Multimedia Appendix 5

DiabetesSistersVoices end of study satisfaction survey.

[DOCX File, 22KB - jmir_v21i5e13312_app5.docx]

Multimedia Appendix 6

Comparison of DiabetesSistersVoices study participants' characteristics--those who only completed baseline social support and quality of life survey (and not end of study survey), N=205, compared with participants who completed both baseline and end of study social support and quality of life survey N=111.

[DOCX File, 15KB - jmir_v21i5e13312_app6.docx]

References

1. Centers for Disease Control and Prevention. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2017. National Diabetes Statistics Report URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2019-04-22] [WebCite Cache ID 77pdsVRLv]
2. Mayer-Davis EJ, Dabelea D, Lawrence JM. Incidence trends of type 1 and type 2 diabetes among youths, 2002-2012. N Engl J Med 2017 Dec 20;377(3):301 [FREE Full text] [doi: 10.1056/NEJMc1706291] [Medline: 28723318]
3. Peters SA, Huxley RR, Sattar N, Woodward M. Sex differences in the excess risk of cardiovascular diseases associated with type 2 diabetes: potential explanations and clinical implications. Curr Cardiovasc Risk Rep 2015;9(7):36 [FREE Full text] [doi: 10.1007/s12170-015-0462-5] [Medline: 26029318]
4. Morris ME, Aguilera A. Mobile, social, and wearable computing and the evolution of psychological practice. Prof Psychol Res Pr 2012 Dec;43(6):622-626 [FREE Full text] [doi: 10.1037/a0029041] [Medline: 25587207]
5. Peters SA, Huxley RR, Woodward M. Diabetes as risk factor for incident coronary heart disease in women compared with men: a systematic review and meta-analysis of 64 cohorts including 858,507 individuals and 28,203 coronary events. Diabetologia 2014 Aug;57(8):1542-1551. [doi: 10.1007/s00125-014-3260-6] [Medline: 24859435]
6. Jones JM, Lawson ML, Daneman D, Olmsted MP, Rodin G. Eating disorders in adolescent females with and without type 1 diabetes: cross sectional study. Br Med J 2000 Jun 10;320(7249):1563-1566 [FREE Full text] [Medline: 10845962]

7. Evers IM, de Valk HW, Visser GH. Risk of complications of pregnancy in women with type 1 diabetes: nationwide prospective study in the Netherlands. Br Med J 2004 Apr 17;328(7445):915 [FREE Full text] [doi: 10.1136/bmj.38043.583160.EE] [Medline: 15066886]
8. Copeland KL, Brown JS, Creasman JM, van den Eeden SK, Subak LL, Thom DH, et al. Diabetes mellitus and sexual function in middle-aged and older women. Obstet Gynecol 2012 Aug;120(2 Pt 1):331-340 [FREE Full text] [doi: 10.1097/AOG.0b013e31825ec5fa] [Medline: 22825093]
9. Phelan S, Kanaya AM, Ma Y, Vittinghoff E, Barrett-Connor E, Wing R, Diabetes Prevention Program Research Group. Long-term prevalence and predictors of urinary incontinence among women in the Diabetes Prevention Program Outcomes Study. Int J Urol 2015 Feb;22(2):206-212 [FREE Full text] [doi: 10.1111/iju.12654] [Medline: 25352018]
10. White M, Dorman SM. Receiving social support online: implications for health education. Health Educ Res 2001 Dec;16(6):693-707. [Medline: 11780708]
11. Feng Y, Xie W. Digital divide 2.0: the role of social networking sites in seeking health information online from a longitudinal perspective. J Health Commun 2015;20(1):60-68. [doi: 10.1080/10810730.2014.906522] [Medline: 25119019]
12. Okun S, Goodwin K. Building a learning health community: by the people, for the people. Learn Health Syst 2017 Jun 6;1(3):e10028. [doi: 10.1002/lrh2.10028]
13. van de Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
14. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
15. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
16. Allen CG, Brownstein N. Capacity building and training needs for community health workers working in health care organizations. J Community Med Health Educ 2016;06(01):-. [doi: 10.4172/2161-0711.1000403]
17. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
18. 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]
19. Schroeder EB, Desai J, Schmittdiel JA, Paolino AR, Schneider JL, Goodrich GK, et al. An innovative approach to informing research: gathering perspectives on diabetes care challenges from an online patient community. Interact J Med Res 2015 Jun;4(2):e13 [FREE Full text] [doi: 10.2196/ijmr.3856] [Medline: 26126421]
20. Litchman ML, Rothwell E, Edelman LS. The diabetes online community: Older adults supporting self-care through peer health. Patient Educ Couns 2017 Sep 6:518-523. [doi: 10.1016/j.pec.2017.08.023] [Medline: 28947360]
21. Fleurence R, Selby JV, Odom-Walker K, Hunt G, Meltzer D, Slutsky JR, et al. How the Patient-Centered Outcomes Research Institute is engaging patients and others in shaping its research agenda. Health Aff (Millwood) 2013 Feb;32(2):393-400 [FREE Full text] [doi: 10.1377/hlthaff.2012.1176] [Medline: 23381533]
22. Lehmann HG, Singerman R, King S, Warren J, May J. Design of a knowledge exchange for community health workers. In: Proceedings of the First African Conference on Human Computer Interaction. 2015 Presented at: AfriCHI'16; November 21-25, 2016; San Francisco, CA p. 14.
23. O'Haire CM, Nakamoto E, LaBrant L, Most C, Lee K, Graham E, et al. Effective Health Care Program. 2011. Methods for Engaging Stakeholders To Identify and Prioritize Future Research Needs. Methods Future Research Needs Report No. 4. (Prepared by the Oregon Evidence-based Practice Center and the Vanderbilt Evidence-based Practice Center under Contract No. 290-2007-10057-I.) AHRQ Publication No. 11-EHC044-EF. Rockville, MD: Agency for Healthcare Research and Quality URL: https://effectivehealthcare.ahrq.gov/search?fB0D=field product typeAresearch report&fB1D=field product typeAsystematic review&fB2D=field product typeAtechnical brief&sort by=field product pub date [accessed 2019-04-22] [WebCite Cache ID 77pq4d1K2]
24. Centers for Disease Control and Prevention. 2013. CDC's Behavioral Risk Factor Surveillance System URL: https://www.health.ny.gov/statistics/brfss/ [accessed 2019-04-22] [WebCite Cache ID 77pqOqlS1]
25. Pew Research Center. The Social Life of Health Information URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2019-04-22] [WebCite Cache ID 77pqcrdEt]
26. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32(6):705-714. [Medline: 2035047]
27. Hays R, Bjorner J, Revicki D, Spritzer K, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009 Sep;18(7):873-880 [FREE Full text] [doi: 10.1007/s11136-009-9496-9] [Medline: 19543809]
28. Hixson JD, Barnes D, Parko K, Durgin T, van Bebber S, Graham A, et al. Patients optimizing epilepsy management via an online community: the POEM Study. Neurology 2015 Jul 14;85(2):129-136 [FREE Full text] [doi: 10.1212/WNL.0000000000001728] [Medline: 26085605]
29. Brownstein CA, Brownstein JS, Williams DS, Wicks P, Heywood JA. The power of social networking in medicine. Nat Biotechnol 2009 Oct;27(10):888-890. [doi: 10.1038/nbt1009-888] [Medline: 19816437]

30. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
31. Jackson ML, Bex PJ, Ellison JM, Wicks P, Wallis J. Feasibility of a web-based survey of hallucinations and assessment of visual function in patients with Parkinson's disease. Interact J Med Res 2014;3(1):e1 [FREE Full text] [doi: 10.2196/ijmr.2744] [Medline: 24394559]
32. DasMahapatra P, Raja P, Gilbert J, Wicks P. Clinical trials from the patient perspective: survey in an online patient community. BMC Health Serv Res 2017 Feb 27;17(1):166 [FREE Full text] [doi: 10.1186/s12913-017-2090-x] [Medline: 28241758]
33. Simacek K, Raja P, Chiauzzi E, Eek D, Halling K. What do ovarian cancer patients expect from treatment?: perspectives from an online patient community. Cancer Nurs 2017;40(5):E17-E27. [Medline: 27454765]
34. Coley HL, Sadasivam RS, Williams JH, Volkman JE, Schoenberger Y, Kohler CL, et al. Crowdsourced peer- versus expert-written smoking-cessation messages. Am J Prev Med 2013 Nov;45(5):543-550 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.004] [Medline: 24139766]
35. Allen C, Vassilev I, Kennedy A, Rogers A. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. J Med Internet Res 2016;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
36. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. Br Med J 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
37. Fisher EB, Boothroyd RI, Elstad EA, Hays L, Henes A, Maslow GR, et al. Peer support of complex health behaviors in prevention and disease management with special reference to diabetes: systematic reviews. Clin Diabetes Endocrinol 2017;3:4 [FREE Full text] [doi: 10.1186/s40842-017-0042-3] [Medline: 28702258]
38. Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. PLoS One 2012;7(12):e53244 [FREE Full text] [doi: 10.1371/journal.pone.0053244] [Medline: 23285271]
39. Peek M, Cargill A, Huang ES. Diabetes health disparities: a systematic review of health care interventions. Med Care Res Rev 2007 Oct;64(5 Suppl):101S-156S [FREE Full text] [doi: 10.1177/1077558707305409] [Medline: 17881626]
40. Blanks S, Treadwell H, Bazzell A, Graves W, Osaji O, Dean J, et al. Community engaged lifestyle modification research: engaging diabetic and Prediabetic African American women in community-based interventions. J Obes 2016;2016:3609289 [FREE Full text] [doi: 10.1155/2016/3609289] [Medline: 27493797]
41. Rosas L, Lv N, Lewis M, Venditti E, Zavella P, Luna V, et al. A Latino patient-centered, evidence-based approach to diabetes prevention. J Am Board Fam Med 2018;31(3):364-374 [FREE Full text] [doi: 10.3122/jabfm.2018.03.170280] [Medline: 29743220]
42. Graffigna G, Barello S, Libreri C, Bosio CA. How to engage type-2 diabetic patients in their own health management: implications for clinical practice. BMC Public Health 2014 Jun 25;14:648 [FREE Full text] [doi: 10.1186/1471-2458-14-648] [Medline: 24966036]
43. Attridge M, Creamer J, Ramsden M, Cannings-John R, Hawthorne K. Culturally appropriate health education for people in ethnic minority groups with type 2 diabetes mellitus. Cochrane Database Syst Rev 2014 Sep 4(9):CD006424. [doi: 10.1002/14651858.CD006424.pub3] [Medline: 25188210]
44. Kumanyika SK, Whitt-Glover MC, Haire-Joshu D. What works for obesity prevention and treatment in black Americans? Research directions. Obes Rev 2014 Oct;15(Suppl 4):204-212. [doi: 10.1111/obr.12213] [Medline: 25196414]
45. Samuel-Hodge CD, Keyserling TC, Park S, Johnston LF, Gizlice Z, Bangdiwala SI. A randomized trial of a church-based diabetes self-management program for African Americans with type 2 diabetes. Diabetes Educ 2009;35(3):439-454. [doi: 10.1177/0145721709333270] [Medline: 19383882]
46. Pew Research Center. 2018 Mar 1. Social media use in 2018 URL: https://www.pewinternet.org/2018/03/01/social-media-use-2018-acknowledgments/ [accessed 2019-04-22] [WebCite Cache ID 77ps0rmEp]
47. Pew Research Center. 2014 Jan 6. African Americans and Technology Use: A Demographic Portrait URL: https://www.pewinternet.org/2014/01/06/african-americans-and-technology-use/ [accessed 2019-04-22] [WebCite Cache ID 77ps8IvyD]
48. Pew Research Center. 2017 Aug 31. Smartphones help blacks, Hispanics bridge some - but not all - digital gaps with whites URL: https://www.pewresearch.org/fact-tank/2017/08/31/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ [accessed 2019-04-22] [WebCite Cache ID 77psNH3D6]
49. Staffileno BA, Zschunke J, Weber M, Gross LE, Fogg L, Tangney CC. The feasibility of using facebook, craigslist, and other online strategies to recruit young African American women for a web-based healthy lifestyle behavior change intervention. J Cardiovasc Nurs 2016 Jul 13:365-371. [doi: 10.1097/JCN.0000000000000360] [Medline: 27428356]
50. Watkins D, Jefferson SO. Recommendations for the use of online social support for African American men. Psychol Serv 2013 Aug;10(3):323-332 [FREE Full text] [doi: 10.1037/a0027904] [Medline: 22924797]

51. Fogel J, Ribisl K, Morgan PD, Humphreys K, Lyons EJ. Underrepresentation of African Americans in online cancer support groups. J Natl Med Assoc 2008;100(6):705-712. [Medline: 18595573]
52. Corbie-Smith G, Thomas SB, St George DM. Distrust, race, and research. Arch Intern Med 2002 Nov 25;162(21):2458-2463. [Medline: 12437405]
53. Williams I, Utz S, Jones R, Hinton I, Steeves R, Alexander G. Recruitment of rural African Americans for research projects: lessons learned. South Online J Nurs Res 2011 Apr 1;11(1):8 [FREE Full text] [Medline: 24791157]
54. Browne JL, Ventura A, Mosely K, Speight J. 'I call it the blame and shame disease': a qualitative study about perceptions of social stigma surrounding type 2 diabetes. BMJ Open 2013 Nov 18;3(11):e003384 [FREE Full text] [doi: 10.1136/bmjopen-2013-003384] [Medline: 24247325]
55. Browne JL, Ventura A, Mosely K, Speight J. 'I'm not a druggie, I'm just a diabetic': a qualitative study of stigma from the perspective of adults with type 1 diabetes. BMJ Open 2014;4(7):e005625 [FREE Full text] [doi: 10.1136/bmjopen-2014-005625] [Medline: 25056982]
56. Browne JL, Ventura AD, Mosely K, Speight J. Measuring the stigma surrounding type 2 diabetes: development and validation of the Type 2 Diabetes Stigma Assessment Scale (DSAS-2). Diabetes Care 2016 Dec;39(12):2141-2148. [doi: 10.2337/dc16-0117] [Medline: 27515964]
57. Holmes-Truscott E, Browne J, Ventura A, Pouwer F, Speight J. Diabetes stigma is associated with negative treatment appraisals among adults with insulin-treated Type 2 diabetes: results from the second Diabetes MILES - Australia (MILES-2) survey. Diabet Med 2018 Dec;35(5):658-662. [doi: 10.1111/dme.13598] [Medline: 29417612]
58. Newton-John T, Ventura A, Mosely K, Browne J, Speight J. 'Are you sure you're going to have another one of those?': A qualitative analysis of the social control and social support models in type 2 diabetes. J Health Psychol 2017 Dec;22(14):1819-1829. [doi: 10.1177/1359105316642005] [Medline: 27106089]
59. Edelmann N. Reviewing the definitions of "lurkers" and some implications for online research. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):645-649. [doi: 10.1089/cyber.2012.0362] [Medline: 23848960]
60. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
61. Gopalsamy R, Semenov A, Pasiliao E, McIntosh S, Nikolaev A. Engagement as a driver of growth of online health forums: observational study. J Med Internet Res 2017 Aug 29;19(8):e304 [FREE Full text] [doi: 10.2196/jmir.7249] [Medline: 28851677]

Abbreviations

IQR: interquartile range
MOS: Medical Outcomes Study
PROMIS: Patient-Reported Outcomes Measurement Information System

Edited by G Eysenbach; submitted 04.01.19; peer-reviewed by E Schroeder, M Litchman, MS Aslam; comments to author 31.01.19; revised version received 29.03.19; accepted 30.03.19; published 10.05.19.

Please cite as:
Han P, Nicholson W, Norton A, Graffeo K, Singerman R, King S, Sundaresan A, Bennett W
DiabetesSistersVoices: Virtual Patient Community to Identify Research Priorities for Women Living With Diabetes
J Med Internet Res 2019;21(5):e13312
URL: https://www.jmir.org/2019/5/e13312/
doi:10.2196/13312
PMID:31094360

Original Paper

Prediction of the 1-Year Risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the State of Maine

Xiaofang Wang[1,2*], PhD; Yan Zhang[3*], MD; Shiying Hao[4,5*], PhD; Le Zheng[4,5], PhD; Jiayu Liao[6,7], PhD; Chengyin Ye[8], PhD; Minjie Xia[9], BS; Oliver Wang[9], BA; Modi Liu[9], BS; Ching Ho Weng[2], MS; Son Q Duong[10], MD; Bo Jin[9], MS; Shaun T Alfreds[11], MBA; Frank Stearns[9], MHA; Laura Kanov[9], BS, MBA; Karl G Sylvester[2], MD; Eric Widen[9], MHA; Doff B McElhinney[4,5], MD; Xuefeng B Ling[2,5*], PhD

[1]Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, China
[2]Department of Surgery, Stanford University, Stanford, CA, United States
[3]Department of Oncology, The First Hospital of Shijiazhuang, Shijiazhuang, China
[4]Department of Cardiothoracic Surgery, Stanford University, Stanford, CA, United States
[5]Clinical and Translational Research Program, Betty Irene Moore Children's Heart Center, Lucile Packard Children’s Hospital, Palo Alto, CA, United States
[6]Department of Bioengineering, University of California, Riverside, CA, United States
[7]West China-California Multiomics Research Center, West China Hospital, Sichuan University, Chengdu, China
[8]Department of Health Management, Hangzhou Normal University, Hangzhou, China
[9]Healthcare Business Intelligence Solutions Inc, Palo Alto, CA, United States
[10]Lucile Packard Children’s Hospital, Palo Alto, CA, United States
[11]HealthInfoNet, Portland, ME, United States
[*]these authors contributed equally

Corresponding Author:
Xuefeng B Ling, PhD
Department of Surgery
Stanford University
S370 Grant Bldg
Stanford, CA,
United States
Phone: 1 650 427 9198
Fax: 1 650 723 1154
Email: bxling@stanford.edu

Abstract

Background: Lung cancer is the leading cause of cancer death worldwide. Early detection of individuals at risk of lung cancer is critical to reduce the mortality rate.

Objective: The aim of this study was to develop and validate a prospective risk prediction model to identify patients at risk of new incident lung cancer within the next 1 year in the general population.

Methods: Data from individual patient electronic health records (EHRs) were extracted from the Maine Health Information Exchange network. The study population consisted of patients with at least one EHR between April 1, 2016, and March 31, 2018, who had no history of lung cancer. A retrospective cohort (N=873,598) and a prospective cohort (N=836,659) were formed for model construction and validation. An Extreme Gradient Boosting (XGBoost) algorithm was adopted to build the model. It assigned a score to each individual to quantify the probability of a new incident lung cancer diagnosis from October 1, 2016, to September 31, 2017. The model was trained with the clinical profile in the retrospective cohort from the preceding 6 months and validated with the prospective cohort to predict the risk of incident lung cancer from April 1, 2017, to March 31, 2018.

Results: The model had an area under the curve (AUC) of 0.881 (95% CI 0.873-0.889) in the prospective cohort. Two thresholds of 0.0045 and 0.01 were applied to the predictive scores to stratify the population into low-, medium-, and high-risk categories. The incidence of lung cancer in the high-risk category (579/53,922, 1.07%) was 7.7 times higher than that in the overall cohort (1167/836,659, 0.14%). Age, a history of pulmonary diseases and other chronic diseases, medications for mental disorders, and social disparities were found to be associated with new incident lung cancer.

Conclusions: We retrospectively developed and prospectively validated an accurate risk prediction model of new incident lung cancer occurring in the next 1 year. Through statistical learning from the statewide EHR data in the preceding 6 months, our model was able to identify statewide high-risk patients, which will benefit the population health through establishment of preventive interventions or more intensive surveillance.

(J Med Internet Res 2019;21(5):e13260) doi:10.2196/13260

KEYWORDS

lung cancer; risk prediction model; electronic health records; prospective study

Introduction

Background

Lung cancer is the most common cancer and leading cause of cancer death worldwide [1,2]. In 2018, the number of new cases of lung and bronchus cancer was estimated to be 234,030 (13.5% of all new cancer cases); an estimated 154,050 people will die of this disease (25.3% of all cancer-related deaths) in the United States alone [3]. Statistics on survival in people with lung cancer vary depending on the stage of the cancer when it is diagnosed. Early captures at stage I have a 56.3% 5-year survival rate, which decreases to 4.7% by stage III in the United States, based on data from the Surveillance, Epidemiology, and End Results Program 18, 2008-2014 [3]. Most people with lung cancer are diagnosed at a late stage when curative treatment is less effective. Therefore, early detection and timely disease intervention play an important role in reducing the mortality rate of lung cancer.

Annual low-dose computed tomography (LDCT) screening is a viable screening tool for early lung cancer detection. The US-based National Lung Screening Trial demonstrated that LDCT screening reduced lung cancer mortality by 20% relative to conventional chest x-ray screening [4]. However, the screening criteria for LDCT are only age (55-74 years) and smoking history (>30 pack-years, <15 years quit time) [5]. Therefore, a lot of patients take unnecessary tests, which is a serious misuse of social resources; at the same time, many people who seem healthy have been missed [6]. However, so far, there is no tool aimed at the whole population. An effective risk prediction model is critically needed for the initial screening of high-risk patients at the population level, which would hold promise for seeking out those high-risk individuals for further LDCT examination, ensuring that resources are focused on those who are most likely to benefit from them.

Accurate lung cancer risk prediction models would facilitate early diagnoses, decrease mortality rates, and reduce overall costs, ultimately benefiting patients, clinicians, and health care providers.

Possible Limitations of Existing Lung Cancer Risk Prediction Models

Many lung cancer risk prediction models have been proposed [7-21]; however, the clinical needs have not been sufficiently addressed [22]. Most of the recent lung cancer risk prediction models were developed (1) with a small number of risk predictors (N<15) [7,9-21], (2) with a small sample size and using data from only a single medical facility [12,17,21], and (3) with a focus on a particular subgroup of the population (eg, age>45 or smokers) [7,9-14,16,18-20] and the lack of generalizability across the heterogenous population [7-20]. Furthermore, most prior studies used smoking status (eg, smoking duration, smoking intensity, and years since cessation) as a risk factor and predictor [10,12,15,17,20,21], which may not be readily available in many of the medical data sources.

The use of electronic health records (EHRs) has increased dramatically in recent years. The large size and high-dimensional clinical patient information captured in EHRs may be more reflective of the characteristics of the general population than those of cohort studies based on a targeted subgroup of limited profiles. EHRs provide a unique opportunity to understand the health care status at the population level [23]. EHR-based models were developed for diseases including but not limited to type 2 diabetes, chronic kidney disease, and hypertension [23-32]; however, no model was reported to predict new incident lung cancer based on statewide EHR data in the United States.

Aim

Our study aimed to prospectively estimate the future 1-year risk of new incident lung cancer in a US state population. The predictive model uses the preceding 6 months' EHR information including current health conditions, diagnosed diseases, symptoms, laboratory tests, medication history, clinical utilization measures, and social determinants. The model outputs a risk score that describes the probability of a diagnosis of new incident lung cancer in the next 1 year. The risk scores stratify patients into low-, medium-, and high-risk categories, by which limited health care resources can be targeted to high-risk groups to allow proactive intervention, which can ultimately allow early detection of cancer and reduction of regional/statewide lung cancer mortality rates.

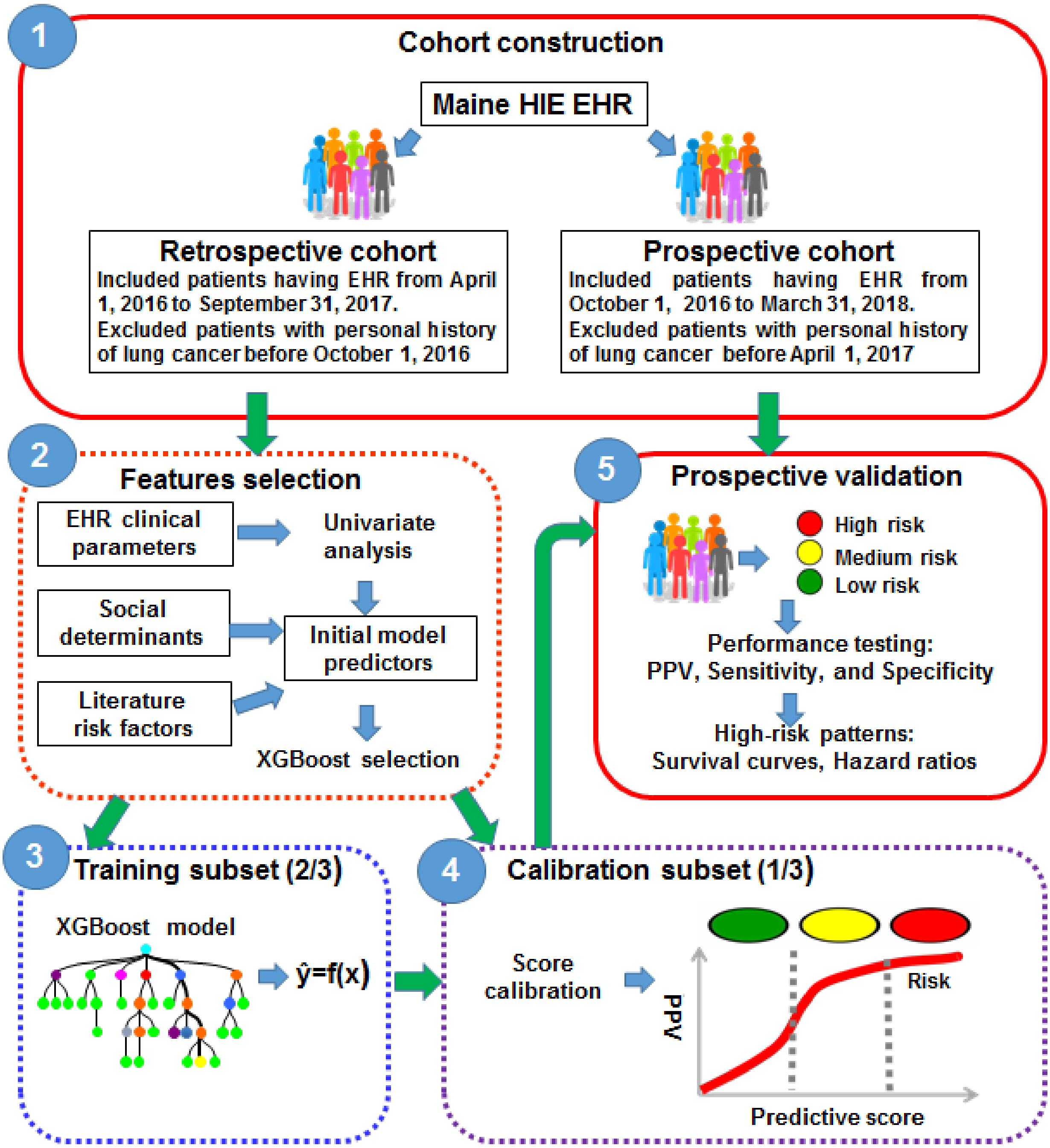
Methods

A workflow to develop the new incident lung cancer risk prediction model is provided in Figure 1. This study includes five steps from cohort construction to prospective validation.

Ethics Statement

Protected personal health information was removed before the process of analysis and publication. Because this study analyzed deidentified data, it was exempted from ethics review by the Stanford University Institutional Review Board (March 20, 2017).

Figure 1. A workflow to develop the new incident lung cancer risk prediction model. EHR: electronic health record; HIE: health information exchange; PPV: positive predictive value. XGBoost: Extreme Gradient Boosting.



Population and Data Sources

The EHRs of Maine Health Information Exchange (HIE; HealthInfoNet) dataset cover records of nearly 95% of the population of the state of Maine [33]. The study included patients who visited any care facility, 35 hospitals, 34 federally qualified health centers, and more than 400 ambulatory practices in the Maine state, from April 1, 2016, to March 31, 2018.

Lung Cancer Definition

Lung cancer in this study was defined using International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) diagnosis codes. The diagnosis codes included category C34 (malignant neoplasm of bronchus and lung), C39 (malignant neoplasm of other and ill-defined sites in the respiratory system and intrathoracic organs), and C46.5 (Kaposi sarcoma of the lung).

Cohort Construction

This study contains a retrospective cohort and a prospective cohort (Figure 1). The retrospective cohort contained 873,598 patients with EHRs from April 1, 2016, to September 31, 2016, and 1091 of them developed lung cancer in the next year (from October 1, 2016, to September 31, 2017). Patients were excluded from the retrospective cohort if there was any record of a lung cancer diagnosis before October 1, 2016. The prospective cohort included 836,659 patients from October 1, 2016, to March 31, 2017, and 1167 of them were diagnosed with lung cancer in the next year (from April 1, 2017, to March 31, 2018). Patients with a history of lung cancer before April 1, 2017, were excluded from the prospective cohort.

Feature Selection

The clinical parameters extracted from EHRs included demographic information, disease diagnoses (primary and secondary), symptoms and procedures (coded using ICD-10-CM), laboratory test results (coded by Logical Observation Identifier Names and Codes and labeled as abnormal or normal according to thresholds provided by each facility participating in the HIE network), clinical utility records, and outpatient medication prescriptions (coded according to the National Drug Code and referred to the number of prescriptions for a particular medicine during the past 6 months). We also extracted a number of accessible social determinants from the US census website [34] using zip code or county name in the Advanced Search of American FactFinder (Multimedia Appendix 1). These social determinants were mapped to the EHR database through a patient's zip code. In addition, features associated with lung cancer identified by previous studies were also extracted as risk factors. Those risk factors included demographics (ie, age and gender), smoking, pulmonary diseases (ie, chronic obstructive pulmonary disease [COPD], chronic bronchitis, emphysema, and pneumonia), symptoms (ie, hemoptysis, cough, and chest pain), and abnormal laboratory test results (ie, C-reactive protein and fibrinogen). Overall, there were 33,788 features in our original data pool.

Given that high-dimensional EHR data are sparse and subject to noisy and missing data, a feature selection process was adopted before model construction. The process had included univariate analysis and XGBoost selection. For EHR clinical parameters, a univariate correlation filtering analysis was adopted to remove features that are not significantly related to lung cancer (*P*>.05). Specifically, the Cochran-Mantel-Haenszel test [35], capable of testing the association between a binary predictor and a binary outcome while taking into account the stratification, was applied to investigate the association between the binary features and the targeted outcome under age-group strata. The Cochran-Armitage trend test [35,36], also called the Chi-square test for trend, was used in the analysis of categorical data to assess the strength of the association between an ordinal variable with *k* categories. Univariate logistic regression [37] was used to assess features that are continuous variables. Social determinants, literature risk factors, and features identified by our univariate analysis were combined (N=346) for the downstream XGBoost analytics. XGBoost adopted the approximate greedy algorithm to split trees by sorting and picking features on each node to optimize purity at each splitting level. The algorithm can output estimates of feature importance after going through the training process [38].

Model Construction

Samples in the retrospective cohort (873,598 patients) were randomly split into 2 subsets for training (582,398 patients) and calibration (291,200 patients) purposes. The model construction, using the retrospective dataset, was accomplished in two phases: (1) the training subset was used to develop the initial model and generate predictive estimates and (2) the calibration subset was used to convert predictive estimates to risk scores for each patient.

Training

XGBoost [38], a gradient tree boosting algorithm, was applied to develop a prediction model. XGBoost algorithm is designed to discover statistical patterns in high-dimensional and multivariate datasets and is able to handle nonlinear correlations and random errors both in input features and the output variable [39]. We used binary classification with logistic objective function for the predictive estimate. The output predictions were probability confidence scores in (0,1), corresponding to the probability of receiving a new diagnosis of incident lung cancer within the next 1 year. The objective was implemented in the "xgboost" package for the R language provided by the creators of the algorithm. The output of the algorithm can be written as

where *F* represents the space of a set of classification trees and *K* is the maximum number of trees (*K*=500 in this study). Each f_k corresponds to an independent tree, and the maximum depth of each tree was set to 5 in this study. The final prediction was calculated by summing up the scores of all the individual trees. To avoid overfitting, the model at the *t*-th iteration was trained to minimize the following item,

where

was the prediction of the *i*-th instance at the *t-1*-th iteration, and *l* is a differentiable convex loss function. The term Ω indicates the penalty of the model complexity and is defined as

where γ and are parameters controlling penalty for the number of leaves *T* and magnitude of leaf weights *w*, respectively. The penalty parameter is selected by cross-validation from the values ranging from $\lambda=10e^{2}$ to $\lambda=10e^{-2}$, essentially covering the full range of scenarios from the null model containing no penalty to the least squares fit.

An approximate algorithm was used to split the finding. It first proposes candidate splitting points according to percentiles of feature distribution, following which splitting points were chosen to optimize purity at the next level.

Calibration

A calibration process was launched to map the predictive estimates of XGBoost to a measure of positive predictive values (PPVs) [40] in the retrospective cohort. It provided a universal, standardized risk measure. A PPV of a corresponding predictive estimate was defined as the proportion of new incident lung cancer events in the cohort with predictive estimates the same as or larger than . Thus, PPVs could be interpreted as risk scores. Following that, we further ranked individuals by their risk scores from low to high, and two risk thresholds were applied to subgroup all patients into low-risk, medium-risk, and high-risk groups.

The scores after calibration were converted to relative risks. The relative risk of each individual was calculated by dividing the score of the individual by the mean score of all patients in the cohort. The relative risk measured the ratio of the probability of having lung cancer to the population baseline. The higher the relative risk, the higher was the probability of receiving a diagnosis of lung cancer in the next year.

Prospective Validation

The model was tested with the prospective cohort (836,659 patients). Performance of the model was investigated within each risk category in terms of PPV, sensitivity, and specificity. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) were also calculated. Relative risk of a subgroup (the ratio of the mean score of the patients in the subgroup to the population mean) was used to measure the increase or decrease in the chance of obtaining a new diagnosis of lung cancer in the next year for patients in the subgroup, compared to the population baseline.

Age- and gender-adjusted odds ratios (ORs) between cases and controls were calculated for top features using logistic regression. Clinical patterns stratified by risk categories were explored and compared. Multivariable Cox regression was used for subpopulation comparison. Spearman rank correlations were performed to assess the correlation between social determinants and the next 1-year risk of lung cancer.

Results

Baseline Characteristics

Baseline demographic and clinical features of the retrospective and prospective cohorts are summarized in Table 1. Most characteristics were similarly distributed between these two cohorts.

Model Performance

By applying the XGBoost algorithm on the EHR-based data, the prediction model reached an AUC of 0.881 (95% CI 0.873-0.889) in the prospective cohort (Figure 2). The model also had effective discriminatory power within patient subgroups: (1) the smoking subgroup (14,248/836,659, 1.7%) with an AUC of 0.865 (95% CI 0.823-0.907), (2) subgroup of age≥65 years (220,702/836,659, 26.4%) with an AUC of 0.755 (95% CI 0.738-0.772), and (3) subgroup of age<45 years (366,752/836,659, 43.8%) with an AUC of 0.880 (95% CI 0.776-0.984; Figure 2). Predictive scores of cases and controls in the prospective cohort were analyzed using the Wilcoxon test [41] (*P*<.001), supporting the statistical importance of our results.

To explore the effectiveness and advantages of our model, we compared the predictive performance of our XGBoost algorithm with a few state-of-the-art existing predictive algorithms in the prospective cohort. Algorithms included RandomForest [42], Boosting [43], Support Vector Machine [44], Lasso [45], and k-nearest neighbors (KNN) [46]. Multimedia Appendix 2 compared the ROC AUCs with the 95% CIs of our model and other existing predictive algorithms to predict the future risk of new incident lung cancer in the next 1 year. Algorithm performances (ROC AUC) were compared and the differences were quantified using the deLong method [47]. For all comparisons, our model outperformed other models, with significantly superior predictive performance (*P*<.001). We also compared our model's predictive performance to other feature selection methods including information gain and Gini index methods. The results were shown in Multimedia Appendix 3. Our comparative results showed that the model predictive performance based on our feature selection method outperformed the performance based on the other methods in terms of the ROC AUC (*P*<.001; *P* values calculated by the deLong method to compare the AUCs).

Table 1. Baseline characteristics of the retrospective cohort (N=873,598) and prospective cohort (N=836,659).

Characteristic	Retrospective cohort, n (%)	Prospective cohort, n (%)
Age (years)		
<45	385,009 (44.1)	366,752 (43.8)
45-54	116,655 (13.4)	109,986 (13.1)
55-64	143,960 (16.5)	139,219 (16.6)
≥65	227,974 (26.1)	220,702 (26.4)
Gender		
Male	386,251 (44.2)	369,022 (44.1)
Female	487,347 (55.8)	467,637 (55.9)
Smoking[a]	16,611 (1.9)	14,248 (1.7)
Other cancer history	59,239 (6.8)	72,039 (8.6)
Pulmonary disease		
COPD[b]	32,180 (3.7)	36,221 (4.3)
Pneumonia	9,896 (1.1)	12,179 (1.5)
Other respiratory disorders[c]	5131 (0.6)	5738 (0.7)
Other chronic disease		
Diabetes	73,854 (8.5)	70,005 (8.4)
CVDs[d]	166,088 (19)	161,685 (19.3)
CKD[e]	18,458 (2.1)	18,912 (2.3)
Symptom		
Cough	26,574 (3)	36,810 (4.4)
Chest pain	32,101 (3.7)	35,057 (4.2)
Hemoptysis	770 (0.1)	981 (0.1)
Dyspnea	4071 (0.5)	3755 (0.5)
Pleural effusion	2024 (0.2)	2356 (0.3)
Abnormal weight loss	6136 (0.7)	5801 (0.7)
Abnormal laboratory test		
C-reactive protein test	11,613 (1.3)	8,517 (1)
Leukocytes count	90,131 (10.3)	71,694 (8.6)
Platelets	69,334 (7.9)	51,477 (6.2)
Glomerular filtration rate	21,446 (2.5)	18,695 (2.2)
Glucose in serum or plasma	137,575 (15.8)	103,671 (12.4)

[a] Smoking was defined with a diagnosis code of Z72_2 from the International Classification of Diseases, 10th Revision, Clinical Modification.

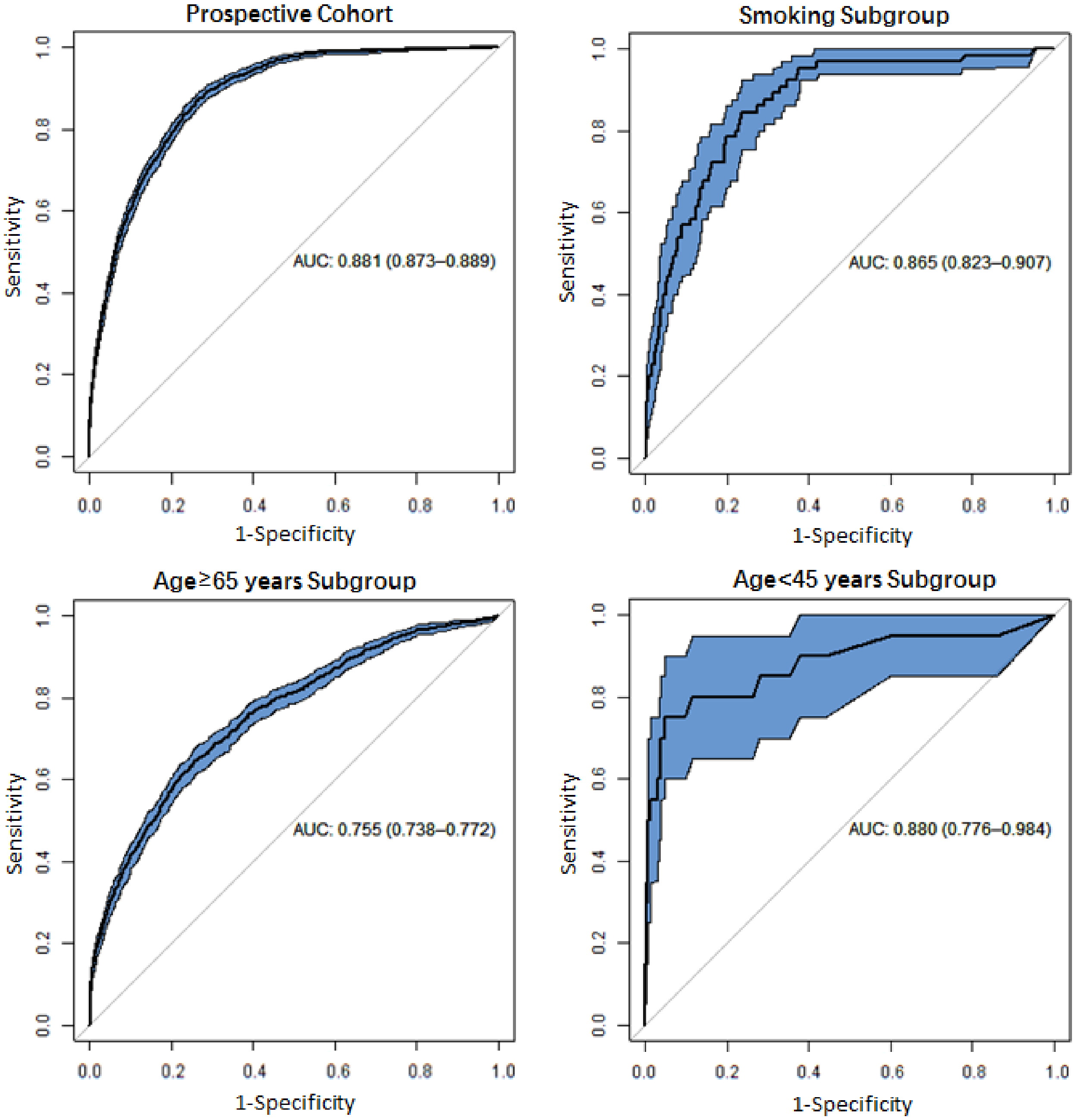
[b] COPD: chronic obstructive pulmonary disease (including chronic bronchitis and emphysema).

[c] Other respiratory disorders were defined with a diagnosis code of J98 from the from the International Classification of Diseases, 10th Revision, Clinical Modification.

[d] CVD: cardiovascular disease (including hypertension, coronary artery disease, peripheral vascular disease, arrhythmia, and abdominal aortic aneurysm).

[e] CKD: chronic kidney disease.

Figure 2. The receiver operating characteristic curves derived from the prospective cohort, smoking subgroup, age≥65 years subgroup, and age<45 years subgroup. The 95% CI of each receiver operating characteristic curve is indicated by the blue shaded area and the AUC (with 95% CI) of each subgroup is listed under each receiver operating characteristic curve. AUC: area under the curve.

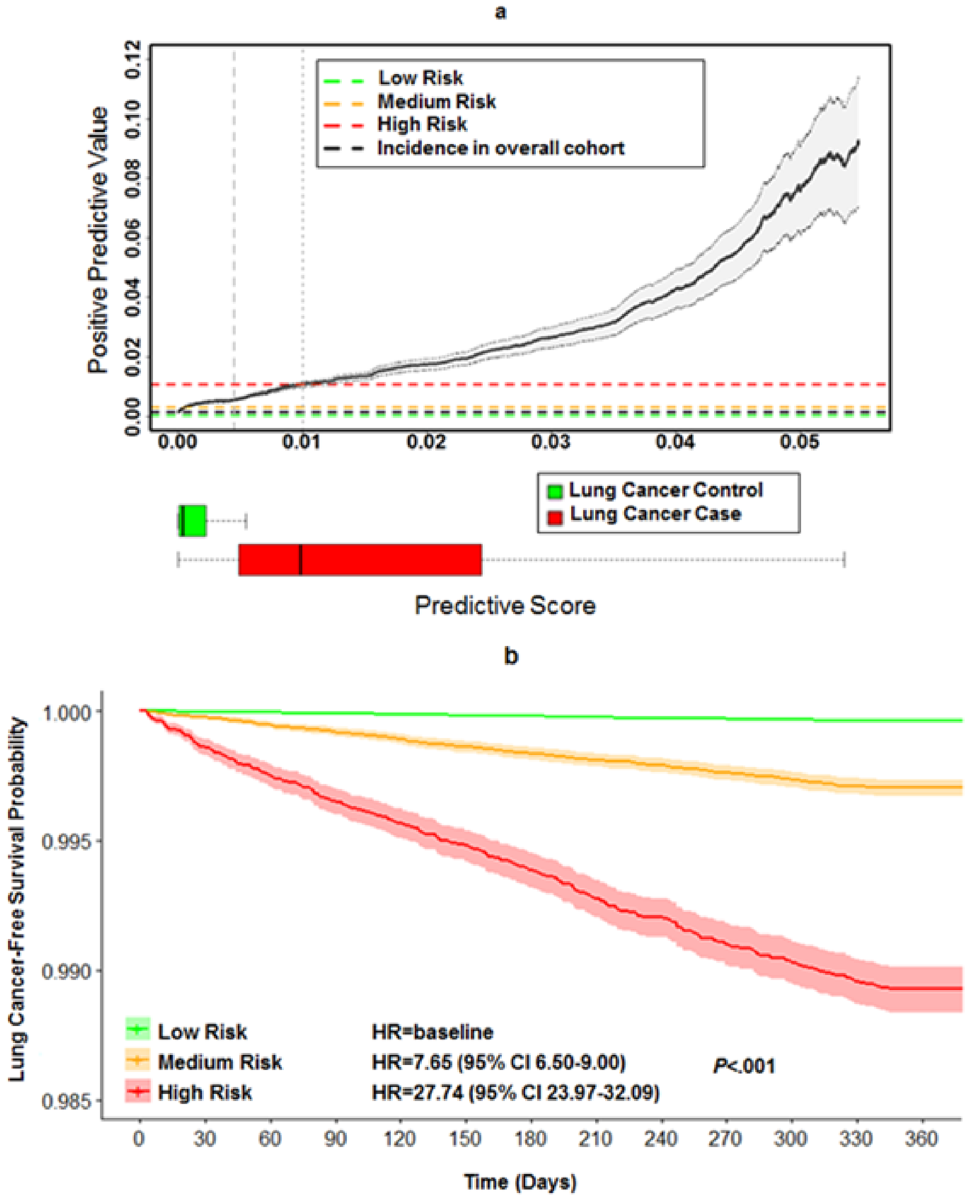


The relationship between the PPVs and predictive scores with the prospective cohort is shown in Figure 3 a. Vertical dashed lines indicate two thresholds of 0.0045 and 0.01 of the predictive scores to group all patients into three risk categories (low, medium, and high). Horizontal dashed lines indicate the incidence of lung cancer in the overall cohort (0.14%, black), low-risk category (0.04%, green), medium-risk category (0.3%, orange), and high-risk category (1.07%, red). The 95% CI of the PPV curve is indicated by the gray shaded area. The box plots at the bottom show the distributions of predictive scores. The performance of our model within each risk category in terms of PPV, sensitivity, specificity, and mean relative risk is shown in Multimedia Appendix 4. Among the 1167 patients in the prospective cohort with confirmed lung cancer in the next 1 year, about half (579/1167, 49.61%) were correctly classified into the high-risk category (with a score≥0.01), and only 22.45% (262/1167) of them were classified into the low-risk category (with a score<0.0045). The relative risk showed a monotonic increase from the low-risk category (0.28) to the high-risk category (7.7). We also calculated the sensitivity and specificity of our model based on the best cut-off threshold for predictive scores, which was defined as the point at which the Youden index (sensitivity+specificity-1) [48] is maximum. After considering the Youden index, the best cut-off point was 0.0029, and the corresponding sensitivity and specificity of our model were 0.8363 and 0.7681, respectively.

A survival analysis using univariable cox regression was performed on each risk category to further evaluate the model performance. Three distinct survival curves stratified patients in terms of lung cancer hazard (*P*<.001), yielding a hazard ratio (HR) as high as 27.74 (95% CI 23.97-32.09) for the high-risk category relative to the low-risk group (Figure 3 b). In addition, our model identified 41.82% (289/691) of high-risk patients 6 months or more prior to assignment of a lung cancer diagnosis code. A total of 68.02% (470/691) of lung cancer cases were identified as high risk at least 3 months before the confirmatory diagnosis was made by physicians.

From the original 33,788 features, 346 features survived from the first step of feature selection process (filtered by univariate analysis) and 118 features were identified by XGBoost algorithm as final predictors of the model (filtered by nonzero weight in algorithm). They consisted of two demographic features, 11 social determinations, 19 diagnostic diseases, 9 clinical symptoms, 28 laboratory tests, 37 medication prescriptions, and 12 clinical utilization measures. The top 60 features with their age-gender adjusted ORs or coefficients are listed in Multimedia Appendix 5. COPD, pneumonia, and other respiratory disorders were recognized as the pulmonary diseases most associated with lung cancer, with ORs of 4.978, 2.790, and 5.484, respectively. Other cancer history, cardiovascular diseases (CVDs), diabetes, and CKD were considered to be chronic diseases most associated with lung cancer, with ORs of 1.899, 1.329, 1.374, and 1.270, respectively. Smoking also had a strong association with lung cancer (OR=4.084). Hemoptysis, pleural effusion, cough, and abnormal weight loss were recognized as symptoms most associated with lung cancer, with ORs of 5.080, 4.130, 2.108, and 2.010, respectively. For abnormal laboratory test results, inflammation marker C-reactive protein was most associated with lung cancer (OR=1.771). Medications for treatment of chronic diseases and mental disorders, clinical utilization, and social determinants were also detected by the model as powerful predictors of incident lung cancer within the next year.

Figure 3. (a) Stratification of patients in the prospective cohort. Positive predictive value was plotted as a function of the predictive score. Two thresholds of 0.0045 and 0.01 were applied to stratify the population into low-, medium- and high-risk categories. (b) Survival curves of the three risk categories. HR: Hazard Ratio.



Clinical Patterns Stratified by Risk Categories

Distribution patterns of impactful risk predictors were explored and compared among different risk categories: low risk (score 0-0.0045; 673,075 patients), medium risk (score 0.0045-0.01; 109,662 patients), and high risk (score 0.01-1; 53,922).

Age

In our study, age was aggregated into four distinct age groups (<45, 45-54, 55-64, and ≥65 years). A significant difference was found in the age distribution between the low- and the high-risk categories (Multimedia Appendix 7). In the low-risk category, younger individuals (<45) accounted for 54.16% (364,545/673,075), whereas older individuals (≥65 years) accounted for only 13.64% (91,815/673,075). In the high-risk category, the group aged ≥65 years constituted the largest subset (83.32%, 44,927/53,922), whereas the group aged <45 years constituted only 0.79% (426/53,922; Multimedia Appendix 6).

Diagnosed Diseases

History of pulmonary diseases and other chronic diseases also differed between low- and high-risk patients: 50.48% (27,219/53,922), 24.29% (13,100/53,922), and 21.62% (11,659/53,922) of the high-risk patients had CVDs, diabetes, and COPD, respectively, while 12.30% (82,757/673,075), 5.04% (33,927/673,075), and 1.98% (13,322/673,075) of the low-risk patients had these diseases, respectively (Multimedia Appendix 6).

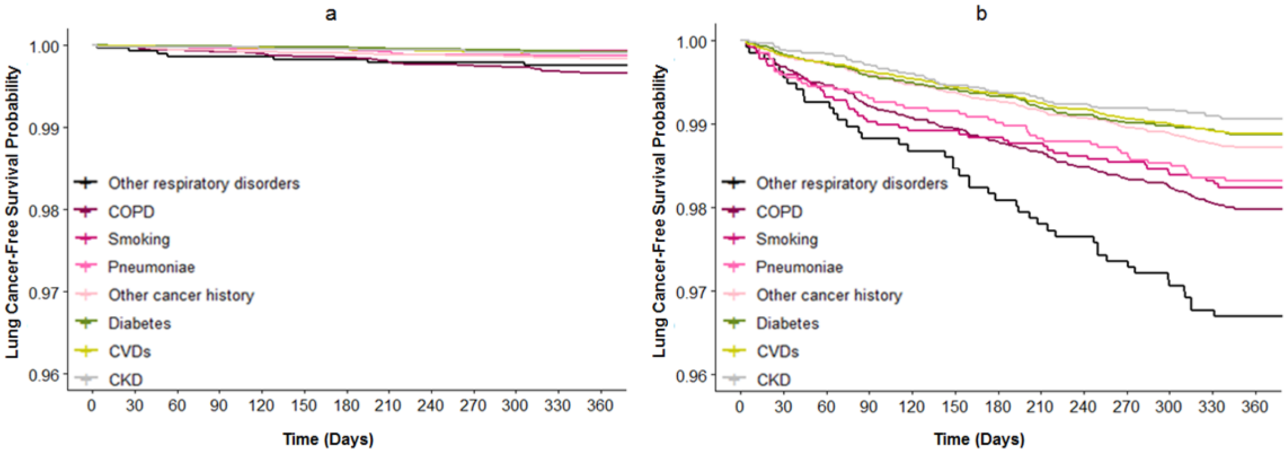
The time-to-diagnosis curves were created using univariable Cox regression to explore lung cancer diagnoses in the high- and low-risk categories in different disease subgroups (ie, COPD, pneumonia, other respiratory disorders, CVDs, diabetes, and CKD), smoking subgroups, and other cancer history subgroups (Figure 4). In the high-risk category, 1.61% (47/2677) of patients with a smoking history and 1.28% (185/14477) of patients with other cancer history received diagnoses of lung cancer in the next 1 year. In addition, 1.68% (46/2734), 2% (235/11,659), and 3.3% (45/1363) of patients with pneumonia, COPD, other respiratory disorders, respectively, received diagnoses of lung cancer in the next 1 year. This probability remained around 1% for patients with CVDs, diabetes, and CKD. These results implied that patients with pulmonary diseases (ie, COPD, pneumonia, or other respiratory disorders) have higher risks for developing lung cancer than patients with other chronic diseases (ie., CVDs, diabetes, and CKD). In the low-risk category, more than 99.6% of the patients were free from development of lung cancer in the next 1 year, and the survival curve dropped faster for patients with pulmonary diseases than for those with other chronic diseases.

We also investigated the time-to-diagnosis curves for patients who only had pulmonary diseases and patients who had pulmonary diseases together with at least one other chronic disease (including diabetes, CVDs, CKD, and other cancer history). Results showed that a history of pulmonary disease together with other chronic diseases increased the risk of incident lung cancer (HR=1.7; Multimedia Appendix 8). In addition, 22.05% (11,890/53,922) of the high-risk patients had pulmonary diseases together with other chronic diseases, while 5.76% (3082/53,922) of the patients had pulmonary diseases only. In the low-risk category, these values were 1.45% (9732/673,075) and 1.95% (13,100/673,075), respectively.

Symptoms and Abnormal Laboratory Test Results

The model recognized 6 symptoms and 10 abnormal laboratory test results as powerful predictors of the 1-year lung cancer risk (Multimedia Appendix 5). Most were acute symptoms related to the respiratory system, and most of the laboratory test results were markers associated with inflammation and chronic diseases. The percentage of patients with these symptoms and abnormal laboratory test results gradually increased from the low- to high-risk category (Multimedia Appendix 6). Furthermore, 23.73% (12,798/53,922) and 39.62% (21,364/53,922) of people in the high-risk category had at least one symptom or one abnormal laboratory test result compared with 9.12% (61,368/673,075) and 11.67% (78,564/673,075) in the low-risk category, respectively.

Figure 4. Time-to-diagnosis curves of the disease subgroup, smoking subgroup, and other cancer history subgroup for the low-risk (a) and high-risk (b) categories of the prospective cohort. Disease subgroups comprised patients who received diagnoses of COPD, pneumonia, other respiratory disorders, diabetes, CVDs, or CKD. CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; CVD: cardiovascular disease.



Mental Disorders

The model identified 15 medications prescribed for mental disorders (eg, depression and anxiety disorders) as impactful features (Multimedia Appendix 5). Mental disease information was unavailable in HIE due to privacy concerns. Our study used mental disorder-related medications to explore the association between mental disorders and lung cancer risk. People with a history of mental disorder-related medications were significantly enriched in the high-risk group (Multimedia Appendix 6), with a total of 22.97% (12,388/53,922) in the high-risk category and 7.36% (49,537/673,073) in the low-risk category.

Survival analysis was performed according to mental health status (mental disorder or no mental disorder) in the subgroup that comprised people with at least one chronic disease diagnosis and the subgroup of patients with no history of chronic diseases (Multimedia Appendix 9). Presence of a mental disorder was associated with an increased risk of lung cancer in both subgroups before and after adjustment for age, gender, and smoking ($P<.001$). Therefore, the presence of mental disorders increased the risk of lung cancer independent of chronic diseases, age, gender, and smoking factors.

Clinical Utilization Indicators

We compared the clinical utilization indicators in our study across three risk categories. These utilization indicators of patients such as outpatient visits, emergency visits, inpatient admissions, inpatient days, clinical cost, and number of chronic diseases in the past 6 months gradually increased from the low- to high-risk categories (Multimedia Appendix 6).

We further compared patients in the high- and low-risk categories by the average clinical costs in the past 6 months in 8 disease subgroups (including COPD, CVDs, pneumonia, other respiratory disorders, diabetes, CKD, mental disorders, and other cancer history) stratified by the average number of chronic diseases (Multimedia Appendix 10). The circles in Multimedia Appendix 10 were formed by 8 disease subgroups under the low-risk (green circle) and high-risk (red circle) categories. The circle size indicates the proportion of the disease subgroup under each risk category. Reference groups consisted of patients with no diagnosis of any of these chronic diseases. It is obvious that the low-risk patients and high-risk patients were separated by the number of the chronic diseases. High-risk patients, in general, had a higher chronic disease burden and thus higher clinical costs.

Social Determinants

In our study, the social determinants were derived from zip code or county-based census and were recognized as community-level social and environmental indicators. Spearman rank correlations were used to investigate the association between social determinants and lung cancer risk (Multimedia Appendix 11). Parameters related to a decreased risk of lung cancer ($\rho<0$) included high education levels (the proportion of people who received college or associate's degree or bachelor's and higher degree), high median household income, high proportion of population within half a mile of a park, and private insurance coverage. The high-risk category had a higher proportion of low-income or low-education populations (the percentage of a combination of the population aged 18-24 years with less than high school graduate diploma education and the population aged ≥25 years with less than 12th grade diploma education in the area) than the low-risk category (Multimedia Appendix 6), indicating a health disparity related to lung cancer.

Discussion

Summary of Main Findings

In this study, we developed and prospectively validated a risk prediction model of the future 1-year incidence of lung cancer using EHR data derived from more than 1.1 million people in the state of Maine. Patients were stratified into three risk categories, ranking the lung cancer risk as low, medium, and high. The model achieved an AUC of 0.881 (95% CI 0.873-0.889) in the prospective cohort, indicating our model's ability to target those most at risk for subsequent prevention management. The incidence of lung cancer in the high-risk category (579/53,922, 1.07%) was 7.7 times higher than that in the overall prospective cohort (1167/836,659, 0.14%). Performance of the model in subgroups (Figure 2), especially those considered low risk by prior models (age<45 years), was fairly good (AUC is 0.880), showing predictive power in patients that traditional models tend to ignore. Key parameters of age, a history of pulmonary diseases and other chronic diseases, medications for mental disorders, and social disparities were found to be significantly associated with incident lung cancer.

Comparison With Prior Work

We compared our model with seven other risk prediction models for incident lung cancer (Multimedia Appendix 12). The models achieved AUCs between 0.57 and 0.87, but with limitations in targeted clinical application. The Bach model [20] and $\text{PLCO}_{\text{m2012}}$ (Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial) model [19] were only applicable to smokers. The Liverpool Lung Project model [12] and two Spitz models [17,21] were developed with case-control matched studies with small sample sizes that were not validated with a general population. The Extended Spitz model [17] required genetic test information, which was unavailable in routine clinical data. The EPIC (European Prospective Investigation into Cancer and Nutrition) [15] and HUNT (Helseundersøkelsen i Nord-Trøndelag) [10] models used smoking status (eg, smoking duration, smoking intensity, and years since cessation) collected from a questionnaire as a predictor, which may not be feasible in a large, general population. Furthermore, our model had a short prediction time frame (1 year) compared with most other studies where the follow-up periods were up to several years. A short prediction time frame resulted in a low incidence (0.12% in the retrospective cohort and 0.14% in prospective cohort), which raised the challenge of prediction. To address this issue, our model adopted more predictors (118 features) than prior risk models (<15 features), making our risk prediction more effective.

Interpretation of Risk Predictors and Implications for Prevention and Early Intervention

Pulmonary Diseases and Inflammation Markers

A total of 27.7% (14,972/53,922) individuals in the high-risk group had one or more pulmonary diseases, which was much higher than the number in the low-risk population (3.39%, 22,832/673,075). The association between the pulmonary diseases and lung cancer was reported in many previous studies [49-52]. Pulmonary diseases induced an inflammatory response in the lung, and inflammation played a critical role in the development of lung cancer [53-56]. The C-reactive protein level and leukocyte count are blood test markers for inflammation. Elevated C-reactive protein levels and leukocytes counts have been found to be associated with pulmonary diseases [57,58] and lung cancer [59], suggesting an etiologic role of pulmonary inflammation in lung cancer pathophysiology [59]. Consistent with these studies, our model recognized abnormal C-reactive protein levels and leukocyte counts as top features (Multimedia Appendix 5). In the high-risk category, 2.52% (1358/53,922) individuals had abnormal C-reactive protein levels and 19.98% (10,774/53,922) had high leukocytes counts compared with 0.7% (4711/673,075) and 6.27% (42,202/673,075) in the low-risk category, respectively (Multimedia Appendix 6).

Combination of Pulmonary Diseases and Other Chronic Diseases

Our study showed that more than 80% (43,165/53,922) of individuals in the high-risk category had at least one chronic disease, and the majority (75.41%, 507,565/673,075) of the low-risk population had no chronic disease diagnosis. Moreover, 22.05% (11,890/53,922) of the high-risk patients had pulmonary diseases together with other chronic diseases (ie, CVDs, CKD, and diabetes), and the risk of incident lung cancer increased among these patients (Multimedia Appendix 8). In addition, the concurrent chronic diseases led to an increased burden of clinical utilization and cost (Multimedia Appendix 10). We also found that incident lung cancer was associated with abnormal results of many chronic disease markers (eg, glomerular filtration rate, glucose level, and platelet count) and a group of previously prescribed medications including drugs for pulmonary diseases (Ipratropium bromide, albuterol sulfate, etc), drugs for diabetes (metformin HCl, glipizide, etc), and drugs for CVDs (amlodipine besylate, diltiazem HCl, valsartan, etc). Such markers and medication histories indicated that patients at risk for or living with diseases might develop lung cancer. These findings were consistent with those of previous studies. High chronic disease burden is a growing concern in the US population. It was reported that 6/10 adult Americans have at least one chronic disease and 4/10 have more than one chronic disease [60,61]. A recent study found that chronic diseases are an overlooked risk factor for cancer, and a substantial cancer risk is associated with a combination of cardiovascular disease markers, diabetes, chronic kidney disease markers, and pulmonary diseases [62], which were found to be linked to the risk of the next-year incident lung cancer by our model.

Mental Disorders

Mental disorders may affect the immune system and endocrine function, thus influencing the body’s susceptibility to cancer [63]. Several studies showed a positive association between mental disorders and the overall risk of cancer [63,64]. For lung cancer, some studies showed an etiological association [65,66], whereas some claimed there was no association between mental disorders and lung cancer [67,68]. We explored the role of this controversial and unclear association in our study.

Due to the data policy of the EHR data on mental illness in the state of Maine, we used the consumption of mental illness–related drugs as a proxy for mental disorders. We found that mental disorders had a positive association with the 1-year lung cancer incidence risk (Multimedia Appendix 9). Patients with no chronic diseases who were undergoing treatment for mental disorders had nearly 2.5 times the risk of incident lung cancer compared to those without any mental disorders, regardless of age, gender, and smoking status. This can be explained by a previous finding that adverse psychological events such as pressure and stress may impair the immune system and cause the development and progression of tumors [52]. A similar correlation was also observed in a study focusing on behavioral immunological activities: Researchers found that unpleasant or hostile emotions could cause immune system disorders, and consequently, the occurrence of tumors [53]. More interestingly, those psychological events were found to be correlated to smoking: People with a diagnosis of posttraumatic stress disorder were found to have a higher rate of smoking (45%) than people without a mental health diagnosis (23%) [69]. Smoking is an important factor of lung cancer. Therefore, close attention should be paid to those with a high-risk mental status to allow prevention of and intervention for lung cancer.

Social Determinants

A few studies have shown that lung cancer risk is inversely associated with socioeconomic status factors such as educational attainment, income, and occupation [70]. Socioeconomic status was found to be linked with health status through multiple pathways such as social resources, physical and psychosocial stressors, and health-related behaviors [71]. Consistent with these studies, our model found that patients with low income or less education had a higher risk of lung cancer (Multimedia Appendix 11). Living distance to parks and coverage by Medicaid were also risk factors in our model. The former may be explained by the fact that physical environment factors such as the concentration of parks in the living area can directly shape peoples’ physical activities and ultimately decrease the lung cancer risk. Low education level could be the causality for future low family income, less access to health care, and attainment of Medicaid health insurance.

Implications of Findings

The predictive model and risk scores can benefit health care organizations at multiple levels. For health care providers, stratifying the population by our risk score will help with budget planning and target intervention. For clinicians, the model can be used as an assistant tool for decision making. Our model can also act as a prescreening tool: High-risk patients identified by our model can be referred to the LDCT screening test to decide whether the patients already have lung cancer.

The ultimate goal of this study is to guide health care providers to make decisions for the prevention and intervention of lung cancer. There are already established guidelines in lung cancer preventive care to address both nonmodifiable and modifiable risk factors. The modifiable risk factors such as concurrent chronic conditions and lifestyles are even more important than nonmodifiable predictors such as age and gender, as they offer an opportunity to both clinicians and patients to proactively manage the disease by implementing interventions before deterioration.

Our model identified 68.02% (470/691) of high-risk patients at least 3 months before the confirmatory diagnosis was made by physicians. This may provide the opportunity of early interventions to prevent or delay the development of lung cancer as well as to reduce corresponding health care expenditures. Early detection of lung cancer can lead to an improved 5-year survival rate [3]. In addition, a recent study revealed that patients who received anti-inflammatory therapy had a marked reduction in the incidence of lung cancer [72]. Another route of intervention is through the management of multiple chronic diseases. Recent studies showed a substantial impact of chronic diseases (eg, cardiovascular diseases, diabetes, chronic kidney diseases, and pulmonary diseases) jointly on cancer risk, which was as important as five lifestyle factors combined (smoking, unhealthy diet, physical inactivity, obesity, and alcohol misuse) [62]. Better management of chronic diseases in primary care is therefore an effective strategy for future cancer prevention. In addition, increasing physical activity is a way to improve lifestyle. Our model found that patients living far away from a park were prone to an elevated risk of lung cancer, indicating that targeting this subgroup of patients with personalized action plans might lead to healthier life styles and a possible reduced risk of lung cancer.

Limitations

Our study has several limitations. First, some data were missing in our dataset. Tobacco use was not fully recorded in the EHR data; occupational exposure and family history of lung cancer were so sparse in the data source that our model did not include them as predictors; and patients with lung cancer might not have any record of this diagnosis, leading to an underestimation of lung cancer prevalence in the study. Second, air quality, cancer biomarker, and some individual-level lifestyle information (eg, diet habit and physical activity) could be potentially useful predictors for development of lung cancer, but in EHRs, these data were not available. Third, the grade and stage of lung cancer were not described in the data source, and the socioeconomic factors were analyzed at a community-level, limiting the findings between the association of the individual socioeconomic status and lung cancer.

Conclusions

A risk prediction model of the future 1-year incidence of lung cancer was developed and prospectively validated using the preceding 6 months' EHR data derived from more than 1.1 million people in the state of Maine. The model was able to assign each individual a risk score and stratified patients into three risk categories of low, medium, and high risk. The model reached an AUC of 0.881 in the prospective cohort. Age, a history of pulmonary diseases and other chronic diseases, medications for mental disorders, and social disparities were found to be associated with new incident lung cancer. Targeting individuals at high risk has the potential to facilitate early intervention and reduce overall costs, which will ultimately benefit patients, clinicians, and health care providers.

Acknowledgments

We would like to thank and express our gratitude to the hospitals, medical practices, physicians, and nurses participating in Maine's HIE. We also thank the biostatistics colleagues at the Department of Health Research and Policy for critical discussions. XW is supported by the China Scholarship Council.

Authors' Contributions

XW, YZ, SH, LZ, CY, MX, OW, ML, and BJ carried out the initial analysis and interpretation of data and drafted the initial manuscript. DBM, KGS, EW, and XBL contributed to the conceptualization of this study. SA, FS, and LK coordinated and supervised data acquisition. JL, SQD, CHW, and XBL contributed to the review and editing of the manuscript. All authors have read and approved this submission for publication. All authors have agreed to be accountable for all aspects of the work.

Conflicts of Interest

KGS, EW, and XBL are cofounders and equity holders of HBI Solutions, Inc, which is currently developing predictive analytics solutions for health care organizations. MX, ML, BJ, OW, FS, LK, and EW are employed by HBI. From the Stanford University School of Medicine, Stanford, California, KS and XBL conducted this research as part of a personal outside consulting arrangement with HBI. The research and research results are not associated with Stanford University in any way. There are no patents, further products in development, or marketed products to declare.

Multimedia Appendix 1

List of social determinant variables downloaded from the US census, with details of the data source and mapping. method.

[PDF File (Adobe PDF File), 99KB - jmir_v21i5e13260_app1.pdf]

Multimedia Appendix 2

Comparative analysis of the model performance, quantified by the area under the curve (AUC) with 95% CI. (1) Our Extreme Gradient Boosting (XGBoost) algorithm, (2) RandomForest, (3) Boosting, (4) Support Vector Machine (SVM), (5) LASSO, and (6) K-Nearest Neighbors (KNN).

[PDF File (Adobe PDF File), 145KB - jmir_v21i5e13260_app2.pdf]

Multimedia Appendix 3

Comparison of the predictive performances to predict future 1-year risk of new incident lung cancer in the prospective cohort, measured by the receiver operating characteristic area under the curve (ROC AUC): Model 1 is based on our feature selection method, Model 2 is with the Gini index, and Model 3 is with information gain feature selection methods.

[PDF File (Adobe PDF File), 132KB - jmir_v21i5e13260_app3.pdf]

Multimedia Appendix 4

The performance of the 1-year lung cancer risk prediction model in the prospective cohort, summarized as positive prediction value, sensitivity, specificity, and mean relative risk.

[PDF File (Adobe PDF File), 86KB - jmir_v21i5e13260_app4.pdf]

Multimedia Appendix 5

The top 60 features selected by our lung cancer prediction model.

[PDF File (Adobe PDF File), 71KB - jmir_v21i5e13260_app5.pdf]

Multimedia Appendix 6

Distribution of top risk predictors across the three risk categories.

[PDF File (Adobe PDF File), 38KB - jmir_v21i5e13260_app6.pdf]

Multimedia Appendix 7

Constituent ratios of age subgroups across the identified three risk categories.

[PDF File (Adobe PDF File), 42KB - jmir_v21i5e13260_app7.pdf]

Multimedia Appendix 8

The time-to-diagnosis curves for patients who only had pulmonary diseases and patients who had pulmonary diseases together with at least one other chronic disease in the prospective cohort.

[PDF File (Adobe PDF File), 56KB - jmir_v21i5e13260_app8.pdf]

Multimedia Appendix 9

Predicted time-to-diagnosis curves for patients who had at least one diagnosis of chronic diseases and those who did not receive a diagnosis of chronic diseases. Curves for both subgroups were stratified by mental health status (Mental Disorder vs No Mental Disorder).

[PDF File (Adobe PDF File), 74KB - jmir_v21i5e13260_app9.pdf]

Multimedia Appendix 10

Graph of patients’ average clinical costs in the past 6 months against the average number of chronic diseases.

[PDF File (Adobe PDF File), 115KB - jmir_v21i5e13260_app10.pdf]

Multimedia Appendix 11

Spearman rank correlation between six social determination features and prospective lung cancer risk scores.

[PDF File (Adobe PDF File), 86KB - jmir_v21i5e13260_app11.pdf]

Multimedia Appendix 12

Characteristics of the compared risk models.

[PDF File (Adobe PDF File), 100KB - jmir_v21i5e13260_app12.pdf]

References

1. Boloker G, Wang C, Zhang J. Updated statistics of lung and bronchus cancer in United States (2018). J Thorac Dis 2018 Mar;10(3):1158-1161 [FREE Full text] [doi: 10.21037/jtd.2018.03.15] [Medline: 29708136]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018 Jan;68(1):7-30 [FREE Full text] [doi: 10.3322/caac.21442] [Medline: 29313949]
3. Cancer Stat Facts: Lung and Bronchus Cancer. URL: https://seer.cancer.gov/statfacts/html/lungb.html [accessed 2018-12-20] [WebCite Cache ID 74oO19kRC]
4. National Lung Screening Trial Research Team, Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, et al. The National Lung Screening Trial: overview and study design. Radiology 2011 Jan;258(1):243-253 [FREE Full text] [doi: 10.1148/radiol.10091808] [Medline: 21045183]
5. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2018: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin 2018 Jul;68(4):297-316 [FREE Full text] [doi: 10.3322/caac.21446] [Medline: 29846940]
6. Tota JE, Ramanakumar AV, Franco EL. Lung cancer screening: review and performance comparison under different risk scenarios. Lung 2014 Feb;192(1):55-63. [doi: 10.1007/s00408-013-9517-x] [Medline: 24153450]
7. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and Validation of Risk Models to Select Ever-Smokers for CT Lung Cancer Screening. JAMA 2016 Jun 07;315(21):2300-2311 [FREE Full text] [doi: 10.1001/jama.2016.6255] [Medline: 27179989]
8. Ten Haaf K, Jeon J, Tammemägi MC, Han SS, Kong CY, Plevritis SK, et al. Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study. PLoS Med 2017 Apr;14(4):e1002277 [FREE Full text] [doi: 10.1371/journal.pmed.1002277] [Medline: 28376113]
9. Charvat H, Sasazuki S, Shimazu T, Budhathoki S, Inoue M, Iwasaki M, JPHC Study Group. Development of a risk prediction model for lung cancer: The Japan Public Health Center-based Prospective Study. Cancer Sci 2018 Mar;109(3):854-862 [FREE Full text] [doi: 10.1111/cas.13509] [Medline: 29345859]
10. Markaki M, Tsamardinos I, Langhammer A, Lagani V, Hveem K, Røe OD. A Validated Clinical Risk Prediction Model for Lung Cancer in Smokers of All Ages and Exposure Types: A HUNT Study. EBioMedicine 2018 May;31:36-46 [FREE Full text] [doi: 10.1016/j.ebiom.2018.03.027] [Medline: 29678673]
11. Samson M, Nomura S, Dash C, Bethea T, Rosenberg L, Adams-Campbell L. Abstract 5252: Prospective cohort study of physical activity and lung cancer risk in the Black Women's Health Study. Cancer Res 2018 Jul 01;78(13 Supplement):5252-5252. [doi: 10.1158/1538-7445.AM2018-5252]
12. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008 Jan 29;98(2):270-276 [FREE Full text] [doi: 10.1038/sj.bjc.6604158] [Medline: 18087271]
13. Cronin KA, Gail MH, Zou Z, Bach PB, Virtamo J, Albanes D. Validation of a model of lung cancer risk prediction among smokers. J Natl Cancer Inst 2006 May 03;98(9):637-640. [doi: 10.1093/jnci/djj163] [Medline: 16670389]
14. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, et al. Development and validation of a lung cancer risk prediction model for African-Americans. Cancer Prev Res (Phila) 2008 Sep;1(4):255-265 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-08-0082] [Medline: 19138969]
15. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Østergaard JN, et al. A risk model for lung cancer incidence. Cancer Prev Res (Phila) 2012 Jun;5(6):834-846 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0237] [Medline: 22496387]
16. Maisonneuve P, Bagnardi V, Bellomi M, Spaggiari L, Pelosi G, Rampinelli C, et al. Lung cancer risk prediction to select smokers for screening CT--a model based on the Italian COSMOS trial. Cancer Prev Res (Phila) 2011 Nov;4(11):1778-1789 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0026] [Medline: 21813406]
17. Spitz MR, Etzel CJ, Dong Q, Amos CI, Wei Q, Wu X, et al. An expanded risk prediction model for lung cancer. Cancer Prev Res (Phila) 2008 Sep;1(4):250-254 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-08-0060] [Medline: 19138968]
18. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, et al. Lung cancer risk prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial models and validation. J Natl Cancer Inst 2011 Jul 06;103(13):1058-1068 [FREE Full text] [doi: 10.1093/jnci/djr173] [Medline: 21606442]
19. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013 Feb 21;368(8):728-736 [FREE Full text] [doi: 10.1056/NEJMoa1211776] [Medline: 23425165]

20. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst 2003 Mar 19;95(6):470-478. [Medline: 12644540]
21. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst 2007 May 02;99(9):715-726. [doi: 10.1093/jnci/djk153] [Medline: 17470739]
22. Gray EP, Teare MD, Stevens J, Archer R. Risk Prediction Models for Lung Cancer: A Systematic Review. Clin Lung Cancer 2016 Mar;17(2):95-106. [doi: 10.1016/j.cllc.2015.11.007] [Medline: 26712102]
23. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208. [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
24. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based Real-Time Case Finding for the Population Health Management of Patients With Diabetes Mellitus: A Prospective Validation of the Natural Language Processing-Based Algorithm With Statewide Electronic Medical Records. JMIR Med Inform 2016 Nov 11;4(4):e37 [FREE Full text] [doi: 10.2196/medinform.6328] [Medline: 27836816]
25. Guo Y, Zheng G, Fu T, Hao S, Ye C, Zheng L, et al. Assessing Statewide All-Cause Future One-Year Mortality: Prospective Study With Implications for Quality of Life, Resource Utilization, and Medical Futility. J Med Internet Res 2018 Jun 04;20(6):e10311 [FREE Full text] [doi: 10.2196/10311] [Medline: 29866643]
26. Hao S, Jin B, Shin AY, Zhao Y, Zhu C, Li Z, et al. Risk prediction of emergency department revisit 30 days post discharge: a prospective study. PLoS One 2014;9(11):e112944 [FREE Full text] [doi: 10.1371/journal.pone.0112944] [Medline: 25393305]
27. Hao S, Wang Y, Jin B, Shin AY, Zhu C, Huang M, et al. Development, Validation and Deployment of a Real Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange. PLoS One 2015;10(10):e0140271 [FREE Full text] [doi: 10.1371/journal.pone.0140271] [Medline: 26448562]
28. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. J Med Internet Res 2018 Jan 30;20(1):e22 [FREE Full text] [doi: 10.2196/jmir.9268] [Medline: 29382633]
29. Hao S, Fu T, Wu Q, Jin B, Zhu C, Hu Z, et al. Estimating One-Year Risk of Incident Chronic Kidney Disease: Retrospective Development and Validation Study Using Electronic Medical Record Data From the State of Maine. JMIR Med Inform 2017 Jul 26;5(3):e21 [FREE Full text] [doi: 10.2196/medinform.7954] [Medline: 28747298]
30. Li L, Cheng W, Glicksberg BS, Gottesman O, Tamler R, Chen R, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 2015 Oct 28;7(311):311ra174 [FREE Full text] [doi: 10.1126/scitranslmed.aaa9364] [Medline: 26511511]
31. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. J Am Med Inform Assoc 2009;16(3):371-379 [FREE Full text] [doi: 10.1197/jamia.M2846] [Medline: 19261943]
32. Amarasingham R, Velasco F, Xie B, Clark C, Ma Y, Zhang S, et al. Electronic medical record-based multicondition models to predict the risk of 30 day readmission or death among adult medicine patients: validation and comparison to existing models. BMC Med Inform Decis Mak 2015 May 20;15:39 [FREE Full text] [doi: 10.1186/s12911-015-0162-6] [Medline: 25991003]
33. HealthInfoNet. URL: http://hinfonet.org [accessed 2019-05-07] [WebCite Cache ID 78C83kJOv]
34. The United States Census Bureau. URL: https://www.census.gov/ [accessed 2018-09-20] [WebCite Cache ID 780hEgsOX]
35. Agresti A. Categorical Data Analysis (wiley Series In Probability And Statistics). Hoboken, NJ: Wiley-Interscience; 2019:2002.
36. Sasieni PD. From genotypes to genes: doubling the sample size. Biometrics 1997 Dec;53(4):1253-1261. [Medline: 9423247]
37. Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. Biometrika 1967 Jun;54(1):167-179. [Medline: 6049533]
38. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13 - 17, 2016; San Francisco, California, USA p. 785-794. [doi: 10.1145/2939672.2939785]
39. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 Jun;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
40. Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. BMJ 2006 Aug 12;333(7563):327 [FREE Full text] [doi: 10.1136/bmj.38870.657917.AE] [Medline: 16815882]
41. Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1945;1(6):80-83.
42. randomForest: Breiman and Cutler's Random Forests for Classification and Regression. URL: https://cran.r-project.org/web/packages/randomForest/index.html [accessed 2018-11-20] [WebCite Cache ID 781QCayMU]
43. gbm: Generalized Boosted Regression Models. URL: https://cran.r-project.org/web/packages/gbm/index.html [accessed 2018-11-20] [WebCite Cache ID 781QSUtmm]

44. kernlab: Kernel-Based Machine Learning Lab. URL: https://cran.r-project.org/web/packages/kernlab/ [accessed 2018-11-20] [WebCite Cache ID 781Pppx4w]
45. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. URL: https://cran.r-project.org/web/packages/glmnet/index.html [accessed 2018-11-20] [WebCite Cache ID 781Qfckq7]
46. FNN: Fast Nearest Neighbor Search Algorithms and Applications. URL: https://cran.r-project.org/web/packages/FNN/index.html [accessed 2018-11-20] [WebCite Cache ID 781RvvHOQ]
47. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988 Sep;44(3):837-845. [Medline: 3203132]
48. Hilden J. The area under the ROC curve and its competitors. Med Decis Making 1991;11(2):95-101. [doi: 10.1177/0272989X9101100204] [Medline: 1865785]
49. Brenner D, McLaughlin J, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. PLoS One 2011 Mar 31;6(3):e17479 [FREE Full text] [doi: 10.1371/journal.pone.0017479] [Medline: 21483846]
50. Zhang X, Jiang N, Wang L, Liu H, He R. Chronic obstructive pulmonary disease and risk of lung cancer: a meta-analysis of prospective cohort studies. Oncotarget 2017 Sep 29;8(44):78044-78056 [FREE Full text] [doi: 10.18632/oncotarget.20351] [Medline: 29100446]
51. Powell H, Iyen-Omofoman B, Baldwin D, Hubbard R, Tata LJ. Chronic obstructive pulmonary disease and risk of lung cancer: the importance of smoking and timing of diagnosis. J Thorac Oncol 2013 Jan;8(1):6-11 [FREE Full text] [doi: 10.1097/JTO.0b013e318274a7dc] [Medline: 23196277]
52. Watanabe S, Saeki K, Waseda Y, Murata A, Takato H, Ichikawa Y, et al. Lung cancer in connective tissue disease-associated interstitial lung disease: clinical features and impact on outcomes. J Thorac Dis 2018 Feb;10(2):799-807 [FREE Full text] [doi: 10.21037/jtd.2017.12.134] [Medline: 29607151]
53. Brenner D, Fanidi A, Grankvist K, Muller D, Brennan P, Manjer J, et al. Inflammatory Cytokines and Lung Cancer Risk in 3 Prospective Studies. Am J Epidemiol 2017 Dec 15;185(2):86-95. [doi: 10.1093/aje/kww159] [Medline: 27998891]
54. Shiels M, Pfeiffer R, Hildesheim A, Engels E, Kemp T, Park JH, et al. Circulating inflammation markers and prospective risk for lung cancer. J Natl Cancer Inst 2013 Dec 18;105(24):1871-1880 [FREE Full text] [doi: 10.1093/jnci/djt309] [Medline: 24249745]
55. Shiels M, Katki H, Hildesheim A, Pfeiffer R, Engels E, Williams M, et al. Circulating Inflammation Markers, Risk of Lung Cancer, and Utility for Risk Stratification. J Natl Cancer Inst 2015 Oct;107(10):1-9 [FREE Full text] [doi: 10.1093/jnci/djv199] [Medline: 26220734]
56. Seijo LM, Zulueta JJ. Understanding the Links Between Lung Cancer, COPD, and Emphysema: A Key to More Effective Treatment and Screening. Oncology (Williston Park) 2017 Dec 15;31(2):93-102 [FREE Full text] [Medline: 28205188]
57. Sprague B, Trentham-Dietz A, Klein B, Klein R, Cruickshanks K, Lee K, et al. Physical activity, white blood cell count, and lung cancer risk in a prospective cohort study. Cancer Epidemiol Biomarkers Prev 2008 Oct;17(10):2714-2722 [FREE Full text] [doi: 10.1158/1055-9965.EPI-08-0042] [Medline: 18843014]
58. Aref H, Refaat S. CRP evaluation in non-small cell lung cancer. Egyptian Journal of Chest Diseases and Tuberculosis 2014 Jul;63(3):717-722. [doi: 10.1016/j.ejcdt.2014.02.003]
59. Chaturvedi A, Caporaso N, Katki H, Wong H, Chatterjee N, Pine S, et al. C-reactive protein and risk of lung cancer. J Clin Oncol 2010 Jun 01;28(16):2719-2726 [FREE Full text] [doi: 10.1200/JCO.2009.27.0454] [Medline: 20421535]
60. RAND Corporation. Multiple Chronic Conditions in the United States URL: https://www.rand.org/pubs/tools/TL221.html [accessed 2018-12-20] [WebCite Cache ID 781cdOZUb]
61. Sambamoorthi U, Tan X, Deb A. Multiple chronic conditions and healthcare costs among adults. Expert Rev Pharmacoecon Outcomes Res 2015;15(5):823-832 [FREE Full text] [doi: 10.1586/14737167.2015.1091730] [Medline: 26400220]
62. Tu H, Wen CP, Tsai SP, Chow W, Wen C, Ye Y, et al. Cancer risk associated with chronic diseases and disease markers: prospective cohort study. BMJ 2018 Dec 31;360:k134 [FREE Full text] [doi: 10.1136/bmj.k134] [Medline: 29386192]
63. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med 2011 Dec;32(4):605-644 [FREE Full text] [doi: 10.1016/j.ccm.2011.09.001] [Medline: 22054876]
64. Gross AL, Gallo JJ, Eaton WW. Depression and cancer risk: 24 years of follow-up of the Baltimore Epidemiologic Catchment Area sample. Cancer Causes Control 2010 Feb;21(2):191-199 [FREE Full text] [doi: 10.1007/s10552-009-9449-1] [Medline: 19885645]
65. Jia Y, Li F, Liu YF, Zhao JP, Leng MM, Chen L. Depression and cancer risk: a systematic review and meta-analysis. Public Health 2017 Aug;149:138-148. [doi: 10.1016/j.puhe.2017.04.026] [Medline: 28641155]
66. Knekt P, Raitasalo R, Heliövaara M, Lehtinen V, Pukkala E, Teppo L, et al. Elevated lung cancer risk among persons with depressed mood. Am J Epidemiol 1996 Dec 15;144(12):1096-1103. [Medline: 8956621]
67. Capo-Ramos DE, Gao Y, Lubin JH, Check DP, Goldin LR, Pesatori AC, et al. Mood disorders and risk of lung cancer in the EAGLE case-control study and in the U.S. Veterans Affairs inpatient cohort. PLoS One 2012;7(8):e42945 [FREE Full text] [doi: 10.1371/journal.pone.0042945] [Medline: 22880133]
68. Liang J, Sun L, Muo C, Sung F, Chang S, Kao C. The analysis of depression and subsequent cancer risk in Taiwan. Cancer Epidemiol Biomarkers Prev 2011 Mar;20(3):473-475 [FREE Full text] [doi: 10.1158/1055-9965.EPI-10-1280] [Medline: 21297039]

69. American Lung Association. Behavioral Health & Tobacco Use URL: https://www.lung.org/stop-smoking/smoking-facts/behavioral-health-tobacco-use.html [accessed 2018-11-20] [WebCite Cache ID 780tnXqxv]
70. Mitra D, Shaw A, Tjepkema M, Peters P. Social determinants of lung cancer incidence in Canada: A 13-year prospective study. Health Rep 2015 Jun;26(6):12-20 [FREE Full text] [Medline: 26086335]
71. Hovanec J, Siemiatycki J, Conway DI, Olsson A, Stücker I, Guida F, et al. Lung cancer and socioeconomic status in a pooled analysis of case-control studies. PLoS One 2018;13(2):e0192999 [FREE Full text] [doi: 10.1371/journal.pone.0192999] [Medline: 29462211]
72. Ridker PM, MacFadyen JG, Thuren T, Everett BM, Libby P, Glynn RJ, CANTOS Trial Group. Effect of interleukin-1β inhibition with canakinumab on incident lung cancer in patients with atherosclerosis: exploratory results from a randomised, double-blind, placebo-controlled trial. Lancet 2017 Oct 21;390(10105):1833-1842. [doi: 10.1016/S0140-6736(17)32247-X] [Medline: 28855077]

Abbreviations

AUC: area under the curve
CKD: chronic kidney disease
COPD: chronic obstructive pulmonary disease
CVD: cardiovascular disease
EHR: electronic health record
EPIC: European Prospective Investigation into Cancer and Nutrition
HIE: health information exchange
HR: hazard ratio
HUNT: Helseundersøkelsen i Nord-Trøndelag
ICD-10-CM: International Classification of Diseases, 10th Revision, Clinical Modification
LDCT: low-dose computed tomography
OR: odds ratio
$\mathbf{PLCO_{m2012}}$: Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial
PPV: positive predictive value

Edited by A Moorhead; submitted 31.12.18; peer-reviewed by R Guan, T Basu, K Pradeep, I Mircheva; comments to author 23.01.19; revised version received 18.04.19; accepted 23.04.19; published 16.05.19.

Please cite as:
Wang X, Zhang Y, Hao S, Zheng L, Liao J, Ye C, Xia M, Wang O, Liu M, Weng CH, Duong SQ, Jin B, Alfreds ST, Stearns F, Kanov L, Sylvester KG, Widen E, McElhinney DB, Ling XB
Prediction of the 1-Year Risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the State of Maine
J Med Internet Res 2019;21(5):e13260
URL: http://www.jmir.org/2019/5/e13260/
doi:10.2196/13260
PMID:31099339

Original Paper

Developing the National Usability-Focused Health Information System Scale for Physicians: Validation Study

Hannele Hyppönen[1], PhD; Johanna Kaipio[2], PhD; Tarja Heponiemi[1], PhD; Tinja Lääveri[3,4], MD; Anna-Mari Aalto[1], PhD; Jukka Vänskä[5], MSocSci; Marko Elovainio[1], PhD

[1]National Institute for Health and Welfare, Helsinki, Finland
[2]Aalto University, Espoo, Finland
[3]Helsinki University Hospital, Helsinki, Finland
[4]University of Helsinki, Helsinki, Finland
[5]Finnish Medical Association, Helsinki, Finland

Corresponding Author:
Tarja Heponiemi, PhD
National Institute for Health and Welfare
PO Box 30
Helsinki, 00271
Finland
Phone: 358 295247434
Email: tarja.heponiemi@thl.fi

Abstract

Background: Problems in the usability of health information systems (HISs) are well acknowledged, but research still lacks a validated questionnaire for measuring and monitoring different dimensions of usability of HISs. Such questionnaires are needed not only for research but also for developing usability of HISs from the viewpoint of end-user experiences.

Objective: This study aimed to develop and test the validity of the questionnaire measuring the National Usability-Focused HIS-Scale (NuHISS) among a nationally representative sample of Finnish physicians.

Methods: We utilized 2 cross-sectional data collected from a random sample of Finnish physicians in 2014 (N=3781; of which 2340 [61.9%] were women) and 2017 (N=4018; of which 2604 [64.8%] were women). Exploratory and confirmatory factor analyses (structural equation modeling [SEM]) were applied to test the structural validity of the NuHISS. As the concurrent validity measure, we used the self-reported overall quality of the electronic health record system (*school grade*) provided by the participants using marginal structural models.

Results: The exploratory factor analyses with Varimax rotation suggested that the 7-factor solution did offer a good fit to the data in both samples (C^2=2136.14 in 2014 and C^2=2109.83 in 2017, both P<.001). Moreover, structural equation modelling analyses, using comparative fit index (CFI), Tucker-Lewis Index (TLI), Normed Fit Index (NFI), root mean squared error of approximation (RMSEA), and Standardized Root Mean square Residual (SRMR), showed that the 7-factor solution provided an acceptable fit in both samples (CFI=0.92/0.91, TLI=0.92/0.91, NFI=0.92/0.91, RMSEA=0.048/0.049, and SRMR=0.040/0.039). In addition, concurrent validity of this solution was shown to be acceptable. *Ease of use*, but also all other dimensions, was especially associated with overall quality reports independent of measured confounders. The 7-factor solution included dimensions of *technical quality*, *information quality*, *feedback*, *ease of use*, *benefits*, *internal collaboration*, and *cross-organizational collaboration*.

Conclusions: NuHISS provides a useful tool for measuring usability of HISs among physicians and offers a valid measure for monitoring the long-term development of HISs on a large scale. The relative importance of items needs to be assessed against national electronic health policy goals and complemented with items that have remained outside the NuHISS from the questionnaire when appropriate.

(J Med Internet Res 2019;21(5):e12875) doi:10.2196/12875

KEYWORDS

physicians; health information systems; questionnaire; validation studies

Introduction

Problems in usability of health information systems (HISs) are well acknowledged in research [1]. The vast investments in the adoption of HISs in the United States as well as in Europe have been driven by expectations reflecting key usability goals, particularly increased effectiveness and efficiency in health care [2-4]. The clinician community has, while considering electronic health record (EHR) systems as an improvement over the paper-based system, expressed frustration with the level of usability of available systems as well as their support for information exchange [1,5]. Moreover, several studies suggest that HISs may cause stress and frustration to clinicians and this appears to have increased recently [6,7]. If clinicians cannot achieve their goals with efficiency, effectiveness, and satisfaction by using the implemented information systems (ISs), they seek alternative solutions to reach their goals, that is, by using paper to document and transfer health information [8]. Declining to use is one important indication that the anticipated benefits are not being realized.

Research has shown that with high-quality management and high perceived usability, HISs could yield significant quality and productivity gains [4]. To show evidence of these gains arising from HISs use, appropriate measures need to be used. However, usability is qualitative and multidimensional by nature, thereby challenging the measuring [9] and thereby accumulation of results [10] on a larger scale. Depending on the definition used, the attributes and metrics of measurement have varied [11-14]. The widely known definition for usability is defined by the International Organization for Standardization (ISO) standard as follows: “usability refers to an extent to which a system, product or service can be used by specified users to achieve specific goals with effectiveness, efficiency and satisfaction in a specified context of use” [13]. The possible measures for each of these 3 aspects of usability are many. On the basis of a review of practices in measuring usability by Hornbæk [9], the challenges include, for example, distinguishing and empirically comparing subjective and objective measures of usability, studying long-term use and usability, extending measures of satisfaction beyond postuse questionnaire, validating and standardizing the host of subjective satisfaction questionnaires used, and studying correlations between usability measures as a means for validation. Alongside with usability, user experience (UX) as a concept has gained interest and is described as a person’s perceptions and responses resulting from the use or anticipated use of a product, system, or service [13]. According to the ISO standard [13], usability criteria can be used to assess aspects of UX as well.

Although governments have pushed the adoption of electronic Health (eHealth) systems and services, they have often lacked knowledge of longer term and larger scale eHealth usability and UX in health care contexts. Both the Organisation for Economic Co-operation and Development (OECD) and European Union (EU) have included information and communication technology (ICT) benchmarking as one of the issues on their policy agendas [15,16]; however, current HIS monitoring on a national level focuses on availability and usage rate of key functionalities of EHRs, personal health record (PHR) systems, health information exchange (HIE), and telemedicine [17-19]. Some HIS-related health care output or efficiency indicators (eg, number of visits saved, impact on length of stay, time saved by system use, and cost saved by system use) have also been defined [20] but rarely used in systematic, large-scale eHealth monitoring.

Validated usability questionnaires (eg, Software Usability Measurement Inventory (SUMI) [21], System Usability Scale (SUS) [22], and Questionnaire for User Interaction Satisfaction (QUIS) [23]) are context and domain independent and focus on evaluating the usability of user interfaces typically after usability testing. Our aim was to study and monitor longer-term experiences on usability and experienced outcomes of complex HISs from the viewpoint of physicians and their clinical tasks in specific contexts of clinical work. In addition, we aimed at addressing the development of the EHR systems. SUMI, SUS, or QUIS questions were therefore not as such regarded suitable as repeated measures of perceived usability of HISs on a national level.

The large-scale eHealth monitoring tools developed by the EU and OECD [17,24] include indicators for availability of key HIS functionalities and information contents, for example, documentation and retrieval of patient data, availability of patient summary data, and medication list. They do not cover end-users’ experiences on usability or experienced benefits of these functionalities or information contents.

The DeLone and McLean ISs success model that was developed in 1992 and revised in 2003 [25,26] is based on vast theoretical and empirical IS research with validation and updates. It has been used in over 30 scientific publications and, more importantly, also applied on a national level in the health care context in the Canada Health Infoway evaluation framework for HISs [20,27-29]. The original model and its application in the eHealth evaluation framework in Canada offer 6 interrelated dimensions on IS success: (1) System quality, (2) Information quality, (3) Service quality, (4) Use, (5) User satisfaction, and (6) Benefits. The framework does not list specific items or measures. The subareas are not specific to key functionalities and information contents needed in clinical work.

The Finnish eHealth strategy from 2015 listed a national-level eHealth usability survey as one strategic means to reach the strategy objectives [30]. In 2012, Nordic eHealth research network was established under the Nordic Council of Ministers to develop common Nordic eHealth indicators. The Finnish HIS usability-focused questionnaire for physicians was adopted in Iceland and Denmark in 2018. The OECD is also following work by the Nordic countries on common eHealth indicators for countries with advanced national ICT infrastructures [30].

The need to collect national-level evidence on usability and UXs of HISs in health care contexts to steer national eHealth policies and implementation coupled with the lack of suitable measures led to development of the Finnish *EHR systems as tools for physicians* questionnaire and within it, the National Usability-Focused HIS Scale (NuHISS) for physicians in 2009. The development team consisted of 2 physician-researchers, 1 usability researcher, and 2 researchers with experience in sociology of technology and national benchmarking studies. The first questionnaire for physicians with NuHISS was conducted in 2010, before implementation of the first national eHealth service—ePrescription. The second data collection round in 2014 was timed so that the national ePrescription service was in full use in the public sector and implementation of the national patient data repository (Kanta) was to begin. In 2017, the data collection was repeated for physicians for the third time and also extended to cover the nurses. The data collection was timed so that the Kanta system was in full use in Finland. The Ministry of Social Affairs and Health has agreed to fund the next data collection round in 2020, extending it further to social workers and asking for a plan beyond 2020.

Establishing the Finnish questionnaire containing the NuHISS as a regular source of evidence for steering national eHealth policies as well as increased international attention about the Finnish questionnaire has added pressure for validation of the key measures used for usability. Thus, the objective of this study was to develop a valid and reliable measure of NuHISS for physicians. The research questions were as follows:

1. What are the most important dimensions of usability and UXs of HISs that the NuHISS for physicians should measure?
2. Does the developed NuHISS show acceptable construct validity (do the selected survey items measure their corresponding dimension and can the dimensions be separated from each other) and concurrent validity (do the dimensions have associations to other factors, such as quality ratings, that they theoretically should) of the scale?

Methods

The Survey and Data Used to Validate the Usability-Focused HIS Scale

The Survey Population

This report is based on the data collected from a random sample of Finnish physicians at 2 data collection points (2 cross-sectional samples) in 2014 (N=18,257) and 2017 (N=18,326). The response rates were 21% in 2014 and 22% in 2017 providing analytic samples of 3781 in 2014 and 4018 in 2017 [5,31]. Owing to missing values in a part of the variables, we imputed both datasets (missing pattern of both the samples is reported in the Multimedia Appendix 1). We used the multiple imputation method with chained equations using Mice R-package. We generated 5 separate imputed datasets for analyses purposes. The study has been approved by the Finnish National Institute for Health and Welfare Ethics Board.

The Survey Questionnaire

Key EHR, HIE, and PHR functionalities and information needed by physicians in everyday clinical work (eg, documentation, viewing and retrieval of patient data, medication list, decision support, HIE, and patient-provided data) were used as a starting point of the questionnaire; see the OECD model survey *The OECD Guide for Measuring ICTs in the Health Sector* [17,18]. The questions were grounded in physicians' key HIS use tasks to increase face validity of the questions and to increase comparability with HIS availability monitoring (including the OECD model survey). The usability-focused questions (including aspects of stability, reaction speed, ease-of-use, recovery from errors, learnability, availability, information quality, and utility of information as well as effects) were modified based on the Finnish Medical Association's information technology (IT)-physicians' listing of key strengths and weaknesses of their HIS, as well as a comprehensive analysis of existing usability constructs [eg, 11-14,20-23,25,26]. The full questionnaire included 17 background questions, 41 usability-focused items, a list of 17 most urgent EHR development needs to select from, a list of 15 best functioning EHR features to select from, and a 9-item module measuring HIS-related well-being. For manager-level physicians, there was an additional 11-item module measuring HIS support for management.

The Usability-Focused HIS-Scale Items Used

Of the 41 usability-focused items, 32 items that were identical in the 2014 and 2017 questionnaires were selected for the NuHISS to measure physicians' experiences on usability and benefits of HISs. The scale items are depicted in Table 1.

Items Used for Concurrent Validity Evaluation

As the concurrent validity measure of the dimensions of our construct, we selected the self-reported overall quality of the EHR system (*school grade*) provided by the participants. The overall quality was rated by a continuous scale from 4=*very poor* to 10=*excellent*. The scale was dichotomized into low (7 or less) and high (more than 7) quality estimates. In 2014, 1095 of 3781 respondents (29%) and in 2017 altogether 1323 of 4018 respondents (33%) rated the system they used as a high-quality system.

In the concurrent validity analyses, gender, the year of birth, work tenure (in years), and whether the respondent was specialized (yes/no) were used as time independent covariates and overall system quality (measured by school grade given to the system) as the outcome variable (Table 2).

Table 1. Measured items in the National Usability-Focused HIS-Scale scale.

Short name	Questionnaire item (with 5-point Likert scale: 1=fully disagree, 5=fully agree)
Logic	The arrangement of fields and functions is logical on computer screen
Terminology	Terminology on the screen is clear and understandable (eg, titles and labels)
Documenting	Entering and documenting patient data is quick, easy, and smooth
Operating info	The systems keep me clearly informed about what it is doing (eg, saving data)
Straightforward tasks	Routine tasks can be performed in a straight forward manner without the need for extra steps using the system
Needed patient data	It is easy to obtain necessary patient information using the EHR[a] system
Nursing record	The information on the nursing record is in easily readable format
Stability	The systems are stable in terms of technical functionality (does not crash and no downtime)
System errors	Faulty system function has caused or has nearly caused a serious adverse event for the patient
Reaction speed	The system responds quickly to inputs
Unexpected actions	In my view, the system frequently behaves in unexpected or strange ways
Missing info	Information entered/documented occasionally disappears from the IS[b]
Medic list quality	The patient’s current medication list is presented in a clear format
Summary view	The EHR system generates a summary view (eg, on a timeline) that helps to develop an overall picture of the patients’ health status
Order completion	The system monitors and notifies when the orders given to nurses have been completed
Patient-provided info	Measurement results provided electronically by the patient (eg, via patient portal) help to improve the quality of care
Collaboration	EHR systems support cooperation and communication between physicians and patients
Suggestion implementation	The system supplier implements suggested corrections and amendments as wished
Vendor interest	The system supplier is interested in feedback from users
Implementation speed	Suggestions for corrections and amendments are implemented sufficiently quickly
Care quality	ISs help to improve quality of care
Care continuity	ISs help to ensure continuity of care
Guideline adherence	ISs support compliance and adherence with the treatment recommendations
Medication errors	ISs help in preventing errors and mistakes associated with medications
Duplicate tests	ISs help to avoid duplicate tests and examinations
Care needs and impacts	The EHR system provides me with information about the need for and effectiveness of treatment of my patients
HIE[c] medication	Information on medications ordered in other organizations is easily available
HIE speed	Obtaining patient information from another organization often takes too much time
HIE data quality	Patient data (also from other organizations) are comprehensive, up-to-date, and reliable
HIE collaboration	EHR systems support cooperation and communication between physicians working in different organizations
Professionals collaboration	EHR systems support cooperation and communication between physicians and nurses
Physician collaboration	EHR systems support cooperation and communication between physicians in your own organization

[a] EHR: electronic health record.

[b] IS: information system.

[c] HIE: health information exchange.

Table 2. Demographics, including means and SDs of overall system quality (school grade) covariates in the 2014 and 2017 data.

Variable	Year of data collection	
	2014	2017
Target population, n	16,350	17,210
Questionnaire sent (target population), %	91	93
Respondents	3781	4018
Gender		
Male, %	38.1	35.1
Female, %	61.9	64.8
Mean (SD)	1.62 (0.49)	1.65 (0.48)
Age (years), %		
<34	16.9	16.7
35-44	21.2	21.9
45-54	28.4	26.7
55>	33.6	34.7
Year of birth, mean	1966	1969
Specialization		
Not specialized (1), %	33.1	32.6
Specialized (2), %	66.9	67.4
Mean (SD)	1.33 (0.47)	1.33 (0.47)
Work tenure (experience, years)		
<10, %	25.7	25.1
>10<20, %	22.0	22.1
>20<30, %	27.1	27.0
30+, %	25.2	25.7
Mean (SD)	19.63 (11.42)	19.71 (11.48)
School grade given to the primary electronic health record system		
10 (excellent)	0.5	0.6
9	5.2	6.3
8	20.9	23.5
7	29.6	32.7
6	24.2	21.9
5	14.7	11.2
4 (fail)	4.9	3.8
Mean (SD)	6.64 (1.27)	6.82 (1.23)

Statistical Analyses Assessing Sample Differences

Means and SDs of the NuHISS items for the 2 time points (2014 and 2017) were calculated and Welch 2-sample *t* test was used to analyze the differences in the mean profile of the responses from the 2 time points. Following the research questions and established psychometric testing procedures we tested the validity of the NuHISS in 2 steps: (1) structural validity test (do the intended dimensions or latent variables explain the covariance of the corresponding items) showing whether the scale measures the defined dimensions and (2) concurrent validity test (are the dimensions associated with the factors they should be associated with).

Assessing Construct Validity of the NuHISS Scale Dimensions

The preliminary structural analyses were conducted by calculating bivariate correlations between the 32 study variables in the 2014 and then in the 2017 data. For grouping the NuHISS scale items into dimensions, we then tested the factor structure and number of dimensions using exploratory factor analyses with eigenvalue 1 and loading structure as a criterion for the appropriate number of factors. The factorial validity of the original scales was tested separately among the 2 samples (time points 2014 and 2017) with exploratory factor analyses (Varimax rotation; [32]) and then structural equation modeling (SEM; confirmatory factor analyses).

Second, we tested the structure using SEM [33] that offers more stringent testing and allowing items to be loaded only to their corresponding latent variables (factors). SEM is a multivariate statistical analysis technique. This technique is a combination of factor analysis and multiple regression analysis, and it is used to analyze the structural relationship between measured variables and latent constructs. Goodness-of-fit of the SEM models was evaluated based on the chi-square test (X^2), RMSEA, CFI, TLI, and Akaike's information criterion (AIC). A nonsignificant chi-square value indicates that the model is a good fit to the data. RMSEA values of less than .05 and .08 suggest a good and reasonable fit, respectively. For CFI and TLI, values above .90 and .95 represent an acceptable and a good fit, respectively. AIC is a measure used to compare any models that have the same set of variables. In such cases, the model with the smaller value of AIC will be preferred [34]. Testing the final structure was done in 3 steps. First, a 1-factor model was estimated where all remaining items were loaded on the same underlying dimension (null model). In the second step, a model representing the original theoretical model was estimated, and, in the final step, the structural invariance (test showing whether the structure can be considered as similar between 2 measurement points) was tested between 2 samples as strong/scalar invariance with factor loadings and intercepts constrained to be similar. Again the same fit indexes were used as in the overall SEM test.

As a final step for structural validation, we assessed internal consistency of each dimension using Cronbach alpha reliability coefficient. The closer Cronbach alpha coefficient is to 1.0, the greater the internal consistency of the items in the dimension.

Assessing Concurrent Validity of the Scale Dimensions

Concurrent validity is the extent to which one measurement is backed up by a related measurement obtained at about the same point in time. In testing, the validity of results obtained from one test (in this case *ease of use*) can often be assessed by comparison with a separate but related measurement (in this case, school grade given to the system) collected at the same point in time [35]. Concurrent validity was tested separately for each dimension of the EHR system with the overall quality (measured by school grade given to the system) evaluation as the criteria. The associations between overall quality evaluation and each dimension were tested using the marginal structural model (MSM) approach proposed by Robins et al [36,37], with inverse probability weights taking into account the effects of potential covariates. The approach produces a pseudopopulation with balance in all included covariates.

Statistical programming language R (version 3.5.1)/R-studio and multiple statistical R-packages (psych, psycho, missForest, mice, miceadds, ggplot2, resahpe2, lavaan, semPlot, semTools, ipw, sandwich, and survey) were used for the statistical analyses.

Results

Differences of Samples From the 2 Time Points

Means and SDs of the measured items in 2014 and 2017 are presented in Figure 1. There were some differences between time points especially in HIE medication (*information on medication prescribed in other organizations is easily accessible*), duplicate tests (*ISs help prevent overlapping examinations*), care needs and impacts (*IS provides me information on the need and efficiency of care*), and B2C collaboration (*ISs support collaboration between physicians and patients*). The mean profiles were, however, relatively similar: Overall correlations between samples were 0.97, suggesting that the means were very similar between time points. The Welch 2-Sample *t* test suggested that the difference of the overall means between samples (in 2014, mean 3.28; in 2017, mean 3.27; difference=0.016) was not significant ($t_{(7757.06)}$=1.14; 95% CI [–0.012 to 0.044]; *P*>.10).

Construct Validity of the NuHISS Scale

Item-to-Item Correlations

Correlation matrixes in both samples are presented in Figures 2 and 3. In 2014 data, items around dimension that we named *technical quality* show strong mutual correlations in both years, also correlating especially with most of the relatively strongly mutually correlating items in dimension that we named *ease of use*. *Ease of use* items also correlate with the dimension named *benefit*. Items in the dimension that we named *cross-organization collaboration* also cluster together, but with less significant correlations (lighter color) and correlations also with *internal collaboration*, *ease of use*, and *benefits* items. Items in the *feedback* dimension show strongest between-item correlation, represented by the darkest color, without strong external correlations. Items in the *info quality* dimension do not cluster together so clearly. *Benefit* dimension items correlate clearly with each other. The 2 *internal collaboration* items correlate with each other and also with some of *ease of use* and *cross-organization collaboration* items.

In the 2017 data, the clearest item-to-item correlations are within the *feedback* dimension items. In addition, *ease of use* items, *technical quality* items, *internal collaboration* items, and most of *benefits* dimension items show clear item-to-item correlations. *Cross-organizational collaboration* items show clearer within-item correlations in 2017 than in 2014 data.

Figure 1. Means and standard deviations of the measured items in 2014 and 2017. HIE: health information exchange.

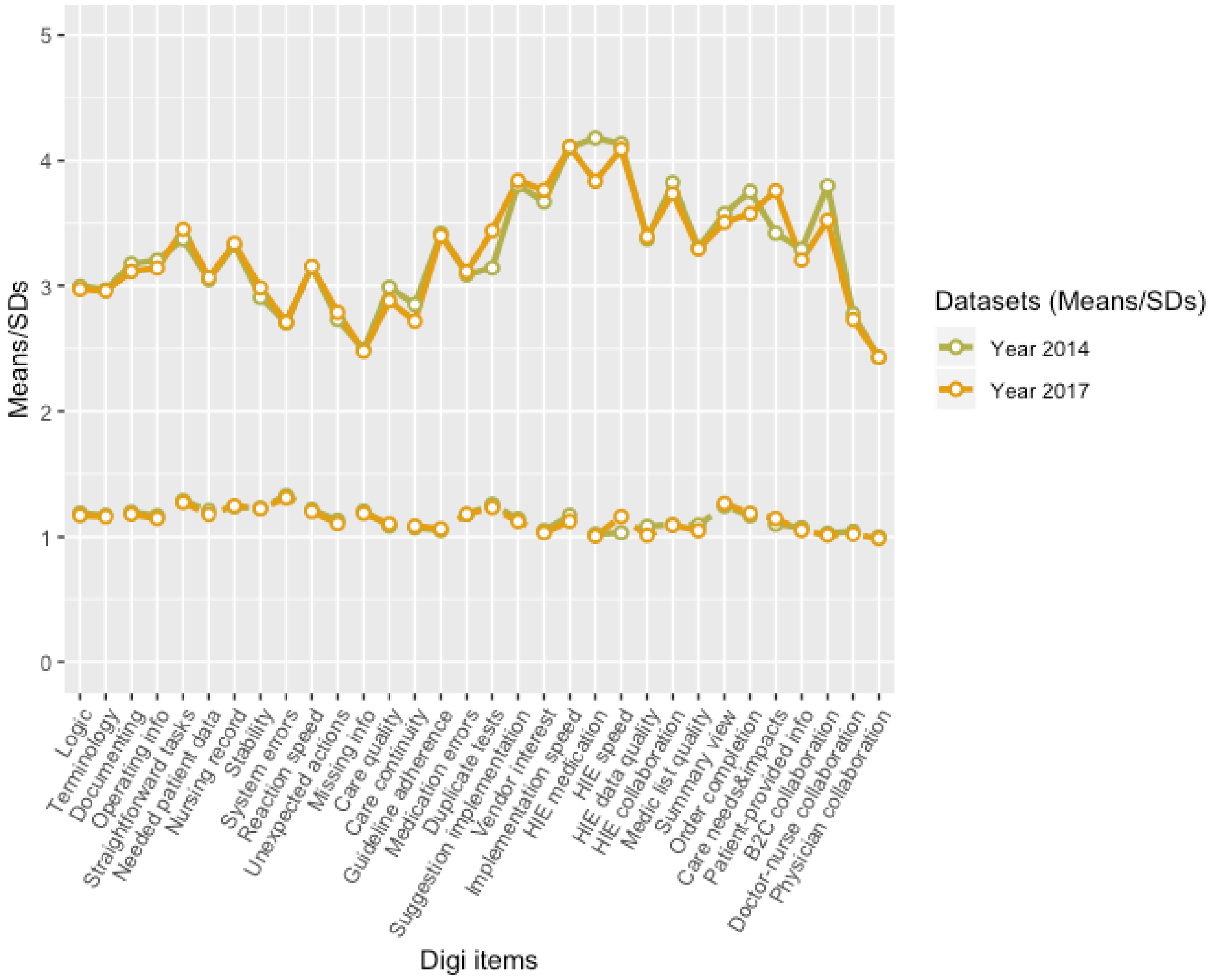


Figure 2. Correlation matrixes among items in 2014.

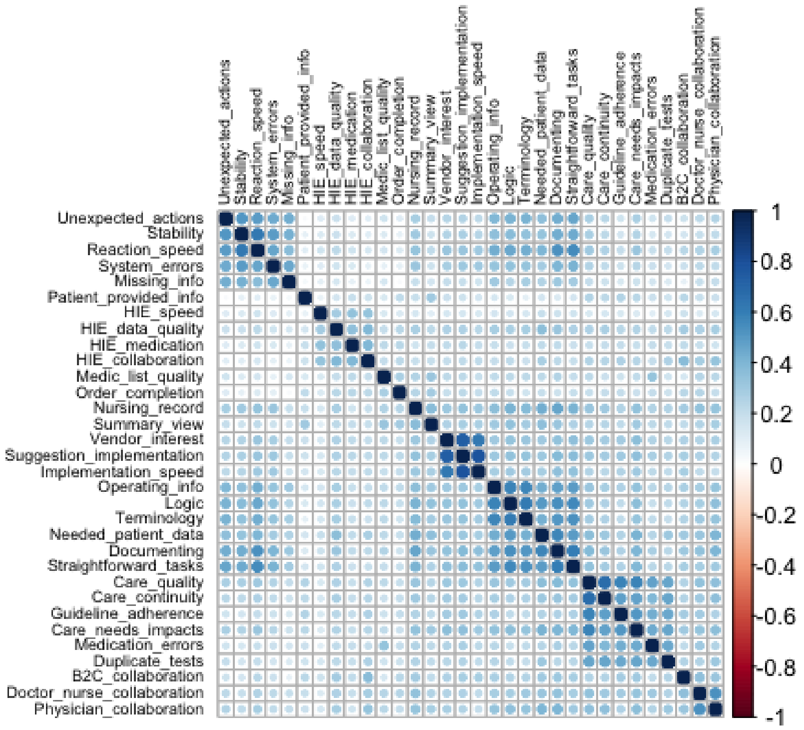
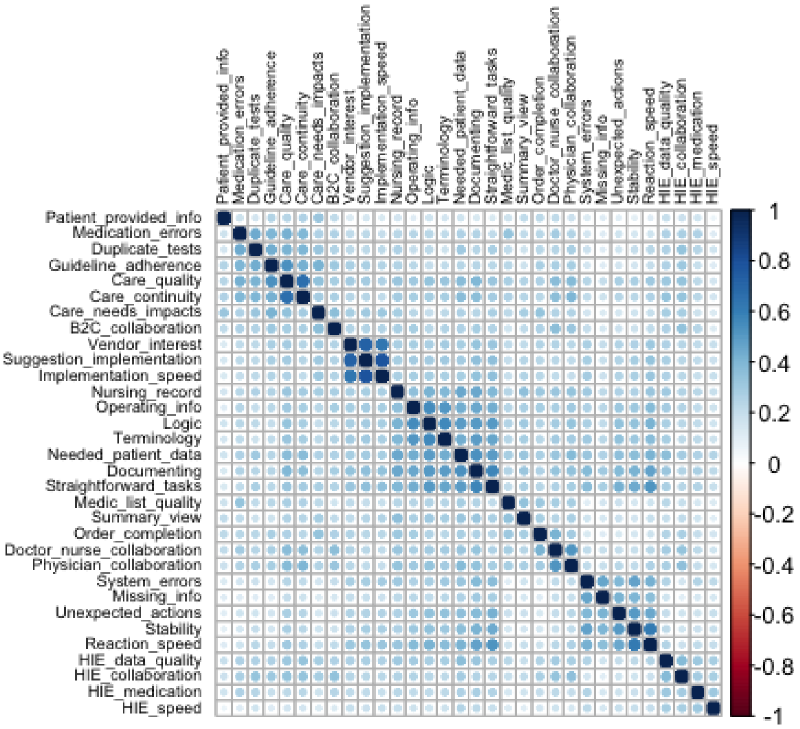


Figure 3. Correlation matrixes among items in 2017.



Scale Structure and Reliability Testing

The exploratory factor analyses with Varimax rotation suggested that the 7-factor solution did offer a good fit to the data in both samples (test of the hypothesis that 7 factors are sufficient, X^2_{293}=2136.14 in 2014 and X^2_{293}=2109.83 in 2017, both P<.001, see also Figures 4 and 5). However, in both samples the choice of 6 factor was supported by 2 (out of 9; 22.22%) methods (Optimal Coordinates and Parallel Analysis) and the choice of 8 factor was supported by 2 (out of 9; 22.22%) methods (Bayesian information criterion [BIC] and Sample Size Adjusted BIC).

In SEM analyses, the null model (without predicted structure) did not provide an acceptable fit to the data in 2014 (CFI=0.64, TLI=0.62, NFI=0.64, RMSEA=0.100, and SRMR=0.081) or in 2017 (CFI=0.66, TLI=0.64, NFI=0.66, RMSEA=0.094, and SRMR=0.074). The original 7-factor solution, however, provided an acceptable fit in both samples (CFI=0.92/0.91, TLI=0.92/0.91, NFI=0.92/0.91, RMSEA=0.048/0.049, and SRMR=0.040/0.039). The final solutions are presented in Figures 6 and 7.

Figure 4. Eigenvalues and explained variance by the number of factors (exploratory factor analyses) in 2014 sample.

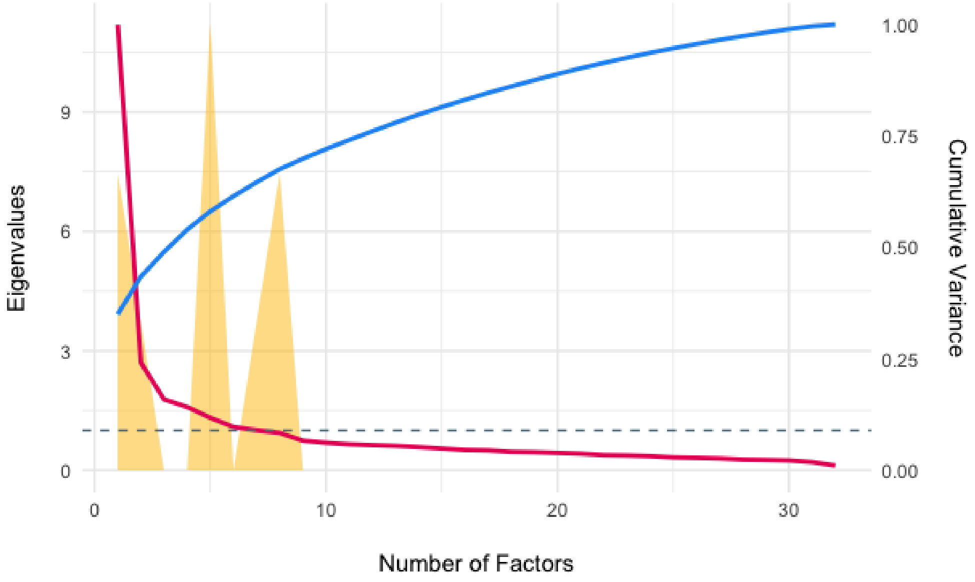


Figure 5. Eigenvalues and explained variance by the number of factors (exploratory factor analyses) in 2017 sample.

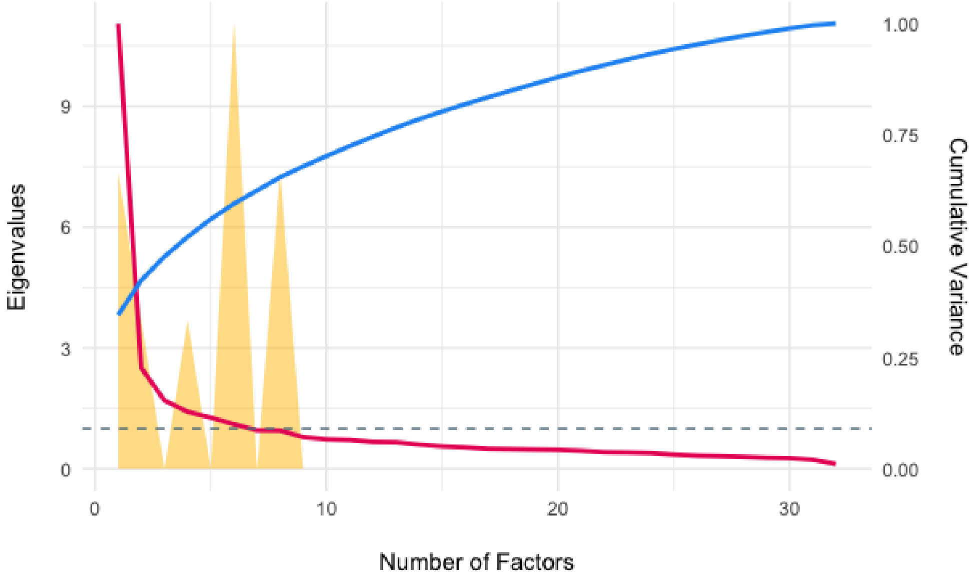


Figure 6. Confirmatory factor analysis in 2014 data.

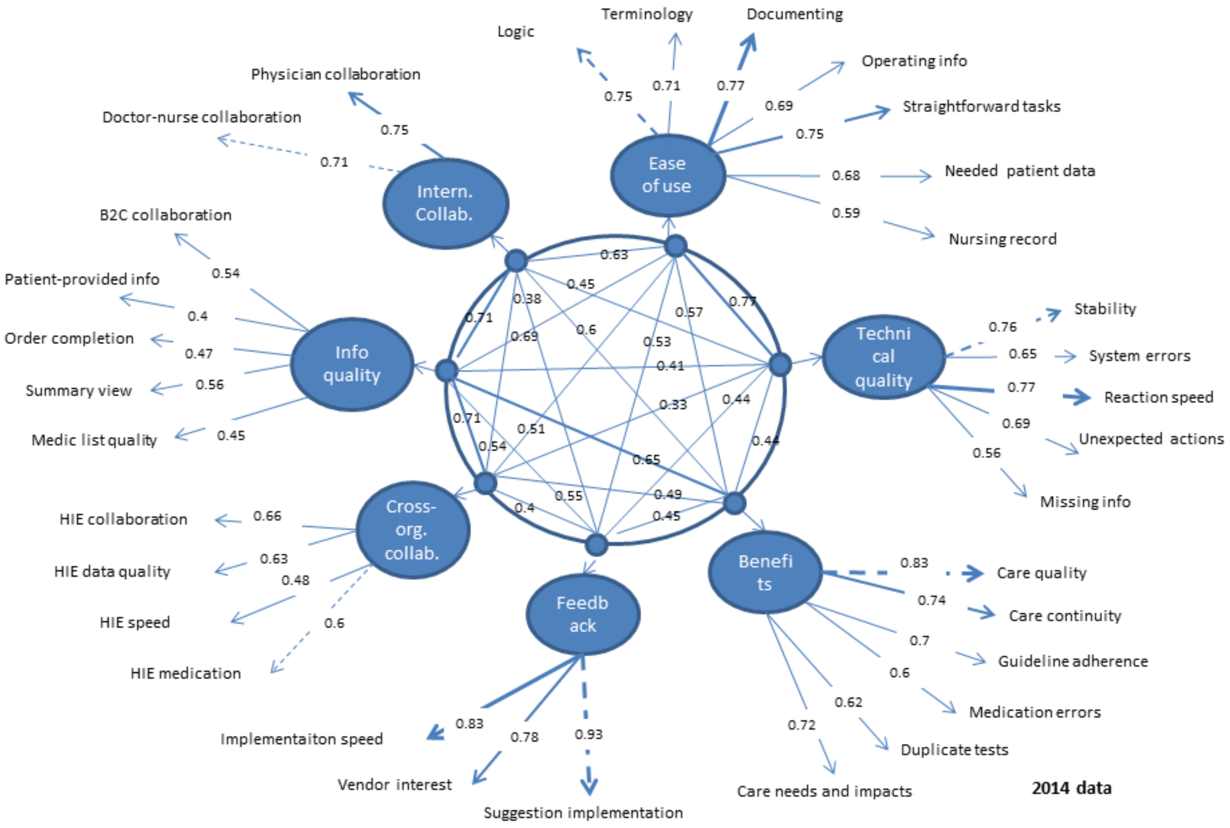
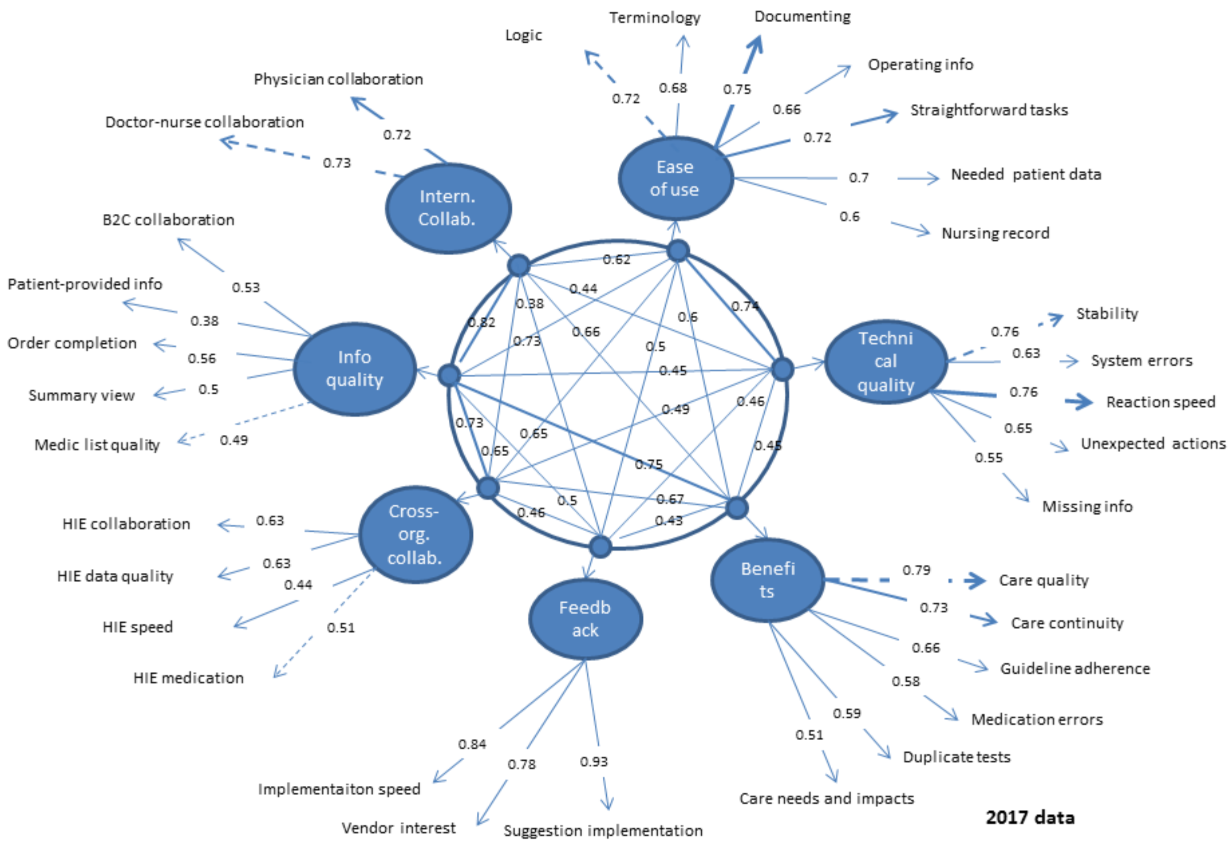


Figure 7. Confirmatory factor analysis in 2017 data.



The strongest 2 factors (measured by the loadings of items) were *feedback* and *internal collaboration*—in both years all the item loadings were over 0.7. All items in *ease of use*, *technical quality*, and *benefits*—factors had factor loadings over 0.5 both years. The weakest factor loadings in 2017 were in *information quality* factor (0.38 in 2014 for Patient-provided information item, 0.45/0.49 for Medication list quality item, and 0.47/0.56 for Order completion item in 2014/2017). In addition, *cross-organizational collaboration* factor loading for 1 item remained under 0.5 in 2017 (0.44 for HIE speed item). In addition, the internal reliability of the final factors (assessed with alpha coefficients) in the 2014/2017 data shows that the first 4 factors were the strongest:

1. Factor 1 *ease of use* (alpha=.87/.86).
2. Factor 2 *technical quality* (alpha=.82/.80).
3. Factor 3 *benefits* (alpha=.85/.81).
4. Factor 4 *feedback* (alpha=.88/.88).
5. Factor 5 *cross-organizational collaboration* (alpha=.69/.64).
6. Factor 6 *information quality* (alpha=.61/.62).
7. Factor 7 *internal collaboration* (alpha=.70/.69).

As the test of structural consistency, we performed the measurement invariance test that assesses the psychometric equivalence of a construct across groups or across time [38]. Results suggest that the final measure provided only minimal invariance between samples. The sequential tests are reported in Table 3. This was expected, because the fit of the final 7-factor solution was relatively modest.

Table 3. Changes of the fit indexes according to the sequential invariance tests.

Invariance tests steps	*Df*	Akaike's information criterion	Bayesian information criterion	Chi-square	Chi-square difference	*Df* difference	*P*-difference (>chi-square)
fit.configural	886	660,523	662,145	8613.6	—[a]	—	—
fit.loadings	911	660,626	662,075	8766.9	153.29	25	<2.2e-16[b]
fit.intercepts	936	661,458	662,733	9648.7	881.78	25	<2.2e-16[b]
fit.residuals	968	661,693	662,747	9947.7	299.00	32	<2.2e-16[b]
fit.means	975	661,848	662,853	10,116.5	168.86	7	<2.2e-16[b]

[a] Not applicable.

[b] *P*<.001.

Concurrent Validity of the Scale Items

In the weighted sample, the covariates had a balanced distribution between those who evaluated the overall quality of the system (*school grade*) low and high in both measurement years (Tables 4 and 5). The distribution suggests that the MSM fits relatively well and is effective in balancing covariates across *Overall system quality* (school grade) evaluation sequences (Table 6).

Table 4. Distribution of covariates between those who evaluated the overall quality of the electronic health record system (*school grade*) low and high in both measurement years from inverse probability weighting in 2014.

Variable	0 (school grade low; n=3780.96[a]), mean (SD)	1 (school grade high; n=3781.33[a]), mean (SD)	Standardized mean difference
Gender	1.62 (0.49)	1.62 (0.49)	0.007
Year of birth	1966.24 (10.93)	1966.22 (10.84)	0.002
Tenure	19.85 (11.46)	19.86 (11.33)	0.001
Specialized	1.33 (0.47)	1.33 (0.47)	0.003

[a] n refers to pseudopopulation samples from inverse probability weighting.

Table 5. Distribution of covariates between those who evaluated the overall quality of the electronic health record system (*school grade*) low and high in both measurement years from inverse probability weighting in 2017.

Variable	0 (school grade low); n=4018.24[a])	1 (school grade high; n=4017.07)	Standardized mean difference
Gender, mean (SD)	1.65 (0.48)	1.65 (0.48)	0.001
Year of birth, mean (SD)	1969.24 (11.16)	1969.21 (10.83)	0.003
Tenure, mean (SD)	19.74 (11.60)	19.77 (11.26)	0.003
Specialized, mean (SD)	1.33 (0.47)	1.32 (0.47)	0.002

[a] n refers to pseudopopulation samples from inverse probability weighting.

Table 6. Controlled direct effect of overall quality evaluation (*school grade*) estimated from the marginal structural model.

Factor	Odds ratio	95% CI
2014		
Ease-of-use	0.29	0.28 to 0.31
Technical quality	0.36	0.34 to 0.38
Benefits	0.49	0.46 to 0.52
Feedback	0.44	0.42 to 0.48
Cross-organizational collaboration	0.59	0.56 to 0.62
Information quality	0.58	0.55 to 0.60
Internal collaboration	0.49	0.46 to 0.52
2017		
Ease-of-use	0.34	0.32 to 0.36
Technical quality	0.39	0.37 to 0.41
Benefits	0.55	0.52 to 0.58
Feedback	0.48	0.45 to 0.51
Cross-organizational collaboration	0.60	0.57 to 0.63
Information quality	0.58	0.55 to 0.60
Internal collaboration	0.54	0.51 to 0.57

The Validated NuHISS Scale

The study resulted in a validated scale including 7 dimensions: the *technical quality* dimension measures reliability and safety aspects of the EHR system, including 5 items. The 5 items in the *information quality* dimension reflect availability and format of key information types of the EHR system. Information quality measures for the HIE functionality are included in the *cross-organizational collaboration* dimension. The *feedback* dimension measures responsiveness of the EHR system vendor to improvement suggestions including 3 items. The *ease of use* dimension consists of 7 items related to the key functionalities of the EHR system (including reading, documenting, and patient data retrieval). The *benefit* dimension covers 6 items measuring UX on overall benefits of HISs. The *internal collaboration* dimension with its 2 measures is actually a specific benefit that measures how well the EHR system supports cooperation and communication between professionals within their own organization. *Cross-organizational collaboration* dimension is another specific benefit dimension with 4 measures on systems support cooperation and communication among professionals in different organizations (focusing on HIE functionalities). Loadings of each of the measures show the strongest and weakest measures within each dimension (Table 7).

Table 7. The validated National Usability-Focused HIS-Scale with dimension reliability and item loadings.

Dimension (reliability 2014/2017) and short name	Item on the questionnaire (with 5-point Likert scale: 1=fully disagree, 5=fully agree)	Factor loading 2014/2017
Technical quality (alpha=.82/.80)		
Stability	The systems are stable in terms of technical functionality (does not crash, no downtime)	0.76/0.76
System errors	Faulty system function has caused or has nearly caused a serious adverse event for the patient	0.65/0.63
Reaction speed	The system responds quickly to inputs	0.77/0.76
Unexpected actions	In my view, the system frequently behaves in unexpected or strange ways	0.69/0.65
Missing info	Information entered/documented occasionally disappears from the IS	0.56/0.55
Information quality (alpha=.61/.62)		
Medic list quality	The patient's current medication list is presented in a clear format	0.45/0.49
Summary view	The EHR system generates a summary view (eg, on a timeline) that helps to develop an overall picture of the patient's health status	0.56/0.5
Order completion	The system monitors and notifies when the orders given to nurses have been completed	0.47/0.56
Patient-provided info	Measurement results provided electronically by the patient (eg, via patient portal) help to improve the quality of care	0.40/0.38
B2C collaboration	EHR systems support co-operation and communication between physicians and patients	0.54/0.53
Feedback (alpha=.88/.88)		
Suggestion implementation	The system supplier implements suggested corrections and amendments as wished	0.93/0.93
Vendor interest	The system supplier is interested in feedback from users	0.78/0.78
Implementation speed	Suggestions for corrections and amendments are implemented sufficiently quickly	0.83/0.84
Ease of use (alpha=.87/.86)		
Logic	The arrangement of fields and functions is logical on computer screen	0.75/0.72
Terminology	Terminology on the screen is clear and understandable (eg, titles and labels)	0.71/0.68
Documenting	Entering and documenting patient data is quick, easy and smooth	0.77/0.75
Operating info	The systems keep me clearly informed about what it is doing (eg, saving data)	0.69/0.66
Straightforward tasks	Routine tasks can be performed in a straight forward manner without the need for extra steps using the system	0.75/0.72
Needed patient data	It is easy to obtain necessary patient information using the EHR system	0.68/0.7
Nursing record	The information on the nursing record is in easily readable format	0.59/0.6
Benefits (alpha=.85/.81)		
Care quality	ISs help to improve quality of care	0.83/0.79
Care continuity	ISs help to ensure continuity of care	0.74/0.73
Guideline adherence	ISs support compliance and adherence with the treatment recommendations	0.7/0.66
Medication errors	ISs help in preventing errors and mistakes associated with medications	0.6/0.58
Duplicate tests	ISs help to avoid duplicate tests and examinations	0.62/0.59
Care needs and impacts	The EHR system provides me with information about the need for and effectiveness of treatment of my patients	0.72/0.51
Cross-organizational collaboration (alpha= .69/.64)		
HIE medication	Information on medications ordered in other organizations is easily available	0.6/0.51
HIE speed	Obtaining patient information from another organization often takes too much time	0.48/0.44
HIE data quality	Patient data (also from other organizations) are comprehensive, up-to-date and reliable	0.63/0.63

Dimension (reliability 2014/2017) and short name	Item on the questionnaire (with 5-point Likert scale: 1=fully disagree, 5=fully agree)	Factor loading 2014/2017
HIE collaboration	EHR systems support co-operation and communication between physicians working in different organizations	0.66/0.63
Internal collaboration (alpha=.70/.69)		
Professionals collaboration	EHR systems support co-operation and communication Between physicians and nurses	0.71/0.73
Physician collaboration	EHR systems support co-operation and communication Between physicians in your own organization	0.75/0.72

Discussion

The main aim of this study was to develop and test the validity of the questionnaire measuring NuHISS among a nationally representative sample of Finnish physicians. The exploratory factor analyses showed that the 7-factor solution did offer a good fit to the data and SEM analyses showed that it provided an acceptable fit. Moreover, concurrent validity of this solution was shown to be acceptable. The 7-factor solution included *technical quality*, *information quality*, *feedback*, *ease of use*, *benefits*, *internal collaboration*, and *cross-organizational collaboration* dimensions. Our results show that NuHISS provides a useful tool for measuring the usability of HISs among physicians and offers a valid measure for monitoring the long-term development of HISs on a large scale.

Differences of Samples

Overall correlations between the 2 samples suggested that the means were very similar between time points. The Welch 2-Sample *t* test verified that the difference was not significant. This generated a good basis for validation of the scale.

Validity and Reliability of the Scale

Construct validity—the degree to which a test measures what it claims to be measuring—was supported by correlation analysis, which revealed the items clustering together, although correlations for some clusters were stronger and for some others weaker and more dispersed. The underlying components were analyzed using factor analysis—principal components analysis (PCA) for data collected with the same scale in 2 time points. PCA led to a 32-item 7-component solution with 65% of the total variance explained in 2014 and 63% in 2017. In confirmatory factor analyses, the original 7-factor solution provided an acceptable, although modest, fit in both samples. Component or factor loadings revealed that the highest loadings were for all items in the *feedback* factor and many items in the *technical quality*, *ease of use*, *benefits*, and *internal collaboration* factors. The *information quality* factor had overall the lowest loadings.

Internal consistency of factors was assessed by reliability (Cronbach alpha). Reliability of all factors exceeded .60, with the highest alphas in *feedback* (over .80), *ease of use*, and *benefit* factors. The *information quality* factor was the weakest in both years (.61/.62). *Cross-organizational collaboration* reliability was also below .70 in both years.

Concurrent validity measures the extent to which results correspond to those of a previously established measurement of the construct. The covariates had a balanced distribution between those who evaluated the overall quality of the system (school grade) low and high in both measurement years, suggesting that the MSM fits relatively well and is effective in balancing covariates across quality evaluation sequences.

Comparison of NuHISS and IS Success Model

The final 7-factor solution was compared with IS success model dimensions and items (Canada Infoway 2012 version, Multimedia Appendix 2). NuHISS covered all but one of the IS success model dimensions (Use) and most of the key subareas. The IS success model dimension of *Use* was out of the scope of our survey (Availability and Usage rate are assessed in Finland with a separate survey to health care providers). In addition, the items were grouped in a slightly different manner. Multimedia Appendix 2 compares the 2 scales, where black font shows the actual items for each factor or dimension and gray font location of the item, if different in the scales.

The comparison shows that overall, the NuHISS-scale items are more grounded to the physician work and more detailed than the IS success model items. The NuHISS scale has a more focused *technical quality* factor than the IS success model *System Quality* dimension. In addition to similar items, the latter includes items that fell into NuHISS *ease of use* and *benefits* factors. NuHISS has a dedicated item for *ease of use*, whereas in the IS success model, corresponding items are listed under *system quality* and *information quality* dimensions. NuHISS *benefits* factor and *collaboration* factors correspond to IS success model *user satisfaction* and *information quality* dimensions. NuHISS has specific factors for *cross-organizational* and *internal collaboration*, unlike the IS success model. NuHISS does not include a Service quality factor, although the full questionnaire contains these items.

On the basis of this comparison, *service quality* factor is a potential new factor to be added in the scale in future. In addition, *information quality* factor has the clearest improvement needs based on the weakest reliability and factor loadings.

Strengths and Limitations

The face validity of our questionnaire is good: experts (physicians using the different HISs in their everyday work) have participated in generating the questionnaire and questions were grounded to the core IT functionalities and information contents needed by physicians. The physicians have evaluated the importance of questions from the viewpoint of actual end-users and long-term use of the systems in clinical environments. The surveys have also been pilot tested on both occasions (before sending out the 2014 and 2017 surveys) on a subset of the intended population. The scale offers a valid tool for measuring usability of HISs among professionals and adds substantially to previous scales such as those focusing on measuring usability, acceptability, and user satisfaction of mobile apps among clients [39-41].

However, the questionnaire may have shortcomings typical to questionnaire-based surveys: it may include questions that are understood differently by different respondents, depending on the experience they have had with different systems and other individual factors. Although trying to use the language familiar to practicing physicians, we have also been obliged to use terminology from user interface and interaction design fields, such as *label*, *input*, or *screen*. In addition, we used the term *EHR* in instances where questions were particularly focused on the respondents' main EHR system, and *ISs* in instances where questions focused not only on EHRs but also related specialty-specific regional or national *ISs* or patient-provided data (either via the EHR or via standalone systems). The wording aimed at relieving respondents from knowing the particulars of system integration, but may have also confused some respondents. As the clinical ICT system environments often include several systems, which are used simultaneously, it is not always clear for the physicians which functionalities and features are related to EHR systems and which to other IT systems.

The total number of respondents in both surveys was about 4000. Response rates remained relatively low (21% to 22%). A small proportion of this is due to the sampling method: We targeted the survey to all physicians in clinical work. The Finnish Medical association's membership register did not allow us to select only physicians in clinical work into the target population. Therefore, the questionnaires were sent to all working aged physicians (ie, to a larger target population, with a cover letter calling for responses from physicians in clinical work). Calculation of the response rate was carried out from the sent questionnaires, which included physicians not in clinical work. Comparison of the respondents with the target population in 2014 [5] and 2017 [31] showed good representativeness of samples in both years.

The questionnaire length is another possible reason for the relatively low response rate. Although the core *NuHISS* for physicians had only 32 items, the full questionnaire had 17 background questions, 41 usability-focused items, a list of 17 most urgent EHR development needs to select from, a list of 15 best functioning EHR features to select from, and a 9-item module measuring HIS-related well-being. For manager-level physicians, there was an additional 11-item module measuring HIS support for management.

All the questionnaire items related to IS success were not included in the validated scale. Our questions on system integration and proficiency of use were not included because of a different scale. In the exploratory factor analysis (Multimedia Appendix 3), our items measuring learnability (*Learning to use the EHR system does not require a lot of training*), recovery from errors (*It is easy to correct mistakes [such as entry errors, ending up in the wrong screen, changing incorrect selections, etc]*), and decision support quality (*The reminders, alerts and warnings provided by the system are useful and are adequate*) loaded on the *ease of use* factor but did not qualify there in the validation. In the exploratory factor analysis, our items *Diagnostic imaging results are easily available on a regional level* and *Laboratory results are easily available and are logically presented on a regional level* loaded on *cross-organizational collaboration* factor but did not qualify for the validated scale. In the exploratory factor analysis, our questions *If I have problems with the system I can easily get help* and *Use of EHR systems frequently takes my attention away from the patient* loaded on *technical quality* factor but did not qualify into the validated scale. In the exploratory factor analysis, our item *I know to whom and how I can send feedback on the system, if I so wish* loaded on the *feedback* factor but did not qualify into the validated scale. All these questions reflect aspects of usability that may still be important to keep in the full questionnaire.

Although the functionalities and information needed by doctors in daily patient care remain relatively constant, development of technologies may enhance some functionalities (ie, artificial intelligence-assisted decision support and improved personalization possibilities). These developments should also be considered in the full questionnaire for physicians in future. In addition, further development of some of the measures in the *information quality* factor may be called for. In the exploratory factor analysis, there was an eighth factor, which we named *Business to Client collaboration*. This did not qualify as an individual factor in the validation, but combining the items with the *information quality* factor provided a satisfactory result. The eighth factor may become valid, if some items are added or rephrased.

Conclusions

To our knowledge, Finland is the first country to have administered regular national monitoring of usability of HISs from the viewpoint of end-user experiences. The introduced tool—NuHISS for physicians—offers a valid measure for monitoring the long-term development of HISs on a large scale. The scale is highly adoptable in other countries—it has already been used in Iceland, Denmark, and Germany in 2018. The relative importance of items needs to be assessed against national eHealth policy goals and complemented with items that have remained outside the NuHISS from the questionnaire when appropriate. Development of HIS functionalities calls for further development of the scale, especially within the *information quality* domain. Similar national-level scales have been developed in Finland for nurses and social workers, based on the physician scale. First data collections were conducted from 2017 to 2018. Validation of these in due course will show generalizability of the physician scale across professional groups.

Acknowledgments

This study was funded by the Finnish Work Environment Fund (project 116104), the Strategic Research Council at the Academy of Finland (project 303607), and the Ministry of Social Affairs and Health (project 112241).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Number of missing values in each item and missing patterns.

[PDF File (Adobe PDF File), 456KB - jmir_v21i5e12875_app1.pdf]

Multimedia Appendix 2

Comparison of National Usability-Focused HIS-Scale and Information System Success model dimensions and items.

[PDF File (Adobe PDF File), 34KB - jmir_v21i5e12875_app2.pdf]

Multimedia Appendix 3

Exploratory factor analysis factor matrix with factor loadings.

[PDF File (Adobe PDF File), 64KB - jmir_v21i5e12875_app3.pdf]

References

1. Roman LC, Ancker J, Johnson S, Senathirajah Y. Navigation in the electronic health record: a review of the safety and usability literature. J Biomed Inform 2017 Dec;67:69-79 [FREE Full text] [doi: 10.1016/j.jbi.2017.01.005] [Medline: 28088527]
2. European Commission. 2012. eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century URL: https://ec.europa.eu/digital-single-market/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century [accessed 2019-03-08] [WebCite Cache ID 76iP86cl0]
3. Sheikh A, Sood HS, Bates D. Leveraging health information technology to achieve the "triple aim" of healthcare reform. J Am Med Inform Assoc 2015 Jul;22(4):849-856 [FREE Full text] [doi: 10.1093/jamia/ocv022] [Medline: 25882032]
4. Jones SS, Heaton P, Rudin R, Schneider E. Unraveling the IT productivity paradox--lessons for health care. N Engl J Med 2012 Jun 14;366(24):2243-2245. [doi: 10.1056/NEJMp1204980] [Medline: 22693996]
5. Kaipio J, Lääveri T, Hyppönen H, Vainiomäki S, Reponen J, Kushniruk A, et al. Usability problems do not heal by themselves: national survey on physicians' experiences with EHRs in Finland. Int J Med Inform 2017 Dec;97:266-281 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.010] [Medline: 27919385]
6. Heponiemi T, Hyppönen H, Vehko T, Kujala S, Aalto AM, Vänskä J, et al. Finnish physicians' stress related to information systems keeps increasing: a longitudinal three-wave survey study. BMC Med Inform Decis Mak 2017 Oct 17;17(1):147 [FREE Full text] [doi: 10.1186/s12911-017-0545-y] [Medline: 29041971]
7. Heponiemi T, Hyppönen H, Kujala S, Aalto AM, Vehko T, Vänskä J, et al. Predictors of physicians' stress related to information systems: a nine-year follow-up survey study. BMC Health Serv Res 2018 Dec 13;18(1):284. [doi: 10.1186/s12913-018-3094-x] [Medline: 29653530]
8. Hyppönen H, Lumme S, Reponen J, Vänskä J, Kaipio J, Heponiemi T, et al. Health information exchange in Finland: Usage of different access types and predictors of paper use. Int J Med Inform 2019 Feb;122:1-6. [doi: 10.1016/j.ijmedinf.2018.11.005] [Medline: 30623778]
9. Hornbaek K. Current practice in measuring usability: challenges to usability studies and research. Int J Hum Comp Stud 2006;64:102. [doi: 10.1016/j.ijhcs.2005.06.002]
10. Scott RE, Saeed A. ResearchGate. 2008. Global eHealth - Measuring Outcomes: Why, What, and How: A Report Commissioned by the World Health Organization's Global Observatory for eHealth URL: https://www.researchgate.net/publication/237279666_Global_eHealth_-_Measuring_Outcomes_Why_What_and_How_A_Report_Commissioned_by_the_World_Health_Organization's_Global_Observatory_for_eHealth [accessed 2019-03-12] [WebCite Cache ID 76pj8AVbG]
11. Nielsen J. Usability Engineering. Boston: Academic Press; 1993.
12. Kurosu M. Revised: What Have We Learnt About Usability Since 1998? Human-Computer Interaction: Design and Evaluation Cham. London: Academic Press; 2015.
13. International Organization for Standardization. 2015. ISO 9241-210:2010: Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems URL: https://www.iso.org/standard/52075.html [accessed 2019-03-08] [WebCite Cache ID 76iRTHG6L]

14. Bevan N. Usability is quality of use. Adv Hum Fact Ergon 1995;20:349-354. [doi: 10.1016/S0921-2647(06)80241-8]
15. Organisation for Economic Co-operation and Development. 2015. OECD Recommendation on Health Data Governance URL: http://www.oecd.org/els/health-systems/health-data-governance.htm [accessed 2019-03-08] [WebCite Cache ID 76iS4HXdH]
16. European Commission. 2017. Multi annual work programme (MWP) 2018-2021 URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20170509_co01_en.pdf [accessed 2019-03-08] [WebCite Cache ID 76iSGlPYO]
17. Organisation for Economic Co-operation and Development. 2015. Draft OECD Guide for Measuring ICTs in the Health Sector URL: http://www.oecd.org/health/health-systems/Draft-oecd-guide-to-measuring-icts-in-the-health-sector.pdf [accessed 2019-03-08] [WebCite Cache ID 76iSSUVpd]
18. Adler-Milstein J, Ronchi E, Cohen G, Winn LA, Jha AK. Benchmarking health IT among OECD countries: better data for better policy. J Am Med Inform Assoc 2014;21(1):111-116 [FREE Full text] [doi: 10.1136/amiajnl-2013-001710] [Medline: 23721983]
19. European Commission. 2014. European Hospital Survey: Benchmarking Deployment of e-Health Services (2012-2013) URL: https://ec.europa.eu/digital-single-market/en/news/european-hospital-survey-benchmarking-deployment-ehealth-services-2012-2013 [accessed 2019-03-13] [WebCite Cache ID 76ppEQiyD]
20. Canada Health Infoway. 2006. Canada Health Infoway Benefits Evaluation Indicators Technical Report URL: https://www.infoway-inforoute.ca/en/ [accessed 2019-03-13] [WebCite Cache ID 76ppkMaqi]
21. Kirakowski J. SUMI Questionnaire. 2019. The use of questionnaire methods for usability assessment, Background Notes on the SUMI Questionnaire 1994 URL: http://sumi.uxp.ie/about/sumipapp.html
22. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan P, Thomas T, Weerdmeester A, editors. Usability Evaluation in Industry. London: Taylor and Francis; 1996.
23. Chin J, Diehl V, Norman K. Development of an instrument measuring user satisfaction of the human–computer interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1988 Presented at: CHI'88; May 15-19, 1988; Washington, DC, USA p. 213-218.
24. Codagnone C, Lupiañez-Villanueva F. European Commission. 2014. Benchmarking Deployment of eHealth among General Practitioners (2013) URL: https://ec.europa.eu/digital-single-market/en/news/benchmarking-deployment-ehealth-among-general-practitioners-2013-smart-20110033 [accessed 2019-03-13] [WebCite Cache ID 76pqWX0o7]
25. DeLone W, McLean E. Information Systems Success: The Quest for the Dependent Variable. Inf Syst Res 1992;3(1):60-95 [FREE Full text] [doi: 10.1287/isre.3.1.60]
26. DeLone WH, McLean ER. The DeLone and McLean Model of Information Systems Success: a ten-year update. J Man Inf Syst 2003;19:9-30. [doi: 10.1080/07421222.2003.11045748]
27. Decima H. Canada Health Infoway. 2012. Canadian Views on Electronic Health Records URL: https://www.infoway-inforoute.ca/en/component/edocman/1687-harris-decima-survey-on-electronic-health-records/view-document?Itemid=0 [accessed 2019-03-11] [WebCite Cache ID 76mxVSugw]
28. Canada Health Infoway. 2012. Canada Health Infoway Benefits Evaluation Indicators Technical Report Version 2.0 URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/450-benefits-evaluation-indicators-technical-report-version-2-0 [accessed 2019-03-13] [WebCite Cache ID 76prFDNej]
29. Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. Healthc Q 2007;10(1):112-6-112-8. [Medline: 17326376]
30. Ministry of Social Affairs and Health. 2016. Information to Support Well-Being And Service Renewal URL: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/74459/URN_ISBN_978-952-00-3575-4.pdf?sequence=1&isAllowed=y
31. Saastamoinen P, Vänskä J, Kaipio J, Hyppönen H, Reponen J, Lääveri T. [Doctors' estimates of patient information systems improved slightly]. Finn Med J 2018;34:1814-1819.
32. Maharee-Lawler S, Rodwell J, Noblet A. A step toward a common measure of organizational justice. Psychol Rep 2010 Apr;106(2):407-418. [doi: 10.2466/PR0.106.2.407-418] [Medline: 20524541]
33. Jöreskog KG. Testing structural equation models. In: Bollen KA, Long JS, editors. Testing Structural Equation Models. Thousand Oaks, CA: Sage Publications; 1993:294-316.
34. Kelloway EK. Using LISREL for Structural Equation Modeling: A Researcher's Guide. Thousand Oaks, CA: Sage Publications; 1998.
35. Nunnally JC. McGraw-Hill series in psychology. Psychometric theory. New York: McGraw-Hill; 1967.
36. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000 Sep;11(5):550-560.
37. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. Epidemiology 2000 Jan;20(1):18-26.
38. Putnick DL, Bornstein M. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Devel Rev 2016;41:71-90.

39. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. The Service User Technology Acceptability Questionnaire: psychometric evaluation of the Norwegian Version. JMIR Hum Factors 2018 Dec 21;5(4):e10255 [FREE Full text] [doi: 10.2196/10255] [Medline: 30578191]
40. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
41. Schnall R, Cho H, Liu J. Health Information Technology Usability Evaluation Scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 5;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]

Abbreviations

AIC: Akaike’s information criterion
BIC: Bayesian information criterion
CFI: comparative fit index
EHR: electronic health record
eHealth: electronic health
EU: European Union
HIE: health information exchange
HIS: health information system
ICT: information and communication technology
IS: information system
ISO: International Organization for Standardization
IT: information technology
MSM: marginal structural model
NFI: Normed Fit Index
NuHISS: National Usability-Focused HIS-Scale
OECD: Organisation for Economic Co-operation and Development
PCA: principal components analysis
PHR: personal health record
QUIS: Questionnaire for User Interaction Satisfaction
RMSEA: root mean squared error of approximation
SEM: structural equation modeling
SRMR: Standardized Root Mean square Residual
SUMI: Software Usability Measurement Inventory
SUS: System Usability Scale
TLI: Tucker-Lewis Index
UX: user experience

Edited by G Eysenbach; submitted 27.11.18; peer-reviewed by L Roa, I Mircheva; comments to author 30.01.19; accepted 18.02.19; published 16.05.19.

Please cite as:
Hyppönen H, Kaipio J, Heponiemi T, Lääveri T, Aalto AM, Vänskä J, Elovainio M
Developing the National Usability-Focused Health Information System Scale for Physicians: Validation Study
J Med Internet Res 2019;21(5):e12875
URL: https://www.jmir.org/2019/5/e12875/
doi:10.2196/12875
PMID:31099336

Original Paper

Evaluation of Different Recruitment Methods: Longitudinal, Web-Based, Pan-European Physical Activity Through Sustainable Transport Approaches (PASTA) Project

Mailin Gaupp-Berghausen[1], Dipl Ing; Elisabeth Raser[1], Dipl Ing; Esther Anaya-Boig[2], MSc; Ione Avila-Palencia[3,4,5], PhD; Audrey de Nazelle[2], PhD; Evi Dons[6,7], PhD; Helen Franzen[8], BA (Hons); Regine Gerike[9], PhD; Thomas Götschi[10], PhD; Francesco Iacorossi[11]; Reinhard Hössinger[1], PhD; Mark Nieuwenhuijsen[3,4,5], PhD; David Rojas-Rueda[3,4,5], PhD; Julian Sanchez[12,13], MSc; Emilia Smeds[14], MSc; Manja Deforth[10], MSc; Arnout Standaert[7], PhD; Erik Stigell[15], PhD; Tom Cole-Hunter[3,16,17], PhD; Luc Int Panis[7,18], PhD

[1]Institute for Transport Studies, University of Natural Resources and Life Sciences, Vienna, Vienna, Austria
[2]Centre for Environmental Policy, Imperial College London, London, United Kingdom
[3]ISGlobal, Barcelona, Spain
[4]Universitat Pompeu Fabra, Barcelona, Spain
[5]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Madrid, Spain
[6]Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium
[7]Flemish Institute for Technological Research (VITO), Mol, Belgium
[8]ICLEI, European Secretariat GmbH, Freiburg, Germany
[9]Chair of Integrated Transport Planning and Traffic Engineering, Technische Universität Dresden, Dresden, Germany
[10]Physical Activity and Health Unit, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland
[11]Agenzia Roma Servizi per la Mobilità, Rome, Italy
[12]London Borough of Newham, London, United Kingdom
[13]The London School of Economics and Political Science, London, United Kingdom
[14]Department of Science, Technology, Engineering and Public Policy, University College London, London, United Kingdom
[15]Trivector Traffic AB, Stockholm, Sweden
[16]International Laboratory for Air Quality and Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
[17]Centre for Air Pollution, Energy and Health Research, Sydney, Australia
[18]Transportation Research Institute (IMOB), Hasselt University, Diepenbeek, Belgium

Corresponding Author:
Mailin Gaupp-Berghausen, Dipl Ing
Institute for Transport Studies
University of Natural Resources and Life Sciences, Vienna
Peter-Jordan-Straße 82
Vienna, 1190
Austria
Phone: 43 1 47654 ext 85628
Fax: 43 1 47654 85609
Email: mailin.gaupp-berghausen@boku.ac.at

Abstract

Background: Sufficient sample size and minimal sample bias are core requirements for empirical data analyses. Combining opportunistic recruitment with a Web-based survey and data-collection platform yields new benefits over traditional recruitment approaches.

Objective: This paper aims to report the success of different recruitment methods and obtain data on participants' characteristics, participation behavior, recruitment rates, and representativeness of the sample.

Methods: A longitudinal, Web-based survey was implemented as part of the European PASTA (Physical Activity through Sustainable Transport Approaches) project, between November 2014 and December 2016. During this period, participants were recruited from 7 European cities on a rolling basis. A standardized guide on recruitment strategy was developed for all cities, to reach a sufficient number of adult participants. To make use of the strengths and minimize weakness, a combination of different opportunistic recruitment methods was applied. In addition, the random sampling approach was applied in the city of Örebro. To reduce the attrition rate and improve real-time monitoring, the Web-based platform featured a participant's and a researchers' user interface and dashboard.

Results: Overall, 10,691 participants were recruited; most people found out about the survey through their workplace or employer (2300/10691, 21.51%), outreach promotion (2219/10691, 20.76%), and social media (1859/10691, 17.39%). The average number of questionnaires filled in per participant varied significantly between the cities (*P*<.001), with the highest number in Zurich (11.0, SE 0.33) and the lowest in Örebro (4.8, SE 0.17). Collaboration with local organizations, the use of Facebook and mailing lists, and direct street recruitment were the most effective approaches in reaching a high share of participants (*P*<.001). Considering the invested working hours, Facebook was one of the most time-efficient methods. Compared with the cities' census data, the composition of study participants was broadly representative in terms of gender distribution; however, the study included younger and better-educated participants.

Conclusions: We observed that offering a mixed recruitment approach was highly effective in achieving a high participation rate. The highest attrition rate and the lowest average number of questionnaires filled in per participant were observed in Örebro, which also recruited participants through random sampling. These findings suggest that people who are more interested in the topic are more willing to participate and stay in a survey than those who are selected randomly and may not have a strong connection to the research topic. Although direct face-to-face contacts were very effective with respect to the number of recruited participants, recruiting people through social media was not only effective but also very time efficient. The collected data are based on one of the largest recruited longitudinal samples with a common recruitment strategy in different European cities.

(J Med Internet Res 2019;21(5):e11492) doi:10.2196/11492

KEYWORDS

longitudinal survey; multicentral; Web-based survey; opportunistic sampling; recruitment; Web-based questionnaire; mobile phone

Introduction

Recruiting participants has become increasingly challenging in the face of privacy concerns, an increasing number of surveys, expectations of rewards for survey completion, and necessary effort to achieve an unbiased representative sample [1]. Sufficient sample sizes and minimal sample bias are core requirements for empirical data collection in order to properly obtain answers to research questions [2]. Meeting these requirements is especially challenging for surveys with high response burdens such as longitudinal studies [3]. Traditional recruitment methods, such as mailed invitations based on samples drawn from population registries, are costly and increasingly yield sample biases due to declining response rates and selectivity effects [4-7], for example, because of an increase in the exclusive use of mobile phones or email rather than traditional mail or landline phone. Opportunistic approaches, such as recruitment through social media, promise cost savings [4,7-12] and a better coverage of person groups that are hard to reach with the traditional recruitment methods [13], like parents of adolescents [8], adolescents themselves [10], people with special conditions [14,15], smokers [16], low-income people [17], or people with a disproportionate risk for poor health outcomes [18]. Social media, like Facebook and Twitter, can potentially have a strong snowballing effect [19], given their intensive use and continuing growth (around 1.94 billion monthly active Facebook users worldwide [20]) and their features that allow information to be shared very easily among networks [19]; therefore, they are able to reach a large number of people in a very short time. Combining opportunistic recruitment with a Web-based survey and data-collection platform comprises additional benefits such as real-time monitoring of recruitment progress and enabling ongoing optimization of recruitment activities [4]. Poor response rates can further be improved by including rewards for participation such as financial compensations [21]. One major drawback of the opportunistic recruitment methods is the concern of sample bias, as the population sampled does not necessarily represent each group of the total population equally well [10]. Specific types of social media might, for example, be preferably used by younger people [3,7,11] and not by the elderly. In this case, the elderly have little chance to be sampled at all. Other, more traditional methods, such as flyers, involve problems of respondents not having a direct and convenient access to the survey and may result in a smaller recruitment rate [22].

The PASTA (Physical Activity through Sustainable Transport Approaches) Project [23,24] used a combination of different opportunistic recruitment methods to utilize strengths and minimize weaknesses. The project study collected data in a longitudinal, Web-based survey with a cohort design to study the effects of active mobility (like walking and cycling) on the overall physical activity and health, crash risks, and exposure to traffic-related air pollution. Data collection was performed in seven European cities: Antwerp, Barcelona, London, Örebro, Rome, Vienna, and Zurich. The target population for the project survey was the entire adult population in each of the seven case study cities, with the aim of oversampling participants who use a bicycle for their daily trips.

The objective of this paper was to report the success of these different methods with regard to obtaining participants' characteristics, participation behavior, recruitment rates, and representativeness of the sample. More specifically, we aimed to (1) describe participant characteristics in the seven European cities in terms of the number of recruited people, gender, age, education level, and employment status; (2) show how participants found out about the survey; (3) illustrate participation behavior by reporting on the number of filled-in questionnaires, attrition, and withdrawal rate; (4) present the effectiveness (the number of predicted participants) and time efficiency of different recruitment approaches; and (5) compare our sample with the general population. Finally, we present our conclusions for the use of recruitment studies in future research on comparable topics.

Methods

Approach

The PASTA project used a longitudinal, Web-based survey between November 2014 and December 2016. After the first questionnaire, which collected baseline information, participants received a follow-up questionnaire every 13 days to collect prospective data on travel behavior, levels of physical activity, and traffic safety incidents [23]. During this period, participants in the seven European cities were recruited on a rolling basis by using different opportunistic approaches. To reach a sufficient number of adult participants, a standardized guide on recruitment strategy was developed for all cities. A common recruitment strategy was important to achieve a similar number of participants and ensure more evenly distributed participation across transport modes and social groups in each city. The strategy for all cities included press releases and editorials; common promotional materials following the same visual identity guidelines; direct targeting of local stakeholders and community groups to distribute survey information through their communication channels (like newsletters, intranet, and webpages); extensive use of social media (each city had its own Facebook and Twitter pages where the link to the survey was regularly posted; Figure 1); and incentivizing for participation (ie, participants entered into a prize draw if they completed a questionnaire). The chance of participants winning increased with each additional completed questionnaire, except for participants in Sweden (Örebro) where the lottery was not allowed by law.

A professional designer created a visual identity to enable clear and easy recognition of the project. All recruitment materials produced under the project were designed in line with the visual identity, which resulted in the creation of a strong brand (Figure 2). The survey as well as different dissemination materials (eg, flyers) were developed first in English and then translated into the local languages (ie, Dutch, Spanish, Catalan, Swedish, Italian, Swizz German, and Austrian German). This guaranteed that the same recruitment materials with the same contents were used in each city. Within this framework, there was flexibility to enable local initiatives and targeted city-specific recruitment, such as promoting recruitment for the project at social and cultural events. Furthermore, the city of Örebro applied an additional random sampling approach by contacting people aged 18-74 years through mail or phone.

To ensure high-data quality, several measures were put in place to reduce attrition rates, such as a user-friendly and custom-made survey platform and the automatic sending of reminder emails. Furthermore, participants were able to log into the platform at any time. There, they received an overview of their personal completed and open questionnaires and were able to complete unfinished questionnaires. Furthermore, they were given the opportunity to withdraw themselves actively from the survey if they did not want to participate any longer. In addition to the participants' user interface, the platform also featured a researchers' user interface and dashboard for real-time monitoring of recruitment and survey data collection. A user engagement strategy was also developed, including regular contact with the respondents, project branding, regular posting on social media, and keeping the project website up-to-date.

Participants were asked in the baseline questionnaire how they found out about the survey. They were given a choice between several different options, ranging from word of mouth to large-scale advertising campaigns. This question was answered by all participants who finalized the baseline questionnaire. At the same time, all city partners kept records on their local recruitment activities to measure invested efforts, including date, category, description, and invested time for each applied recruitment activity. Different categories were classified as follows: collaboration with local administration or organization (eg, survey link on webpages, newsletters, and intranet); handout of flyers at specific locations or specific events; display of posters at specific locations; use of mailing lists; advertisement in Web-based media or print media; papers in Web-based media, print media, or magazines; oral presentations for recruitment purposes; radio or television spots; Facebook; Twitter; street recruitment; and use of random sampling.

Ethics approval was obtained from the local ethics committees in the countries where the work was conducted and sent to the European Commission before the start of the survey.

Statistical Analysis

Standard descriptive statistics outlined overall participant characteristics and were stratified by city, gender, age, education level, and employment status. To assess participation behavior, we tested the number of filled-in questionnaires by different sociodemographic characteristics using the nonparametric Kruskal-Wallis rank-sum test. Each significant result ($P<.05$) was followed by a Dunn test to account for significant differences within a variable. The sociodemographic characteristics of the sample (age and gender) were compared with each city's census data by applying the Pearson chi-square test and size effect calculations.

Figure 1. Cities' own Twitter and Facebook accounts to recruit and inform people. Left: A Twitter page from London (Source: PASTA consortium, 2016, Project profile [Twitter]). Right: A Facebook page from Antwerp (Source: PASTA consortium, 2016, Project profile [Facebook]).

PASTA LONDON
@PASTALondon · Folgt dir
London case study for @EUPASTA research looking at #walking #cycling & #health. By @IC_CEP @NewhamLondon. Help & take part in online survey via link below.
London UK
survey.pastaproject.eu/london
Beigetreten September 2014

PASTA LONDON @PASTALondon · 13. Jan
Today is the last day of the @EUPASTA Survey. A big THANK YOU for your support over the past two years. #activetravel #eu
Original (Englisch) übersetzen

PASTA Antwerpen
@PASTAAntwerpen

Is fietsen in Antwerpen gezond?
Zijn Antwerpenaars fitter dan inwoners van London of Rome?
Zijn automobilisten in hun vrije tijd sportiever dan fietsers?

PASTA is een wetenschappelijk onderzoeksproject dat een antwoord zoekt op deze vragen. Het project loopt over een periode van vier jaar in zeven Europese steden (Antwerpen, Barcelona, Londen, Örebro, Rome, Wenen, Zürich).

Deelnemen is eenvoudig: surf naar survey.pastaproject.eu en registreer je. Binnenkort krijg je een uitnodiging om een online vragenlijst in te vullen over jouw verplaatsingsgedrag, fysieke activiteit en ongevallen. Of stel je kandidaat voor een experiment waarbij je zelf je blootstelling aan luchtvervuiling meet! Iedereen kan meedoen en er vallen mooie prijzen te winnen.

PASTA zoekt nog steeds deelnemers. Vul de vragenlijs...

Invite your friends to like this Page
105 people like this
102 people follow this

Figure 2. Recruitment material. Left: T-shirts worn by recruiters during outreach activities in Barcelona. Right: Registration postcard distributed in the city of Vienna.

Find us on Twitter and Facebook
@PastaWien

Prizes in the value of 10.000€ are raffled among all participants

Take part in our survey:
survey.pastaproject.eu/wien

PASTA is a research project investigating the effects of transportation on people and their city. Therefore, a survey among residents and commuters in Vienna is carried out these days. Make a contribution to the life quality in your city and participate in the study.

This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under grant agreement no 602624-2. The sole responsibility for the content of this publication lies with the PASTA project. It does not necessarily reflect the opinion of the European Union which is not responsible for any use that maybe made of the information contained therein.

Printed according to the guideline "printed material" of the Austrian eco-label, Druck & Medienwerk GmbH, UW-No. 845

To estimate the effectiveness (number of participants who started the baseline survey) of different recruitment approaches, we developed a recruit-prediction model in the form of a nonlinear least squares model in cities that provided the most comprehensive and detailed information on their local recruitment activities (Antwerp, Barcelona, and Vienna). The prediction model is based on the assumption that each recruitment activity generates an effect on the number of participants, which resembles a density function of a log-normal distribution, that is, a steep increase and a flat decrease:

Where r_i denotes the predicted number of responses on day i; the index c=1, 2... C refers to different categories of recruitment activities (such as Facebook, Twitter, and flyer); the index a=1, 2... A refers to particular activities of category C; e_c is the intrinsic effectiveness of an activity of category c; i_{ca} is the intensity of a particular activity a of category c; d is the number of days elapsed since the start of the activity; μ and σ are the location and dispersion parameter of the log-normal density function, respectively. The curve characteristics of the density function were assumed to be the same for all recruitment categories: Only one set of parameters μ and σ was estimated for all categories. The intensity was assumed to vary as follows: each category has its intrinsic (baseline) effectiveness e_c; within a given category, the intensity varies according to the invested effort i_{ca} (indicated by the reported number of working hours); and in some cases, with strong peaks of recruited people, which could unambiguously be assigned to particular recruitment activities with exceptional success, the intensity parameter was manually increased to capture the success of this activity adequately.

To estimate the time efficiency of different recruitment categories, we divided the number of predicted participants by the number of invested hours for each recruitment category. This was possible because the hours of work for project members to apply each activity (except for Facebook and Twitter posts) was recorded in detail. For social media activities, an average time value of 5 minutes per post was used. All statistical analyses were performed using R (version 3.4.0; The R Foundation for Statistical Computing, Vienna, Austria). Values given throughout the text are mean (SE).

Results

Sociodemographic Characteristics of Participants

A total of 10691 participants were recruited over a period of 27 months in the seven European cities, ranging from 1844 (17.25%) individuals in Rome to 1356 (12.68%) in Zurich. In all cities, except Rome, more women than men were recruited, with an average age of 41.9 (SE 0.19) years for men and 40.0 (SE 0.17) years for women. Most participants who stated their educational level were highly educated, with 6180 participants (of 8525, 72.49%) possessing a university degree and 2217 participants (of 8525, 26.01%) possessing a secondary education. In addition, 5993 participants (of 9892, 60.58%) were full-time employees, followed by 1662 (of 9892, 16.80%) part-time employees and 1361 (of 9892 13.76%) students; furthermore, 876 (of 9892, 8.86%) participants had home duties, were retired, or were unemployed (Table 1). Regarding nationality, 3560 participants (of 8586, 41.46%) were from Western Europe (predominantly in Antwerp, Vienna, and Zurich), 2840 (of 8586, 33.08%) were from Southern Europe (Barcelona and Rome), and 1858 (of 8586, 21.64%) were from Northern Europe (Örebro and London; according to the Geographic Regions listed by the United Nations).

Table 1. Respondent demographic information by recruitment city.

Characteristics	Antwerp	Barcelona	London	Örebro	Rome	Vienna	Zurich	Total
Participants in the sample, n (%)	1445 (13.52)	1727 (16.15)	1446 (13.53)	1401 (13.10)	1844 (17.25)	1472 (13.77)	1356 (12.68)	10,691 (100)
Gender, n (%)								
Male	689 (47.68)	706 (40.88)	600 (41.49)	530 (37.83)	1130 (61.28)	682 (46.33)	585 (43.14)	4922 (46.04)
Female	756 (52.32)	1021 (59.12)	846 (58.51)	871 (62.17)	714 (38.72)	790 (53.67)	771 (56.86)	5769 (53.96)
Age in years, mean (SE)								
Overall	43.5 (0.32)	37.3 (0.30)	40.4 (0.34)	45.2 (0.40)	40.3 (0.26)	39.7 (0.35)	40.6 (0.34)	40.9 (0.13)
Male	44.8 (0.48)	38.2 (0.48)	41.3 (0.54)	46.3 (0.67)	40.6 (0.35)	41.2 (0.52)	43.0 (0.53)	41.9 (0.19)
Female	42.4 (0.43)	36.8 (0.38)	39.8 (0.44)	44.6 (0.50)	40.0 (0.41)	38.4 (0.46)	38.8 (0.44)	40.0 (0.17)
Education level, n (%)	N=1245	N=1366	N=1033	N=1036	N=1548	N=1167	N=1130	N=8525
University education	1044 (83.86)	1100 (80.53)	875 (84.70)	691 (66.70)	1015 (65.57)	756 (64.78)	699 (61.86)	6180 (72.49)
Secondary education	191 (15.34)	249 (18.23)	151 (14.62)	319 (30.79)	530 (34.24)	371 (31.79)	406 (35.93)	2217 (26.01)
Primary education	6 (0.48)	11 (0.81)	1 (0.10)	26 (2.51)	0 (0)	37 (3.17)	22 (1.95)	103 (1.21)
No degree	4 (0.32)	6 (0.44)	6 (0.58)	0 (0)	3 (0.19)	3 (0.26)	3 (0.27)	25 (0.29)
Employment status, n (%)	N=1378	N=1595	N=1300	N=1257	N=1680	N=1377	N=1305	N=9892
Full-time	951 (69.01)	978 (61.32)	813 (62.54)	815 (64.84)	1151 (68.51)	639 (46.41)	646 (49.50)	5993 (60.58)
Part-time	281 (20.39)	200 (12.54)	180 (13.85)	116 (9.23)	183 (10.89)	277 (20.12)	425 (32.58)	1662 (16.80)
Student	31 (2.25)	285 (17.87)	156 (12.00)	142 (11.38)	278 (16.55)	302 (21.93)	167 (12.80)	1361 (13.76)
Home duties, retired, or unemployed	115 (8.35)	132 (8.28)	151 (11.62)	184 (14.64)	68 (4.05)	159 (11.55)	67 (5.13)	876 (8.86)

How Participants Found Out About the Survey

Table 2 shows that for all 10,691 participants, the three main sources of finding out about the survey were workplaces or employers (2300, 21.51%), outreach promotion (2219, 20.76%) such as presenting the project at different events or street recruitment, and social media (1859, 17.39%). The results varied across cities. In Antwerp, London, and Zurich, the highest share of respondents were reached through workplaces, whereas respondents in Rome were primarily reached through social media announcements and those in Barcelona, Örebro, and Vienna were reached through outreach promotion. In terms of gender, men were more likely to be recruited through outreach promotion (976/4922, 19.83%) or social media (942/4922, 19.14%), while women were most likely to be recruited through their workplace (1399/5769, 24.25%). Although participants aged 30-60 years could be best reached through their workplace, those aged 20-29 and >60 years were reached most often through outreach activities. In addition, students and participants without employment could best be reached through outreach activities. In case of respondents without a school-leaving qualification, recruitment through workplace or outreach activities was also the most successful approach.

Table 2. Participants' responses to the question, "How did you find out about this survey?" during the baseline questionnaire.

Characteristics	Work	Word of mouth	Other organizations	Outreach activities	News	Social media[a]	Public notice	Random sampling	Other	Don't know
Participants (N=10,691), n (%)	2300 (21.51)[b]	1219 (11.40)	1358 (12.70)	2219 (20.76[b])	800 (7.48)	1859 (17.39[b])	186 (1.74)	360 (3.38)	368 (3.44)	22 (0.21)
Gender, n (%)										
Male (N=4922)	901 (18.31)	576 (11.70)	681 (13.84)	976 (19.83[b])	416 (8.45)	942 (19.14[b])	113 (2.30)	158 (3.21)	151 (3.07)	8 (0.16)
Female (N=5769)	1399 (24.25[b])	643 (11.15)	677 (11.73)	1243 (21.55)	384 (6.66)	917 (15.90)	73 (1.27)	202 (3.50)	217 (3.76)	14 (0.24)
City, n (%)										
Antwerp (N=1445)	376 (26.02[b])	195 (13.49)	289 (20.00)	132 (9.13)	98 (6.78)	311 (21.52)	13 (0.90)	N/A[c]	28 (1.94)	3 (0.21)
Barcelona (N=1727)	199 (11.52)	369 (21.37)	113 (6.54)	665 (38.51[b])	41 (2.37)	220 (12.74)	19 (1.10)	N/A[c]	95 (5.50)	6 (0.35)
London (N=1446)	317 (21.92[b])	146 (10.10)	282 (19.50)	198 (13.69)	110 (7.61)	243 (16.80)	22 (1.52)	N/A[c]	125 (8.64)	3 (0.21)
Örebro (N=1401)	343 (24.48)	15 (1.07)	36 (2.57)	550 (39.26[b])	51 (3.64)	33 (2.36)	9 (0.64)	360 (25.70)	0 (0)	4 (0.29)
Rome (N=1844)	337 (18.28)	244 (13.23)	141 (7.65)	158 (8.57)	253 (13.72)	533 (28.90[b])	90 (4.88)	N/A[c]	88 (4.77)	0 (0)
Vienna (N=1472)	233 (15.83)	155 (10.53)	315 (21.40)	329 (22.35[b])	109 (7.40)	275 (18.68)	27 (1.83)	N/A[c]	25 (1.70)	4 (0.27)
Zurich (N=1356)	495 (36.50[b])	95 (7.01)	182 (13.42)	187 (13.79)	138 (10.18)	244 (17.99)	6 (0.44)	N/A[c]	7 (0.52)	2 (0.15)
Age (years), n (%)										
20-29 (N=2339)	378 (16.16)	334 (14.28)	239 (10.22)	618 (26.42[b])	138 (5.90)	393 (16.80)	75 (3.21)	58 (2.48)	100 (4.28)	6 (0.26)
30-39 (N=2339)	721 (23.84[b])	415 (13.72)	358 (11.84)	528 (17.46)	197 (6.51)	590 (19.51)	37 (1.22)	80 (2.65)	96 (3.17)	2 (0.07)
40-49 (N=2339)	560 (24.39[b])	231 (10.06)	298 (12.98)	400 (17.42)	199 (8.67)	435 (18.95)	28 (1.22)	63 (2.74)	79 (3.44)	3 (0.13)
50-59 (N=2339)	474 (24.73[b])	151 (7.88)	302 (15.75)	383 (19.98)	145 (7.56)	295 (15.39)	28 (1.46)	63 (3.29)	69 (3.60)	7 (0.37)
>60 (N=2339)	153 (15.09)	75 (7.40)	146 (14.40)	266 (26.23[b])	109 (10.75)	132 (13.02)	13 (1.28)	96 (9.47)	20 (1.97)	4 (0.39)
Education level, n (%)										
University education (N=2339)	1405 (22.70)[b]	784 (12.67)	781 (12.62)	1193 (19.28)	449 (7.25)	1086 (17.55)	89 (1.44)	187 (3.02)	208 (3.36)	7 (0.11)
Secondary education (N=2339)	454 (20.48)[b]	212 (9.56)	288 (12.99)	465 (20.97)[b]	180 (8.12)	392 (17.68)	53 (2.39)	106 (4.78)	64 (2.89)	3 (0.14)
Primary education (N=2339)	17 (16.50)	6 (5.83)	12 (11.65)	29 (28.16)[b]	14 (13.59)	13 (12.62)	0 (0)	9 (8.74)	3 (2.91)	0 (0)
No degree (N=2339)	7 (28.00)[b]	3 (12.00)	2 (8.00)	6 (24.00)	0 (0)	5 (20.00)	0 (0)	0 (0)	2 (8.00)	0 (0)
Employment status, n (%)										

Characteristics	Work	Word of mouth	Other organizations	Outreach activities	News	Social media[a]	Public notice	Random sampling	Other	Don't know
Full-time (N=2339)	1633 (27.25[b])	644 (10.75)	738 (12.31)	1042 (17.39)	443 (7.39)	1023 (17.07)	61 (1.02)	191 (3.19)	211 (3.52)	7 (0.12)
Part-time (N=2339)	408 (24.55[b])	170 (10.23)	252 (15.16)	307 (18.47)	124 (7.46)	301 (18.11)	17 (1.02)	38 (2.29)	43 (2.59)	2 (0.12)
Student (N=2339)	123 (9.04)	201 (14.77)	145 (10.65)	413 (30.35[b])	83 (6.10)	248 (18.22)	67 (4.92)	32 (2.35)	48 (3.53)	1 (0.07)
Home duties, retired, or unemployed (N=2339)	17 (1.94)	110 (12.56)	137 (15.64)	266 (30.37[b])	84 (9.59)	137 (15.64)	18 (2.05)	79 (9.02)	26 (2.97)	2 (0.23)

[a] Social media refers to Facebook and Twitter.

[b] These values refer to the highest share of participants reached through different recruitment activities.

[c] N/A: Not applicable.

Participation Rates and Behavior

A total of 12,825 people registered for the survey; however, 2134 never started the baseline questionnaire (attrition rate 16.64%). From the remaining 10,691 participants who started the baseline questionnaire, 8567 finalized it (additional attrition rate 19.87%). The attrition rates between people who registered, started, and finalized the baseline questionnaire varied across cities, with the lowest rates in Antwerp and the highest rates in Örebro and London (Figure 3).

The number of filled-in questionnaires per participant varied significantly across the cities (*P*<.001), with the highest number in Zurich (11.0, SE 0.33) and the lowest in Örebro (4.8, SE 0.17). In almost all cities, women filled in fewer questionnaires than men (7.7 [SE 0.1] vs 8.6 [SE 0.2]); furthermore, younger people or students tended to fill in fewer questionnaires than people aged 30-80 years or employees and people with home duties (*P*<.001). In addition, the way people were informed about the survey had a significant impact on the number of filled-in questionnaires (*P*<.001). Most questionnaires were filled in when people found out about the survey through other organizations (9.6, SE 0.3) or the news (9.5, SE 0.4; Table 3). In total, 12.17% (1301/10,691) participants withdrew from the survey, with the highest share in Örebro (311/1401, 22.20%) and the lowest share in Rome (82/1844, 4.45%).

Figure 3. Number of participants who registered for the survey and who started and finalized the baseline questionnaire. (a) Attrition rate between the number of registrations and the number of started baseline questionnaires. (b) Attrition rate between the number of started and finalized baseline questionnaires.

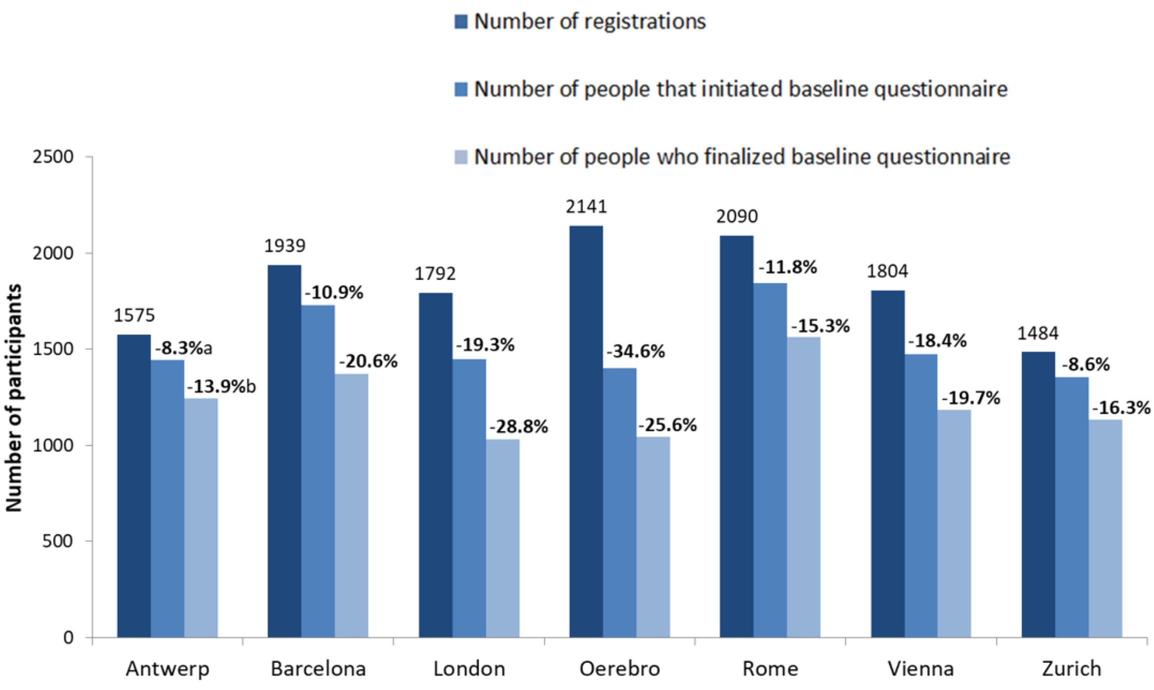


Table 3. Number of filled-in questionnaires per participant by city, age groups, employment status, and mode through which participants found out about the survey.

Characteristics	Mean (SE)	Kruskal-Wallis test value (post hoc test: Dunn test)	*P* value
City		478.95	<.001[a]
Antwerp	10.5 (0.32)[b]		
Barcelona	8.2 (0.25)[c]		
London	5.3 (0.20)[d]		
Örebro	4.8 (0.17)[d]		
Rome	7.4 (0.23)[e]		
Vienna	10.0 (0.32)[f]		
Zurich	11.0 (0.33)[b]		
Age (years), n (%)		91.679	<.001[a]
<20	3.14 (0.39)[b]		
20-29	6.85 (0.20)[c]		
30-39	8.13 (0.19)[d]		
40-49	8.80 (0.23)[d,e]		
50-59	9.11 (0.26)[d,e]		
≥60	8.33 (0.34)[d]		
Employment status		90.077	<.001[a]
Full-time	8.9 (0.1)[b]		
Part-time	9.2 (0.3)[b]		
Student	6.7 (0.3)[d]		
Home duties, retired, or unemployed	8.2 (0.4)[c]		
Found out about survey through		137.1	<.001
Work	8.5 (0.2)[b]		
Word of mouth	7.9 (0.3)[c]		
Other organizations	9.6 (0.3)[b]		
Outreach activities	6.9 (0.2)[d]		
News	9.5 (0.4)[b]		
Social media	8.6 (0.3)[b]		
Public notice	5.5 (0.6)[e]		
Random sampling	7.1 (0.5)[b,c]		
Other	6.2 (0.2)[c,d]		

[a] Significant effect.

[b,c,d,e,f] Values with different letters differ significantly from each other (eg, mean values marked with “b” differ significantly from each mean value marked with “c,” “d,” “e,” or “f”).

Table 4. Effectiveness and time efficiency of different recruitment categories.

Recruitment categories	Effectiveness (n participants predicted)	t [a]	P value [a]	Time-efficiency (n participants predicted per working hour invested)
Antwerp				
Facebook	402[b]	9.26	<.001[b]	41.2[b]
Mailing lists	909[b]	80.48	<.001[b]	60.6[b]
Collaborations with local administrations	9	1.17	.24	0.8
Collaboration with local organizations	6	0.48	.63	0.2
Flyer	19	2.18	.03[b]	1.1
Poster	4	0.56	.58	1.0
Radio	32	2.09	.04[b]	10.7
Web-based advertisement	64[b]	8.67	<.001[b]	64.0[b]
Barcelona				
Facebook	442[b]	10.08	<.001[b]	36.8[b]
Mailing lists	66	5.55	<.001[b]	1.8
Street recruitment	1002[b]	8.49	<.001[b]	0.9
Print media	9	0.93	.35	3.0
Vienna				
Facebook	534[b]	8.34	<.001[b]	33.0[b]
Web-based media	39	4.08	<.001[b]	3.9
Mailing list	21	2.32	.02[b]	2.3
Collaboration with local administrations	149[b]	8.63	<.001[b]	2.3
Collaboration with local organization	406[b]	7.27	<.001[b]	1.2
Flyer	197	7.52	<.001[b]	2.1
Poster	18	1.90	.06	3.6
Street recruitment	86[b]	6.65	<.001[b]	2.7

[a] Columns 3 and 4 show the *t* and *P* values of the average parameter of a respective recruitment category; parameters of exceptionally successful activities (if available for a given category) are not shown (these parameters are always highly significant, eg, peak in Figure 4; the bottom graph). Residual SE: Antwerp, 4.552 (*df*=782); Barcelona, 3.028 (*df*=725); Vienna, 2.795 (*df*=719).

[b] Significant effect.

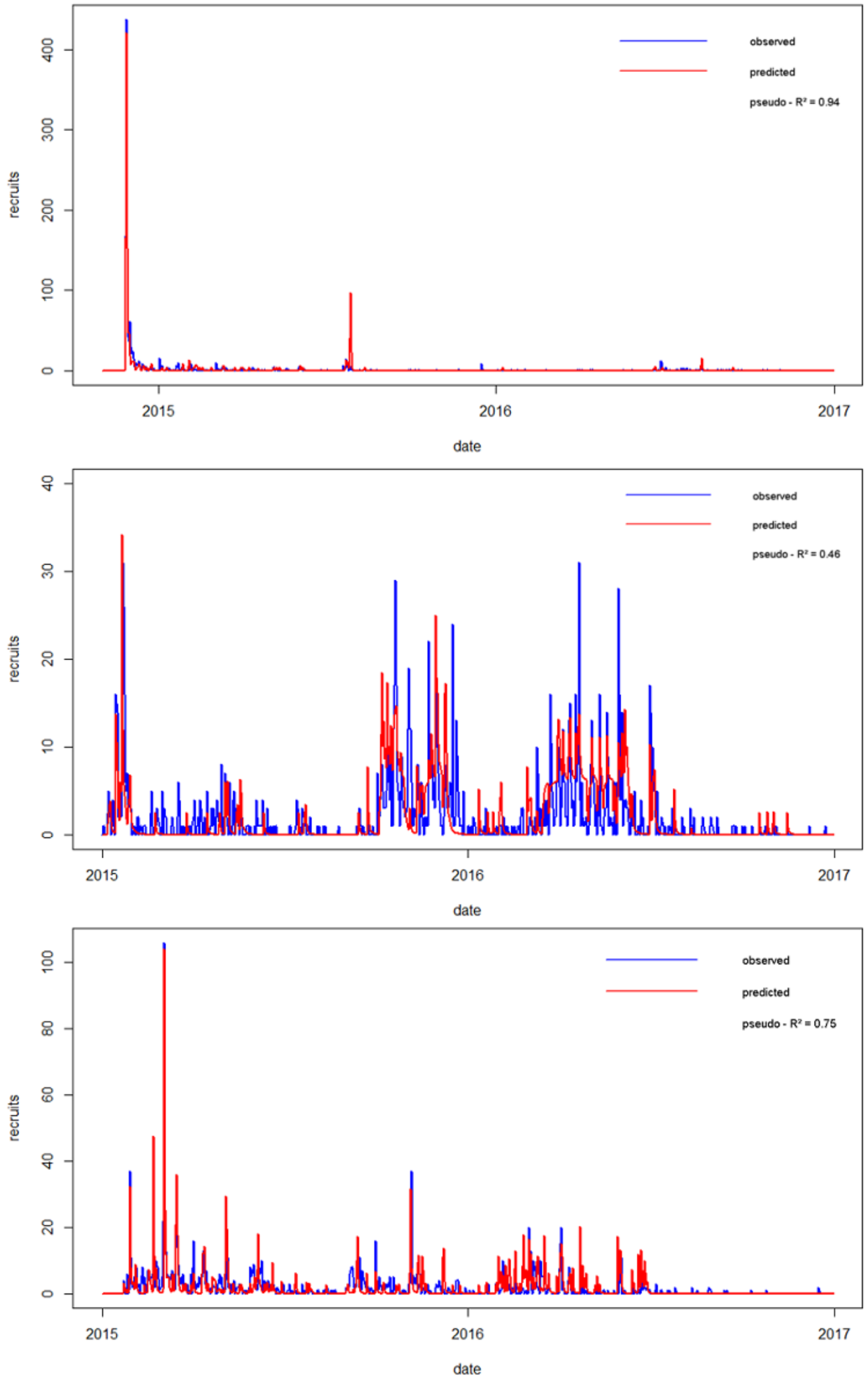
Effectiveness and Efficiency of Different Recruitment Approaches

Table 4 gives an overview of only effective recruitment categories, that is, only activities that were able to recruit participants according to the model. The categories in different cities were as follows: (1) Antwerp: Facebook, mailing lists, collaboration with local administrations and organization, use of flyers and posters, radio spots, and Web-based advertisement; (2) Barcelona: Facebook, mailing lists, street recruitment, and print media; and (3) Vienna: Facebook, Web-based media, mailing lists, collaboration with local administrations and organizations, use of flyers and posters, and street recruitment. One of the most effective approaches in all 3 cities was Facebook, with >400 predicted participants in Antwerp and Barcelona and >500 predicted participants in Vienna (P<.001;Table 4). In Antwerp, most people could be reached through different mailing lists (>900 participants) and in Barcelona, through a range of street recruitment activities (>1000 participants). In Vienna, especially, collaborations with local organizations (like the local bike sharing provider) were very effective in reaching a high share of predicted participants (eg, peak in Figure 4, the bottom graph). Considering the invested working hours, one of the most time-efficient categories in all 3 cities was Facebook, with >30 participants per invested working hour. Although mailing lists or Web-based advertisements were also effective in Antwerp, reaching approximately 60 participants per invested working hour, only 1 participant could be reached per working hour in Barcelona through street recruitment. In Figure 4, lines represent the sum of different recruitment activities. The nonlinear model does not provide an R^2, because R^2 is not defined for nonlinear models. We defined a pseudo R^2 value, because the model is reasonably close to a linear model and the sample size is sufficiently large.

Representativeness of the Sample

Compared with the cities' census data, study participants in all cities, except Rome, were broadly representative in terms of gender distribution (P>.05). This was mainly the case if participants were informed about the survey by news, through word of mouth (friends, neighbors, or relatives) or social media (Multimedia Appendix 1). The main difference was that our recruited sample was on an average younger than the general population (high deviation from census data within the age class >60 years; Multimedia Appendix 2).

Figure 4. Number of recruited (observed) and predicted participants in Antwerp (top graph), Barcelona (middle graph), and Vienna (bottom graph; based on a nonlinear model).



Discussion

The main finding of this study was that Facebook was one of the most effective approaches in reaching a high share of participants. Considering the invested working hours, it was also one of the most time-efficient recruitment methods. Recruited participants were representative of the population in terms of gender but younger than the general population.

Principal Results and Participation Behavior

The main source of information about the survey was workplaces or employers. Collaborations with different organizations who forwarded the survey information to their employees (eg, through their intranet or regular newsletters) was fundamental for raising awareness to drive recruitment. Outreach promotion by project members (eg, direct face-to-face recruitment at different events or on public places) and the use of social media channels (Facebook and Twitter) were the next most informative activities. This was especially the case in Rome, where social media was based on an account of the city council with a lot of followers.

One-third of people who registered for the survey did not complete the baseline questionnaire, with the highest attrition rate in Örebro and the lowest rate in Antwerp. In addition, Örebro had the lowest number of filled-in questionnaires per participant and the highest share of people who actively deregistered from the survey. One explanation may be the different approach adopted in Örebro, which also recruited participants through random sampling. The findings suggest that people who are more interested in the topic (in this case, active mobility research) are more willing to participate in a survey and more likely to stay in the study than those who are selected randomly and may not have a strong connection to the research topic [4-7]. Therefore, a (costly) random selection may eventually still lead to a biased sample that was to be avoid. However, poor response rates can also be improved by including rewards for participation [21]. The high attrition and deregistration rate in Örebro may also be (partly) caused by the fact that participation in Örebro was not rewarded.

Comparative Effectiveness and Representativeness of Recruitment Methods

Although direct face-to-face contacts (eg, street recruitment) were very effective in terms of the number of recruited participants, recruiting people through social media (mainly Facebook) was not only effective but also very time efficient. Similar results were observed by others [4,10,11], who applied Web-based sampling in their research. By using the same medium (ie, internet), such sampling reduces the burden of participants because they can easily reach the survey by only clicking on a provided link. Regarding the near ubiquity of the internet, it has become easier for people to engage in surveys [25], as it can overcome barriers such as physical distance, transportation, and limited time [14]. Nevertheless, the effectiveness and time efficiency must be balanced with how representative the resulting sample is of the target population. Compared with the general population, study participants in almost all cities were broadly representative in terms of gender distribution. In particular, reaching people through news, word of mouth, or social media were the most successful options in recruiting a gender-balanced sample that represented a city's population. There was, however, an age bias among the applied strategies compared with the city's census data. Although the cities applied different strategies to include older people (eg, by visiting seniors who recently completed computer courses), most strategies attracted a higher proportion of younger people. Several studies, however, report successfully recruiting people who reflect the demographic spread of the general population using opportunistic sampling approaches [9,11,19,26]. People in our sample were further highly educated (6180/8525, 72.49% participants possessed a university degree), which is a common occurrence for survey research [9,27,28]. Nevertheless, this study found that targeting people without a school-leaving qualification through their workplaces or outreach activities was most promising; as such, some recruitment activities are better suited than others to attract hard-to-reach groups.

Limitations and Strengths

Although this study represents a comprehensive examination of different recruitment approaches in a longitudinal, Europe-wide, Web-based survey, there were some limitations that could not be addressed by the research design. First, the sociodemographic characteristics of participants recruited by the discussed methods may be different for topics other than active mobility and the corresponding health aspects. The study population in our sample was highly educated and younger than the general population. This may be because our recruitment partly addressed cyclists, and the subject may hold particular interest to those with higher education. Second, participants needed to have access to the internet to participate in the Web-based survey, which could also explain the high proportion of young participants in this study. Finally, the sample was limited to the adult population. The strategies used may be effective in recruiting children or adolescents.

These limitations are offset by several strengths. First, we were able to recruit one of the largest longitudinal samples in different European cities with a common recruitment strategy. Second, we were able to shed new light on the effectiveness and time efficiency of different recruitment approaches. We now have a large and very detailed database on response behavior as per the recruitment method for seven different cities in Europe. We observed that offering a mixed recruitment approach was very effective in reaching a high participation rate. The resulting database can answer research questions and analyze the effects of active mobility on people's health, crash risks, and exposure to traffic-related air pollution (eg, for cyclists) because of its size and composition. Thus, overall, the use of a mixed-methods approach has been successful.

Acknowledgments

This work was supported by the European PASTA project. PASTA is a 4-year project funded by the European Union’s Seventh Framework Program under European Commission (Grant Agreement #602624). ED is supported by a postdoctoral scholarship from FWO (Research Foundation Flanders). We would like to thank all recruiters in the seven European cities and Astrid Gühnemann for proofreading the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic characteristic (gender) by recruitment strategy of participants compared with each city's census.

[PDF File (Adobe PDF File), 293KB - jmir_v21i5e11492_app1.pdf]

Multimedia Appendix 2

Demographic characteristic (age groups) by recruitment strategy of participants compared with each city's census.

[PDF File (Adobe PDF File), 358KB - jmir_v21i5e11492_app2.pdf]

References

1. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
2. Fekete C, Segerer W, Gemperli A, Brinkhof MW, SwiSCI Study Group. Participation rates, response bias and response behaviours in the community survey of the Swiss Spinal Cord Injury Cohort Study (SwiSCI). BMC Med Res Methodol 2015 Oct 08;15:80 [FREE Full text] [doi: 10.1186/s12874-015-0076-0] [Medline: 26450702]
3. Mishra GD, Hockey R, Powers J, Loxton D, Tooth L, Rowlands I, et al. Recruitment via the Internet and social networking sites: the 1989-1995 cohort of the Australian Longitudinal Study on Women's Health. J Med Internet Res 2014 Dec 15;16(12):e279 [FREE Full text] [doi: 10.2196/jmir.3788] [Medline: 25514159]
4. Batterham PJ. Recruitment of mental health survey participants using Internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191. [doi: 10.1002/mpr.1421] [Medline: 24615785]
5. van Gelder MMHJ, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 01;172(11):1292-1298. [doi: 10.1093/aje/kwq291] [Medline: 20880962]
6. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. Am J Epidemiol 2006 Feb 01;163(3):197-203. [doi: 10.1093/aje/kwj036] [Medline: 16339049]
7. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. Am J Epidemiol 2015 May 15;181(10):737-746. [doi: 10.1093/aje/kwv006] [Medline: 25883155]
8. Gilligan C, Kypri K, Bourke J. Social networking versus facebook advertising to recruit survey respondents: a quasi-experimental study. JMIR Res Protoc 2014 Sep 17;3(3):e48 [FREE Full text] [doi: 10.2196/resprot.3317] [Medline: 25230740]
9. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Feb 01;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
10. Hildebrand J, Burns S, Zhao Y, Lobo R, Howat P, Allsop S, et al. Potential and Challenges in Collecting Social and Behavioral Data on Adolescent Alcohol Norms: Comparing Respondent-Driven Sampling and Web-Based Respondent-Driven Sampling. J Med Internet Res 2015 Dec 24;17(12):e285 [FREE Full text] [doi: 10.2196/jmir.4762] [Medline: 26704736]
11. Nelson EJ, Hughes J, Oakes JM, Pankow JS, Kulasingam SL. Estimation of geographic variation in human papillomavirus vaccine uptake in men and women: an online survey using facebook recruitment. J Med Internet Res 2014 Sep 01;16(9):e198 [FREE Full text] [doi: 10.2196/jmir.3506] [Medline: 25231937]
12. Reaves AC, Bianchi DW. The role of social networking sites in medical genetics research. Am J Med Genet A 2013 May;161A(5):951-957. [doi: 10.1002/ajmg.a.35903] [Medline: 23554131]
13. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res 2016 Dec 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
14. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May 01;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]

15. Williams MT, Proetto D, Casiano D, Franklin ME. Recruitment of a hidden population: African Americans with obsessive-compulsive disorder. Contemp Clin Trials 2012 Jan;33(1):67-75 [FREE Full text] [doi: 10.1016/j.cct.2011.09.001] [Medline: 21983626]
16. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook Recruitment of Young Adult Smokers for a Cessation Trial: Methods, Metrics, and Lessons Learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
17. Lohse B. Facebook is an effective strategy to recruit low-income women to online nutrition education. J Nutr Educ Behav 2013;45(1):69-76. [doi: 10.1016/j.jneb.2012.06.006] [Medline: 23305805]
18. Hudnut-Beumler J, Po'e E, Barkin S. The Use of Social Media for Health Promotion in Hispanic Populations: A Scoping Systematic Review. JMIR Public Health Surveill 2016 Jul 11;2(2):e32 [FREE Full text] [doi: 10.2196/publichealth.5579] [Medline: 27400979]
19. Brickman BC. Not by the Book. Sociological Methods & Research 2012 Mar 21;41(1):57-88. [doi: 10.1177/0049124112440795]
20. STATISTA. 2019. Number of monthly active Facebook users worldwide as of 4th quarter 2018 (in millions) URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ [accessed 2019-02-26] [WebCite Cache ID 729bc6JZS]
21. Parkinson A, Jorm L, Douglas KA, Gee A, Sargent GM, Lujic S, et al. Recruiting general practitioners for surveys: reflections on the difficulties and some lessons learned. Aust J Prim Health 2015;21(2):254-258. [doi: 10.1071/PY13129] [Medline: 24581281]
22. Nolte M, Shauver M, Chung K. Analysis of four recruitment methods for obtaining normative data through a Web-based questionnaire: a pilot study. Hand (N Y) 2015 Sep;10(3):529-534 [FREE Full text] [doi: 10.1007/s11552-014-9730-y] [Medline: 26330790]
23. Dons E, Götschi T, Nieuwenhuijsen M, de Nazelle A, Anaya E, Avila-Palencia I, et al. Physical Activity through Sustainable Transport Approaches (PASTA): protocol for a multi-centre, longitudinal study. BMC Public Health 2015 Nov 14;15:1126 [FREE Full text] [doi: 10.1186/s12889-015-2453-3] [Medline: 26577129]
24. Gerike R, de Nazelle A, Nieuwenhuijsen M, Panis L, Anaya E, Avila-Palencia I, PASTA Consortium. Physical Activity through Sustainable Transport Approaches (PASTA): a study protocol for a multicentre project. BMJ Open 2016 Jan 07;6(1):e009924 [FREE Full text] [doi: 10.1136/bmjopen-2015-009924] [Medline: 26743706]
25. Zhang D, Bi P, Lv F, Tang H, Zhang J, Hiller J. Internet use and risk behaviours: an online survey of visitors to three gay websites in China. Sex Transm Infect 2007 Dec;83(7):571-576 [FREE Full text] [doi: 10.1136/sti.2007.026138] [Medline: 17971376]
26. Thornton L, Batterham P, Fassnacht D, Kay-Lambkin F, Calear A, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
27. Brown WJ, Bryson L, Byles JE, Dobson AJ, Lee C, Mishra G, et al. Women's Health Australia: recruitment for a national longitudinal cohort study. Women Health 1998;28(1):23-40. [Medline: 10022055]
28. Strandhagen E, Berg C, Lissner L, Nunez L, Rosengren A, Torén K, et al. Selection bias in a population survey with registry linkage: potential effect on socioeconomic gradient in cardiovascular risk. Eur J Epidemiol 2010 Mar;25(3):163-172. [doi: 10.1007/s10654-010-9427-7] [Medline: 20127393]

Abbreviations

PASTA: Physical Activity through Sustainable Transport Approaches

Edited by G Eysenbach; submitted 05.07.18; peer-reviewed by P Batterham, J Kapp; comments to author 08.08.18; revised version received 19.09.18; accepted 22.09.18; published 06.05.19.

Please cite as:
Gaupp-Berghausen M, Raser E, Anaya-Boig E, Avila-Palencia I, de Nazelle A, Dons E, Franzen H, Gerike R, Götschi T, Iacorossi F, Hössinger R, Nieuwenhuijsen M, Rojas-Rueda D, Sanchez J, Smeds E, Deforth M, Standaert A, Stigell E, Cole-Hunter T, Int Panis L
Evaluation of Different Recruitment Methods: Longitudinal, Web-Based, Pan-European Physical Activity Through Sustainable Transport Approaches (PASTA) Project
J Med Internet Res 2019;21(5):e11492
URL: https://www.jmir.org/2019/5/e11492/
doi:10.2196/11492
PMID:31066715

Original Paper

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

Susan Fletcher[1], PhD; Janine Clarke[2,3], MPsychol (Clin), PhD; Samineh Sanatkar[2,3], PhD; Peter Baldwin[2,3], PhD; Jane Gunn[1], MBBS, PhD; Nick Zwar[4], MBBS, PhD; Lesley Campbell[5], MBBS; Kay Wilhelm[3], MBBS; Mark Harris[6], MBBS, MD; Helen Lapsley[3], PhD; Dusan Hadzi-Pavlovic[3], BSc, MPsychol; Judy Proudfoot[2,3], PhD

[1]Department of General Practice, University of Melbourne, Carlton, Australia
[2]Black Dog Institute, Sydney, Australia
[3]School of Psychiatry, University of New South Wales Sydney, Sydney, Australia
[4]School of Medicine, University of Woollongong, Woollongong, Australia
[5]Diabetes and Metabolism Division, Garvan Institute of Medical Research, Sydney, Australia
[6]Centre for Primary Health Care and Equity, University of New South Wales Sydney, Sydney, Australia

Corresponding Author:
Susan Fletcher, PhD
Department of General Practice
University of Melbourne
Ground Floor
200 Berkeley Street
Carlton, 3052
Australia
Phone: 61 390354872
Email: susanlf@unimelb.edu.au

Abstract

Background: E-mental health (eMH) interventions are now widely available and they have the potential to revolutionize the way that health care is delivered. As most health care is currently delivered by primary care, there is enormous potential for eMH interventions to support, or in some cases substitute, services currently delivered face to face in the community setting. However, randomized trials of eMH interventions have tended to recruit participants using online recruitment methods. Consequently, it is difficult to know whether participants who are recruited online differ from those who attend primary care.

Objective: This paper aimed to document the experience of recruiting to an eMH trial through primary care and compare the characteristics of participants recruited through this and other recruitment methods.

Methods: Recruitment to the SpringboarD randomized controlled trial was initially focused on general practices in 2 states of Australia. Over 15 months, we employed a comprehensive approach to engaging practice staff and supporting them to recruit patients, including face-to-face site visits, regular contact via telephone and trial newsletters, and development of a Web-based patient registration portal. Nevertheless, it became apparent that these efforts would not yield the required sample size, and we therefore supplemented recruitment through national online advertising and promoted the study through existing networks. Baseline characteristics of participants recruited to the trial through general practice, online, or other sources were compared using the analysis of variance and chi square tests.

Results: Between November 2015 and October 2017, 780 people enrolled in SpringboarD, of whom 740 provided information on the recruitment source. Of these, only 24 were recruited through general practice, whereas 520 were recruited online and 196 through existing networks. Key barriers to general practice recruitment included perceived mismatch between trial design and diabetes population, prioritization of acute health issues, and disruptions posed by events at the practice and community level. Participants recruited through the 3 different approaches differed in age, gender, employment status, depressive symptoms, and diabetes distress, with online participants being distinguished from those recruited through general practice or other sources. However, most differences reached only a small effect size and are unlikely to be of clinical importance.

Conclusions: Time, labor, and cost-intensive efforts did not translate into successful recruitment through general practice in this instance, with barriers identified at several different levels**.** Online recruitment yielded more participants, who were broadly similar to those recruited via general practice.

(J Med Internet Res 2019;21(5):e12793) doi:10.2196/12793

KEYWORDS

e-mental health; primary care; patient recruitment; depression; type 2 diabetes; learning

Introduction

The potential for e-mental health (eMH) interventions to address many of the challenges faced by health care systems globally has seen them receive increasing attention from researchers and policy makers alike. eMH interventions provide an opportunity for individuals affected by mild-to-moderate symptoms of mental health disorders to access low-cost evidence-based treatments and aim to reduce the burden on providers while maintaining patients' connection to the broader health system.

The value of eMH interventions may be particularly pertinent to primary care clinicians who are responsible for the majority of mental health care [1]. However, although the efficacy of eMH interventions is now well established [2], their implementation into routine care remains limited [3].

Of the 13 eMH randomized controlled trials (RCTs) included in a recent meta-analysis [2], only one recruited participants through primary care [4]. Some of this evidence gap is attributed to difficulties in conducting research in this setting, which are by no means unique to trials of eMH interventions. Primary care–based research accounts for a disproportionately small proportion of all health care research [5], and only one-third of RCTs recruit to target [6]. Documented barriers at the patient, practitioner, and organizational levels contribute to these difficulties [7]. As such, many randomized trials of eMH interventions to date have instead sourced participants online [2]. This approach is generally considered effective and efficient, but there is also evidence to the contrary [8-10], and the degree to which these samples are representative of the broader population has been questioned [11,12].

There is a need to correct the dearth of primary care–based research and for greater transparency and reporting of the issues researchers are facing in this setting. Our recent SpringboarD RCT [13,14] provides an opportunity to examine these issues in relation to an eMH intervention for people with mild-to-moderate depressive symptoms and type 2 diabetes (T2D)—2 highly prevalent and commonly comorbid conditions in primary care [15,16].

This paper has described an intensive and ultimately unsuccessful approach to primary care recruitment and the recruitment strategies that were required to supplement it and reports the outcomes of each. It then explored how participants recruited through primary care compare with those recruited through other avenues on key demographic and clinical characteristics. Such data help identify the tensions inherent in primary care that impact on recruitment.

Methods

Study Overview

The SpringboarD RCT (trial registration ACTRN12615000931572) [13,14] examined the effectiveness of a Web-based cognitive behavioral therapy–based self-help program (myCompass, Black Dog Institute) for improving work and social functioning and depressive symptoms in people with depressive symptoms and T2D at 3, 6, and 12 months compared with a placebo control program.

Our joint focus on eMH and primary care led us to design our recruitment strategy accordingly. First, we aimed to recruit through general practices in the 2 most populous states of Australia (New South Wales and Victoria). To supplement this approach, we also recruited through national online advertising and promoted the study through existing clinical and research registries. All promotional materials invited interested individuals to visit the SpringboarD website to learn more about this *wellbeing project for people with T2D*. After reading the trial information, interested participants provided informed consent and completed a screening questionnaire to determine eligibility. Those eligible then completed a baseline assessment and indicated how they learned about SpringboarD by selecting a referral source from a drop-down menu. They were then randomized to use the myCompass program or an active placebo control program for 12 weeks. The trial was approved by the University of New South Wales Human Research Ethics Committee (HREC; No. 15090) and registered with the University of Melbourne HREC (No. 1545422).

Participants

The target sample size was 600 participants at baseline (300 in each arm). Australian residents were eligible for the trial if aged 18 to 75 years, reported having T2D diagnosed by a health professional, screened positive for depressive symptoms, and had access to an internet-enabled device. Exclusion criteria were as follows: inability to read English; presence of severe depressive symptoms; probable psychosis; high suicide risk; current participation in face-to-face psychotherapy for depression; recent (<2 months) change to antidepressant medication; and previous use of myCompass.

Stage I: Agreement in Principle to Participate

Practice recruitment took place in Victoria and New South Wales between September 2015 and February 2016. We conducted a multipronged approach to recruiting primary care professionals and their practices, including the following:

1. Emailing members of relevant professional organizations (eg, Australian Association of Practice Managers and

Australian Diabetes Educators Association) and primary care providers subscribed to the Black Dog Institute's eMH in practice training suite.
2. Promoting SpringboarD at relevant conferences (eg, Australian Primary Health Care Nurses Association).
3. Directly contacting 281 general practices with existing links to the University of New South Wales or the University of Melbourne via mail or phone.

Interested practices were offered a visit from a member of the research team (SF [Victoria] or JC [NSW]) at a time convenient to them to discuss the trial in more detail. Practices were advised that by consenting to take part in SpringboarD they also agreed to (1) invite patients with T2D to visit the trial website, by informing clients verbally and providing an information brochure during consultations, as well as by conducting a mailout to all patients of the practice aged 18 to 75 years with T2D (materials and postage provided) and (2) nominate a member of their team as the point of contact for the study (typically the practice manager or diabetes educator). In return, practices were offered Aus $50 per enrolled participant as recompense for time and effort and a report at the end of the project providing aggregated, deidentified data comparing their patients with those from other practices.

Recruitment

General Practice Recruitment

As described by Bower et al [17], recruiting to primary care research can be considered across 4 stages. This paper has focused on the first 3 of these, in which (1) practices or primary care providers are initially approached and consent to take part in the research and (2) subsequently invite their patients who (3) then provide agreement to take part.

Stage II: Agreement to Recruit Patients

The nominated SpringboarD coordinators at each practice searched the practice database for patients with T2D aged between 18 and 75 years. Through this process, a total of 4569 patients were identified (on average, 169.2 per practice). Between November 2015 and March 2016, each of these patients was mailed a letter from the practice (using the practice letterhead), introducing the study and encouraging the patient to consider participating by visiting the trial website. Also enclosed was a letter of invitation from the SpringboarD research team, again providing the Web address and information about the option of entering a prize draw (Aus $100 grocery shopping voucher or tablet device) at each study time point, as recompense for the time and effort involved in study participation. All participating practices were provided with professionally produced posters (see Figure 1) and brochures to display in the waiting and consulting rooms.

Throughout 2016, recruitment proceeded slowly. We maintained regular contact including joint teleconferences with practice staff to identify and address barriers to patient recruitment. These discussions led to a series of practice engagement initiatives and reinforcing activities including the following:

1. The development of a recruitment Web portal for use during consultations where general practitioners (GPs) and diabetes educators could enter the patient's name and email address to generate an automated email to the patient from the SpringboarD website with a link to the study page.
2. Development of a simple 3-step guide to introduce the study to patients using the Web portal above and laminated tags with the portal's Web address that could be stuck on desktop computers as a prompt to navigate to it.
3. Provision of materials to remind practitioners about the trial, including trial-branded pens and magnets.
4. Offering alternative ways for the practice to promote the study, including providing text to upload on their website, or include in-patient newsletters, follow-up letters, or diabetes cycle of care reminders (including texts, emails, and phone calls) or having a research assistant located in the practice to invite patients to the trial as they presented for consultations.
5. World Diabetes Day (WDD) packs for patients, with WDD information, SpringboarD brochure, and a small gift (herbal tea and balloons), and suggestions about other ways the practice could link Springboard with their WDD activities.
6. A webinar for practice nurses to share experiences of introducing SpringboarD to patients and hear from a guest speaker (diabetes expert and practicing academic GP) about the latest developments in diabetes care and how these align with the goals of SpringboarD (ie, greater focus on impact and management of mental health comorbidities).
7. Scripts for practice staff to use in discussing the trial with potential candidates.

We also sent regular newsletters in both hard and soft copy to the nominated SpringboarD contact in each practice. These newsletters (see example in Figure 2) covered topics such as the rationale for and importance of study, provided reminders on using the Web portal, reiterated the offers of support in promoting the study to patients, introduced some of the SpringboarD practices and their staff (including a question and answer session about the benefits they perceived from taking part in the trial), and provided testimonials from participants.

Figure 1. Poster for display in general practitioners (GPs) waiting rooms.

SpringboarD

A wellbeing project for people with type 2 diabetes

Your GP clinic and doctor have agreed to help us with this project.

We are inviting people with type 2 diabetes aged between 18 and 75 years to take part.

We hope that you will help too.

To find out more or to take part, go to https://springboard.blackdoghealth.org.au or email us at Springboard@unsw.edu.au

Spring BOARD

LIVING WELL WITH DIABETES

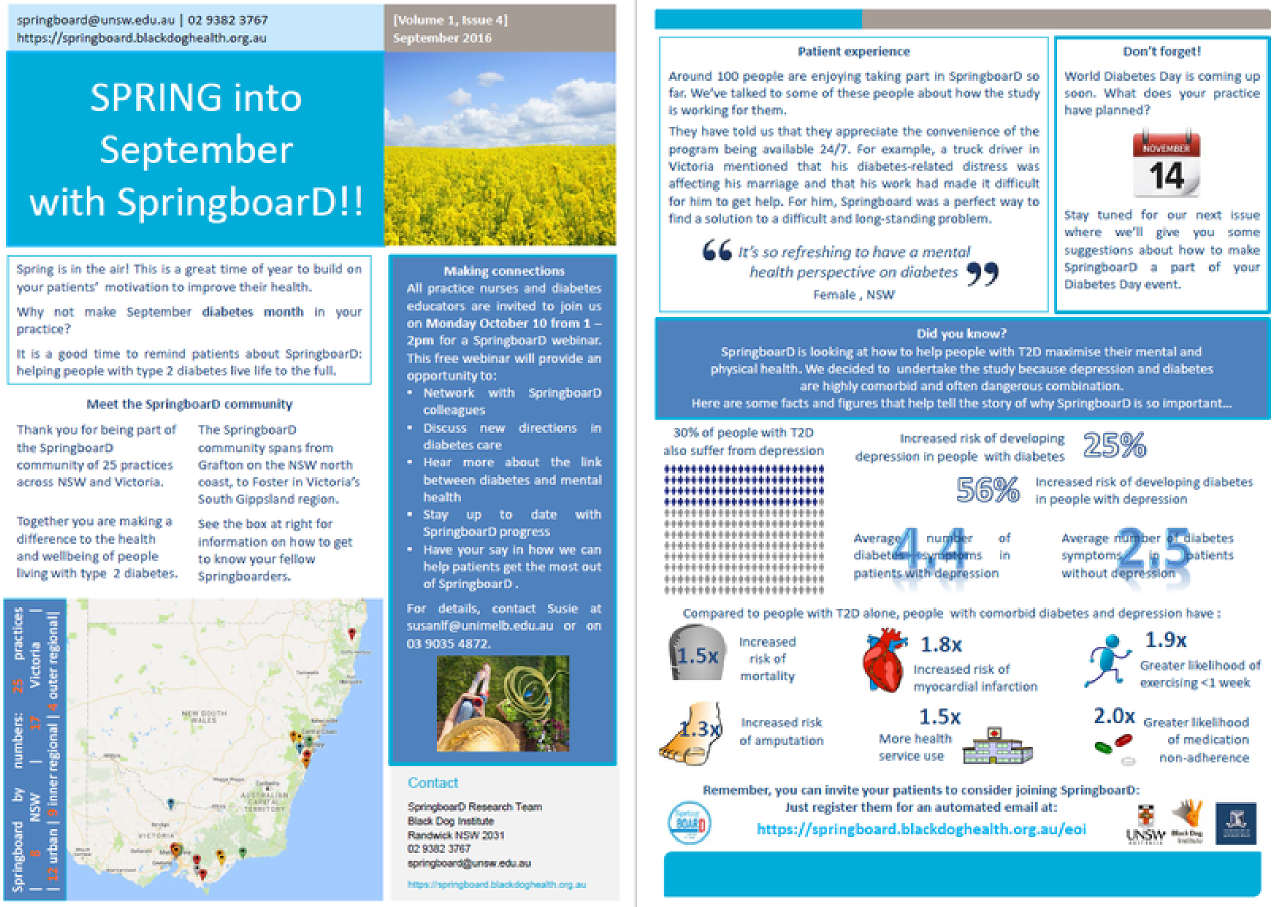
This project has been approved by UNSW HREC 15090.

Black Dog Institute

UNSW THE UNIVERSITY OF NEW SOUTH WALES

THE UNIVERSITY OF MELBOURNE

Figure 2. Example of newsletter sent to SpringboarD practices.



springboard@unsw.edu.au | 02 9382 3767
https://springboard.blackdoghealth.org.au

[Volume 1, Issue 4]
September 2016

SPRING into September with SpringboarD!!

Spring is in the air! This is a great time of year to build on your patients' motivation to improve their health.

Why not make September **diabetes month** in your practice?

It is a good time to remind patients about SpringboarD: helping people with type 2 diabetes live life to the full.

Meet the SpringboarD community

Thank you for being part of the SpringboarD community of 25 practices across NSW and Victoria.

Together you are making a difference to the health and wellbeing of people living with type 2 diabetes.

The SpringboarD community spans from Grafton on the NSW north coast, to Foster in Victoria's South Gippsland region.

See the box at right for information on how to get to know your fellow Springboarders.

| Springboard by numbers: 25 practices | 8 NSW | 17 Victoria | 12 urban | 9 inner regional | 4 outer regional |

Making connections

All practice nurses and diabetes educators are invited to join us on **Monday October 10 from 1 – 2pm** for a SpringboarD webinar. This free webinar will provide an opportunity to:

- Network with SpringboarD colleagues
- Discuss new directions in diabetes care
- Hear more about the link between diabetes and mental health
- Stay up to date with SpringboarD progress
- Have your say in how we can help patients get the most out of SpringboarD .

For details, contact Susie at susanlf@unimelb.edu.au or on 03 9035 4872.

Contact

SpringboarD Research Team
Black Dog Institute
Randwick NSW 2031
02 9382 3767
springboard@unsw.edu.au
https://springboard.blackdoghealth.org.au

Patient experience

Around 100 people are enjoying taking part in SpringboarD so far. We've talked to some of these people about how the study is working for them.

They have told us that they appreciate the convenience of the program being available 24/7. For example, a truck driver in Victoria mentioned that his diabetes-related distress was affecting his marriage and that his work had made it difficult for him to get help. For him, Springboard was a perfect way to find a solution to a difficult and long-standing problem.

> *It's so refreshing to have a mental health perspective on diabetes*
>
> Female , NSW

Don't forget!

World Diabetes Day is coming up soon. What does your practice have planned?

Stay tuned for our next issue where we'll give you some suggestions about how to make SpringboarD a part of your Diabetes Day event.

Did you know?

SpringboarD is looking at how to help people with T2D maximise their mental and physical health. We decided to undertake the study because depression and diabetes are highly comorbid and often dangerous combination.

Here are some facts and figures that help tell the story of why SpringboarD is so important…

30% of people with T2D also suffer from depression

Increased risk of developing depression in people with diabetes 25%

56% Increased risk of developing diabetes in people with depression

Average number of diabetes symptoms in patients with depression 4.4

Average number of diabetes symptoms in patients without depression 2.5

Compared to people with T2D alone, people with comorbid diabetes and depression have :

1.5x Increased risk of mortality

1.8x Increased risk of myocardial infarction

1.9x Greater likelihood of exercising <1 week

1.3x Increased risk of amputation

1.5x More health service use

2.0x Greater likelihood of medication non-adherence

Remember, you can invite your patients to consider joining SpringboarD:
Just register them for an automated email at:
https://springboard.blackdoghealth.org.au/eoi

Online Recruitment

Our online recruitment activities were mostly Facebook advertising campaigns that commenced in November 2015 and intensified from mid-2016 as it became clear that our efforts to recruit through general practice would not achieve the required sample size. We selected Facebook as it is the most commonly used social media platform in Australia, with around 60% of the population holding an account (14 million users in November 2015) and half of all Australians accessing Facebook at least once a day [18]. Recognizing the possible risks associated with Facebook advertising for potentially sensitive topics such as mental health [19], our recruitment approach included the following:

1. The use of a static advertisement that appeared on the right-hand side of the screen for people in the target demographic (located in Australia, aged between 18 and 65 or older, interested in T2D awareness or type 2 diabetes mellitus awareness).
2. Posting links to the study on the Black Dog Institute Facebook page and asking other organizations (eg, Diabetes Australia) to do the same.

Over the course of the study, we ran 10 Facebook advertising campaigns, each building on the learnings of the last (Table 1). Through this iterative process, we identified the features that were most successful in encouraging link clicks and developed our advertisements accordingly (see examples in Figure 3). This resulted in an average click-through rate (ie, number of link clicks/number of unique users who had a SpringboarD advertisement appear on the screen) of 6.92% for campaigns 4 to 7, which compares favorably with the average click-through rate of 0.90% for Facebook advertisements generally and 0.83% for those that are health-related [20].

In addition to Facebook, we also advertised SpringboarD on the Black Dog Institute website and asked relevant organizations to do the same.

Other Recruitment Strategies

Our final approach toward recruitment involved specifically targeting populations in which we expected to have a high likelihood of recruitment success, *warm* populations with known T2D and an interest in research or diabetes care. This included sending email invitations and promoting the study at public diabetes forums and exhibitions, recruiting through diabetes organizations, word of mouth, and contacting members of 2 existing research cohorts (Table 2). The *diamond* cohort [21], recruited in 2005 through 30 general practices across Victoria, comprised 789 patients with depressive symptoms. Of these, 75 had reported having diabetes and consented to be contacted about relevant research. The 45 and Up study commenced in New South Wales in 2006 and recruited over 250,000 people aged over 45 years [22]. In total, 13,245 of these were identified as being potentially eligible for SpringboarD, of whom 4175 were randomly selected to be contacted by the Sax Institute on behalf of the SpringboarD team.

Table 1. Overview of Facebook recruitment campaigns.

Campaign number	Dates	Target audience	Budget (Aus$)	Reach[a]	Clicks	Click-through rate[b] (%)	Comments
1	13 November-11 December 2015	Location: New South Wales, Victoria; Interests: T2D[c] awareness or type 2 diabetes mellitus awareness	$420	23,047	1034	4.49	—[d]
2	4 April-30 June 2016	—	$500	9030	529	5.86	Small campaign (Aus $500 over 3 months) provided room to better define a target audience.
3	10-24 July 2016	Location: Australia; Interests: Type 2 diabetes mellitus awareness, T2D awareness, or diabetes awareness	$96	6510	254	3.90	Pilot testing new target audience.
4	1 August-30 September 2016	As per campaign 3	$904	29,912	2674	8.94	Increased budget and longer time frame allowed for monitoring and better analysis of which advertisements were performing the best during the campaign. This allowed us to pause posts with lower click-through rates and increased budget to higher performing advertisements, resulting in a proportional increase in daily click-throughs over time (from <20 per day early in the campaign to >100 in the second half).
5	23 October-22 November 2016	As per campaign 3	$500	20,764	1761	8.40	Nothing was changed from the previous campaign except that only the best performing posts were used. This infers that a working formula for recruitment for SpringboarD on Facebook had been achieved by this stage; continued through subsequent campaigns.
6a	12-30 April 2017	As per campaign 3	$440	39,868	2453	6.15	—
6b	12-19 April 2017	Location: High population areas of T2D in Sydney	$60	2233	103	4.61	Ceased owing to slow performance (ie, lower click-through rate) compared with concurrent campaign 6a
7	8 May-30 June 2017	As per campaign 3	$1200	87,981	5727	6.51	—
8	12-31 July 2017	As per campaign 3	$982	74,724	3157	3.66	—
9	1 August-30 August 2017	As per campaign 3	$4463	222,495	8314	3.28	—
10	1 September-30 September 2017	As per campaign 3	$5999	59,442	3486	4.98	—

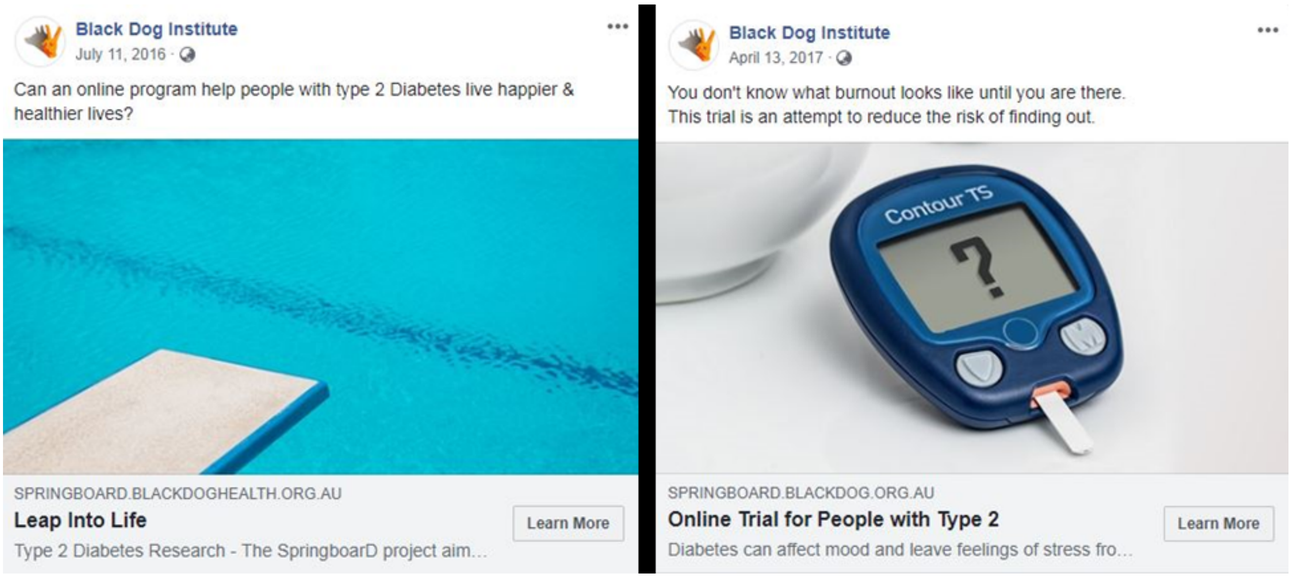
[a] Reach: the number of unique users who had a SpringboarD advertisement appear on their screen.

[b] Click-through rate: link clicks/reach.

[c] T2D: type 2 diabetes.

[d] No additional comments provided.

Figure 3. Examples of Facebook advertisements (Left: campaign 2; Right: campaign 4).



Black Dog Institute
July 11, 2016

Can an online program help people with type 2 Diabetes live happier & healthier lives?

SPRINGBOARD.BLACKDOGHEALTH.ORG.AU
Leap Into Life
Type 2 Diabetes Research - The SpringboarD project aim...

Black Dog Institute
April 13, 2017

You don't know what burnout looks like until you are there.
This trial is an attempt to reduce the risk of finding out.

SPRINGBOARD.BLACKDOG.ORG.AU
Online Trial for People with Type 2
Diabetes can affect mood and leave feelings of stress fro...

Table 2. Recruitment campaigns through existing registries.

Registry name	Number of invitations sent	Date
Black Dog Institute Volunteer Research Registry	2378	November 2016
diamond[a] cohort	75	May 2017
45 and Up cohort	4175	June 2017-September 2017
Members of diabetes organizations	25,550	April-June 2016

[a] Study name italicized as per [21].

Data Collection

The data presented here were collected through the SpringboarD website as part of study enrollment procedures. Demographic information included age, gender, highest level of education, relationship status, and current employment.

Clinical Characteristics

Regardless of the referral source, all participants completed the same baseline questionnaire that assessed mental health and diabetes management. A series of forced choice items (yes/no) asked whether participants had ever sought professional help for their emotional health, had done so in the past 6 weeks, had ever been diagnosed with a mental illness, were currently taking medication for mental illness, had visited a GP in the past 6 weeks for diabetes, had been in hospital in the past 6 weeks for diabetes, or had been referred for a blood test (for glycated hemoglobin [HbA_{1c}]) in the past 3 months. Participants were also asked to report the results of their most recent HbA_{1c} blood test, if known.

A number of validated measures in the screening and baseline questionnaires assessed mental health symptom severity and functioning. These included the Patient Health Questionnaire 9-item version (PHQ-9) [23] that assesses depressive symptom severity; the Generalized Anxiety Disorder 7-item scale (GAD-7) [24], a measure of anxiety symptom severity; the Work and Social Adjustment Scale (WSAS) [25] to assess the impact of current symptoms on functioning; and the Diabetes Distress Scale (DDS) [26] to assess distress attributed to diabetes. In all 4 measures, higher scores indicate poorer health and functioning.

Analysis

All analyses were conducted using SPSS version 24 [27]. Information on the recruitment source was collapsed into 3 categories: (1) general practice; (2) online (social media, search, and link from another website); and (3) other (research registries, diabetes forums and exhibitions, diabetes organizations, and word of mouth). We grouped word of mouth together with research registries and diabetes organizations as we considered all 3 to be reaching potentially *warm* populations, likely to have a particular interest in the research.

Average scores and prevalence of participant characteristics were compared across the 3 groups using analysis of variance (ANOVA) for continuous data, with effect sizes calculated using eta squared (η^2). Preliminary analyses revealed that variance in age was nonhomogenous across groups, and we therefore employed Welch's ANOVA for this variable [28], with post hoc differences examined using the Games-Howell test. For all other continuous variables, differences between groups were compared using standard ANOVA and post hoc differences were examined using Tukey HSD (which, in SPSS, automatically implements the Tukey-Kramer modification to account for unequal group sizes). For categorical data, we employed Fisher exact test owing to its robustness to small sample sizes; where overall results indicated a significant difference, pairwise z-tests were conducted to identify where differences lay. Effect sizes for categorical analyses are reported using Cramer V. For all analyses, Bonferroni corrections were applied for multiple testing, and missing data were handled using a pairwise deletion strategy to maximize the number of participants retained.

Results

Between November 2015 and October 2017, a total of 888 people met the eligibility criteria. Of these, 780 completed the baseline questionnaire and enrolled in SpringboarD. In total, 40 participants did not provide information on the recruitment source and were excluded from further analysis; hence, the sample reported here comprises 740 participants. Figure 4 provides an overview of the recruitment of these participants over time.

Outcomes of General Practice Recruitment

In total, our general practice recruitment efforts resulted in 27 practices consenting to take part in the trial. Despite our intensive approach to engaging these sites in patient recruitment, only 24 trial participants indicated they had learned about SpringboarD through their GP (0.53% of those who were sent an invitation letter from their GP). Most enrollments occurred soon after the initial invitation letter was mailed, and only 3 GP enrollments were recorded after April 2016.

Interviews with nominated SpringboarD coordinators identified several barriers to inviting patients to the trial, including the following:

1. Perception by some practice staff of a mismatch between trial participation requirements and the diabetes population. Some practice staff reported that because many of their patients with diabetes were older and therefore possibly less computer literate, they chose not to mention the study to them. In rural areas, some of our coordinators considered low socioeconomic status and internet or mobile phone access to be additional barriers to participation in the trial.
2. Practice nurses reported that because acute health issues must take priority in consultations, they often felt it impractical or inappropriate to bring up research (even when it was relevant to the patient's presentation).
3. Cycle of care reminders were not used in some practices, meaning that the SpringboarD prompts offered by researchers to include in text, mail, or phone contacts were not taken up.
4. Some SpringboarD coordinators reported that patients had previously been unhappy with the practice sending third party emails, which made several practices reluctant to use the Web portal to send automated emails despite this being done in a GP or practice nurse consultation, with patient consent.
5. Community or practice events occasionally made it difficult for staff to focus on SpringboarD. This included routine events, such as staff leave, and unexpected events, for example, a dairy crisis in a region of Victoria affecting the entire community, including the local practice's day-to-day operations.

Outcomes of Online Recruitment

Online recruitment was relatively slow through our early campaigns, as we tested different target audiences and nomenclature. In campaign 4, we identified that promoting the study as one trying to reduce *burnout related to diabetes* attracted most Facebook users and continued this approach through subsequent campaigns. By the end of recruitment, Facebook accounted for the majority of participants (462/740; 62.4%). Also included in the group of participants who found out about SpringboarD online were 58 individuals (7.8% of the total sample [58/740]) who came to the trial through an internet search or link from another website.

Outcomes of Other Recruitment Strategies

Recruitment through other sources yielded just over a quarter of the sample overall (196/740; 26.5%), with marked variation in effectiveness and efficiency. Sending invitations to members of existing research registries was most successful, accounting for 75.5% of participants in the *other* recruitment category (148/196; 20.0% of the overall sample). On the contrary, word of mouth generated very few participants (12/740; 1.6%).

Table 3 shows the number of participants indicating each recruitment source and the estimated number of people reached by each recruitment campaign. Note that higher numbers of recruited participants are associated with a larger population of people who received a study invitation or were targeted by an online advertisement. Overall, the conversion rate from potential exposure to promotional materials to trial enrollment was less than 1%, ranging from a low of 0.09% (520/573,333) from online recruitment and 0.61% (196/32,178) from other sources. These figures should be interpreted in light of the fact that online figures represent the total reach across multiple Facebook campaigns and individuals may be counted more than once. However, other potential denominators (such as the number of link clicks) do not translate across the recruitment source, and thus, we considered the total reach of each recruitment strategy to be the most appropriate point of comparison.

Figure 4. Cumulative enrolment in SpringboarD by recruitment source. GP: general physician.

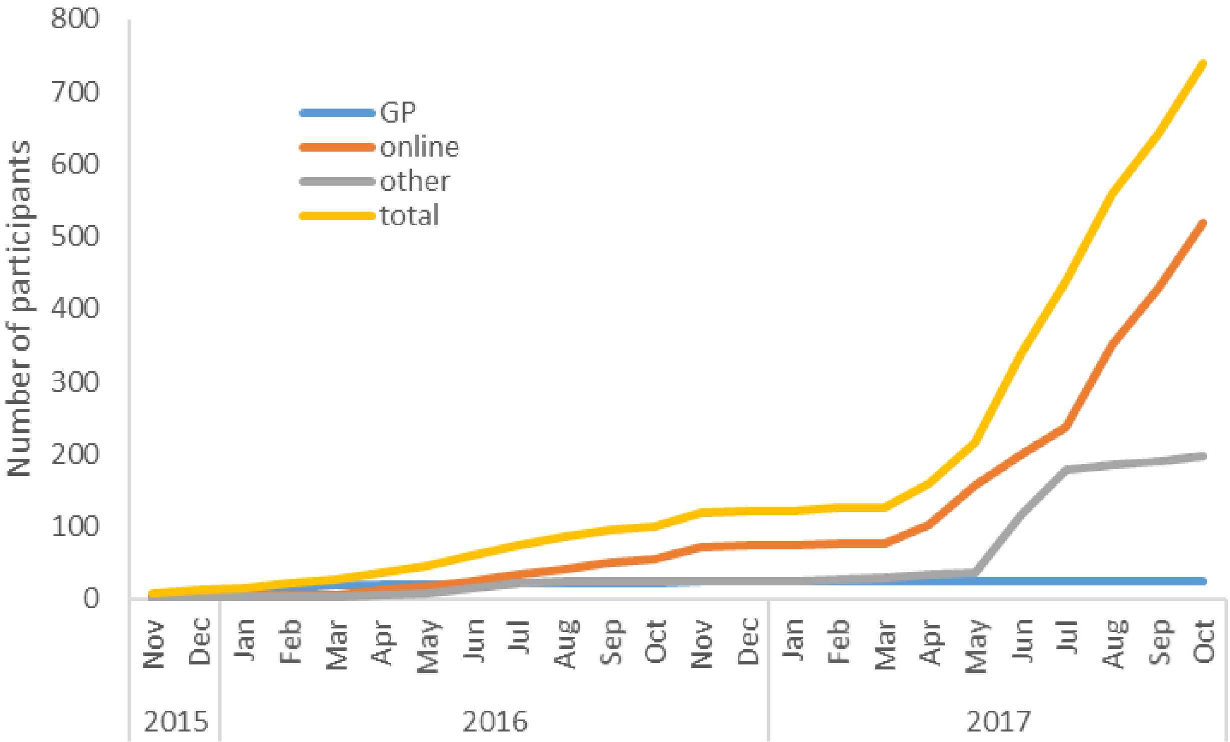


Table 3. Number of participants recruited by source.

Recruitment source (general and specific)	Participants recruited[a], n (%)	Estimated reach[b]	Participants as percentage of estimated reach
Online	**520 (70.3)**	**573,733**	**0.09**
Facebook	462 (62.4)	573,733[c]	0.08
Internet search/link from other website	58 (7.8)	—[c]	—
Other	**196 (26.5)**	**32,128**	**0.61**
Research registries	148 (20.0)	6628	2.23
Diabetes organization	36 (4.9)	25,550	0.14
Word of mouth	12 (1.6)	—	—
General practice	24 (3.2)	4569	0.53
Total	740 (100.0)	610,480	0.12

[a] Excludes 40 participants who did not indicate a recruitment source.

[b] Note that individuals may be exposed to multiple recruitment campaigns, both within and across sources.

[c] Unknown.

Characteristics of Participants by Recruitment Source

A comparison of participant characteristics at baseline suggested that our 3 different approaches to participant recruitment yielded groups that were broadly similar. Overall differences were identified on only 5 of the 15 variables considered, including 3 demographic and 2 clinical characteristics, as described below.

In terms of demographic variables, overall group differences were evident in gender, age, and employment. Pairwise testing revealed that a significantly lower proportion of participants recruited online were male compared with both the GP and other groups that did not differ from each other (Table 4). Looking at age, post hoc testing indicated that the overall significant but weak effect was because of online participants being significantly younger than those recruited through either general practice or other sources (Table 5). Potentially as an artifact of this age difference, participants recruited online were also more likely to be employed than those recruited through other sources, with no difference in employment rates between the online and GP groups or the GP and other groups. There was also a nonsignificant trend toward educational differences between groups, driven by a tendency for participants recruited online to be less likely to have a university degree than those recruited through other sources.

Table 4. Baseline characteristics of participants recruited through general practice, online, and other sources (categorical).

Characteristics	General practice (N=24), n (%)	Online (N=520), n (%)	Other (N=196), n (%)	*P* value	Effect size (Cramer V)
Gender (male)	16 (66)	141 (27.1)	92 (46.9)	.00[a,b]	0.224
Relationship status				**.23**	**0.070**
Married/de facto	16 (66)	332 (63.8)	139 (70.9)	—[c]	—
Separated/divorced	4 (16)	85 (16.3)	30 (15.3)	—	—
Single	3 (12)	83 (16.0)	17 (8.7)	—	—
Widowed	1 (4)	20 (3.8)	10 (5.1)	—	—
Employed (vs not)	9 (37)	283 (54.4)	67 (34.2)	<.001[b]	0.182
Highest level of education				**.006**	**0.108**
High school	10 (41)	170 (32.7)	44 (22.4)	—	—
University	8 (33)	146 (28.1)	82 (41.8)	—	—
Apprentice/trade	2 (8)	34 (6.5)	15 (7.7)	—	—
Diploma/other	4 (16)	170 (32.7)	55 (28.1)	—	—
Mental health care					
Sought professional help for emotional health in the past 6 weeks	4 (16)	77 (14.8)	33 (16.8)	.78	0.025
Current medication for mental illness	5 (20)	178 (34.2)	60 (30.6)	.29	0.060
Diabetes care					
Visited GP[d] in the past 6 weeks for diabetes	14 (58)	299 (57.5)	120 (61.2)	.66	0.033
Been to the hospital in the past 6 weeks for diabetes	1 (4)	23 (4.4)	6 (3.1)	.65	0.030
Had HbA_{1c}[e] test in the past 3 months	18 (75)	394 (75.8)	164 (83.7)	.04	0.075

[a] Significant difference between general practice and online group (Bonferroni-corrected pairwise comparison).

[b] Significant difference between online and other group (Bonferroni-corrected pairwise comparison).

[c] Posthoc analysis not conducted due to nonsignificant main effect.

[d] GP: general practice.

[e] HbA_{1c}: glycated hemoglobin.

Looking at clinical characteristics, no differences between groups were found in any variables related to mental health care or diabetes management (Table 4). Participants recruited online reported higher PHQ-9 and DDS scores (Table 5) and showed a trend toward higher GAD-7 scores than those in the *other* group, although these differences should be interpreted in light of several caveats. First, in all cases, the effect size was small. Second, a cutoff of 10 is typically used on the PHQ-9 to delineate clinically significant depressive symptoms [23] and both the online and other groups were just above this cutoff, whereas the GP group was below it. Finally, all groups fell below the cutoff of 3 on the DDS, indicating mild distress not requiring clinical attention. Therefore, differences between the online and other groups in these measures may be statistically but not clinically significant.

Table 5. Baseline characteristics of participants recruited through general practice, online, and other sources (continuous).

Characteristics	General practice (N=24), mean (SD)	Online (n=520), mean (SD)	Other (n=196), mean (SD)	Significance		Effect size (eta-squared [η²])
				F test value (*df*)	*P* value	
Age	62.99 (8.58)	55.32 (10.41)	63.07 (7.90)	59.49 (2, 734)	<.001[a,b]	0.012
PHQ-9[c]	8.71 (4.51)	11.51 (3.94)	10.08 (4.02)	13.46 (2, 737)	<.001[b]	0.035
GAD-7[d]	6.33 (4.78)	7.79 (4.19)	6.68 (9.98)	5.97 (2, 737)	.003	0.016
WSAS[e]	10.42 (7.28)	13.29 (7.68)	12.79 (8.82)	1.64 (2, 737)	.20	0.004
DDS[f]	2.03 (0.85)	2.71 (0.95)	2.24 (0.88)	22.04 (2, 737)	<.001[b]	0.056
HbA_{1c}[g]	7.00 (1.04)	8.11 (5.82)	7.31 (4.18)	0.76 (2, 289)	.469	0.005

[a] Significant difference between general practice and online group (Bonferroni-corrected pairwise comparison).

[b] Significant difference between online and other group (Bonferroni-corrected pairwise comparison).

[c] PHQ-9: Patient Health Questionnaire 9-item.

[d] GAD-7: Generalized Anxiety Disorder 7-item.

[e] WSAS: Work and Social Adjustment Scale.

[f] DDS: Diabetes Distress Scale.

[g] HbA_{1c}: glycated hemoglobin.

Discussion

Despite considerable time, labor, and financial investment, as well as employing many recruitment strategies previously reported to improve the likelihood of success, we were unable to recruit successfully through general practice to a trial of a Web-based mental health intervention for people with T2D and mild-to-moderate depression. The characteristics of the small group of participants who were recruited to the trial via general practice were broadly comparable with the larger sample, suggesting that difficulties in primary care recruitment are not a significant threat to internal validity. However, as is common in research, across all recruitment sources, only a small proportion of people who were exposed to SpringboarD recruitment materials accepted the invitation to take part—and thus the broader challenge of research generalizability remains.

Comparison With Previous Research

Our findings that online recruitment yielded the greatest number of participants are consistent with some previous research showing this approach to be more time- and/or cost-effective than more traditional methods of RCT recruitment [10,29]. However, other authors have reported comparable or even lower efficiency of online recruitment compared with other sources [9,30,31]. Similarly, the evidence as to how online recruitment affects sample characteristics remains limited and equivocal. Published reviews comparing the characteristics of participants recruited via social media and other sources found no consistent trend in the effect of recruitment source on age, with social media variously associated with younger, older, or same aged samples [31,32]. More consistent was the tendency to find no difference in gender, level of education, and relevant clinical characteristics, although exceptions in both directions have also been identified [30-36]. Thus, our finding that participants recruited online are slightly more likely to be younger, female, employed, and more distressed supports some, but not all, previous reports. The jury appears to still be out when it comes to identifying meaningful patterns in the effect of recruitment source and sample characteristics. It is worth noting that much of our understanding of this relationship relates to cohort studies rather than RCTs and to studies of smoking cessation more so than other health behaviors [30,31]. This study therefore builds the evidence related to randomized trial recruitment while extending previous research to a new population (people with depressive symptoms and diabetes).

The Challenge of General Practice Recruitment: There Is No Right Way

General practice recruitment proved difficult despite an intensive effort, including involvement of GPs early in the trial design. This may reflect the fact that eMH has not yet been normalized into everyday practice, despite significant investment in raising GP awareness of these interventions (including nation-wide training for GPs—eMHPrac—funded by the Australian Government Department of Health and development of guidelines on when, how, and why to refer patients to these programs [37]). Alternatively, efforts to engage GPs in eMH may not have translated to engaging other key primary health professionals, such as diabetes educators. The SpringboarD trial was not designed to explore attitudes toward eMH among GPs and other primary care providers, and as far as we are aware, this topic has not been investigated previously. It therefore represents an important avenue for future enquiry.

Of course, our limited success in recruiting through general practice may not have been only due to barriers at the provider level, with patient response to the initial mailout also limited. This echoes our experience with people with type I diabetes, suggesting that diabetes itself may pose a barrier to research participation for some people. Participants in a study by Clarke et al [38] expressed limited awareness of the link between mental health and diabetes and spoke about the *double stigma* of chronic mental and physical comorbidity. This suggests that providing further education around the bidirectional relationship between mental and physical health may have benefits in terms of both patient outcomes and the future of primary care research for people with T2D.

There is limited evidence for how to optimize recruitment in primary care, but the general agreement is that a multifaceted approach is best [7]. However, identifying the most successful strategies is more complicated than it might appear. As noted by Graffy et al [39], both trials that successfully recruit, and those that do not, often report using many of the same recruitment strategies. Our SpringboarD experience certainly attests to this. Our first approach to recruitment through general practice was to send invitation letters to all patients with T2D aged between 18 and 75 years. We have previously used this approach with success in both randomized trials [40] and cohort studies [21], where 28.57% (5742/20,100) and 43.12% (7667/17,780) of patients, respectively, responded to a mailed invitation letter. Similarly, our recruitment approach shared many similarities with Reed et al’s [5] successful recruitment of older adults to a trial of chronic disease self-management, where 38.12% (634/1663) of patients who were mailed a letter of invitation expressed interest in participating in the trial. In the current trial, however, the number of participants who indicated they found out about SpringboarD through their general practice was less than 1% of the number of invitations mailed.

The general consensus is that primary care recruitment is most successful when patient eligibility criteria are simple and practice staff are not expected to spend time managing patient consent [6,7]. We designed our recruitment approach with this in mind and asked practice staff only to invite patients with T2D to visit the study website; all eligibility screening was completed online. To further reduce the burden on practice staff, we offered practices the option of having a research assistant located in the practice to invite patients directly, an approach we have previously found effective [41]. Our experience suggests that although this approach can help to achieve sample size targets, practices can feel uncomfortable with a shift to researcher-led recruitment midtrial if this is not initially agreed to.

Who Takes Part in an e-Mental Health Trial?

SpringboarD participants from all recruitment sources were generally similar to each other. This is perhaps not surprising when we consider that the majority of Australians can be categorized as both GP patients and Facebook users; the two are far from mutually exclusive and it is possible that internet searchers were looking for SpringboarD after hearing about it through our active recruitment strategies. Just as eMH interventions are designed to supplement traditional mental health care, participants recruited online should not be considered to exist outside the health care system. Over 80% of Australians visit a GP once per year, and this number rises to 95% among people with a long-term health condition [42]. Thus, Facebook users with T2D may be considered a subset of the general practice population, who happen to enter trials such as SpringboarD via a different path.

One key difference between our recruitment groups was that a greater proportion of the participants recruited online were female. This gender difference is consistent with findings that women are more likely to engage in help-seeking behaviors, generally, and searching for health information online, specifically [43]; they may have been more likely to find and follow the SpringboarD invitation link. The population prevalence of diabetes is similar in men and women overall (6% and 5%, respectively), but among older age groups (55 and older), the condition is more common in men [44]. One of the often-cited benefits of online recruitment is that it can reach populations that are traditionally hard to engage. When it comes to mental health research, men are one such population, but in this study, online recruitment did little to overcome this gap. The significance of the potential to reach men through GP or other avenues of recruitment should not be dismissed. Particularly where the research question is especially important to men’s health, the investment in time and resources required to recruit through primary care may be justified, if, of course, recruitment targets can be achieved within the study timeline and funding.

That different recruitment pathways yielded mostly similar groups provides reassurance that the effect of recruitment source on sample characteristics is negligible. However, questions about selection effects remain. The individuals who enrolled in SpringboarD and are reported here represent less than 1% of people who were potentially exposed to one or more invitations or advertisements. Our findings support those of previous authors, who reported similar biases in both online and offline recruitment methods [36], and similar characteristics of participants who enroll in trials of eMH and traditional psychological and pharmacological treatments [45]. In other words, people who take part in mental health research differ from people who do not, regardless of the recruitment approach or intervention being tested. Thus, although recruiting online may be a solution to the (significant) challenge of obtaining an adequate sample size, it does not address the broader issue of overcoming selection bias in community-based recruitment. As the saying goes, everything old is new again, and in the new era of Web-based research, we continue to face the old problem of representativeness.

Limitations

The findings reported above should be considered in light of certain limitations. First, it should be noted that information on the recruitment source was only collected for patients who were eligible for the trial. In total, 6145 people visited the SpringboarD landing page and 2849 of those completed the screening questionnaire but were ineligible; whether these individuals were spread evenly across recruitment sources is unclear. In addition, we did not collect information from participating general practices about all patients who were mailed an invitation letter and it is therefore unclear how participants and nonparticipants compare. Also note that although we base our uptake calculations on the estimated *reach* of each campaign, we have no evidence that Facebook users actually saw the advertisement or that invitation letters reached or were opened by GP patients and research registry participants. Our calculations therefore potentially overestimate the number of potential participants and therefore underestimate the proportion of those who enrolled in the trial. It is also possible that this over or underestimate differs by recruitment source;

potential participants may be more likely to open a letter from their GP than notice an advertisement on Facebook. Finally, this study relates to recruitment only; retention is a further significant challenge in all trials (particularly those conducted online). The interested reader is referred to Clarke et al [14] for further information on participant retention in the SpringboarD RCT.

Conclusions

Consistent with previous research, this study showed that recruiting to an RCT through general practice was difficult and ultimately unsuccessful. Our findings support previous recommendations that working with practice staff to identify the best approach early on may improve the likelihood of success. Furthermore, ensuring a match between recruitment strategies and the nature of the trial or intervention being tested may be beneficial. Nonetheless, trial participants recruited via general practice were not markedly different from those recruited online or through other means. This suggests that regardless of how people came into the trial, recruitment materials attracted those with similar characteristics. Thus, the threat of different recruitment sources to internal validity appears minimal. GPs can be reassured that current evidence for the effectiveness of eMH interventions, frequently generated from samples recruited online, is likely similar to that obtained if the sample was obtained entirely through general practice.

Acknowledgments

The SpringboarD RCT was funded by the National Health and Medical Research Council (Grant ID 1083116). The authors gratefully acknowledge the dedication and input of all participating general practices and trial participants. They also thank Ms Amy Coe for her contribution to early versions of this paper and Mr Dean Winder for his assistance in analyzing and interpreting Facebook advertising campaign data.

Conflicts of Interest

None declared.

References

1. Burgess PM, Pirkis JE, Slade TN, Johnston AK, Meadows GN, Gunn JM. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2009 Jul;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
2. 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]
3. Granja C, Janssen W, Johansen M. 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]
4. [No authors listed]. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2016 Jan 12;352:i195 [FREE Full text] [doi: 10.1136/bmj.i195] [Medline: 26759375]
5. Reed RL, Barton CA, Isherwood LM, Baxter JM, Roeger L. Recruitment for a clinical trial of chronic disease self-management for older adults with multimorbidity: a successful approach within general practice. BMC Fam Pract 2013 Aug 28;14(1):125. [doi: 10.1186/1471-2296-14-125]
6. Bower P, Wilson S, Mathers N. Short report: how often do UK primary care trials face recruitment delays? Fam Pract 2007 Dec;24(6):601-603. [doi: 10.1093/fampra/cmm051] [Medline: 17872907]
7. Ngune I, Jiwa M, Dadich A, Lotriet J, Sriram D. Effective recruitment strategies in primary care research: a systematic review. Qual Prim Care 2012;20(2):115-123. [Medline: 22824564]
8. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015 Jul 22;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
9. Juraschek S, Plante T, Charleston J, Miller E, Yeh H, Appel L, et al. Use of online recruitment strategies in a randomized trial of cancer survivors. Clin Trials 2018 Dec;15(2):130-138 [FREE Full text] [doi: 10.1177/1740774517745829] [Medline: 29361843]
10. Kayrouz R, Dear B, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. Internet Interv 2016 May;4:1-10 [FREE Full text] [doi: 10.1016/j.invent.2016.01.001] [Medline: 30135786]
11. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009 Apr;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
12. Mohr D, Weingardt K, Reddy M, Schueller S. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
13. 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]

14. Clarke J, Sanatkar S, Baldwin P, Fletcher S, Gunn J, 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): results of a randomised controlled trial. J Med Internet Res (forthcoming) 2019:e (forthcoming).
15. Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and publication frequency. Aust Fam Physician 2013;42(1-2):65-68 [FREE Full text] [Medline: 23529466]
16. Roy T, Lloyd CE. Epidemiology of depression and diabetes: a systematic review. J Affect Disord 2012 Oct;142 Suppl:S8-21. [doi: 10.1016/S0165-0327(12)70004-6] [Medline: 23062861]
17. Bower P, Wallace P, Ward E, Graffy J, Miller J, Delaney B, et al. Improving recruitment to health research in primary care. Fam Pract 2009 Oct;26(5):391-397. [doi: 10.1093/fampra/cmp037] [Medline: 19549623]
18. Social Media News. 2018. Social media statistics Australia URL: https://www.socialmedianews.com.au/ [accessed 2019-04-04] [WebCite Cache ID 77Nk9UhJ0]
19. Tarzia L, Valpied J, Koziol-McLain J, Glass N, Hegarty K. Methodological and ethical challenges in a web-based randomized controlled trial of a domestic violence intervention. J Med Internet Res 2017 Dec 28;19(3):e94 [FREE Full text] [doi: 10.2196/jmir.7039] [Medline: 28351830]
20. Irvine M. Word Stream. 2019. Facebook ad benchmarks for YOUR industry URL: https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks [accessed 2019-04-04] [WebCite Cache ID 77NLLcv03]
21. Gunn J, Gilchrist GP, Chondros P, Ramp M, Hegarty KL, Blashki GA, et al. Who is identified when screening for depression is undertaken in general practice? Baseline findings from the Diagnosis, Management and Outcomes of Depression in Primary Care (diamond) longitudinal study. Med J Aust 2008 Jun 16;188(12 Suppl):S119-S125. [Medline: 18558911]
22. 45Up Study Collaborators, Banks E, Redman S, Jorm L, Armstrong B, Bauman A, et al. Cohort profile: the 45 and up study. Int J Epidemiol 2008 Oct;37(5):941-947 [FREE Full text] [doi: 10.1093/ije/dym184] [Medline: 17881411]
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] [Medline: 11556941]
24. 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]
25. Mundt J, Marks I, Shear M, Greist J. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180:461-464. [Medline: 11983645]
26. Polonsky W, Anderson B, Lohrer P, Welch G, Jacobson A, Aponte J, et al. Assessment of diabetes-related distress. Diabetes Care 1995 Jun;18(6):754-760. [Medline: 7555499]
27. IBM. Armonk, NY: IBM Corp; 2016. IBM SPSS Statistics for Windows, Version 24 URL: https://www.ibm.com/ [accessed 2019-04-04] [WebCite Cache ID 77NlMJ6ik]
28. Shieh G, Jan SL. Optimal sample size allocation for Welch's test in one-way heteroscedastic ANOVA. Behav Res Methods 2015 Jun;47(2):374-383. [doi: 10.3758/s13428-014-0477-8] [Medline: 24903689]
29. Koziol-McLain J, McLean C, Rohan M, Sisk R, Dobbs T, Nada-Raja S, et al. Participant recruitment and engagement in automated eHealth trial registration: challenges and opportunities for recruiting women who experience violence. J Med Internet Res 2016 Dec 25;18(10):e281 [FREE Full text] [doi: 10.2196/jmir.6515] [Medline: 27780796]
30. Watson N, Mull K, Heffner J, McClure J, Bricker J. 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]
31. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Dec 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
32. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81. [doi: 10.1016/j.invent.2016.02.001]
33. Vrangalova Z, Savin-Williams RC. Mostly heterosexual and mostly gay/lesbian: evidence for new sexual orientation identities. Arch Sex Behav 2012 Feb;41(1):85-101. [doi: 10.1007/s10508-012-9921-y] [Medline: 22327566]
34. Carlini BH, Safioti L, Rue TC, Miles L. Using internet to recruit immigrants with language and culture barriers for tobacco and alcohol use screening: a study among Brazilians. J Immigr Minor Health 2015 Apr;17(2):553-560. [doi: 10.1007/s10903-013-9934-1] [Medline: 24563138]
35. Carter-Harris L, Bartlett Ellis R, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. J Med Internet Res 2016 Dec 15;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
36. Batterham PJ. Recruitment of mental health survey participants using internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191. [doi: 10.1002/mpr.1421] [Medline: 24615785]
37. Royal Australian College of General Practitioners. e-Mental health: A guide for GPs. Melbourne: RACGP; 2015. URL: https://www.racgp.org.au/FSDEDEV/media/documents/Clinical%20Resources/Guidelines/Mental%20health/e-mentalhealth-guide.pdf [accessed 2019-04-04] [WebCite Cache ID 77NlwVDFK]

38. Clarke J, Proudfoot J, Vatiliotis V, Verge C, Holmes-Walker D, Campbell L, et al. Attitudes towards mental health, mental health research and digital interventions by young adults with type 1 diabetes: A qualitative analysis. Health Expect 2018 Dec;21(3):668-677 [FREE Full text] [doi: 10.1111/hex.12662] [Medline: 29319923]
39. Graffy J, Bower P, Ward E, Wallace P, Delaney B, Kinmonth AL, et al. Trials within trials? Researcher, funder and ethical perspectives on the practicality and acceptability of nesting trials of recruitment methods in existing primary care trials. BMC Med Res Methodol 2010 Apr 30;10(1):38 [FREE Full text] [doi: 10.1186/1471-2288-10-38] [Medline: 20433728]
40. Hegarty K, O'Doherty L, Taft A, Chondros P, Brown S, Valpied J, et al. Screening and counselling in the primary care setting for women who have experienced intimate partner violence (WEAVE): a cluster randomised controlled trial. Lancet 2013 Jul 20;382(9888):249-258 [FREE Full text] [doi: 10.1016/S0140-6736(13)60052-5] [Medline: 23598181]
41. Gunn J, Wachtler C, Fletcher S, Davidson S, Mihalopoulos C, Palmer V, et al. Target-D: a stratified individually randomized controlled trial of the diamond clinical prediction tool to triage and target treatment for depressive symptoms in general practice: study protocol for a randomized controlled trial. Trials 2017 Dec 20;18(1):342 [FREE Full text] [doi: 10.1186/s13063-017-2089-y] [Medline: 28728604]
42. Australian Bureau of Statistics. Canberra: ABS; 2016. Patient experiences in Australia: Summary of findings, 2016-17 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4839.0Main+Features12016-17 [accessed 2019-04-04] [WebCite Cache ID 77NWEbhNJ]
43. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
44. Australian Institute of Health and Welfare.: AIHW; 2018. Diabetes compendium URL: https://www.aihw.gov.au/reports/diabetes/diabetes-snapshot/contents/how-many-australians-have-diabetes [accessed 2019-04-04] [WebCite Cache ID 77NWj1CPM]
45. Lorenzo-Luaces L, Johns E, Keefe J. The generalizability of randomized controlled trials of self-guided internet-based cognitive behavioral therapy for depressive symptoms: systematic review and meta-regression analysis. J Med Internet Res 2018 Nov 09;20(11):e10113 [FREE Full text] [doi: 10.2196/10113] [Medline: 30413400]

Abbreviations

ANOVA: analysis of variance
DDS: Diabetes Distress Scale
eMH: e-mental health
GAD: Generalized Anxiety Disorder
GPs: general practitioners
HbA $_{1c}$: glycated hemoglobin
HREC: Human Research Ethics Committee
PHQ: Patient Health Questionnaire
RCT: randomized controlled trial
T2D: type 2 diabetes
WDD: World Diabetes Day
WSAS: Work and Social Adjustment Scale

Edited by G Eysenbach; submitted 09.12.18; peer-reviewed by A Nomura, L Zhou, K Reuter; comments to author 31.01.19; revised version received 18.03.19; accepted 24.03.19; published 24.05.19.

Please cite as:
Fletcher S, Clarke J, Sanatkar S, Baldwin P, Gunn J, Zwar N, Campbell L, Wilhelm K, Harris M, Lapsley H, Hadzi-Pavlovic D, Proudfoot J
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;21(5):e12793
URL: https://www.jmir.org/2019/5/e12793/
doi:10.2196/12793
PMID:31127718

Review

Consumer Evaluation of the Quality of Online Health Information: Systematic Literature Review of Relevant Criteria and Indicators

Yalin Sun[1], MS; Yan Zhang[1], PhD; Jacek Gwizdka[1], PhD; Ciaran B Trace[1], PhD

School of Information, The University of Texas at Austin, Austin, TX, United States

Corresponding Author:
Yan Zhang, PhD
School of Information
The University of Texas at Austin
1616 Guadalupe Street Suite #5.202
Austin, TX, 78701
United States
Phone: 1 5124719448
Email: yanz@ischool.utexas.edu

Abstract

Background: As the quality of online health information remains questionable, there is a pressing need to understand how consumers evaluate this information. Past reviews identified content-, source-, and individual-related factors that influence consumer judgment in this area. However, systematic knowledge concerning the evaluation process, that is, why and how these factors influence the evaluation behavior, is lacking.

Objective: This review aims (1) to identify criteria (rules that reflect notions of value and worth) that consumers use to evaluate the quality of online health information and the indicators (properties of information objects to which criteria are applied to form judgments) they use to support the evaluation in order to achieve a better understanding of the process of information quality evaluation and (2) to explicate the relationship between indicators and criteria to provide clear guidelines for designers of consumer health information systems.

Methods: A systematic literature search was performed in seven digital reference databases including Medicine, Psychology, Communication, and Library and Information Science to identify empirical studies that report how consumers directly and explicitly describe their evaluation of online health information quality. Thirty-seven articles met the inclusion criteria. A qualitative content analysis was performed to identify quality evaluation criteria, indicators, and their relationships.

Results: We identified 25 criteria and 165 indicators. The most widely reported criteria used by consumers were trustworthiness, expertise, and objectivity. The indicators were related to source, content, and design. Among them, 114 were positive indicators (entailing positive quality judgments), 35 were negative indicators (entailing negative judgments), and 16 indicators had both positive and negative quality influence, depending on contextual factors (eg, source and individual differences) and criteria applied. The most widely reported indicators were site owners/sponsors; consensus among multiple sources; characteristics of writing and language; advertisements; content authorship; and interface design.

Conclusions: Consumer evaluation of online health information is a complex cost-benefit analysis process that involves the use of a wide range of criteria and a much wider range of quality indicators. There are commonalities in the use of criteria across user groups and source types, but the differences are hard to ignore. Evidently, consumers' health information evaluation can be characterized as highly subjective and contextualized, and sometimes, misinformed. These findings invite more research into how different user groups evaluate different types of online sources and a personalized approach to educate users about evaluating online health information quality.

(J Med Internet Res 2019;21(5):e12522) doi:10.2196/12522

KEYWORDS

health information quality; health information seeking; consumer health informatics; online health information

Introduction

More than 70% of US adults search online for health information [1]. The information found online shapes and influences consumers' health beliefs, intentions, health behaviors, and health care decision making [2-5]. Since the inception of the internet, the quality of health information has been a source of concern for stakeholders due to the unregulated nature of the medium [6]. This concern is furthered by the fast growth of social media and user-generated content and corroborated by more than 200 evaluation studies conducted by subject experts, which collectively suggest that the quality of consumer-oriented health information on the internet varies greatly and that the overall quality was low [7] and remains low [8].

Making decisions based on low-quality health information (eg, information that is inaccurate, incomplete, or biased) may lead to harmful consequences, such as delayed treatment or extreme anxiety [9], and subsequently increase consumer vulnerability [10,11]. Nevertheless, evaluating the quality of information has been a major challenge for online health consumers [12-14]. For example, some consumers are uncertain about the accuracy, completeness, and validity of the information they encounter [15,16]; some cannot differentiate between scientific facts, empirical factors, and personal opinions [17]; and others suffer from information overload and subsequently lack the confidence and ability to evaluate information [18-21]. Studies have found that compared to health care providers or information professionals, consumers tend to give higher quality ratings to health information from both traditional health websites [22] and social media sites [23].

The ability to critically evaluate the quality of health information is an important component of health literacy [10], which is an important determinant of health [24]. To enhance this ability (and related skills), it is necessary to understand how consumers evaluate the quality of health information in the current internet environment. Consumer evaluation is subjective, driven by one's information needs. Therefore, as a starting point, we adopted a broad conceptualization that defines quality through "fitness for use" [25]: Information is of good quality when it serves users' needs. It is worth noting that this concept of quality is described using different terms in the existing literature, including but not limited to quality, credibility, trust, reliability, believability, and usefulness. In this review, we included articles using all these terms. We chose to be inclusive, because we want to achieve a comprehensive view of the assessments that consumers perform in the process of determining whether they would be willing to use a piece of information.

Guided by this understanding of quality, three recent systematic reviews were identified as relevant to our current research: One review focused on identifying factors that impact consumer judgment of trustworthiness and credibility of online health information [11], the second one identified the antecedents of trust in health information websites [27], and the final one reviewed the association between low health literacy and perceived quality and trust in online health information and low literacy consumers' ability to evaluate information quality [10]. These reviews revealed that consumers' quality evaluation is influenced by both source- and content-related factors [10,11,27]. Examples of source-related factors are website design (eg, layout, visual design, and interactive features), loading speed, and the authority of the owner or sponsor [11,27-30]. Examples of content-related factors are the authority of the author, content readability, content organization, use of evidence and citations, and the appearance of advertisements [11,27,31-33]. Additionally, a number of individual characteristics were identified as influencers, including demographics (eg, age, gender, and educational attainment), perceived health status, knowledge about the content, health beliefs, and level of health literacy [10,11,22,27,31].

These reviews provide an informative overview of factors that influence consumer online health information evaluation behavior but shed limited light on why and how these factors influence the evaluation behavior. From the perspective of information seeking, evaluation of information is a judgment and decision-making process that precedes users' acceptance or rejection of received information [34]. Judgment and decision making involve applying certain criteria, principles, or standards to form evaluations [35]. Thus, to understand consumer quality evaluation behavior, it is necessary to understand the criteria used to guide the evaluation. Among the previously mentioned systematic reviews, only one [10] summarized the evaluation criteria reported in five studies on consumers with low health literacy. A more comprehensive understanding of the evaluation criteria is needed. This review intends to fill this gap.

Evaluation of the quality of online health information is a process of applying criteria to evaluate information. Thus, in addition to applying criteria, we need a better understanding of how consumers perceive online information. To achieve this goal, we deliberately differentiate between two concepts: criteria and indicators. Criteria are rules or filters that people apply to an information object to assess its value or worth [36]. Indicators, also termed cues or markers [37], are perceivable elements associated with an information object that allow people to reflect on the quality of the object [8]. Criteria are abstract, reflecting one's values and preferences and mediating information selection decisions. Indicators are affordances of information objects that trigger or support the application of the criteria. Criteria are comparatively stable, whereas indicators are amenable to change. New indicators could emerge, and old ones could disappear with the development of new technologies and design preferences.

In this article, we focus on the following research questions: (1) What criteria do consumers use to evaluate the quality of online health information? (2) What elements of information objects do consumers use as quality indicators? (3) Which indicators convey positive evaluations and which convey negative evaluations? (4) What is the relationship between indicators and criteria, that is, what criteria do each indicator correspond to? We argue that a more comprehensive understanding of criteria used in the evaluation process can bring some clarity to the dimensions of quality perceived by online health consumers as well as their quality evaluation process. By explicating the relationship between indicators and criteria and identifying positive and negative judgments that indicators convey, the results can also inform the design of more user-friendly health information content and information systems.

Methods

Search Strategies

Seven online databases, including PubMed, Web of Science, PsycINFO, CINAHL (Cumulative Index to Nursing and Allied Health), Cochrane Library, Library and Information Science Source, and Communication and Mass Media Complete, were searched in July 2017 to obtain relevant journal articles. These databases were chosen because they cover major academic disciplines that study consumer online health information search, including health, information and library science, psychology, and mass communication. Keywords, including *quality*, *credibility*, *trust*, *reliability*, *accuracy*, *readability*, *relevance*, and *usefulness* were used in combination with the keywords *consumer or patient* and *online health information evaluation or online health information assessment*. After the searches, we manually screened the references to identify relevant articles and further examined the reference lists of these articles. Additionally, we examined the references cited in the three systematic reviews mentioned above and articles that cited these reviews (using Google Scholar’s “cited by” function).

Inclusion/Exclusion Criteria

Articles meeting the following criteria were eligible for inclusion in this review: (1) The study primarily focused on consumer evaluation of health information on the internet. Health consumers include patients, caregivers, and the general public who sought or were interested in seeking health information. This focus differentiates this review from prior reviews of health care professionals or expert evaluation of online health information for consumers [7,8]. Articles that focus on media other than the internet (eg, TV and radio) were excluded. (2) The study was empirical and based on direct inquiries with health consumers where criteria were described by participants and not imposed by researchers. Articles that used only predefined evaluation criteria to survey consumers or analyze their responses without allowing new criteria to emerge were excluded. We also excluded correctional studies that focus on identifying factors (eg, source expertise) influencing consumer evaluation behavior but do not provide additional results on how quality evaluation is performed. (3) The article was published after 2002, when research on consumer evaluation of online health information began to emerge. (4) The article was written in the English language. (5) The article was published in a peer-reviewed journal.

Study Identification

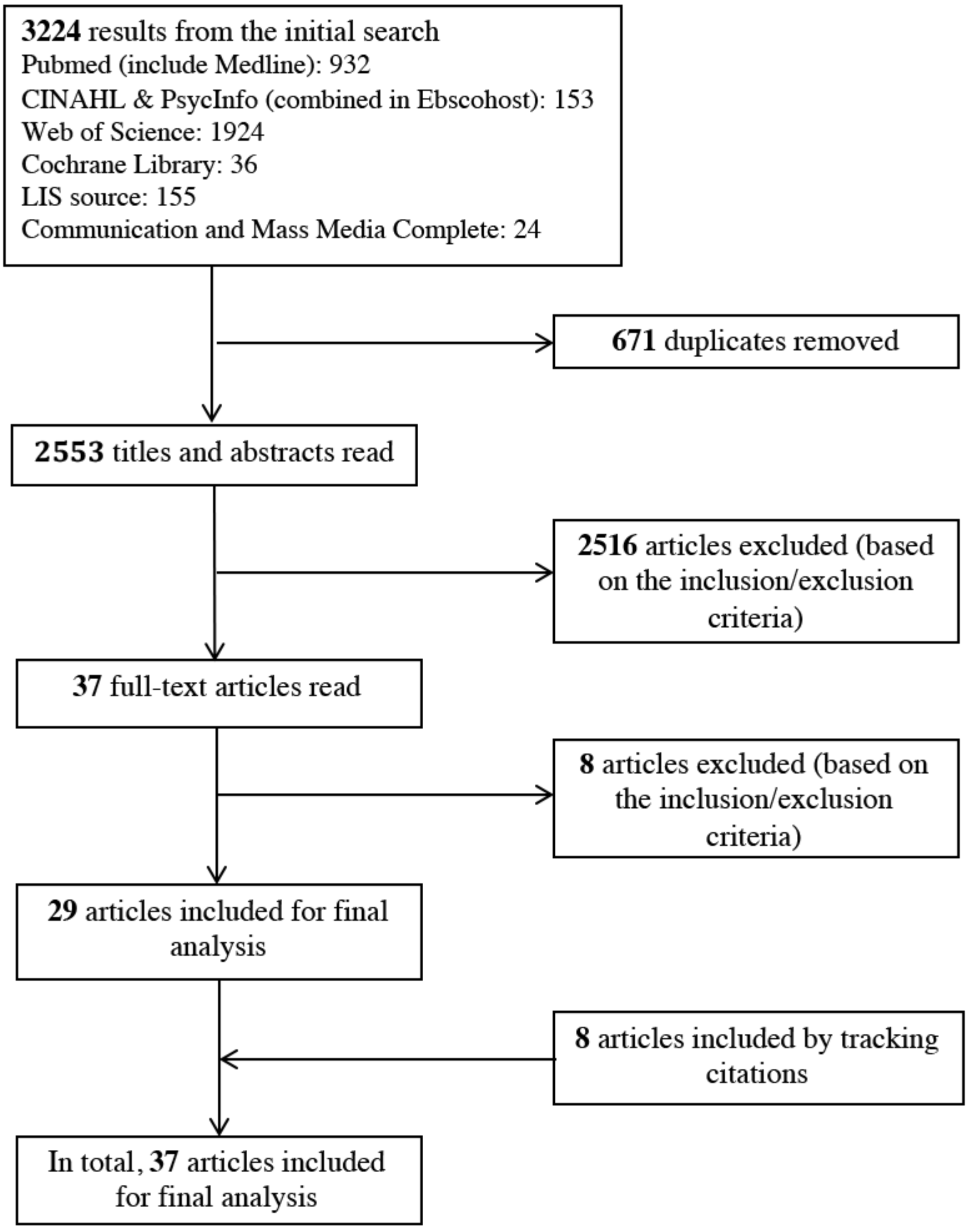
Figure 1 shows the process involved in identifying eligible studies. Three authors (YS, YZ, and JG) reviewed a subset of the search results by reading titles and abstracts. YS and YZ both reviewed 10% of the records (256 records in total) to check the intercoder agreement in filtering potentially relevant articles (Cohen kappa=0.83). Both YS and YZ screened the full-text articles. When there was uncertainty involved in excluding a full-text article, the other two authors provided their input.

Data Extraction and Analysis

Full text of the 37 selected articles was imported into MAXQDA 12 (VERBI Software GmbH, Berlin, Germany) for analysis. We extracted the following information: basic characteristics of the articles (eg, year of publication, country of origin, health topics, and aims of the study), research methods, sampling techniques, participant characteristics (eg, demographics and disease experiences), source studied (eg, the internet or specific health websites), and characteristics of the search tasks (eg, self-generated vs assigned) when search tasks were involved. Guided by their corresponding definitions, indicators and the corresponding criteria were extracted from the results and discussion reported in the original papers. When no clear relationships were reported (in most of such cases, indicators were reported without mentioning the criteria. For example, .com was reported as a negative indicator of quality, but criteria by which this judgment was reached were not reported), the authors of the review derived the relationships from the participants’ direct quotes reported in the original papers, the original authors’ discussion of the results, or the interpretation of the authors of the review. Indicators were further coded into positive (+, entailing positive quality judgment), negative (–, entailing negative judgment), or both (±, entailing both positive and negative judgments). When participants commented on the absence of an indicator (eg, no author credential or no advertisements), it was coded as positive if the absence implies low quality and as negative if the absence implies high quality. The criteria were also coded into the three categories based on their correspondence with indicators.

We analyzed the basic characteristics of the included studies using descriptive statistics. The qualitative content analysis method [26] was used to identify themes and build categories based on the extracted information concerning criteria and indicators in an iterative manner. YS coded all the articles. YZ validated the results by comparing each assigned code to the full-text of the articles. A number of group meetings were held to discuss the codes, especially relationships between indicators and criteria. Discrepancies were discussed among all authors.

Figure 1. Article screening process.



Results

Basic Characteristics of the Included Articles

The 37 articles included in the review were published between 2002 and 2017. They originated from 8 countries, primarily United Kingdom (N=12), United States (N=11), Australia (N=4), and the Netherlands (N=3). The characteristics of each included article are summarized in Table 1.

Focus groups (n=17), interviews (n=16), and observations of participants performing predefined (n=11) or self-generated (n=6) search tasks were the primary research methods used in the selected articles. Observations were often used with other methods including think aloud, guided interviews, focus groups, or diaries. Fourteen articles used multiple research methods.

Twenty-one articles focused on information on a specific health condition or issue (eg, HIV prevention, diabetes, disabilities, and chronic diseases), and the remaining articles did not specify a subject focus. Twelve studies recruited patients with a specific condition, and the others recruited people who had searched online for health information (n=6) or had a strong interest in their health or a particular condition (n=5). Twenty-eight articles involved adult participants (≥18 years old), of which 10 articles also involved older adults (>64 years old). Four studies included adolescents aged 11-17 years. The number of participants ranged from 5 to 188 (median=21). In terms of sampling technique, 26 articles used purposive sampling, five used convenience sampling, and the remaining six did not report the sampling methods.

Table 1. Characteristics of the included articles.

Articles	Health topics	Source studied	Sampling method	Participants			Data collection methods
				N	Age range (years)	Disease experience	
Eysenbach and Köhler [30]	Not specified	The internet, in general	Purposive	21	19-71 (mean=37)	Healthy volunteers who had searched online for health information	Focus groups, naturalistic observation of consumers searching predefined search tasks, and follow-up interviews
Frisby et al [39]	Smoking cessation	The internet, in general	Convenience	13	19-64	Smokers from a smoking cessation campaign	Interviews, observations of participants searching both predefined and self-selected search tasks, and think aloud
Peterson et al [40]	Medicines/drugs	The internet, in general	Purposive	46	18-67 (mean=41.7)	People who had searched online for health information	Focus groups
Williams et al [41]	Not specified	The internet, in general	Purposive	42	30-49	People who had searched online for health information	Open-question survey
Bernhardt and Felter [42]	Pre- and postnatal health	The internet, in general	Purposive	20	22-42 (mean=34.5)	Mothers of young children	Focus groups
Childs [43]	Not specified	The internet, in general	Not reported	35	Not reported	Parents and caregivers of children with rare diseases	Focus groups
Adam et al [44]	Not specified	The internet, in general	Purposive	18	20-60	People who had searched online for health information	Observation of participants searching both self-generated and predefined tasks, and semistructured qualitative interviews
Crystal and Greenberg [45]	Not specified	The internet, in general	Purposive	12	Not reported	People who have a strong interest in their health	Observation of participants searching self-generated search tasks, think aloud, and guided interviews
Kerr et al [46]	Chronic conditions (eg, Alzheimer disease)	Interactive health communication application	Purposive	40	30-79	Chronic disease patients and caregivers	Focus groups
Marshall and Williams [18]	Not specified	Preselected websites	Purposive	32	Not reported	Patients with various conditions and care givers	Information review groups
Hoffman-Goet and Friedman [47]	Breast cancer information	Preselected websites	Convenience	25	50-71 (mean=59.2)	Canadian aboriginal senior women	Interview
Sillence and Briggs [48]	Not specified	The internet, in general	Not reported	42	22-68	Internet users interested in their health	Focus groups
Sillence et al [32]	Menopause	The internet, in general, and preselected websites	Not reported	15	41-60 (mean=49)	Women faced with decisions concerning menopause and hormone replacement therapy	Observation of participants searching predefined and self-generated search tasks with think aloud and guided focus groups, and free search with diary keeping
Sillence et al [49]	Hypertension	The internet, in general, and preselected websites	Not reported	13	33-68	Hypertension patients	Observation of participants searching self-selected and predefined search tasks, with think aloud and guided focus groups, and free search with diary keeping
Buhi et al [50]	Sexual health	The internet, in general	Purposive	24	Not reported	First-year undergraduate students	Observation of participants searching predefined search tasks and think aloud

Articles	Health topics	Source studied	Sampling method	Participants			Data collection methods
				N	Age range (years)	Disease experience	
Freeman and Spyridakis [51]	Diabetes	The CDC[a] website	Convenience	188	Mean=21	University students	Controlled experiment with open-ended questions in a questionnaire
Mackert et al [52]	Childhood obesity and nutrition	The internet, in general	Purposive	43	≥18	Parents with low health literacy	Focus groups
Marton [53]	Mental health	The internet, in general	Convenience	5	Not reported	Chronic mental health patients	Interviews
Kim et al [54]	Preconception nutrition	The internet, in general	Purposive	11	20-22	University students	Observation of participants searching predefined search tasks with guided interviews
Feufel and Stahl [13]	Not specified	The internet, in general	Purposive	22	>50 or <30 (mean for older cohort=65, mean for younger cohort=23)	Older vs younger cohorts (with different health literacy skills)	Observation of participants searching predefined search tasks and concurrent talk-aloud
Henderson and Eccleston [38]	Pain problem	The internet, in general	Purposive	13	12-17 (mean=14.38)	Adolescent users of online content for pain	Online focus groups
Colombo et al [55]	Multiple sclerosis	The internet, in general	Purposive	60	18-60	Multiple sclerosis patients and their family members	Offline/online focus groups
Lederman et al [56]	Not specified	Online forums	Purposive	16	≥18	Consumers who had searched online for health information	Interviews
McPherson et al [57]	Chronic conditions	Preselected websites	Purposive	6	11-23 (mean=16.7)	Children and young people with chronic conditions	Focus groups
Payton, et al [58]	HIV prevention	The NIH[b] website	Not reported	40	18-24	Black female college students	Focus groups
Briones [59]	Not specified	The internet, in general	Purposive	50	18-25	University students	Interviews
Rennis et al [60]	Not specified	The internet, in general	Convenience	14	Mean=25.71	Urban community college students	Focus groups
Santer et al [20]	Childhood eczema	The internet, in general	Purposive	28	26-46 (median=36)	Parents of children with eczema	Interviews
Subramaniam et al [61]	Obesity and other general health issues	The internet, in general, and preselected obesity websites	Purposive	30	10-15 (mean=12.8)	Adolescents from low socioeconomic status and minority family	Participants searching self-selected health topics followed by search log analysis, and interviews of preselected websites
Cunningham and Johnson [62]	Not specified	Patients.co.uk	Not reported	11	Not reported	General public	Observation of participants searching predefined search tasks and concurrent talk-aloud
Diviani et al [63]	Not specified	The internet, in general	Purposive	44	Mean=37	Italian-speaking adults with different health literacy levels	Interviews
Sillence et al [64]	“Raw” milk	Pre-selected raw milk websites	Purposive	41	24-85 (mean=48)	Milk consumers	Observation of participants searching predefined search tasks, log analysis, and guided group discussion
Alsem et al [65]	Physical disabilities	The internet, in general	Purposive	15	26-58	Parents of children with physical disabilities	Interviews

Articles	Health topics	Source studied	Sampling method	Participants			Data collection methods
				N	Age range (years)	Disease experience	
Champlin et al [66]	Not specified	The internet, in general	Purposive	40	Mean=39	People with different health literacy levels	Interviews
Cusack et al [67]	Not specified	The internet, in general	Purposive	27	12-15	Students in grades 7-9	Interviews
Peddie and Kelly-Campbell [68]	Hearing health	The internet, in general	Purposive	11	44-84 (median=70)	Hearing-impaired patients	Observation of participants searching predefined search tasks, think aloud, and guided interviews
Scantlebury et al [69]	Not specified	The internet, in general	Purposive	14	21-70	People who had searched online for health information	Focus groups

[a] CDC: Centers for Disease Control and Prevention.

[b] NIH: National Institutes of Health.

Regarding evaluation of internet sources, 28 articles did not specify a scope. The remaining nine articles specified or preselected sources for evaluation (eg, pediatric sun protection websites, the National Institutes of Health website, the Centers for Disease Control and Prevention website, patients.co.uk, and online forums).

Quality Evaluation Criteria Used by Consumers

Twenty-five criteria were identified (Table 2). The definitions were derived from the codes or drawn directly from the included studies.

Among these criteria, trustworthiness, expertise, and objectivity were reported most often in the articles, followed by transparency, popularity, and understandability. Eight criteria including relevance, familiarity, accessibility, identification, believability, accuracy, readability, and currency were reported in 10-15 articles. The remaining 11 criteria appeared in less than 10 articles.

Quality Indicators Used by Consumers

Indicators used by consumers to evaluate the quality of online health information were related to three aspects of online information: source, content, and design. Table 3 shows their distribution across the three categories.

About 52% of the indicators were content related, followed by design (25%) and source factors (23%); 69% of the indicators were associated with positive quality judgment, 21% were associated with negative quality judgment, and 10% could lead to both positive and negative judgment.

Table 2. Criteria used by consumers to evaluate the quality of online health information.

Criterion	Definition	Articles reporting the criterion, n (%)
Trustworthiness	Whether a source or information is honest or truthful and can be trusted	31 (84)
Expertise	Whether a source or author has a sufficient level of subject-related knowledge	31 (84)
Objectivity	Whether a source or information presents facts that are not influenced by personal feelings or commercial interests	30 (81)
Transparency	Whether important information that influences a user’s ability to make informed choices (eg, motivation of a site or owner contact information) are disclosed	21 (57)
Popularity	Whether a source or information appears in multiple venues or is received or accepted by a large number of people (eg, ranked high in search engines or followed or accepted by the crowd in social media)	19 (51)
Understandability	Whether a source or information is in appropriate depth, quantity, and specificity and error free	18 (49)
Relevance	Whether information is relevant to the topic of interest or to information seekers’ situation and background	15 (41)
Familiarity	How familiar the source is to an individual	14 (38)
Accessibility	Whether a source is easy to access and stable	14 (38)
Identification	Whether a source or information conforms to an individual’s identity, goals, styles, arguments, or objectives [62].	13 (35)
Believability	Whether information is logical and can be believed	12 (32)
Accuracy	Whether a source or information is consistent with agreed-upon scientific findings	12 (32)
Readability	Whether information is presented in a form that is easy to read (eg, concise and clear layout)	10 (27)
Currency	Whether a source or information is up to date	10 (27)
Navigability	Whether a source or information is organized in a way that is easy to navigate	9 (24)
Aesthetics	Whether the appearance of the interface is visually pleasing	9 (24)
Interactivity	Whether a source offers sufficient functions to allow users to interact with the source	9 (24)
Comprehensiveness	Whether a source or information covers a wide range of topics or offers different interaction features (eg, shopping, socializing, and researching)	8 (22)
Practicality	Whether information can be readily applied by an individual (eg, personal advice and experience)	8 (22)
Completeness	Whether necessary or expected aspects of a subject/topic are provided	7 (19)
Usefulness	Whether the amount, depth, or specificity of a source or information are at an appropriate level that can be used by an individual	7 (19)
Balanced	Whether different perspectives concerning a topic or both pros and cons concerning a treatment are provided	6 (16)
Anonymity	Whether a source can be used without forcing users to provide personal information	3 (8)
Security	Whether a source is able to prevent malicious attacks (eg, virus)	2 (5)
Learnability	Whether information can satisfy different learning needs (eg, people with different levels of knowledge)	2 (5)

Table 3. Distribution of quality indicators used by consumers to evaluate the quality of online health.

Indicators	Positive, n (%)	Negative, n (%)	Positive and negative, n (%)	Total, n (%)
Source	24 (63)	5 (13)	9 (24)	38 (23)
Content	62 (73)	17 (20)	6 (7)	85 (52)
Design	28 (67)	13 (31)	1 (2)	42 (25)
Total	114 (69)	35 (21)	16 (10)	165 (100)

Source

Source is the entity that creates, hosts, or distributes content. A source can be a website or the owner, creator, or sponsor of the site. Six categories of source-related quality indicators were identified: site owners/sponsors, site types, disclosures, third-party accreditations, recommendations from other systems or users, and website scope. More detailed indicators reported in the included articles, their direction of influence on quality judgment (positive, negative, or both), the corresponding criteria that guide the consumers' appraisal of the indicators, as well as the value of the criteria (positive, negative, or both) are shown in Table 4. The indicators in the tables are self-explanatory; therefore, we focus on describing the most frequently appearing indicators in the included studies and indicators that can lead to both positive and negative judgments.

The Most Frequently Mentioned Indicators

The most frequently mentioned source-related indictors were site owners/sponsors, with sites run by reputable organizations, educational and academic institutions [18,40,41,46, 52,59], and medical experts and health institutions [32,39,44,46,51, 54,55,57,59,65] being considered more trustworthy and offering higher levels of expertise. The second most frequently reported indicators were about disclosure. Sites that disclose their motivations were highly valued [40,42,48,64,66,67], whereas a lack of a clear statement of purpose and motivation damaged trust [49]. The third mostly frequently reported indicators were recommendations from other systems or users. High ranks in search engines [13,52,63] and a large number of visitors or followers [61,63] were viewed as indicators of high site popularity, and subsequently, high quality. In addition, sites linked from or recommended by a trusted website [30,43,53] or trusted others (eg, health care providers, families, and friends) [20,41,47,61,67] were considered trustworthy.

Indicators With Both Positive and Negative Influences on Evaluation

Mixed attitudes were found toward some indicators representing site owners/sponsors. First, most participants believed that government websites (eg, National Health Service and Centers for Disease Control and Prevention) reflect high levels of expertise and good intentions [39-41,46,53,54,58,59,64,65]; however, some consumers suspected that the information on government websites is biased due to their agendas [41,52], and some, particularly younger generations, did not identify themselves with government sources, considering them "less cool" and not relatable [58]. Second, most people considered sites operated by local health societies to have a high level of expertise; however, some minorities and people from nonmainstream cultures (eg, aboriginal communities) were likely to question the relevance and accuracy of the information from these sites [47]. Third, people usually considered websites owned by commercial companies less objective [42,46,48] and trusted more websites with no commercial interests [18,20,42,55,67]; nevertheless, popular commercial websites such as BabyCenter.com, ParentsPlace.com, and WebMD.com were favored by some people for their expertise and comprehensiveness [42]. Fourth, a few people viewed information from pharmaceutical company websites as "official" [40], whereas others considered their information biased due to the financial interests involved [32,40,46,48].

Consumers had mixed attitudes toward the website types, particularly social media sites. Some consumers favored online discussion groups, chat rooms, and listservs because they offered first-person narratives and practical information and support from peers with whom they could identify (ie, those who have similar conditions) [46,53,57], but some disliked such sites for their lack of objectivity and expertise [13,53,59]. Concerning Wikipedia, some people questioned its objectivity because information can be edited by anyone on the Web [50,58,61], but some consumers were attracted to its encyclopedic nature and comprehensiveness [63].

Consumers also had different opinions regarding sites recommended by others. Some trusted a site recommended by trusted others (eg, health care providers, families, and friends) [20,41,47,61,67]; however, some consumers recognized that recommendations from other individuals may not be relevant to their situation [67].

Table 4. Evaluation of the source.

Indicators	Criteria
Site owners/sponsors (n=30)	
Site name (n=4) [32,40,63,64]	
Inappropriate or weird site names (–[a])	Believability (–)
Domain type (n=5)[40,42,50,57,61]	
.com (–)	Objectivity (–)
.org (+[b])	Trustworthiness (+)
.gov (±[c])	Expertise (+), Trustworthiness (±)
.edu (+)	Expertise (+), Trustworthiness (+)
Owner identity (n=26)[18,20,30,32,32,39-42,44,46,48-55,57-59,64,65,67,68]	
Individual sponsor (–)	Objectivity (–)
Private sites (–)	Objectivity (–)
Reputable organizations (+)	Trustworthiness (+)
Educational and academic institutions (+)	Expertise (+)
Medical or health institutions/experts (+)	Expertise (+)
Scientific publisher (+)	Expertise (+)
Patients’ organization (+)	Trustworthiness (+)
Well-known news sites (+)	Trustworthiness (+)
Government institutions (±)	Expertise (+), Trustworthiness (±), Identification (–)
Local cancer society (±)	Expertise (+), Relevance (–), Accuracy (–)
Commercial sponsor (±)	Objectivity (–), Expertise (+)
No financial gain to the owner (+)	Objectivity (+)
Pharmaceutical industry (±)	Expertise (+), Objectivity (–)
Site types (n=9) [13,46,50,53,57-59,61,63]	
Online peer support and discussion groups (+)	Identification (+), Practicality (+)
Chatrooms (+)	Identification (+), Practicality (+)
Forums (–)	Objectivity (–), Expertise (–)
Personal blogs/websites (±)	Objectivity (–), Expertise (–), Identification (+)
Listservs (±)	Objectivity (–), Expertise (–), Identification (+)
Wikipedia (±)	Objectivity (–), Expertise (–), Comprehensiveness (+)
Disclosure (n=13) [30,40-43,48,49,51,55,63,64,66,67]	
Disclosure of the site owner (+)	Transparency (+)
Age of a website (+)	Transparency (+), Trustworthiness (+)
Picture of the site owner (+)	Transparency (+)
Contact information (+)	Transparency (+)
Motivation of the site (+)	Transparency (+)
Explicit disclaimer and alert (+)	Transparency (+)
Third party accreditation (n=4) [30,43,46,63]	
Quality certificates, seals, stamps, or kitemarking (+)	Accuracy (+)
Recommendations from other systems or users (n=12) [13,20,30,41,43,47,52,53,61,63,67]	
Rank in search engine results (+)	Popularity (+)
Number of site visitors or followers (+)	Popularity (+)

Indicators	Criteria
Titles and excerpts in search engine results (+)	Relevance (+)
Linked from a trustworthy site (+)	Trustworthiness (+)
Recommended by other people (±)	Trustworthiness (+), Relevance (–)
Website scope (n=7) [32,42,44,49,53,54,58]	
A wide range of topics in a site (+)	Comprehensiveness (+)
Multiple functions in a site (+)	Comprehensiveness (+)

[a] – indicates a negative evaluation of quality or that a criterion is judged negatively.

[b] + indicates a positive evaluation of quality or that a criterion is judged positively.

[c] ± could indicate both positive and negative evaluations or a criterion could be judged both positively and negatively.

Content

Content refers to the information contained in a source as well as the presentation of the information. Eight categories of content-related indicators were identified: substance, writing and language, presentation, references, authorship, audience, date/updating, and advertisements. Table 5 shows these indicators, the corresponding criteria that guide the consumers' appraisal of the indicators, and their influence on quality judgment.

The Most Frequently Mentioned Indicators

The most frequently reported content indicators were about consensus among sources. Content that appears in multiple sources, be it online sources, sources in other media (eg, newspaper, television, books, and academic journals), or health care professionals, is trusted by consumers. Writing- and language-related factors were the second most frequently reported content indicators. Consumers expect high-quality information to be error free in spelling and grammar, use straightforward language, and have a clear layout. The third most frequently reported indicators were advertisements. Consumers expect quality websites to neither depend on advertisements [63] nor seek to make a profit [62]. Therefore, sites with advertisements were considered less objective [46,49,54,63,64], be it in the form of commercial links [46], advertisement banners [30,32,55], popups [32,40], or other formats.

Indicators With Both Positive and Negative Influences on Evaluation

Consumers had mixed attitudes toward two content types: objective facts and personal experiences. Some consumers value objective facts [43,45,62], whereas some dissatisfy with information that contained solely objective facts, considering it unbalanced [69]. With regard to personal experiences, some consumers favored first-hand experiences, stories, and advice (eg, recommendations for medical gadgets, meal planning and exercising, and advice on completing medical benefit forms) from people with similar conditions for their practicality [40,41,45,46,49,63,64], but some had concerns that personal experiences lack objectivity, especially when it merely is a personal opinion [30,69].

Celebrity endorsement was also an indicator with both positive and negative influences on quality evaluation. Some trusted the endorsed information but others question its objectivity due to the potential financial interest involved [61,67].

The use of medical and technical vocabularies raised contention among consumers as well. For some consumers, high-quality information was easy to understand, that is, it exhibited less use of professional medical vocabularies [30,43,49,64,69] or provided easy-to-understand definitions of medical jargon [46,49], especially for educational and government sites [18,42,52]; however, for others, the use of technical vocabularies demonstrated expertise and was highly valued [51,60,63].

Some consumers doubted information (especially diagnosis and treatment information) authored by other unknown consumers [42,53], but others tended to trust content written by their peers because of similar demographic or health characteristics [32].

For health interventions, some consumers examined the release time and perceived newer interventions to have higher quality than the existing ones because the new intervention would have undergone more testing and research, whereas others were uncertain about the reliability of newer interventions [67].

Design

Design refers to the appearance of a website or an app and the interactions that it affords. Four categories of design-related quality indicators were identified: interface design, interaction design, navigation design, and security settings. Table 6 shows the specific indicators, the corresponding criteria that guide the consumers' appraisal of the indicators, and their influence on quality judgment.

Table 5. Evaluation of content.

Indicators	Criteria
Substance (n=31)	
Content types (n=15)	
Factual information (n=14) [18,30,40,41,43,45,46,49,53,62-64,67,69]	
Evidence based (+[a])	Objectivity (+)
Clinically proven (+)	Objectivity (+)
Statistics and numbers (+)	Objectivity (+)
Concrete examples (+)	Practicality (+)
Objective facts (±[b])	Objectivity (+), Balanced (–)
Ideological and magical information (–[c])	Accuracy (–)
Unproven and uncertain scientific information (–)	Accuracy (–), Objectivity (–)
Personal experiences (n=9) [30,40,41,45,46,49,63]	
First hand (+)	Accuracy (+)
Practical advice (+)	Practicality (+)
Personal experiences (±)	Objectivity (–), Practicality (+), Identification (+)
Personal opinion (–)	Objectivity (–), Expertise (–)
Content attributes (n=29)	
Balance (n=6) [43,46,48,63,64,67]	
Alternative medicine (+)	Balanced (+)
Conflicting views (+)	Balanced (+)
Both professional and patient viewpoints (+)	Balanced (+)
Potential side effects (+)	Complete (+), Transparency (+)
Depth (n=5) [18,46,49,51,62]	
At the right level of complexity and depth (+)	Understandability (+)
Increasing in depth overtime (+)	Usefulness (+)
In-depth information (+)	Expertise (+)
Quantity (n=5) [18,46,62,67]	
The right amount (+)	Understandability (+)
Too much text (–)	Understandability (–)
Specificity (n=5) [18,46,47,49,67]	
Various levels of detail for different needs (+)	Usefulness (+)
Specific and detailed (+)	Understandability (+)
Overall and general information (–)	Usefulness (–)
Consensus among sources (n=20) [13,18,30,38-43,55-57,59-61,63,65-67,69]	
Reaching agreement among media sources (+)	Popularity (+)
Verified by general practitioners or other health professionals (+)	Expertise (+)
Crowd consensus (+)	Popularity (+)
Endorsed by celebrities (±)	Trustworthiness (+), Objectivity (–)
Specific content elements (n=3) [47,61,67]	
Natural ingredients (+)	Trustworthiness (+)
Amount of investment on an intervention (+)	Trustworthiness (+)
Copyright information (+)	Trustworthiness (+)

Indicators	Criteria
Local support and contact information (+)	Usefulness (+), Relevance (+)
Argument strength (n=6) [30,32,41,48,56,69]	
Reasonable (+)	Believability (+)
Sound plausible and scientific (+)	Believability (+)
Biased or misleading (–)	Objectivity (–)
Writing and language (n=19) [18,30,32,42,43,46,48,49,51,52,56-58,60,62-64,67,69]	
Spelling and grammar errors (–)	Understandability (–), Expertise (–), Trustworthiness (–)
Long sentences (–)	Readability (–)
Professional writing (+)	Expertise (+)
Concise (+)	Comprehensiveness (+), Readability (+)
Use simple, plain, straightforward, and clear language (+)	Understandability (+)
Familiar sounding and inclusive language (+)	Understandability (+), Identification (+)
Sensational (–)	Objectivity (–)
Patronizing tone (–)	Identification (–)
Use of professional medical terms and technical vocabularies (±)	Understandability (–), Expertise (+)
Easy reading level (–)	Expertise (–)
Presentation of content (n=12)	
Organization (n=10) [18,32,44,45,49,54,59,62,66,68]	
Clear layout and organization (+)	Readability (+)
An overview of the information on a site (+)	Readability (+)
Use of bolding and shading (+)	Readability (+)
Bulleted points (+)	Readability (+)
Headings (+)	Readability (+)
Prioritizing content (+)	Understandability (+)
Structure of scientific papers: general approaches and research design (+)	Expertise (+)
Structure of scientific papers: presence of variables or factors (+)	Expertise (+)
Structure of scientific papers: research purposes (+)	Expertise (+)
Labeling (n=2) [43,63]	
Presence of an informative title (+)	Understandability (+)
Clearly marked personal experience (+)	Transparency (+)
References (n=10) [30,39,43,45,56-58,63,64,69]	
Links to original documents (+)	Transparency (+)
Number of references included (+)	Trustworthiness (+), Expertise (+)
Reference to scientific publications (+)	Objectivity (+), Expertise (+)
Reference to a credible person (+)	Trustworthiness (+), Expertise (+)
Reference to a specific project or institution (+)	Transparency (+), Trustworthiness (+)
Authorship (n=16) [30,32,40,42,45,60,62-64,66,67]	
Explicitly listing authors and author's credentials (+)	Transparency (+)
Reference to previous work or curriculum vitae (+)	Trustworthiness (+), Expertise (+)
Picture of the author (+)	Trustworthiness (+), Transparency (+)
Health professionals (+)	Expertise (+)
Journalists (+)	Expertise (+)
Consumers (±)	Practicality (+), Identification (+), Expertise (–), Objectivity (–)

Indicators	Criteria
Economic gains for its authors (–)	Objectivity (–)
Religious figures (–)	Objectivity (–)
Audience (n=11) [32,45-48,51,53,57,58,61,63]	
Targeted to geographical location (+)	Relevance (+)
Translated information (+)	Understandability (+), Accessibility (+)
Tailored and personalized information (+)	Usefulness (+)
Targeted to minority women (+)	Identification (+)
Targeted to professions (+)	Relevance (+)
Targeted to age group (+)	Relevance (+)
Seeing a face that looked similar to theirs (+)	Identification (+)
Written for the most educated audience (+)	Expertise (+)
Aimed at younger children (–)	Relevance (–), Accuracy (–)
Date/updating (n=12) [30,39,41,43,44,46,53,59,60,63,67,69]	
The appearance of publication date (+)	Transparency (+)
Access all the latest research (+)	Currency (+), Completeness (+)
New interventions (±)	Currency (+), Accuracy (–)
Up to date (+)	Currency (+)
Regular updating (+)	Transparency (+), Currency (+)
Advertisements (n=17) [30,32,39,40,42,43,46,49,54,55,59,62-64,66-68]	
Presence of ads (–)	Objectivity (–)
Pushing to sell something (–)	Objectivity (–)
The appearance of commercial links (–)	Objectivity (–)

[a] + indicates a positive evaluation of quality or that a criterion is judged positively.

[b] ± indicates both positive and negative evaluations or a criterion could be judged both positively and negatively.

[c] – indicates a negative evaluation of quality or that a criterion is judged negatively.

Table 6. Evaluation of design.

Indicators	Criteria
Interface design (n=16)	
Overall appearance (n=9) [30,32,39,44,46,49,59,62,64]	
Boring and bland design (–[a])	Aesthetics (–)
Commercial nature/feel (–)	Objectivity (–)
Modern look (+[b])	Aesthetics (+), Identification (+)
Professional (+)	Expertise (+), Trustworthiness (+)
High visual quality (+)	Trustworthiness (+), Aesthetics (+)
Soft colors (+)	Aesthetics (+)
Graphics (n=9) [18,32,40,44,49,54,64,66,67]	
Too many graphics (–)	Aesthetics (–)
Use of flash (–)	Aesthetics (–), Accessibility (–)
Poor graphics (–)	Aesthetics (–)
Inappropriate graphics (–)	Relevance (–), Trustworthiness (–)
The existence of brand logo (+)	Trustworthiness (+)
Relevant illustrations (+)	Relevance (+)
Font (n=5) [18,32,44,49,68]	
Large font size (+)	Accessibility (+)
Font color low contrast (–)	Accessibility (–)
Interaction design (n=14)	
Links (n=4) [45,48,49,55]	
Link to other websites (+)	Trustworthiness (+), Interactivity (+)
Plenty of links (+)	Interactivity (+)
Broken links (–)	Accessibility (–), Trustworthiness (–)
Easy access to further details and sources (+)	Accessibility (+)
Downloadable PDF documents for bibliographies and laws (+)	Accessibility (+)
Interactive functions (n=7) [30,32,44,46,49,61,62]	
Search capabilities (+)	Interactivity (+)
Places to interact and share with other site visitors (+)	Interactivity (+)
“Ask experts” (+)	Interactivity (+), Expertise (+)
Self-management and assessment tools (±[c])	Usefulness (+), Accuracy (–), Objectivity (–)
Other interactive features (n=9) [30,32,40,46,48,61,63,68,69]	
Slow loading time (–)	Accessibility (–)
Required login (–)	Accessibility (–), Anonymity (–)
Absence of pop-ups (+)	Accessibility (+)
Multimedia feature (+)	Interactivity (+), Learnability (+)
Navigation design (n=9) [30,32,44,46,49,54,58,68,69]	
Relevant info on home page (+)	Navigability (+)
Clear entry point (+)	Navigability (+), Accessibility (+)
Easy return to home page (+)	Navigability (+)
Navigation aids (+)	Navigability (+)
Navigation links (+)	Navigability (+)

Indicators	Criteria
Site map (+)	Navigability (+)
Side tool bars (+)	Navigability (+)
Different ordering structures (+)	Navigability (+)
Clear indication when taken offsite (+)	Navigability (+), Transparency (+)
Easy transition between two or more sites (+)	Navigability (+)
“Back” button as the only way to exit (–)	Navigability (–), Accessibility (–)
Heavily relied on dropdown menu (–)	Navigability (–)
Continually sending users offsite (–)	Interactivity (–), Trustworthiness (–)
Security settings (n=2) [39,57]	
Secure sites (+)	Security (+)
Recognized by antivirus software (+)	Security (+)

[a] – indicates a negative evaluation of quality or that a criterion is judged negatively.

[b] + indicates a positive evaluation of quality or that a criterion is judged positively.

[c] ± could entail both positive and negative evaluations or a criterion could be judged both positively and negatively.

The Most Frequently Mentioned Indicators

The most frequently reported design indicators were related to interface design, mostly visual factors, including the overall appearance of a site, the graphics it includes, and font size. Interaction design features, including links, interactive functions, and other interactive features (eg, loading time and login requirement), were the second most frequently mentioned quality indicators. Sites with robust search capabilities (eg, easy to locate and diverse search entrance), offering useful tools (eg, self-management tools), and rendering smooth user-system interaction (eg, providing links to additional relevant sources and not having pop-ups) were perceived as high quality. Navigation-related indicators such as navigation aids and site maps were the third most frequently mentioned quality indicators.

Indicators With Both Positive and Negative Influences on Evaluation

Mixed opinions existed concerning the interactive functions of self-management and assessment tools (eg, health calculators). Some consumers valued tailored results and advice [46,49], but some questioned the accuracy and objectivity of the information generated [46,62].

Individual Factors Influencing Quality Judgment

In addition to source-, content-, and design-related factors, the evaluation of online health information quality was also affected by individual factors including individuals’ personal situation, prior knowledge or experience of a source, personal knowledge and beliefs, and intuition. Table 7 shows the specific factors, the corresponding criteria that guide the consumers’ appraisal, and their influence on quality judgment.

The Most Frequently Mentioned Factors

Individuals’ prior knowledge and experience of a source were mentioned most frequently as factors that influence quality judgment. Consumers tended to trust sites that they had experience with [49,63], because they may already know the source to be credible [13,18,39,50,53,59,65], have had positive experiences with it [42,54,67], have seen it from advertisements on other media (eg, television and magazine) [42,58], or are familiar with the organization behind the source [18,42].

The category of personal situation was the second most frequent factor. Information relevant to individuals’ search topics (eg, hormone replacement therapy) [32,45], needs and goals (eg, offering easy reading level message for younger people) [40,54,57], specific circumstances (eg, localization) [40,45,62,64], and experiences and symptoms [18,56,62,64] was considered to be of high quality.

The other two categories of individual factors were mentioned with the same frequency. One category is personal knowledge and beliefs. Consumers highly valued information consistent with their own beliefs and knowledge [18,32,41,56,63,64]. The other category is intuition. Some consumers undertook “subconscious filtering” to filter out potential political and gender biased information [69], and some consumers relied on common sense [39], sensation [63], instinct, or “gut feelings” [55,65,66] to evaluate information.

Table 7. Individual factors.

Factors	Criteria
Individuals' personal situation (n=9) [18,32,40,45,54,56,57,62,64]	
Relevant topics (+[a])	Relevance (+)
Information relevant to one's needs and search goal (+)	Relevance (+)
Information relevant to one's circumstance and applicable (+)	Relevance (+) Usefulness (+)
Information related to one's experiences and symptoms (+)	Identification (+)
Prior knowledge and experience of a source (n=14) [13,18,39,42,49,50,53,54,58,59,61,63,65,67]	
Known credible websites (+)	Familiarity (+) Expertise (+)
Positive previous experience (+)	Familiarity (+) Trustworthiness (+)
Websites advertised in other media (+)	Familiarity (+)
Familiar organization (+)	Familiarity (+)
Personal knowledge and beliefs (n=7) [13,18,32,41,56,63,64]	
Consistency with one's own beliefs and knowledge (+)	Identification (+)
Intuition (n=7) [39,55,57,63,65,66,69]	
Subconscious (+)	Believability (+)
Common sense (+)	Believability (+)
Instinct/sensation/gut feeling (+)	Believability (+)

[a] + indicates a positive evaluation of quality or that a criterion is judged positively.

Discussion

In this article, we reviewed 37 empirical studies that reported consumers' accounts of how they evaluate the quality of online health information. This review extends the existing literature by making two major conceptual contributions. First, it offers a clear conceptual understanding of the dimensions of quality of online health information perceived by consumers by differentiating criteria from indicators. Second, it explicates the relationship between webpage quality indicators (webpage elements) and the quality judgment by differentiating positive and negative influences that indicators have on judgment. In this section, we discuss each contribution and then outline practical implications and limitations of this review.

Dimensions of Online Health Information Quality

In the existing literature, quality was often defined and assessed differently. We guided the article selection for the review using a general conceptualization that defines quality as "fitness for use" [25]. Other authors have offered more specific conceptualizations. For example, Rieh [70] assessed quality as the extent to which users think that the information is useful, good, current, and accurate. Bates et al [71] measured health information quality in terms of its trustworthiness, truthfulness, readability, and completeness. Benotsch et al [22] rated the quality of health websites on five dimensions: accuracy, amount of detail, trustworthiness-credibility, relevance, and usefulness. Eastin [72] rated the credibility of health information on three dimensions: accuracy, believability, and factualness. The lack of consistency in measuring online health information quality suggests that there is a lack of clear conceptual understanding of what information quality means to online health consumers. By clearly differentiating quality judgment criteria (rules that reflect notions of value and worth) and indicators (properties of information objects to which criteria are applied to form judgments) reported in the included studies, this review identified 25 dimensions (criteria) along which consumers evaluate the quality of online health information (Table 2). Because the included articles differ on aspects such as health issues of concern, participant demographics, and sources examined, this wide range of criteria reported and the uneven distribution of the criteria across the included articles suggest that consumer evaluation of online health information may be influenced by contextual factors such as user characteristics, health conditions, and online sources. In addition to these factors, the current review, consistent with prior reviews [10,11,27], also identified a range of individual factors that influence quality judgment behavior, such as prior experience with a source and personal knowledge and beliefs. Therefore, future studies should attempt to identify the most influential contextual factors (including individual factors) that affect consumers' application of quality criteria to further enhance the theoretical understanding of this behavior. Empirical studies of consumer online health information evaluation should also consider these contextual factors in research design.

Despite the wide range, however, three criteria (trustworthiness, expertise, and objectivity) were reported in 31 articles, indicating that they are used consistently across user groups, source types, and health conditions and that they constitute core dimensions of online health information quality as perceived by consumers. The fact that trustworthiness and expertise are primary dimensions is consistent with general media source credibility research [73]. It is not surprising that objectivity, that is, whether a source or information presents objective factual or evidence-based information, is also important for health information. Three additional criteria—transparency (reported in 21 articles), popularity (reported in 19 articles), and understandability (reported in 18 articles)—are also commonly reported and could be viewed as secondary dimensions of online health information quality. These findings imply that consumers' perceived online health information quality could be reasonably measured by a small set of core dimensions.

Relationship Between Quality Indicators and Quality Judgment

Previous reviews summarized indicators used by consumers to evaluate the quality of online information [10,37]. Sbaffi and Rowley [11] further reported the direction of the effect (ie, positive vs negative) of the (design and content) indicators. However, the situational nature of the relationship between indicators and quality judgment, that is, the fact that their relationship is not one-on-one, but dependent on users' values and the criteria applied, was not explicitly discussed. For example, government institutions, usually associated with high level of expertise and authority, are perceived by some consumers as biased sources with which they have a hard time relating. The other example is that consumer-generated content (eg, personal blogs and listserves) indicates low objectivity and low level of expertise to some consumers, but to others, it is considered highly practical and relatable. Thus, a unique contribution of this review is that it clearly maps out the direction of the impact (ie, positive or negative) of a number of indicators on quality judgment and the underlying reasons (ie, criteria) for the impact.

Practical Implications

The identification and differentiation of positive and negative indicators provide clear guidance for online health information designers. They can incorporate positive indicators (eg, offering authors' credentials and presenting information in a clear and organized way) and avoid negative indicators (eg, dead links and flash media format) to offer users better information seeking experiences. The fact that the same indicator (eg, government institutions as the source owner) can lead to different quality judgment for different people suggests that designers should also carefully investigate target users' values and the corresponding criteria that they use to evaluate health information. This calls for active user research and user involvement in the design process.

The results of the review also have implications for consumer education. The review revealed a wide range of criteria that consumers use to evaluate the quality of online health information. Many of the criteria, such as familiarity, identification, relevance, practicality, and usefulness, are highly subjective and situational, influenced by factors such as information needs, online information search experience, and personal beliefs. In some cases, consumers assign such criteria higher priority than more objective ones such as expertise [57]. The review also revealed that consumers use a diverse set of quality indicators. The implications of some of the indicators are not well understood. For example, some consumers believe that the appearance of copyright information or the word "clinical" indicates high information quality [61,67]. Some consumers view the fact that a website passes the screening of virus/security software as an indicator of high quality [57]. There are also consumers assuming that third-party accreditations are indicators of information accuracy, when, in fact, the guidelines that these accreditations follow do not really check for information accuracy [74,75]. Consumers need education to use more objective criteria to evaluate online health information and understand the implications of a number of quality indicators.

Limitations

This review has several limitations. First, we selected only studies where consumers explicitly described their quality evaluation behavior. These studies tend not to ask consumers to rate criteria or indicators; thus, we could not identify the importance of each indicator or criterion in quality judgment. Future reviews are needed to fill this gap. Second, we did not differentiate and compare results based on observations and results drawn from verbal inquiries as few included studies did. Eysenbach and Kohler [30] reported discrepancies between participants' verbal accounts of what they do to evaluate health information and what they actually did in performing search tasks (based on observations). Thus, future empirical studies are needed to shed light on this gap. Third, in the coding process, we used criterion and indicator terms from the original papers, where feasible. In cases where we needed to infer criteria from indicators, we followed the mostly commonly recognized categorization by referring to prior empirical research and reviews or inferred the criteria from participants' quotes. However, due to the different perspectives of the authors of the original papers and the inherent overlap between terms, such as comprehensiveness and completeness, our syntheses are inevitably affected by a certain degree of subjectivity. Fourth, because most studies treated the internet as one source of information without differentiating source types (eg, regular websites and social media), we were not able to identify whether the use of evaluation criteria and indicators differs by source type.

Conclusions

The quality of online health information is a complex concept involving more than two dozen dimensions, as perceived by consumers. Although a set of core dimensions can be identified, the diversity involved in consumers' use of criteria is too obvious to ignore. Further examination of contextual factors (eg, different source and user characteristics) that influence consumers' application of quality criteria will bring further clarity to the concept. The review identified 165 indicators, to which criteria are applied to reach a quality judgment. Indicators could be source, content, or design related; they can have a positive or negative impact on quality judgment, contingent on situations and users' values and beliefs. The identification and differentiation of positive and negative indicators along with their respective criteria can provide clearer guidance for designers of online health websites and educational interventions. Compared to experts' evaluation, consumers' evaluation of online health information relies heavily on peripheral cues and is influenced by various contextual factors (eg, personal beliefs and information needs). This finding suggests that current quality evaluation checklists, which are mostly based on experts' view of quality, may not effectively serve the needs of consumers. Consumer behavior needs to be considered in the design of interventions that intend to promote quality evaluation in online searches. At the same time, it is worth noting that criteria and indicators used by consumers merit critical evaluation, as some criteria are overly subjective and the implications of some indicators are not well understood. User education is needed to address user misconceptions and the associated suboptimal evaluation behavior.

Acknowledgments

The project is supported by funding from the Portuguese Foundation for Science and Technology and the Digital Media Program at UT-Austin and the UT School of Information Governor Bill Daniel Fellowship.

Conflicts of Interest

None declared.

References

1. Fox S, Duggan M. Pew Research Center Internet & Technology. 2013. Health Online 2013 URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-04-18] [WebCite Cache ID 73FSIZi6p]
2. Kealey E, Berkman CS. The relationship between health information sources and mental models of cancer: findings from the 2005 Health Information National Trends Survey. J Health Commun 2010;15 Suppl 3:236-251. [doi: 10.1080/10810730.2010.522693] [Medline: 21154096]
3. Fox S. Pew Research Center. 2011. The Social Life of Health Information 2011 URL: http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2018-10-16] [WebCite Cache ID 73FTvWOhT]
4. 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 Dec;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
5. McKinley CJ, Wright PJ. Informational social support and online health information seeking: Examining the association between factors contributing to healthy eating behavior. Computers in Human Behavior 2014 Aug;37:107-116. [doi: 10.1016/j.chb.2014.04.023]
6. Eng TR, Maxfield A, Patrick K, Deering MJ, Ratzan SC, Gustafson DH. Access to Health Information and Support:A Public Highway or A Private Road? JAMA 1998 Oct 21;280(15):1371-1375. [doi: 10.1001/jama.280.15.1371]
7. Eysenbach G, Powell J, Kuss O, Sa E. Empirical Studies Assessing the Quality of Health Information for Consumers on the World Wide Web. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691]
8. 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]
9. Cline R, Haynes K. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16:671-692. [doi: 10.1093/her/16.6.671]
10. Diviani N, van den Putte B, Giani S, van Weert JCM. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
11. Sbaffi L, Rowley J. Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
12. Arora NK, Hesse BW, Rimer BK, Viswanath K, Clayman ML, Croyle RT. Frustrated and confused: the American public rates its cancer-related information-seeking experiences. J Gen Intern Med 2008 Mar;23(3):223-228 [FREE Full text] [doi: 10.1007/s11606-007-0406-y] [Medline: 17922166]
13. Feufel MA, Stahl SF. What do web-use skill differences imply for online health information searches? J Med Internet Res 2012 Jun;14(3):e87 [FREE Full text] [doi: 10.2196/jmir.2051] [Medline: 22695686]
14. Car J, Lang B, Colledge A, Ung C, Majeed A. Interventions for enhancing consumers' online health literacy. Cochrane Database Syst Rev 2011(6):CD007092. [doi: 10.1002/14651858.CD007092.pub2] [Medline: 21678364]
15. Sweet SN, Perrier M, Podzyhun C, Latimer-Cheung AE. Identifying physical activity information needs and preferred methods of delivery of people with multiple sclerosis. Disabil Rehabil 2013;35(24):2056-2063. [doi: 10.3109/09638288.2013.800915] [Medline: 23763468]
16. Unruh HK, Bowen DJ, Meischke H, Bush N, Wooldridge JA. Women's approaches to the use of new technology for cancer risk information. Women Health 2004;40(1):59-78. [doi: 10.1300/J013v40n01_04] [Medline: 15778132]
17. Gray NJ, Klein JD, Cantrill JA, Noyce PR. Adolescent girls' use of the Internet for health information: issues beyond access. J Med Syst 2002 Dec;26(6):545-553. [Medline: 12385536]
18. Marshall LA, Williams D. Health information: does quality count for the consumer? Journal of Librarianship and Information Science 2016 Jun 30;38(3):141-156. [doi: 10.1177/0961000606066575]

19. Anderson JG. Consumers of e-Health. Social Science Computer Review 2016 Aug 18;22(2):242-248. [doi: 10.1177/0894439303262671]
20. Santer M, Muller I, Yardley L, Burgess H, Ersser SJ, Lewis-Jones S, et al. 'You don't know which bits to believe': qualitative study exploring carers' experiences of seeking information on the internet about childhood eczema. BMJ Open 2015;5(4):e006339 [FREE Full text] [doi: 10.1136/bmjopen-2014-006339] [Medline: 25854963]
21. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
22. Benotsch EG, Kalichman S, Weinhardt LS. HIV-AIDS patients' evaluation of health information on the internet: the digital divide and vulnerability to fraudulent claims. J Consult Clin Psychol 2004 Dec;72(6):1004-1011. [doi: 10.1037/0022-006X.72.6.1004] [Medline: 15612847]
23. Oh S, Worrall A. Health answer quality evaluation by librarians, nurses, and users in social Q&A. Library & Information Science Research 2013 Oct;35(4):288-298. [doi: 10.1016/j.lisr.2013.04.007]
24. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
25. Gryna F. Juran's quality control handbook. New York: McGraw-Hill College Division; 1988.
26. Zhang Y, Wildemuth BM. Qualitative analysis of content. In: Wildemuth B, editor. Applications of Social Research Methods to Questions in Information and Library Science (2nd ed.). Westport, CT: Libraries Unlimited; 2017:318-329.
27. 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. [doi: 10.1177/1460458214559432] [Medline: 25518944]
28. Fogg B, Soohoo C, Danielson D, Marable L, Stanford J, Tauber E. How Do Users Evaluate the Credibility of Web Sites?: A Study with over 2,500 Participants. 2003 Presented at: Proc Conf Des User Exp; June 06-07, 2003; San Francisco, California p. 1-15. [doi: 10.1145/997078.997097]
29. Robins D, Holmes J, Stansbury M. Consumer health information on the Web: The relationship of visual design and perceptions of credibility. J Am Soc Inf Sci 2009 Oct 02;61(1):13-29. [doi: 10.1002/asi.21224]
30. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [Medline: 11884321]
31. Metzger MJ, Flanagin AJ, Eyal K, Lemus DR, Mccann RM. Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. Annals of the International Communication Association 2016 May 18;27(1):293-335. [doi: 10.1080/23808985.2003.11679029]
32. 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]
33. Hong T. The influence of structural and message features on Web site credibility. J Am Soc Inf Sci 2005 Jan 01;57(1):114-127. [doi: 10.1002/asi.20258]
34. Belkin NJ. Intelligent information retrieval: Whose intelligence? Intelligent information retrieval: Whose intelligence? Proc Fifth Int Symp Inf Sci Konstanz: Universitatsverlag Konstanz; 1996 Presented at: Proc Fifth Int Symp Inf Sci; 1996; Konstanz p. 25-31.
35. Brunswik E. The Conceptual Framework of Psychology. Chicago, IL: University of Chicago Press; 1962.
36. Wang P, Soergel D. A cognitive model of document use during a research project. Study I. Document selection. J Am Soc Inf Sci 1998 Feb;49(2):115-133. [doi: 10.1002/(sici)1097-4571(199802)49:2<115::aid-asi3>3.0.co;2-t]
37. Choi W, Stvilia B. Web credibility assessment: Conceptualization, operationalization, variability, and models. J Assn Inf Sci Tec 2015 May 13;66(12):2399-2414. [doi: 10.1002/asi.23543]
38. Henderson EM, Eccleston C. An online adolescent message board discussion about the internet: Use for pain. J Child Health Care 2015 Sep;19(3):412-418. [doi: 10.1177/1367493513509420] [Medline: 24270989]
39. Frisby G, Bessell TL, Borland R, Anderson JN. Smoking cessation and the Internet: a qualitative method examining online consumer behavior. J Med Internet Res 2002 Nov;4(2):E8 [FREE Full text] [doi: 10.2196/jmir.4.2.e8] [Medline: 12554555]
40. Peterson G, Aslani P, Williams KA. How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. J Med Internet Res 2003 Dec 19;5(4):e33 [FREE Full text] [doi: 10.2196/jmir.5.4.e33] [Medline: 14713661]
41. Williams P, Nicholas D, Huntington P. Health information on the Internet: a qualitative study of NHS Direct Online users. Aslib Proceedings 2003 Dec;55(5/6):304-312. [doi: 10.1108/00012530310498879]
42. Bernhardt J, Felter E. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 01;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
43. Childs S. Developing health website quality assessment guidelines for the voluntary sector: outcomes from the Judge Project. Health Info Libr J 2004 Sep;21 Suppl 2:14-26 [FREE Full text] [doi: 10.1111/j.1740-3324.2004.00520.x] [Medline: 15317572]

44. Adams S, de Bont A, Berg M. Looking for answers, constructing reliability: an exploration into how Dutch patients check web-based medical information. Int J Med Inform 2006 Jan;75(1):66-72. [doi: 10.1016/j.ijmedinf.2005.07.036] [Medline: 16137921]
45. Crystal A, Greenberg J. Relevance criteria identified by health information users during Web searches. J Am Soc Inf Sci 2006 Aug;57(10):1368-1382. [doi: 10.1002/asi.20436]
46. Kerr C, Murray E, Stevenson F, Gore C, Nazareth I. Internet interventions for long-term conditions: patient and caregiver quality criteria. J Med Internet Res 2006 Jul 28;8(3):e13 [FREE Full text] [doi: 10.2196/jmir.8.3.e13] [Medline: 16954123]
47. Hoffman-Goetz L, Friedman DB. A qualitative study of Canadian Aboriginal women's beliefs about. J Cancer Educ 2007;22(2):124-128. [doi: 10.1080/08858190701372950] [Medline: 17605628]
48. Sillence E, Briggs P. Please advise: using the Internet for health and financial advice. Computers in Human Behavior 2007 Jan;23(1):727-748. [doi: 10.1016/j.chb.2004.11.006]
49. Sillence E, Briggs P, Harris P, Fishwick L. Health Websites that people can trust – the case of hypertension. Interacting with Computers 2007 Jan;19(1):32-42. [doi: 10.1016/j.intcom.2006.07.009]
50. Buhi ER, Daley EM, Fuhrmann HJ, Smith SA. An observational study of how young people search for online sexual health information. J Am Coll Health 2009 Sep;58(2):101-111. [doi: 10.1080/07448480903221236] [Medline: 19892646]
51. Freeman KS, Spyridakis JH. Effect of Contact Information on the Credibility of Online Health Information. IEEE Trans Profess Commun 2009 Jun;52(2):152-166. [doi: 10.1109/TPC.2009.2017992]
52. Mackert M, Kahlor L, Tyler D, Gustafson J. Designing e-health interventions for low-health-literate culturally diverse parents: addressing the obesity epidemic. Telemed J E Health 2009 Sep;15(7):672-677 [FREE Full text] [doi: 10.1089/tmj.2009.0012] [Medline: 19694596]
53. Marton C. How Women With Mental Health Conditions Evaluate the Quality of Information on Mental Health Web sites: A Qualitative Approach. Journal of Hospital Librarianship 2010 Jul 28;10(3):235-250. [doi: 10.1080/15323269.2010.491422]
54. Kim H, Park S, Bozeman I. Online health information search and evaluation: observations and semi-structured interviews with college students and maternal health experts. Health Info Libr J 2011 Sep;28(3):188-199 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00948.x] [Medline: 21831218]
55. Colombo C, Mosconi P, Confalonieri P, Baroni I, Traversa S, Hill SJ, et al. Web search behavior and information needs of people with multiple sclerosis: focus group study and analysis of online postings. Interact J Med Res 2014 Jul 24;3(3):e12 [FREE Full text] [doi: 10.2196/ijmr.3034] [Medline: 25093374]
56. Lederman R, Fan H, Smith S, Chang S. Who can you trust? Credibility assessment in online health forums. Health Policy and Technology 2014 Mar;3(1):13-25. [doi: 10.1016/j.hlpt.2013.11.003]
57. McPherson AC, Gofine ML, Stinson J. Seeing is believing? A mixed-methods study exploring the quality and perceived trustworthiness of online information about chronic conditions aimed at children and young people. Health Commun 2014 Oct;29(5):473-482. [doi: 10.1080/10410236.2013.768325] [Medline: 24099647]
58. Payton FC, Kvasny L, Kiwanuka-Tondo J. Online HIV prevention information: How black female college students are seeking and perceiving. Internet Research 2014 Jul 29;24(4):520-543. [doi: 10.1108/IntR-09-2013-0193]
59. Briones R. Harnessing the Web: How E-Health and E-Health Literacy Impact Young Adults' Perceptions of Online Health Information. Med 2 0 2015 Dec 31;4(2):e5 [FREE Full text] [doi: 10.2196/med20.4327] [Medline: 26721292]
60. Rennis L, McNamara G, Seidel E, Shneyderman Y. Google it!: Urban community college students? use of the Internet to obtain self-care and personal health information. Coll Stud J 2015;49(3):414-426.
61. Subramaniam M, St Jean B, Taylor NG, Kodama C, Follman R, Casciotti D. Bit by bit: using design-based research to improve the health literacy of adolescents. JMIR Res Protoc 2015 May 29;4(2):e62 [FREE Full text] [doi: 10.2196/resprot.4058] [Medline: 26025101]
62. Cunningham A, Johnson F. Exploring trust in online health information: a study of user experiences of patients.co.uk. Health Info Libr J 2016 Dec;33(4):323-328. [doi: 10.1111/hir.12163] [Medline: 27870321]
63. Diviani N, van den Putte B, Meppelink CS, van Weert JC. Exploring the role of health literacy in the evaluation of online health information: Insights from a mixed-methods study. Patient Education and Counseling 2016 Jun;99(6):1017-1025. [doi: 10.1016/j.pec.2016.01.007]
64. Sillence E, Hardy C, Medeiros LC, LeJeune JT. Examining trust factors in online food risk information: The case of unpasteurized or 'raw' milk. Appetite 2016 Apr 01;99:200-210. [doi: 10.1016/j.appet.2016.01.010] [Medline: 26792772]
65. Alsem MW, Ausems F, Verhoef M, Jongmans MJ, Meily-Visser JMA, Ketelaar M. Information seeking by parents of children with physical disabilities: An exploratory qualitative study. Res Dev Disabil 2017 Jan;60:125-134. [doi: 10.1016/j.ridd.2016.11.015] [Medline: 27914304]
66. Champlin S, Mackert M, Glowacki EM, Donovan EE. Toward a Better Understanding of Patient Health Literacy: A Focus on the Skills Patients Need to Find Health Information. Qual Health Res 2017 Jul;27(8):1160-1176. [doi: 10.1177/1049732316646355] [Medline: 27179023]
67. Cusack L, Desha LN, Del Mar CB, Hoffmann TC. A qualitative study exploring high school students' understanding of, and attitudes towards, health information and claims. Health Expect 2017 Dec;20(5):1163-1171 [FREE Full text] [doi: 10.1111/hex.12562] [Medline: 28475266]

68. Peddie KA, Kelly-Campbell RJ. How people with hearing impairment in New Zealand use the Internet to obtain information about their hearing health. Computers in Human Behavior 2017 Aug;73:141-151. [doi: 10.1016/j.chb.2017.03.037]
69. Scantlebury A, Booth A, Hanley B. Experiences, practices and barriers to accessing health information: A qualitative study. Int J Med Inform 2017 Dec;103:103-108. [doi: 10.1016/j.ijmedinf.2017.04.018] [Medline: 28550995]
70. Rieh SY. Judgment of information quality and cognitive authority in the Web. J Am Soc Inf Sci 2002;53(2):145-161. [doi: 10.1002/asi.10017]
71. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the Internet. Med Inform Internet Med 2006 Mar;31(1):45-52. [doi: 10.1080/14639230600552601] [Medline: 16754366]
72. Eastin M. Credibility Assessments of Online Health Information: The Effects of Source Expertise and Knowledge of Content. J Comput-Mediat Commun 2006 Jun 23;6(4). [doi: 10.1111/j.1083-6101.2001.tb00126.x]
73. Hovland C, Janis I, Kelley H. Communication and persuasion; psychological studies of opinion change. New Haven, CT: Yale University Press; 1953.
74. Burkell J. Health Information Seals of Approval: What do they Signify? Information, Communication & Society 2004 Jan;7(4):491-509. [doi: 10.1080/1369118042000305610]
75. Allam A, Sak G, Diviani N, Schulz PJ. Do quality markers for health websites affect the perception of vaccination webpages? Computers in Human Behavior 2017 Feb;67:273-281. [doi: 10.1016/j.chb.2016.11.003]

Abbreviations

CDC: Centers for Disease Control and Prevention
CINAHL: Cumulative Index to Nursing and Allied Health
NIH: National Institutes of Health

Edited by G Eysenbach; submitted 19.10.18; peer-reviewed by L Sbaffi, A Jadhav, B van den Putte; comments to author 07.02.19; revised version received 19.03.19; accepted 08.04.19; published 02.05.19.

Please cite as:
Sun Y, Zhang Y, Gwizdka J, Trace CB
Consumer Evaluation of the Quality of Online Health Information: Systematic Literature Review of Relevant Criteria and Indicators
J Med Internet Res 2019;21(5):e12522
URL: https://www.jmir.org/2019/5/e12522/
doi:10.2196/12522
PMID:31045507

Review

Credibility, Accuracy, and Comprehensiveness of Internet-Based Information About Low Back Pain: A Systematic Review

Giovanni Ferreira[1*], MPhil; Adrian C Traeger[1*], PhD; Gustavo Machado[1*], PhD; Mary O'Keeffe[1*], PhD; Christopher G Maher[1*], DSc

Institute for Musculoskeletal Health, Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[*]all authors contributed equally

Corresponding Author:
Giovanni Ferreira, MPhil
Institute for Musculoskeletal Health, Sydney School of Public Health, Faculty of Medicine and Health
The University of Sydney
Level 10 North, King George V Building
Sydney, 2050
Australia
Phone: 61 0286279961
Email: giovanni.ferreira@sydney.edu.au

Abstract

Background: Low back pain (LBP) affects millions of people worldwide, and misconceptions about effective treatment options for this condition are very common. Websites sponsored by organizations recognized as trustworthy by the public, such as government agencies, hospitals, universities, professional associations, health care organizations and consumer organizations are an important source of health information for many people. However, the content of these websites regarding treatment recommendations for LBP has not been fully evaluated.

Objective: This study aimed to determine the credibility, accuracy, and comprehensiveness of treatment recommendations for LBP in noncommercial, freely accessible websites.

Methods: We conducted a systematic review of websites from government agencies, hospitals, universities, professional associations, health care organizations and consumer organizations. We conducted searches on Google. Treatment recommendations were coded based on the 2016 National Institute for Health and Care Excellence (NICE) guidelines and the 2017 American College of Physicians guideline on LBP. Primary outcomes were credibility of the website (4-item Journal of the American Medical Association benchmark), accuracy (proportion of website treatment recommendations that were appropriate), and comprehensiveness of website treatment recommendations (proportion of guideline treatment recommendations that were appropriately covered by a website).

Results: We included 79 websites from 6 English-speaking countries. In terms of credibility, 31% (25/79) of the websites clearly disclosed that they had been updated after the publication of the NICE guidelines. Only 43.28% (487/1125) website treatment recommendations were judged as accurate. Comprehensiveness of treatment recommendations correctly covered by websites was very low across all types of LBP. For acute LBP, an average of 28% (4/14) guideline recommendations were correctly covered by websites. Websites for radicular LBP were the least comprehensive, correctly covering an average of 16% (2.3/14) recommendations.

Conclusions: Noncommercial freely accessible websites demonstrated low credibility standards, provided mostly inaccurate information, and lacked comprehensiveness across all types of LBP.

(J Med Internet Res 2019;21(5):e13357) doi:10.2196/13357

KEYWORDS

medical informatics; low back pain; patient portals; systematic review; consumer health information

Introduction

Background

Low back pain (LBP) is the condition that accounts for the greatest burden of disability worldwide [1]. The global burden of LBP is expected to rise with an ageing population, leading to increased pressure on health systems [2]. The literature on LBP is vast, with a considerable amount of evidence on risk factors [3], prognosis [4], and effectiveness of treatment for LBP [5,6]. However, misconceptions in the general population and among health professionals [7] about LBP are very common [8,9]. A possible solution to reduce the inconsistency between beliefs of the general public and evidence-based information is to disseminate appropriate information through easily accessible means, such as the internet.

The internet has rapidly become an important source of health information [10]. In fact, there is evidence that more people seek health information first on the internet than with health care providers [11]. Recent data indicate that 78% [12] of Australians and 61% of people in the United Kingdom [13] used the internet to obtain health information in the previous 12 months [12,13]. People engaging with internet-based health information do so for several reasons, including making treatment decisions, supplementing information provided by a health professional, and self-managing a health condition [14]. The ready availability of internet-based health information may have both favorable and unfavorable consequences: although it empowers people to actively participate in their health care, the poor quality of information across many health conditions [15-17] calls into question its usefulness.

Government agencies and health care and professional organizations now sponsor websites containing health information for a broad range of health topics, including LBP. Websites hosted by those organizations are usually seen as more credible than other sources of internet-based information for people with LBP [18]. Previous studies examining the content of internet-based information about LBP have limitations, such as investigating limited samples and including websites with commercial purposes [19-21].

Objectives

Our objective was to assess the credibility, accuracy, and comprehensiveness of treatment recommendations for LBP given by websites sponsored by sources typically recognized as trustworthy by the public.

Methods

Eligibility Criteria

We reported this review following guidance provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines where possible [22]. We identified trustworthy websites to be those from government agencies, nonprofit nongovernmental organizations, hospitals, professional societies, universities, and consumer organizations. We sought websites from 6 major English-speaking countries: Australia, Canada, New Zealand, South Africa, the United Kingdom, and the United States. Websites had to present content about treatments for either acute, persistent, or radicular LBP. Websites that had links to other forms of content presentation, such as booklets, leaflets, or brochures, were also included. We excluded websites published in languages other than English, websites only containing information about aspects other than treatment of LBP (eg, anatomy of the lumbar spine), and those requiring any type of password or membership. We excluded websites that did not provide at least 1 clear recommendation for either acute, persistent, or radicular LBP.

Search Strategy

We used Google to search for noncommercial freely accessible websites presenting LBP information. We conducted searches on 5 February and updated it on July 20, 2018. We used Google AdWords to identify the most popular terms on Google related to LBP worldwide. To increase search specificity, we conducted searches on each country's dedicated Google website (Multimedia Appendix 1). Furthermore, 1 researcher (GF) performed the initial screening of websites using the Google Chrome Web browser. Before starting every new search, we cleared the browsing data. We screened the first 50 records from each search by looking at titles, snippets that followed each title, and URLs for each retrieved Web page. All links deemed relevant by the first reviewer were collated to a Microsoft Excel sheet and then screened for eligibility by 1 of the 3 independent reviewers (AT, CM, and GM), with all discrepancies resolved by discussion.

Data Extraction and Coding

A reviewer (GF) extracted data into a spreadsheet and 1 of the 3 other reviewers (AT, GM, and MOK) cross-checked the data. We extracted information on characteristics of the website, such as the host, country, and type of LBP covered. Websites treatment recommendations were coded according to the recommendations from the 2016 National Institute for Health and Care Excellence (NICE) guidelines [5] and the 2017 American College of Physicians [6] guidelines for the management of LBP with or without sciatica. Furthermore, 3 authors (GF, AT, and CM) coded each guideline recommendation as having been (1) endorsed by at least 1 guideline; (2) dismissed by at least 1 guideline; and (3) subject to conflicting positions between the 2 guidelines. The accuracy of treatment recommendations given by websites was judged by concordance with the guideline recommendations and coded as follows:

- Appropriate endorsement: A recommendation given by a website to use a treatment that was endorsed by at least 1 guideline.
- Appropriate dismissal: A recommendation given by a website to avoid a treatment that was dismissed by at least 1 guideline.
- Inappropriate endorsement: A recommendation given by a website to use a treatment that was dismissed by at least 1 guideline.
- Inappropriate dismissal: A recommendation given by a website to avoid a treatment that was endorsed by at least 1 guideline.

- Endorsed: A recommendation given by a website to use a treatment that was not mentioned in either guideline.
- Dismissed: A recommendation given by a website to avoid a treatment that was not mentioned in either guideline.
- Unclear: A recommendation given by a website that was not clearly targeted to a specific LBP condition (eg, a recommendation to use skeletal muscle relaxants for LBP) or when a recommendation was vague in the description of the treatment (eg, *spinal injections* rather than *epidural corticosteroid injection* or *facet joint corticosteroid injection*).

Outcomes

Our primary outcomes were credibility, accuracy, and comprehensiveness of each website.

Credibility: We used the 4-item Journal of the American Medical Association (JAMA) benchmark [23] to assess the credibility of each website. The JAMA benchmark addresses 4 elements: (1) currency of information, (2) declaration of authorship, (3) presentation of a list of references, and (4) disclosure of any conflict of interest, funding, or sponsorship. Each item was categorized as yes, no, or not reported. We considered a website to be up-to-date if its date of publication or last update had been subsequent to the publication date of the 2016 NICE guidelines [5]. We considered authorship to be declared when single or multiple authors were listed or when authorship was attributed to a working group or an entity.

Accuracy: We defined accuracy as the number and proportion of website recommendations that were judged clear and accurate. Accurate recommendations were those coded as being appropriate endorsements, appropriate dismissals, and treatments dismissed by the website and not listed in either guideline. Inaccurate recommendations were those coded as being inappropriate endorsements, inappropriate dismissals, and treatments endorsed by the website but not listed in either guideline.

Comprehensiveness: We defined comprehensiveness as the number and proportion of guideline recommendations that were appropriately covered by a website. The comprehensiveness of a website was given by the ratio between the sum of appropriate endorsements and dismissals and the total number of recommendations in the guidelines for each different type of LBP.

Data Analysis

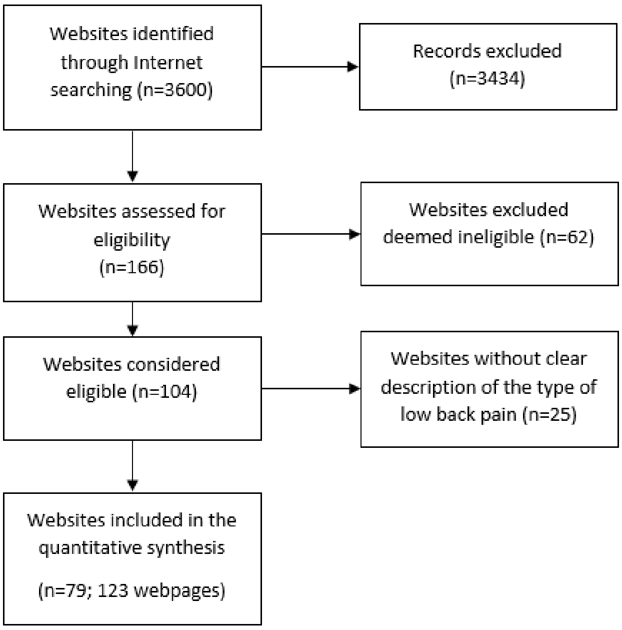
We presented data for acute, persistent, and radicular LBP separately. We used descriptive statistics to summarize credibility, accuracy, and comprehensiveness across the websites. Each item of the JAMA benchmark was presented individually. Data on accuracy were presented as the number and proportion of clear accurate recommendations, the number of clear accurate recommendations to use a treatment, and the number of clear accurate recommendations to avoid a treatment. Data on comprehensiveness were presented as the mean (SD) number of guideline recommendations correctly covered by websites and as the average proportion of website recommendations, as well as the mean (SD) number of guideline recommendations correctly covered by websites to use and to avoid a treatment.

Results

Selected Websites

We conducted 72 searches on Google, resulting in 3600 records to be screened. We excluded 3434 records by reading the titles on Google and retained 166 websites for eligibility assessment. Among these, 87 were ineligible. We, therefore, included 79 unique websites. As some websites had information for more than 1 type of LBP, a total of 123 Web pages were included in the final analysis (Figure 1).

Figure 1. Study flow diagram.



Characteristics of the Included Websites

Table 1 outlines the characteristics of the included websites. 53% (42/79) of the websites were from the United States, 24% (19/79) were from the United Kingdom, 17% (13/79) were from Australia, 4% (3/79) were from Canada, 1% (1/79) were from New Zealand, and 1% (1/79) were from South Africa. 43% (34/79) of websites were owned by hospitals, followed by websites from government agencies, which represented 25% (20/79) of websites. More information about the included websites is shown in Multimedia Appendix 2. Detailed information about the treatments covered by the websites is listed in Multimedia Appendix 3 to Multimedia Appendix 6.

Credibility of Websites

A total of 43 websites (54.4%, 43/79) disclosed their creation date or last update, and only 25 (31%, 25/79) of these websites had been updated after the publication of the NICE guidelines. Only 18 (22%, 18/79) websites provided a declaration of authorship. Only 21 (26%, 21/79) websites presented a list of references. Disclosure of any potential conflict of interest, funding source, or any form of sponsorship was only provided by 5 (6%, 5/79) websites (Table 1). Details on the assessment of credibility are shown in Multimedia Appendix 2.

Table 1. Characteristics of websites and credibility data (N=79).

Descriptive and credibility variables	Statistics, n (%)
Country	
Australia	13 (16.4)
Canada	3 (3.8)
New Zealand	1 (1.2)
South Africa	1 (1.2)
United Kingdom	19 (24.1)
United States	42 (53.3)
Type of website	
Consumer organization	2 (2.5)
Government	20 (25.3)
Hospital	34 (43.0)
Nongovernmental organization	8 (10.1)
Professional association or society	8 (10.1)
University	7 (8.8)
Type of low back pain[a]	
Acute	55 (69.6)
Persistent	29 (36.7)
Radicular	39 (49.3)
Updated after 2016 National Institute for Health and Care Excellence guidelines	
No	19 (24.1)
Yes	25 (31.6)
Not reported	35 (44.3)
Declaration of authorship	
No	61 (77.2)
Yes	18 (22.8)
List of references	
No	58 (73.4)
Yes	21 (26.6)
Disclosure of any conflict of interest, funding, or sponsorship	
Yes	5 (6.3)
Not reported	74 (93.7)

[a] The total is greater than 79 as some websites presented information for more than 1 type of low back pain (LBP).

Table 2. Accuracy of website treatment recommendations.

Condition	Number of recommendations	Number of unclear recommendations, n (%)	Number of accurate recommendations, n (%)	Number of clear accurate recommendations to use a treatment, n (%)	Number of clear accurate recommendations to avoid a treatment, n (%)
Acute LBP	452	98 (21.7)	228 (50.4)	187 (61.7)	41 (80.3)
Persistent LBP	402	105 (26.1)	154 (38.3)	116 (45.8)	38 (86.3)
Radicular LBP	271	28 (10.3)	105 (38.7)	86 (38.7)	17 (84.4)
Total	1125	231 (20.53)	487 (43.28)	389 (50.00)	96 (82.7)

Table 3. Comprehensiveness of website treatment recommendations. Recommendations that were conflicting between guidelines were not included in the assessment of comprehensiveness.

Condition	Number of recommendations in the guidelines	Guideline recommendations correctly covered by websites, mean (SD; percentage of total number of recommendations in guidelines)	Guideline recommendations to use a treatment correctly covered by websites, mean (SD; percentage of total number of recommendations in guidelines)	Guideline recommendations to avoid a treatment correctly covered by websites, mean (SD; percentage of total number of recommendations in guidelines)
Acute LBP[a]	14	4.0 (1.5; 28.6)	3.4 (1.3; 24.3)	0.6 (0.6; 4.3)
Persistent LBP	25	4.5 (2.8; 18.0)	4.0 (2.3; 16.0)	0.5 (1.1; 2.0)
Radicular LBP	14	2.3 (1.9; 16.4)	2.2 (1.9; 15.7)	0.1 (0.3; 0.7)

[a] LBP: low back pain.

Accuracy of Website Recommendations

Data for accuracy are presented in Table 2. In total, websites provided 1125 recommendations, with 487 (43.3%) being accurate, 407 (36.2%) inaccurate, and 231 (20.5%) unclear. Websites provided many more recommendations to use rather than to avoid a treatment (778 vs 116). Acute LBP had the highest number of recommendations among all the types of LBP. The proportion of accurate recommendations for acute LBP (50.4%) was higher than the proportion of accurate recommendations for persistent LBP (38.3%) and radicular LBP (38.7%). Advice to stay active was the most common treatment recommendation given by websites for acute LBP, having been endorsed by 45 out of 55 (81%) websites. Moreover, 28 out of 55 (50%) websites inappropriately endorsed paracetamol for acute LBP, and 6 out of 29 (20%) websites inappropriately endorsed opioids for persistent LBP (Multimedia Appendices 3-5).

Comprehensiveness of Website Recommendations

Details on the comprehensiveness of website recommendations are shown in Table 3. The proportion of guideline recommendations correctly covered by websites was higher for acute LBP (28.6%) compared with persistent LBP (18.0%) and radicular LBP (16.4%). Websites for radicular LBP were the least comprehensive, covering on average 2.3 (1.9) guideline recommendations. Across all types of LBP, the most comprehensive websites correctly covered only about 50% of the guideline recommendations—see Multimedia Appendix 6.

Discussion

Principal Findings

Treatment recommendations for LBP in websites from trustworthy sources failed to meet our benchmarks for credibility, provided a high proportion of inaccurate or unclear recommendations, and lacked comprehensiveness. In general, websites did not provide adequate resources for people to independently verify the truthfulness of the information provided. The accuracy of treatment recommendations was generally low across all different types of LBP. The lack of comprehensiveness across all websites was even more pronounced when covering guideline recommendations to avoid an ineffective treatment.

Comparison With Previous Work

Previous research had surveyed the accuracy of information only for a specific type of LBP (eg, acute LBP) [20], had not distinguished between types of LBP when analyzing recommendations from guidelines [19,21], and examined a limited sample of websites [19-21,24], including commercial websites [19,21]. We presented results for acute, persistent, and radicular LBP separately, given that some treatments known to be effective for 1 type of LBP might be ineffective for another or evidence might be currently lacking or be of questionable quality. That is the case for skeletal muscle relaxants, where evidence supports their use for acute LBP but benefits for persistent and radicular LBP are still uncertain [25].

In contrast to previous studies [19-21,26], we surveyed only websites from sources recognized as trustworthy by patients with LBP [18], such as government agencies, health care organizations, and universities. On the basis of the findings of previous studies showing that commercial websites were mostly of poor quality [19,21], our a priori hypothesis was that noncommercial freely accessible websites would have more comprehensive and accurate information compared with commercial websites [19,21]. This was not the case for many guideline-endorsed treatments. For example, although advice to stay active was recommended for acute LBP by 81% (45/55) of the websites, only 41% (12/29) and 46% (18/39) of websites recommended it for radicular and persistent LBP, respectively. Our findings suggest that websites of trusted sources are failing not only at conveying the accurate message on the benefits of guideline-endorsed [5,6] and first-line treatment [27] recommendations but also at dismissing ineffective treatment options such as bed rest for acute LBP, for which evidence on its ineffectiveness has long been known [28].

The large number of recommendations that were inaccurate and unclear found in our review supports findings from previous studies that people cannot obtain appropriate information about LBP on the internet [18]. For example, more than half of all treatment recommendations given by websites in our review were either inaccurate or unclear, which risks misleading the public [29]. In addition, evidence from the general population suggests that seeking health-related information on the internet is associated with increased health care utilization [30-32]. The fact that people often seek health information on the internet for LBP combined with the large amount of inaccurate and unclear information contained in these websites can potentially be driving people to seek unnecessary or ineffective care. On the contrary, credible and accurate internet-based information may contribute to reducing health care utilization. A recent study has attributed a 12% decrease in overall health care utilization, including LBP, to the launching of a website containing guideline-endorsed information maintained by the Dutch College of General Practitioners [33]. If one of the goals of internet-based health information is to reduce unnecessary consultations in primary care [33], these sources of internet-based information about LBP must provide patients the necessary means to make informed decisions about their health care. For this to happen, improving the credibility standards, as well as providing accurate and comprehensive treatment recommendations is necessary. When recommending treatments for LBP, websites must rely extensively on the evidence provided by high-quality clinical practice guidelines such as the NICE guidelines [5] and American College of Physicians [6] guidelines.

Limitations

The strengths of this study include using Google AdWords to develop search terms individuals actually use to find internet-based information about treatments for LBP. Searching the first 50 records of each search can also be considered a strength, as people searching for internet-based information usually do not look past the 10 or 20 first hits [34]. One potential limitation of our study was the use of Google as the sole Web browser to screen for websites. Nevertheless, our choice was based on the fact that Google is the most used search engine worldwide, Google has the best search validity (ie, returns links to websites that can be opened), and results from other engines usually highly overlap with those from Google [34]. Another limitation involves a very small number of inconsistencies (3 recommendations out of a total of 56, 5%) between the 2 guidelines. Nevertheless, that likely represents the uncertainty around the evidence for treatments for which there is currently no consensus, such as acupuncture [35]. We dealt with this limitation by choosing a more conservative approach and classified all endorsements and dismissals of treatments with inconsistent recommendations as being inappropriate.

Conclusions

Websites from government agencies, consumer organizations, hospitals, nongovernmental organizations, professional associations, and universities demonstrated low credibility standards, provided mostly inaccurate information, and lacked comprehensiveness across all types of LBP. Our findings highlight the need for these organizations to reformulate their treatment recommendations to reflect current evidence in the management of LBP.

Acknowledgments

GF is supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), Brazil, with a PhD scholarship. AT is supported by an National Health and Medical Research Council (NHMRC) Early Career Fellowship. GM is supported by an NHMRC Early Career Fellowship. MOK is supported by a European Commission Horizon 2020 Marie Skłodowska Curie post-doctoral fellowship. CM is supported by an NHMRC Principal Research Fellowship and NHMRC Program Grant.

Authors' Contributions

The corresponding author (GF) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms employed on Google.

[PDF File (Adobe PDF File), 76KB - jmir_v21i5e13357_app1.pdf]

Multimedia Appendix 2

Additional characteristics of the included websites.

[PDF File (Adobe PDF File), 106KB - jmir_v21i5e13357_app2.pdf]

Multimedia Appendix 3

Frequency (%) of websites endorsing or dismissing treatments mentioned in guidelines for acute low back pain (N=55).

[DOCX File, 15KB - jmir_v21i4e13357_app3.docx]

Multimedia Appendix 4

Frequency (%) of websites endorsing or dismissing treatments mentioned in guidelines for persistent low back pain (N=28).

[PDF File (Adobe PDF File), 68KB - jmir_v21i5e13357_app4.pdf]

Multimedia Appendix 5

Frequency (%) of websites endorsing or dismissing treatments mentioned in guidelines for radicular low back pain (N=39).

[PDF File (Adobe PDF File), 76KB - jmir_v21i5e13357_app5.pdf]

Multimedia Appendix 6

Accuracy and comprehensiveness of website recommendations.

[PDF File (Adobe PDF File), 175KB - jmir_v21i5e13357_app6.pdf]

References

1. GBD 2016 Disease Injury Incidence 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. Hurwitz EL, Randhawa K, Yu H, Côté P, Haldeman S. The Global Spine Care Initiative: a summary of the global burden of low back and neck pain studies. Eur Spine J 2018 Sep;27(Suppl 6):796-801. [doi: 10.1007/s00586-017-5432-9] [Medline: 29480409]
3. Parreira P, Maher CG, Steffens D, Hancock MJ, Ferreira ML. Risk factors for low back pain and sciatica: an umbrella review. Spine J 2018 Dec;18(9):1715-1721. [doi: 10.1016/j.spinee.2018.05.018] [Medline: 29792997]
4. Costa LD, Maher C, McAuley J, Hancock MJ, Herbert RD, Refshauge KM, et al. Prognosis for patients with chronic low back pain: inception cohort study. Br Med J 2009 Oct 06;339:b3829 [FREE Full text] [doi: 10.1136/bmj.b3829] [Medline: 19808766]
5. National Institute for Health and Care Excellence. National Institute for Health and Care Excellence. Manchester: National Institute for Health and Care Excellence; 2016. Low back pain and sciatica in over 16s: assessment and management URL: https://www.nice.org.uk/guidance/ng59 [accessed 2019-03-11] [WebCite Cache ID 76mpq301K]
6. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of Physicians. fNoninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017 Apr 04;166(7):514-530. [doi: 10.7326/M16-2367] [Medline: 28192789]
7. Darlow B, Fullen BM, Dean S, Hurley DA, Baxter GD, Dowell A. The association between health care professional attitudes and beliefs and the attitudes and beliefs, clinical management, and outcomes of patients with low back pain: a systematic review. Eur J Pain 2012 Jan;16(1):3-17 [FREE Full text] [doi: 10.1016/j.ejpain.2011.06.006] [Medline: 21719329]
8. Gross DP, Ferrari R, Russell AS, Battié MC, Schopflocher D, Hu RW, et al. A population-based survey of back pain beliefs in Canada. Spine (Phila Pa 1976) 2006 Aug 15;31(18):2142-2145. [doi: 10.1097/01.brs.0000231771.14965.e4] [Medline: 16915103]
9. Klaber Moffett JA, Newbronner E, Waddell G, Croucher K, Spear S. Public perceptions about low back pain and its management: a gap between expectations and reality? Health Expect 2000 Sep;3(3):161-168 [FREE Full text] [Medline: 11281925]
10. Fox S, Duggan M. Pew Research Center. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-03-11] [WebCite Cache ID 1552285578408097]
11. Volkman JE, Luger T, Harvey K, Hogan TP, Shimada SL, Amante D, et al. The National Cancer Institute's Health Information National Trends Survey [HINTS]: a national cross-sectional analysis of talking to your doctor and other healthcare providers

for health information. BMC Fam Pract 2014 Jun 06;15:111 [FREE Full text] [doi: 10.1186/1471-2296-15-111] [Medline: 24906558]

12. Public Opinion Polling. Research Australia. Australia Speaks! URL: https://researchaustralia.org/wp-content/uploads/2017/08/2017-Opinion-Poll-Digital.pdf [accessed 2019-03-11] [WebCite Cache ID 1552285420363140]
13. TNS Political & Social. European Comission. 2015 Mar 30. EUROPEAN CITIZENS' DIGITAL HEALTH LITERACY URL: http://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_404_en.pdf [accessed 2019-03-12] [WebCite Cache ID 76oWMtntF]
14. Ramsey I, Corsini N, Peters MD, Eckert M. A rapid review of consumer health information needs and preferences. Patient Educ Couns 2017 Sep;100(9):1634-1642. [doi: 10.1016/j.pec.2017.04.005] [Medline: 28442155]
15. Fisher JH, O'Connor D, Flexman AM, Shapera S, Ryerson CJ. Accuracy and reliability of internet resources for information on idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2016 Dec 15;194(2):218-225. [doi: 10.1164/rccm.201512-2393OC] [Medline: 26849779]
16. Meric F, Bernstam EV, Mirza NQ, Hunt KK, Ames FC, Ross MI, et al. Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites. Br Med J 2002 Mar 09;324(7337):577-581 [FREE Full text] [Medline: 11884322]
17. Storino A, Castillo-Angeles M, Watkins AA, Vargas C, Mancias JD, Bullock A, et al. Assessing the accuracy and readability of online health information for patients with pancreatic cancer. JAMA Surg 2016 Dec 01;151(9):831-837. [doi: 10.1001/jamasurg.2016.0730] [Medline: 27144966]
18. Nielsen M, Jull G, Hodges PW. Information needs of people with low back pain for an online resource: a qualitative study of consumer views. Disabil Rehabil 2014;36(13):1085-1091. [doi: 10.3109/09638288.2013.829532] [Medline: 24001260]
19. Butler L, Foster NE. Back pain online: a cross-sectional survey of the quality of web-based information on low back pain. Spine (Phila Pa 1976) 2003 Feb 15;28(4):395-401. [doi: 10.1097/01.BRS.0000048497.38319.D3] [Medline: 12590218]
20. Hendrick PA, Ahmed OH, Bankier SS, Chan TJ, Crawford SA, Ryder CR, et al. Acute low back pain information online: an evaluation of quality, content accuracy and readability of related websites. Man Ther 2012 Aug;17(4):318-324. [doi: 10.1016/j.math.2012.02.019] [Medline: 22464886]
21. Li L, Irvin E, Guzmán J, Bombardier C. Surfing for back pain patients: the nature and quality of back pain information on the Internet. Spine (Phila Pa 1976) 2001 Mar 01;26(5):545-557. [Medline: 11242383]
22. 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 healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
23. 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. J Am Med Assoc 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
24. Samanci Y, Celik SE. Low back pain and internet: infopollution. Turk Neurosurg 2017;27(5):804-808 [FREE Full text] [doi: 10.5137/1019-5149.JTN.18521-16.1] [Medline: 27858395]
25. Abdel Shaheed C, Maher CG, Williams KA, McLachlan AJ. Efficacy and tolerability of muscle relaxants for low back pain: systematic review and meta-analysis. Eur J Pain 2017 Dec;21(2):228-237. [doi: 10.1002/ejp.907] [Medline: 27329976]
26. Black NM, Sullivan SJ, Mani R. A biopsychosocial understanding of lower back pain: content analysis of online information. Eur J Pain 2018 Dec;22(4):728-744. [doi: 10.1002/ejp.1158] [Medline: 29280231]
27. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Dec 09;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]
28. Malmivaara A, Häkkinen U, Aro T, Heinrichs ML, Koskenniemi L, Kuosma E, et al. The treatment of acute low back pain--bed rest, exercises, or ordinary activity? N Engl J Med 1995 Feb 09;332(6):351-355. [doi: 10.1056/NEJM199502093320602] [Medline: 7823996]
29. Benigeri M, Pluye P. Shortcomings of health information on the Internet. Health Promot Int 2003 Dec;18(4):381-386. [Medline: 14695369]
30. Hone T, Palladino R, Filippidis FT. Association of searching for health-related information online with self-rated health in the European Union. Eur J Public Health 2016 Dec;26(5):748-753. [doi: 10.1093/eurpub/ckw022] [Medline: 26944894]
31. Suziedelyte A. How does searching for health information on the internet affect individuals' demand for health care services? Soc Sci Med 2012 Nov;75(10):1828-1835. [doi: 10.1016/j.socscimed.2012.07.022] [Medline: 22884947]
32. Ariza M, Guerrisi C, Souty C, Rossignol L, Turbelin C, Hanslik T, et al. Healthcare-seeking behaviour in case of influenza-like illness in the French general population and factors associated with a GP consultation: an observational prospective study. BJGP Open 2018 Jan;1(4):- [FREE Full text] [doi: 10.3399/bjgpopen17X101253] [Medline: 30564694]
33. Spoelman WA, Bonten TN, de Waal MW, Drenthen T, Smeele IJ, Nielen MM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Dec 09;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]

34. Wang LP, Wang JX, Wang M, Li Y, Liang YC, Xu D. Using internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012 May 16;14(3):e74 [FREE Full text] [doi: 10.2196/jmir.1943] [Medline: 22672889]
35. O'Connell NE, Cook CM, Wand BM, Ward SP. Clinical guidelines for low back pain: a critical review of consensus and inconsistencies across three major guidelines. Best Pract Res Clin Rheumatol 2016 Dec;30(6):968-980. [doi: 10.1016/j.berh.2017.05.001] [Medline: 29103554]

Abbreviations

JAMA: Journal of the American Medical Association
LBP: low back pain
NICE: National Institute for Health and Care Excellence

Edited by G Eysenbach; submitted 10.01.19; peer-reviewed by M Rethlefsen, M Thomas; comments to author 27.02.19; revised version received 01.03.19; accepted 03.03.19; published 07.05.19.

Please cite as:
Ferreira G, Traeger AC, Machado G, O'Keeffe M, Maher CG
Credibility, Accuracy, and Comprehensiveness of Internet-Based Information About Low Back Pain: A Systematic Review
J Med Internet Res 2019;21(5):e13357
URL: https://www.jmir.org/2019/5/e13357/
doi:10.2196/13357
PMID:31066689

Original Paper

Case of Paradoxical Cultural Sensitivity: Mixed Method Study of Web-Based Health Informational Materials About the Human Papillomavirus Vaccine in Israel

Nour Abed Elhadi Shahbari[1], BSc, MHA; Anat Gesser-Edelsburg[1], PhD; Gustavo S Mesch[2], PhD

[1]School of Public Health, University of Haifa, Haifa, Israel
[2]Department of Sociology, University of Haifa, Haifa, Israel

Corresponding Author:
Nour Abed Elhadi Shahbari, BSc, MHA
School of Public Health
University of Haifa
199 Aba Khoushy Ave
Mount Carmel
Haifa, 3498838
Israel
Phone: 972 504847217
Email: nabede01@campus.haifa.ac.il

Abstract

Background: Designing web-based informational materials regarding the human papillomavirus (HPV) vaccine has become a challenge for designers and decision makers in the health authorities because of the scientific and public controversy regarding the vaccine's safety and effectiveness and the sexual and moral concerns related to its use.

Objective: The study aimed to investigate how cultural sensitivity (CS) is articulated in the explanatory informational materials on the HPV vaccine that are posted on the websites of the Israeli health authorities. In addition, the study examined the effect of transparency on the expression of CS in the informational materials.

Methods: The study employed a quantitative and qualitative content analysis of the texts of explanatory informational materials published on the Arabic and Hebrew websites of the Israel Ministry of Health and the Clalit health maintenance organization (HMO).

Results: The findings revealed the differences in the dimensions of CS (based on the CS model by Resnicow) between the informational materials targeting the majority Jewish population and those targeting the minority Arab population. Indeed, the research findings point to a paradox. On the one hand, the materials appealing to the conservative Arab population exhibited CS, in that the sexual context of the vaccine was missing. On the other hand, analysis of Resnicow's deep dimensions showed that disregarding the sexual context does not allow the relevant target audience to reflect on the barriers and concerns. In addition, the way the information was provided exhibited a lack of transparency regarding the CS dimensions (surface and deep).

Conclusions: The public health authorities have 2 main objectives in the context of vaccinations. One is to raise the vaccination rates and the other is to provide full and culturally sensitive information to give the public the tools to make intelligent decisions. The findings of this study indicated that despite the high uptake rate for HPV vaccination in the Arab population, the health authorities did not exercise full transparency and CS in transmitting the association between engaging in sexual relations and the necessity of the vaccination. Thus, the major challenge for the health authorities is to find ways to implement the objective of communicating information about the vaccination in a way that is transparent and culturally sensitive, even if this raises questions and fears among the public deriving from their culture.

(J Med Internet Res 2019;21(5):e13373) doi:10.2196/13373

KEYWORDS

web-based health informational materials; HPV vaccine; Israel; cultural sensitivity; transparency; sexuality; quantitative analysis; qualitative content analysis; Hebrew and Arabic

Introduction

Health Promotion and Cultural Sensitivity

Individual and cultural differences (attitudes, intentions, beliefs, subjective norms, and the like) moderate the effects of health promotion interventions and influence how individuals make decisions [1].

Segmentation and determination of a specific target audience aim to identify the characteristics, needs, social structures, and perceived costs and benefits of the audience relative to specific health behavior. Identifying all these factors in the target audience can contribute to the development of an effective health communication [2].

Studies show that messages are more thoroughly processed and defined as convincing and effective if their content is adapted to the cognitive, emotional, and motivational characteristics that in most cases are determined by the cultural background of the target audience [3].

Culturally sensitive materials based on the target population’s cultural background can affect the understanding and effectiveness of health communication [4]. Moreover, culturally sensitive materials increase the awareness of recommended behavior, improve decision making, make messages more persuasive, and lead to deeper processing and understanding in the context of health promotion. Sensitive cultural communication reduces gaps in health outcomes by making messages equally meaningful and effective [5].

When the information provided matches cultural values, the content is likely to lead to deeper reflection on the potential benefits and adverse effects, in turn leading to more appropriate and informed choices in the decision-making process [6].

Despite the emphasis on the importance of audience segmentation and theory-driven development of messages (cultural sensitivity [CS] in this study), health organizations do not always take the dimensions of CS into consideration when designing interventions and formulating messages and informational materials [7,8].

Resnicow’s Model of Cultural Sensitivity

Culture is a typical mediator in the decision-making process of individuals with respect to the recommended health behavior [9-11]. CS is defined as the extent to which the ethnic, cultural, and experiential norms, values, behavior patterns, and beliefs of the target population, as well as the historical, environmental, and social factors are expressed in the design, planning, and evaluation of the informational materials and programs for health promotion [12].

Many theories and models have discussed cultural competence and CS to various ethnic minority groups in different intervention programs [7,13-16]. The cultural competence or CS of materials used for health education and publicity is measured by different tools focusing on the use of culturally adapted language, graphics, and illustrations as well as the use of personages from the community [9,17].

Over the years, many studies have shown that CS is not limited to language or illustrations, suggesting that deep dimensions require additional measures. According to Resnicow's definition [15], CS is conceptualized in terms of 2 main measures: *deep dimensions* and *surface dimensions*.

Surface dimensions include the use of people, places, language, music, food, and clothing familiar to the target audience. The measure also includes channel identification (media), level of literacy, relevance of recommended behavior to the target audience, marketing sites (eg, schools and churches) appropriate for transmitting messages and intervention programs, and the use of a suitable communicator [12].

These surface dimensions are intended to make communication more effective by increasing its attractiveness, relevance, and familiarity as well as by increasing source reliability with respect to the target population [1].

Deep dimensions, in contrast, examine the environmental, cultural, social, psychological, and historical factors related to the target audience that influence their health behavior. The measure makes specific reference to the arts, values, shared experiences, moral concerns, cultural barriers, perceptions regarding those who determine behavior, and perceptions regarding recommended behavior [15].

These dimensions are related to the understanding and interpretation of health communication, which depend on the models that explain the health and illness status of the target population [15].

Cervical Cancer and the Human Papillomavirus Vaccine

The human papillomavirus (HPV) is one of the most common sexually transmitted infections worldwide [18,19]; this contrasts with the situation in Israel, where morbidity and mortality rates caused by HPV infections are the lowest in the West [20]. Some studies indicate that ongoing infection with high-risk HPV strains can develop into cervical cancer. HPV can also cause other cancers (cancers of the vagina, vulva, penis, anus, rectum, and oropharynx) and genital warts [21-23]. High-risk HPV strains include HPV 16 and 18, which cause approximately 70% of cervical cancers [22,24].

In addition to the Papanicolaou (Pap) smear screening procedure for cervical cancer that has significantly reduced the morbidity and mortality rates for cervical cancer in many western countries [25], the HPV vaccine is the first vaccine targeted at preventing cervical cancer [26].

According to the World Health Organization (WHO), since 2006, around 100 nations worldwide have approved the HPV vaccine, and as of 2012, approximately 40 countries have included the vaccine in their immunization programs [27]. HPV vaccine programs vary from country to country with respect to the type of vaccine, implementation strategies, compliance rates, and payment policies [28-30].

According to recent data from the Israel Ministry of Health [20], the age-adjusted incidence rates (per 100,000) of invasive cervical cancer in 2012 were 5.08 in Jewish women and 2.97 in Arab women. The age-adjusted incidence rates (per 100,000) of stage 0 cervical cancer in 2012 were 22.5 in Jewish women and 4.2 in Arab women. The age-adjusted mortality rates (per 100,000) from cervical cancer in 2012 were 1.50 in Jewish women and 0.94 in Arab women. The data show that the rates of HPV infections are even lower.

Starting in 2007, the HPV vaccine was available in Israel in exchange for payment. In December 2012, Israeli health authorities decided to include the HPV vaccine in the medical services basket of medical services provided free of charge. The vaccine included in the basket is for women, boys, and men up to the age of 26 years, who constitute the population at risk for HPV infection [31]. Starting in 2013, the vaccine was included as a part of the routine vaccines given in school to girls in the eighth grade and was later extended to include boys [31,32]. According to the Ministry of Health [20], in the 2016 school vaccination campaign, the rate of HPV vaccine uptake among the Arab population reached 84% (96% in the Bedouin sector) compared with 40% for the Jewish population.

Transparency Regarding the Human Papillomavirus Vaccine

Experts from across the globe recommend additional monitoring of HPV vaccine effectiveness and safety [33]. This recommendation derives from the scientific controversy regarding the effectiveness and safety of this vaccine since its inclusion in vaccination programs worldwide. On the one hand, a large body of studies indicate that the vaccine is effective. For example, various meta-analyses and research studies have attested to the vaccine's high immunological effectiveness, approaching 93% to 100%, against the papilloma virus and other precancerous cervical lesions [28,34-38]. In addition, recent research emphasizes the effectiveness of the HPV vaccine in reducing preinvasive cervical disease and increasing clinical herd protection in unvaccinated women [39]. Yet, the literature still includes studies pointing to a variety of factors that can affect the vaccine's effectiveness, such as age, the type of virus strain, and the individual's immune system [40-44].

On the other hand, other studies point to negative consequences after vaccine uptake. Similar to any other vaccine, the HPV vaccine can cause major pain at the injection site as well as fainting and dizziness [45-47]. However, there have also been reports of more serious side effects [48-59]. With the rising number of these reports, experts around the world have been calling for more studies and improved postmarketing surveillance [56,59].

Balanced health information regarding the HPV vaccine should include 3 main components: (1) completeness—description of the risk of cervical cancer and the vaccine's advantages and disadvantages; (2) transparency—presentation of all the vaccine-related risks in absolute numbers and not as relative figures; and (3) evidence-based correctness—providing information based on science [60].

Transparency is an essential component in effective health communication and in designing messages and explanatory materials intended to promote various health behaviors [61-64]. Studies that examined transparency in providing information on the HPV vaccine to target populations were conducted mainly in Germany [60,65]. These studies emphasized that the German health agencies and the German media as well provided only partial information regarding the risks of cervical cancer and the papilloma virus and used statistics that were not transparent in describing the vaccine's effectiveness and safety. The statistics in the marketing materials provided inaccurate numbers regarding the vaccine's effectiveness, showing absolute effectiveness. Moreover, most of the marketing leaflets did not mention the vaccine's side effects [60,65]. Most studies examining decision making in the context of the HPV vaccine emphasize the importance of transmitting information that is fully transparent and understandable to enable parents to sign the informed consent forms for the vaccination [66-68]. Some studies indicate that people do not always read the forms to the end or do not understand what they read, so, in essence, they sign the forms without their informed consent [66,69,70]. Moreover, most of the scientific literature indicates that in the case of the HPV vaccine, the parents indicate that their knowledge about the vaccine is limited [71-74]. This makes the informed-consent decision-making process to the HPV vaccine more challenging [71,75-77]. The studies show that people often make decisions about HPV vaccination without having sufficient knowledge as required to make age-appropriate health decisions [74,77-82].

Furthermore, providing complete, comprehensive, and transparent information about HPV vaccination helps parents in making a decision about whether to vaccinate their children [66,67]. Parental decision making regarding the vaccination is dependent upon perceptions of the risks with respect to their children's sexual activity [72,74,82,83]. Studies indicate that parents who perceived the risk of sexual activity among their children to be high were more willing to give their children the vaccination. In contrast, parents who perceived the risk of sexual activity among their children to be low because of their cultural or religious background were less willing to give their children the vaccine [72,73,75,79].

Human Papillomavirus Vaccine and Sexuality

The HPV vaccine is intended to protect the individual from contracting the papilloma virus, which is transmitted through sexual intercourse [19]. The *Food and Drug Administration* recommends vaccinating boys and girls against the virus during adolescence, before they become sexually active [84]. Research studies have raised parental concerns and fears regarding the vaccine from different population groups worldwide because of their belief that the vaccine legitimizes sexual activity and promotes sexual abuse [85-87]. Some of these concerns were derived from media coverage that represented giving the vaccine as an act that encourages sexual relations, sexual promiscuity, and sexual abuse [88,89]. Despite these parental concerns, several scientific research studies showed that the HPV vaccine has no impact on the sexual behavior of adolescents [90-93]. Nonetheless, the issue of sexual relations in the context of giving the vaccine increases sensitivity regarding the vaccine [94], particularly among conservative minority population groups such as the Arab minority in Israel [95,96].

Arab Population in Israel and the Context of Sexuality

Constituting around 20% of the general population, the Arab population in Israel is a minority with unique national, religious, linguistic, and cultural attributes. Relations between Arabs and Jews are mainly minority-majority relations. Israel must deal with the internal problems of minority groups for whom issues of justice and equality are central to their relations within the state [97].

The socioeconomic disparities between the Jews and the Arabs are the result of a historical-geographic reality in which most of the Arab population lives in the spatial periphery, thus limiting their access to the country's main labor markets and services. These disparities find expression in unequal allocation of resources for economic development and infrastructure in the local authorities and unequal employment opportunities for Arabs [97].

In addition, the culture of Arab society differs from that of Jewish society. These cultural differences are related to language, religion, nationality, cultural heritage, family structure, values, and primary lifestyles. Traditionalism dominates Arab society, in that the religion is a central component of cultural life (religious self-definition and prayer frequency). Moreover, the Arab population still lives within the framework of a conservative family and community and is characterized by less openness, autonomy, development of personal inclinations, freedom, and the like. This is not the case among the predominantly secular Jewish population, which espouses individualism, competitiveness, achievements, ambitions, career development, materialism, and consumerism [97].

Even though the Arabs in Israel usually live in separate localities marked by lower educational and socioeconomic levels, their exposure to Jewish society is high, which has led to social, cultural, and economic processes of change that have made Arab society more modern. This modernization is characterized by a gradual shift from a collectivist to an individualist orientation, manifested in a rise in the age at first marriage, a drop in the number of children, an increase in the educational level, use of preventive health services, and more years of education among women than among men [98].

Yet, Arab culture is still conservative regarding sexuality and sees sexual relations before marriage as improper and prohibited. Indeed, in most Arab communities, such relations are taboo [99-101]. One of the typical barriers preventing parents from the Arab minority in Western countries from vaccinating their daughters against the papilloma virus is their concern of encouraging unlimited sexual behavior and sexual promiscuity [102-104].

To the best of our knowledge, no studies have examined the topic of CS in explanatory materials geared at promoting the administration of HPV vaccine. Therefore, the purpose of this study was to investigate the following question: How is CS articulated in the explanatory informational materials about the HPV vaccine posted on the websites of Israeli health authorities?

To investigate how CS is articulated in the explanatory informational materials about the HPV vaccine posted on the websites of the Israeli health authorities, we examined 3 specific questions:

1. How are the surface and deep dimensions of CS from Resnicow's model articulated in the explanatory informational materials posted in both languages (Hebrew and Arabic)?
2. Are the surface and deep dimensions of CS articulated differently in materials intended for the Jewish population than in those intended for the Arab population?
3. Is there a difference between the materials written in Arabic and those written in Hebrew with respect to transparency in transmitting complete information?

Methods

Study Sample

The research method entailed quantitative and qualitative content analysis of the texts of explanatory informational materials published on the Arabic and Hebrew websites of the Israel Ministry of Health and the Clalit HMO.

The research was conducted under the approval of the Faculty of Social Welfare and Health Sciences Ethics Committee for research with human subjects at the University of Haifa (approval no.118/16). The researchers analyzed these materials for 6 months (October 2017-March 2018).

The sample comprised 18 instances of explanatory materials published on the Hebrew and Arabic websites of the Clalit HMO and the Ministry of Health and updated after the vaccine was included in the basket of health services (after 2012). These instances had identical content and headings in Hebrew and in Arabic (9 in each language).

Excluded from the study were materials published on sites that did not have an Arabic website, websites that had not been updated after the vaccine was included in the Israeli basket of health services (2012) and those that had original materials from the WHO that were not translated into Hebrew or Arabic and were omitted because of the Arab population's low level of literacy in English [105].

The rationale behind selecting all the explanatory informational materials for the HPV vaccine after it was included in the medical services basket is as follows. Although the vaccine is free and available in school settings, parents are still aware of the scientific controversy surrounding the vaccine's efficacy and effectiveness as publicized in the media. Therefore, they still seek information about the vaccine, its efficacy, and its safety in deciding whether to vaccinate their children [33].

The following health authority websites and the HPV vaccine materials were retrieved:

1. Ministry of Health: Human Papilloma Virus Vaccine, Information Sheet Before and After Administering HPV Vaccine, HPV Vaccine for Eighth Grade Girls, HPV Vaccine for Eighth Grade Boys, Vaccine to Protect Against Cervical Cancer Caused by Papilloma Virus, PowerPoint Presentation about HPV Vaccine.

2. Clalit HMO: Cervical Cancer (first document), Cervical Cancer (second document).

The following keywords were examined: cervical cancer, vaccine against cervical cancer, vaccine against the human papilloma virus, human papilloma virus, condyloma, and genital warts.

It is important to clarify that these sites address most of the Jews and Arabs living in Israel. Most of the Jewish population is secular and, therefore, tends toward more liberal views, except for the conservative ultraorthodox Jews who are in the minority [106]. Similarly, the majority of the Arabs in Israel are Muslim (97% of the Arabs in Israel), who are indeed defined as conservative [107].

Analysis

The explanatory materials were analyzed using 2 parallel methods: quantitative content analysis based on index-coding criteria and qualitative content analysis to interpret and explain the index score.

Cultural Sensitivity Index Coding

Quantitative content analysis is based on coding. According to this method, the analysis begins by identifying relevant concepts based on a theory, model, or previous studies that form the basis of the index-coding scheme. The next step is to determine the index-coding scheme by developing classification rules to assign coding units to specific categories or concepts. The coding guide ensures systematic encoding and allows data recovery. In addition to the coding guide, a coding form is also used to record the details of the codes that apply to the data during the coding process [108].

Building the Cultural Sensitivity Index

On the basis of Resnicow's theoretical definitions of 2 dimensions, deep dimensions and surface dimensions [15], we constructed an index of appropriate categories that served in analyzing the CS of the explanatory materials intended for the Jewish and Arab populations in Israel.

The index was constructed using 2 dimensions: deep and surface dimensions. For each of these dimensions, we designed nominal categorical indicators that can take 1 of 2 values: yes/no (see Textboxes 1 and 2). If the explanatory materials do not provide full and transparent information with respect to a specific indicator—whether for surface dimensions or for deep dimensions—the indicator is scored as *no*. Conversely, if the explanatory materials provide full and transparent information, the indicator is scored as *yes*.

Textbox 1. Surface dimension index: Used to analyze the explanatory health informational materials of the human papillomavirus vaccine intended for the Arab and Jewish population in Israel with nominal categorical indicators (yes/no).

Surface dimension categories:

- Is the target population mentioned in the title or the content, either directly or indirectly?
- Do the materials use people who are appropriate for, familiar to, and acceptable to the target population?
- Do the materials use the mother tongue of the target population?
- Are the materials formulated in a manner appropriate to the literacy level of the target population?
- Do the materials explain the relevance of cervical cancer to the target population?
- Do the materials explain the effectiveness of the human papillomavirus (HPV) vaccine against cervical cancer in the target population (safety, immunity, etc)?
- Do the materials explain the implications of the HPV vaccine against cervical cancer in the target population (side effects, complications, etc)?
- Are appropriate media channels used that are familiar to the target population?
- Are the materials distributed in places most suitable for the target population (churches, schools, medical clinics versus websites)?
- Do the materials use a reliable communicator who is appropriate for the target population?
- Was the organization that transmitted the information from the same ethnic group (Arab/Jewish population)?

Textbox 2. Deep dimension index: Used to analyze the explanatory health informational materials of the human papillomavirus vaccine intended for the Arab and Jewish population in Israel with nominal categorical indicators (yes/no).

Deep dimension categories:

- Do the materials make reference to the moral concerns of the target population regarding the human papillomavirus (HPV) vaccine (increased sexual relations, encouraging promiscuity)?
- Do the materials make reference to the cultural concepts regarding cervical cancer and the HPV vaccine (preference for women practitioners for Pap smears or other gynecological exams)?
- Do the materials make reference to social concepts regarding cervical cancer and the HPV vaccine (cervical cancer is a female disease, reasons for giving the vaccine to males)?
- Do the materials refer to environmental concepts (influence of the Jewish/Arab population, socioeconomic status)?
- Do the materials make reference to value concepts (religion, fatalism, and sexual relations before marriage)?

Cultural Sensitivity Index Scores

The specific dimension score (%) was calculated by the ratio between the number of *yes* answers on each specific measure and the total indicators for each dimension. For example, to calculate the score for the surface dimension, the number of *yes* answers is divided by 11 (the total number of surface dimension indicators).

To calculate the score for the deep dimension, the number of *yes* answers is divided by 5 (the total number of deep dimension indicators).

The overall CS index score (%) is calculated by the average sum of the surface dimension and the deep dimension score (CS=1/2 × surface dimension score + deep dimension score).

The index yields 2 scores: a specific index score and a general score. These 2 scores facilitate direct comparison of the explanatory materials in terms of CS scores and diagnosis of level of cultural appropriateness of materials targeting the Arab and Jewish populations. A score greater than 50% is considered an index of culturally sensitive materials, whereas a score less than 50% was considered an index of materials that are not culturally sensitive to the Arab or Jewish populations in Israel.

For the purpose of statistical comparison of CS between the materials written in Hebrew and those written in Arabic, each type of material underwent a separate chi-square test of independence. In addition, the chi-square test of independence was conducted to compare between the CS of the materials written in the 2 languages.

Qualitative Content Analysis

The second analysis method entailed qualitative content analysis [109,110], which focuses on the language characteristics and the content or contextual meaning of the text and provides knowledge and understanding of the phenomenon under study.

Qualitative content analysis relies on deciding what is going to be analyzed, forming the dataset to be collected, defining the unit or theme of analysis, classifying the content according to ideas that can be words or sentences, developing categories and creating subcategories, linking the interpretations with the basic theories of the study, defining categories with examples, drawing inferences on the basis of themes, and presenting the results [109,110].

In presenting the results related to each theme, we supported these ideas with typical secondary quotes that appeared frequently in most of the explanatory materials analyzed by the quantitative coding index.

In this study, qualitative content analysis is intended to explain and reinforce the researchers' coding and to interpret the absence of indicators in the quantitative CS index.

Moreover, this method is designed to view the subjective interpretation of the content through a systematic classification process of identifying themes and categories that may affect or intensify the issue of CS.

The content analysis was based on comparison with the scientific literature on cervical cancer and the HPV vaccine. In addition, the materials were analyzed with reference to concepts from the field of health and risk communication: *CS* [12,15] and *transparency* in communicating texts about health issues [111].

Reliability and Validity

This study is based on a variety of methods used to analyze rhetorical and aesthetic elements and scientific facts. The use of a combination of several research methods increases the reliability and validity of the findings [112-115] The CS index constructed by the researchers was based on Resnicow's model of CS [12,15], a valid theoretical model. The researchers, among them experts in health communication and CS worked together to investigate the cultural attributes of the Arab Jewish populations in Israel in the context of the HPV vaccine, adapt these attributes to the index, define the dimensions, and formulate the questions. The researchers are able to read and analyze texts written in Hebrew and in Arabic. One of the researchers is a member of the Arab minority group, Arabic is her native language, and she can distinguish cultural uses of the language. The coders performed their coding separately, then compared their coding and found it to be almost identical. The third researcher reviewed all the findings after coding and gave his analysis of CS. Coding that was not equivalent was resolved through discussions between the research team members. A high level of interrater reliability (98%) emerged. An intercoder agreement coefficient exceeding 0.7 suggests an acceptable level of reliability [116].

Furthermore, Scott's pi indicator was calculated by the reliability calculation for the masses, to reinforce the indicator of intercoder reliability. Scott's pi value was 0.9685, thus ensuring valid internal coding reliability [117].

In the qualitative analysis of the texts, the researchers reviewed materials posted on the websites of the Ministry of Health and the Clalit HMO. Together, they determined the inclusion criteria and decided which materials would be included in the research. The major categories and themes emerging from the web-based explanatory materials were discussed and agreed upon [109,110].

Results

Cultural Sensitivity Index Implementation

This section describes the CS index implementation scores for the HPV vaccine materials published in Arabic and Hebrew on the websites of the Ministry of Health and the Clalit HMO. Implementation of the index included coding the nominal categorical indicators of each dimensions (surface and deep) with 1 of 2 values: yes/no (Multimedia Appendix 1) and calculating 2 scores: specific scores for the deep and surface dimensions and an overall CS index score (Multimedia Appendix 2).

Surface Dimension–Specific Score

In most (78%) of the Arabic explanatory materials assessed, the *surface dimension* score was lower than 50%. Surface dimension score was higher than 50% for only 2 out of the 9 materials assessed. Conversely, surface dimension score was equal to or higher than 50% for all the materials (100%) written in Hebrew, starting from 73% up to 91% in some cases (33%).

The majority of the Arabic and Hebrew materials included some surface dimension indicators, such as mentioning the target audience (Arab population), use of suitable people (photos of Arab physicians recommending and explaining about the vaccine), language appropriate to the target audience (Arabic), compatibility with the literacy level (minimal use of scientific terminology), and presentation of the effectiveness and advantages of the recommended behavior (uptake of HPV vaccine).

Furthermore, most of the Arabic materials did not refer to the indicators of appropriate media channels (Arab-language television and radio stations or local websites targeting the Arab population), appropriate communicators (key Arab health and medical personnel), places suitable for transmitting the message (mosques, churches, and schools), and organizations acceptable to the minority target audience (educational originations and the Child and Family Care centers). In contrast, as mentioned, these indicators appeared in materials written in Hebrew, resulting in a surface dimension score equal to or higher than 70%. Moreover, most of the materials written in both languages, Hebrew and Arabic, did not address all the implications of vaccine uptake or the relevance of the vaccine or cervical cancer to both populations.

A chi-square test of independence of the surface dimensions reveals significant interaction between the explanatory materials written in Arabic and those written in Hebrew with respect to the following items: HPV vaccine, HPV vaccine for eighth-grade girls, vaccine to protect against cervical cancer caused by the papilloma virus, and a PowerPoint presentation about HPV vaccine.

Deep Dimension–Specific Score

All the Arabic explanatory materials assessed scored below 50% on the *deep dimension***.** Most of the materials did not refer to the social or normative values or to the environmental or cultural-ethical indicators.

Most of the materials that scored 20% on the deep dimension included 1 deep dimension indicator: either normative or social indicators. The normative indicator referred to Arab women's preference for female gynecologists, particularly for Pap smear. The social indicator referred to providing explanations for giving the vaccine to males. These indicators were also referenced in some of the Hebrew informational materials.

Moreover, it is important to note that sexuality is related to the HPV vaccine even though the topics of sexuality and sexual discourse are known to be taboo among the Arab population. The data show that the Arabic explanatory materials made no reference whatsoever to the population's moral concerns regarding sexual relations and promiscuity or to values and perceptions related to sexual relations before marriage.

In contrast, these indicators were referred to and mentioned in Hebrew explanatory materials, resulting in a deep dimension score exceeding 50% in most (78%) cases. Environmental factors, such as the impact of the sexual behavior of the nation's majority population on the sexual concepts of the minority population, were absent from the informational materials in both languages.

A chi-square test of independence of the deep dimensions revealed significant interaction between the explanatory materials written in Arabic and those written in Hebrew with respect to the following items: information sheet before and after administering HPV vaccine, HPV vaccine, HPV vaccine for eighth-grade boys, and vaccine to protect against cervical cancer caused by the papilloma virus.

Overall Cultural Sensitivity Index Score

The overall CS index score calculated for all the explanatory materials targeting the Arab minority points to a low level of CS, in contrast to the CS expressed in materials targeting the Jewish population. The low score reflects only a partial and superficial degree of CS and refers only to a minor portion of CS indicators. The overall CS index score based on the surface indicators, among them using Arabic, mentioning the Arab population, using Arab figures, and referring to the population's low level of literacy, was higher than the score for the deep dimension indicators. The overall CS index score for deep dimension indicators included at most 1 deep indicator, without any reference to normative, social, cultural, traditional, and ethical indicators of the Arab minority population in Israel. Related issues affecting the decision-making process of this population regarding vaccines in general were not addressed, and reference to the issue of sexuality associated with the HPV vaccine was notably absent.

For the overall CS index score, a chi-square test of independence reveals significant interaction between the materials written in Arabic and those written in Hebrew on all the items except for HPV vaccine.

Moreover, a chi-square test of independence between CS and language (Hebrew vs Arabic) for all the explanatory materials revealed significant interaction on all the dimensions: deep dimensions, surface dimensions, and overall CS index score.

Qualitative Content Analysis of the Explanatory Materials on the Topic of the Human Papillomavirus Vaccine

The findings of the content analysis are presented in the form of a comparison between information given in the explanatory materials and information in the scientific literature. The findings reveal 2 main themes: (1) CS and the absence of any reference to the connection between sexuality and the HPV vaccine in the materials written in Arabic targeting the Arab minority, in contrast to those written in the Hebrew and (2) a lack of transparency in transmitting the information in both languages.

Cultural Sensitivity and the Absence of Reference to Connection Between Sexuality and Human Papillomavirus Vaccine in Explanatory Materials Targeting the Minority Population

The informational materials written in Arabic referred to only 2 indicators of deep CS: a normative indicator, referring to women's preference for female gynecologists, particularly for Pap smear, and a social indicator explaining why a *female* vaccine is administered to males. The following main themes emerged: (1) women prefer female physicians to carry out tests, (2) the HPV vaccine prevents cancer among males, and (3) the HPV vaccine does not legitimize sexuality.

Women's Preference for Female Physicians to Carry Out Tests

Women in the traditional conservative Arab society prefer not to be exposed to male doctors and would rather be examined by female physicians, especially in cases of infertility and gynecology that entail a variety of invasive tests [118]. The findings indicate that the explanatory informational materials referred to this normative issue by informing women they can ask for a female gynecologist to perform the Pap smear:

> *If you want to do the Pap smear, go to a female gynecologist in the clinic nearest to your home.* [Clalit HMO, Arabic]

The Human Papillomavirus Vaccine Prevents Cancer Among Males

The scientific literature indicates that some males are at high risk for developing various types of cancer because of infection by the papilloma virus, for example, anal cancer and other types of cancer more prevalent among men who have sex with men [119-121]. Men and women who are infected with HIV or other diseases marked by suppressed immune systems are also at increased risk of developing cancer related to the papilloma virus [122]. Yet, the explanatory materials did not specify a particular population of males at risk but rather explained why the vaccine is given to males in the context of protection against cancer:

> *The papilloma virus is liable to cause diseases (warts and cancerous growths) in men as well. Vaccinating men achieves two objectives: protecting them from diseases and protecting their sexual partners.* [Clalit HMO, Arabic]
>
> *The vaccine, now included in the basket of health services, is intended for women as well as for men and provides protection against cancer of the penis, anal cancer, vaginal cancer and throat cancer.* [Clalit HMO, Arabic]

Moreover, as mentioned above, the results show that the explanatory materials made no reference whatsoever to the connection between sexuality and the HPV vaccine. The Arabic explanatory materials described the vaccine against the papilloma virus as a vaccine that prevents cervical cancer and other types of cancer among males and females without any mention of the fact that the virus is transmitted only through sexual relations:

> *Why are boys vaccinated? For the same reason that girls are vaccinated: to protect girls and boys from cancerous diseases and from genital warts caused by the virus and to prevent passing the virus from one person to another* [Ministry of Health, Hebrew, Arabic]

The Human Papillomavirus Vaccine Does Not Legitimize Sexuality

The transmission of HPV by means of sexual contact aroused controversy regarding the vaccine. Many parents expressed concerns that the vaccine would raise the likelihood of sexual activity or would promote promiscuity [85-87]. The findings show that only the explanatory materials published in Hebrew refer to these concerns regarding sexuality among adolescents in the context of the vaccine. Most of the current materials explain that there is no relationship between the vaccine and adolescent sexual activity and that the vaccine is intended to promote sexual health among adolescents before they become sexually active:

> *The risk of uncontrolled sexual activity among young girls who receive the HPV vaccine is no greater than among young girls or women who do not receive the vaccine.* [Clalit HMO, Hebrew]
>
> *Research studies show that the vaccine does not encourage earlier commencement of sexual activity.* [Ministry of Health, Hebrew]

Cultural Sensitivity and Lack of Transparency Regarding the Controversy Surrounding the Human Papillomavirus Vaccine

The findings show that the explanatory informational materials written in both languages, Hebrew and Arabic, state that the HPV vaccine is totally safe and effective but make no reference to a variety of other factors affecting safety and effectiveness. Moreover, the findings show that the explanatory informational materials attempt to depict cervical cancer as a serious disease that endangers public health, whereas the HPV vaccine is portrayed as the ultimate solution for cervical cancer without any suggested alternatives.

The following main themes emerged: (1) total effectiveness of the HPV vaccine, (2) absolute safety and minor side effects of the HPV vaccine, (3) the HPV vaccine defeats cervical cancer, (4) there is no suitable alternative for the vaccine, (5) cervical cancer is quite prevalent, and (6) the papilloma virus definitely causes cervical cancer.

Total Effectiveness of the Human Papillomavirus Vaccine

Numerous clinical trials have proven the effectiveness of the HPV vaccine (87%-100%) [34,35,123]. Nevertheless, the literature still contains studies pointing to a variety of factors that can affect vaccine effectiveness, such as age, type of virus strain, and the individual's immune system [40-44]. Analysis of the explanatory materials from the Ministry of Health and Clalit HMO websites indicates that none refers to these factors and all report that the HPV vaccine is "extraordinarily effective":

> *The Gardasil vaccine is 100% effective in preventing cervical cancer and in preventing genital warts caused by the papilloma virus. In addition, the vaccine is 99% effective in preventing precancerous lesions caused by four strains of the papilloma virus (6, 11, 16 and 18).* [Clalit HMO, Hebrew]
>
> *Evidence from 20 research studies in nine countries: The HPV vaccine provides 100% protection from*

> *infection by the strains covered by the vaccine. The vaccine reduced the rate of virus infection among girls aged 14-19 by 64%. The vaccine reduced the rate of genital warts among boys aged 14-19 by 31%.* [Ministry of Health, Hebrew/Arabic]

Absolute Safety and Minor Side Effects of the Human Papillomavirus Vaccine

Numerous studies have pointed to a spectrum of side effects of the HPV vaccine, ranging from common mild vaccine effects such as pain at the injection site [46] to more severe effects such as chronic pain syndrome and chronic fatigue [59]; Guillain-Barré syndrome, spinal cord inflammation, and venous blood clots [55]; autoimmune diseases such as ovarian failure [56-58]; and postural orthostatic tachycardia syndrome with chronic pain [59].

The web-based explanatory materials noted mild side effects that are not critical or severe and that pass within a short period of time as well as a limited number of more serious side effects that were described as rare. The following quotations describe the vaccine’s mild side effects:

> *The vaccine can be accompanied by side effects that pass quickly… redness, pain and swelling at the injection site… fever, general sick feeling, muscle and joint pains, and digestive disturbances.* [Clalit HMO, Hebrew]

> *Local side effects are redness, pain and swelling at the injection site. General effects include fainting, dizziness, nausea, headaches, fever. Diarrhea and vomiting are also possible.* [Ministry of Health, Hebrew/Arabic]

In the following quotations, more serious side effects are described as rare. For example, *anaphylaxis* is described as follows:

> *Infrequently, a rapid and severe allergic reaction known as anaphylaxis is liable to occur. This reaction is characterized by shortness of breath, overall itching or rash, drop in blood pressure, rapid pulse, dizziness, abdominal pain, or vomiting and diarrhea… This reaction is rare and comes on quickly, which is the reason for the 15-minute wait in the clinic after the injection. If necessary, the clinic is equipped with an effective treatment kit for emergency treatment.* [Clalit HMO, Hebrew]

Moreover, *complex regional pain syndrome* and *chronic fatigue syndrome* are depicted as effects that are unrelated to the HPV vaccine:

> *No causal relation has been found between the HPV vaccine and these syndromes. The rate of these syndromes is 150 in every 1,000,000 girls aged 10-19 and was no higher among girls and women who received the vaccine than among those that did not.* [Ministry of Health, Hebrew/Arabic]

The Human Papillomavirus Vaccine Defeats Cervical Cancer

There is research evidence for skepticism regarding the long-term preventive effectiveness of the HPV vaccine [124]. Yet despite this, all the explanatory materials portrayed the vaccine as an effective defense that defeats the silent illness and helps prevent it. For example:

> *The papilloma vaccine: the effective and safe way to prevent cancer.* [Ministry of Health, Hebrew/Arabic]

> *The HPV vaccine protects against cervical cancer and also against genital warts.* [Ministry of Health, Hebrew/Arabic]

There Is No Suitable Alternative for the Vaccine

Many research studies pointed to ambiguous and vague reports regarding alternatives for the HPV vaccine and raised concerns regarding the effectiveness of these alternatives in preventing cervical cancer, including Pap smears [94] and condoms [125]. The findings of this study also indicate that most of the explanatory materials portray the HPV vaccine as the “exclusive preventive measure” against cervical cancer caused by the virus, with no reference to any alternatives. For example, the materials raise doubts about the effectiveness of condom use:

> *Condoms do not prevent infection with this disease. Sometimes the virus is found in warts on the genitals, so that infection can be caused by direct contact with the skin containing warts or Condyloma.* [Clalit HMO, Hebrew]

> *Using condoms during sexual relations can help prevent infection with the papilloma virus. But because condoms do not cover the entire genital area and are usually put on after sexual contact has begun, they do not constitute a guarantee against HPV infection.* [Ministry of Health, Hebrew/Arabic]

In addition, despite the need for ongoing Pap smear after receiving the vaccine [25,126,127] and despite the typical effectiveness of the Pap smear in reducing cervical cancer morbidity worldwide [128,129], the explanatory materials played down this information as follows:

> *Pap smear reveal precancerous cervical lesions, not to prevent the disease.* [Clalit HMO, Hebrew]

Cervical Cancer Is Quite Prevalent

The incidence and mortality rates for cervical cancer vary significantly in different countries worldwide, with the highest rates of incidence and mortality specific to third-world countries [130]. The findings show that the explanatory materials describe the morbidity and mortality rates in vague terms and not specific to any particular country or region. Thus, the high incidence rates shown are more relevant to developing nations and not to developed countries such as Israel [130,131] or other Western countries where cervical cancer is practically nonexistent [132]. It is important to emphasize that the prevalence of cervical cancer in Israel is lower than that in other Western countries [133,134], a fact that most of the explanatory materials do not mention. Indeed, the materials describe cervical cancer as a “common threat,” as in the following quotations:

> *Cervical cancer kills 290,000 women worldwide.* [Clalit HMO, Arabic]
>
> *According to the World Health Organization, around 300 million people worldwide are infected with HPV each year, 490,000 women contract cervical cancer and around 230,000 die from cervical cancer each year.* [Clalit HMO, Hebrew]

The explanatory materials also provide specific figures for cervical cancer in Israel. Yet, most provide morbidity rates that have not been updated for many years:

> *Between 1990 and 2004, 150 new cases of cervical cancer were diagnosed in Israel each year. Sixty women in Israel die of the disease each year.* [Clalit HMO, Hebrew]
>
> *Around 180 women are diagnosed each year in Israel. Every year around 80 women die of cervical cancer.* [Ministry of Health, Hebrew/Arabic]

The Papilloma Virus Definitely Causes Cervical Cancer

The scientific literature indicates that infection with the papilloma virus does not indubitably cause cancer of the cervix. Studies have shown that some papilloma virus infections go away on their own within a period of a year or more without any long-term consequences. Chronic HPV infections raise the chances of developing cervical cancer, though only a small portion of these chronic infections ultimately develop into cancer [135-137]. The study's findings show that most of the explanatory materials of all the health authorities depicted infection with the papilloma virus as the main and primary factor that promotes the immediate and certain development of cervical cancer. For example:

> *99.7% of all cases of cervical cancer are related to HPV… The papilloma virus is responsible for 90% of cases of anal cancer in men and women, 70% of cases of cancers of the mouth and throat, 70% of vaginal and genital cancers, and 50% of cases of cancer of the penis.* [Ministry of Health, Hebrew/Arabic]

Discussion

The research findings related to the web-based explanatory materials that were designed to promote the use of HPV vaccine as well as those that appealed to 2 population groups: Arabs and Jews. The analysis of CS in the explanatory materials is based on the definitions of the surface and deep dimensions proposed by Resnicow et al [12,15].

A comparison between the explanatory informational materials targeting the Jewish population and those targeting the Arab population reveals differences in the expression of CS dimensions. With respect to surface dimension, the results of the study reveal more CS in the explanatory materials targeting the Jewish population. Informational materials written in Hebrew are more culturally sensitive to the target population based on the following indicators: (1) appropriate media channel—the websites of the examined health authorities are considered an appropriate media channel for the Jewish population because the Jewish population has greater access to the internet than the Arab population [138,139], (2) place and organizations suitable for transmitting messages—the Ministry of Health and the Clalit HMO are regarded as suitable because they deliver information in the target audience's language [140], and (3) reliable suitable communicators—the Jewish population places more trust in the Ministry of Health than the Arab population [141]. In contrast, the findings show that the explanatory materials written in Arabic demonstrated only partial CS to the Arab minority, mainly with respect to the use of persons and language appropriate for this population. These findings are in line with those of other studies that examined CS in materials intended to influence other health behaviors among other minority populations in Western countries [142,143].

In addition, the informational materials written in the different languages treated the literacy level in their target audiences differently. Resnicow contends that culturally adapting materials to a specific population group requires taking into consideration that group's ability to search for, read, and understand information [12,15]. The level of health literacy is known to be related to many factors, among them education, socioeconomic status, marital status, and others [144,145]. Nevertheless, every population group, even those where the literacy level is not high, can make a decision if the informational materials are transmitted in a transparent and culturally sensitive manner [15]. This study found that the lack of consideration of the literacy level of the Arab population (which is lower than that of the Jewish population [146] can block the population from obtaining, understanding, and using information, thus having a negative impact on the decision-making process regarding HPV vaccination.

With respect to the deep dimension, a limited portion of the explanatory materials written in Arabic made reference to the deep dimensions of the cultural attributes of the Arab population, such as the preference for female gynecologists in performing Pap smears [118] and the explanations for giving males a vaccine intended to prevent cervical cancer [147,148].

One interesting finding is that all the materials written in Hebrew refer to the sexual context associated with the HPV vaccine, representing high deep CS. Nevertheless, the lack of references in the materials targeting the Arab population to the socially and culturally sensitive topic of the connection between engaging in sexual relations and giving the HPV vaccine is paradoxical. The topic of sexuality is seemingly considered taboo in conservative Arab society [100]. Arab society worldwide [99,101] and in Israel [149] does not consider it legitimate either for men or for women to engage in sexual relations before marriage. In patriarchal societies, it is not acceptable for women to have sexual aspirations or desires before marriage [150,151]. Hence, the designers of the explanatory materials targeting women positioned the vaccine in the context of cancer, without mentioning sexual relations [152,153]. Similarly, the appeal to men in these materials was made in the context of protection against various forms of cancer, without reference to any sexual context. It is reasonable to assume that the designers of the explanatory materials chose to disregard the sexual context to avoid undermining the taboo and to adapt the materials to the cultural and religious norms of the target population [154]. Hence, it can be argued that the explanatory materials demonstrated sensitivity to the culture of the target population.

Yet, conversely, based on analysis of the deep dimensions of the explanatory materials, it is possible to claim that this sensitivity is only superficial and that, in practice, no reference is made to the cultural attributes of the target population. According to the deep dimensions as defined in the model developed by Resnicow et al [12,15], such materials should refer to ethical considerations and to the concerns of the population regarding carrying out behavior that is not in keeping with their cultural and normative customs. On the basis of this definition, the explanatory materials designed for the Arab population should have referred to the ethical concerns of this conservative population regarding the vaccine and its implications with respect to encouraging sexual relations and sexual promiscuity before marriage [102,155,156]. As noted above, this was missing from the materials in practice. This may accentuate the lack of transparency in the transmission of full information about the deep dimension indicators.

This unique finding of this study can explain the high rate of HPV vaccination compliance among the Arab minority population, a rate that is more than double the rate among the Jewish population [157]. One reason for this discrepancy may be related to the lack of CS in the informational materials. Lack of information about the relationship between sexuality and vaccination and the absence of references to moral concerns about sexual behavior following vaccination result in an uninformed and unwise decision-making process. Hence, the Arab population remains unaware that the HPV virus is transmitted through sexual relations, a fact that is liable to lower their compliance rate [72,73,75,79].

Moreover, the textual analysis in this study indicates that the texts are not fully transparent regarding the surface dimensions of the CS index in the context of the vaccine’s effectiveness and safety or the relevance of cervical cancer to the target population. Providing transparent health information regarding the HPV vaccine, including a comprehensive description of the scientific controversy surrounding the vaccine’s safety and effectiveness [50,127,158,159] and figures showing the low prevalence of the disease in Israel relative to the rest of the world [131,133], would enable each individual to make an independent and intelligent decision based on informed consent. Furthermore, transparent health information about cervical cancer, the relevance of this illness to the population, and the vaccine’s effectiveness is likely to reduce individuals’ perceived risk of developing cervical cancer and decrease their intentions to obtain the HPV vaccine [160].

This finding may be another reason for the high compliance rate among the Arab population that is related to the lack of transparency regarding the scientific controversy surrounding the vaccine. The fact that the Jewish population is more exposed to the scientific controversy surrounding the vaccine’s side effects through social networking discussions and searching scientific articles [161] can explain the compliance gap between the 2 population groups.

The current literature on risk communication clearly shows that when the media and the health authorities do not provide complete and transparent information in response to the public’s fears and moral concerns, this can have a boomerang effect [33,162], leading the public to doubt the information’s reliability or to suspect that the authorities are hiding information. Moreover, in the age of new media, the public is active and exposed to a great deal of information, underscoring the importance of CS, transparency, and providing full information. If the public does not receive what it seeks from the official authorities, it will turn to other sources of information.

Therefore, the criterion of transparency is not only a criterion for *high level health communication* but also a criterion used in this paper to reveal the lack of full CS in promoting the HPV vaccine among the conservative Arab population. Hence, to achieve full CS, it is important to treat the target audience with respect and to provide them with all the information relevant to the vaccine to enable them to make culturally appropriate decisions [163,164].

Limitations

This study examined web-based explanatory materials published on the websites of official health authorities in Israel. The research did not include flyers, pamphlets, posters, or information brochures that might be distributed in health clinics or other health institutions. Nevertheless, today, the HPV vaccine is not given in clinics or other health institutions because it is part of the routine school vaccine program. Therefore, it is reasonable to assume that the study included most of the existing informational materials.

As stated, the study included only the informational materials appearing on websites. Considering the 70% internet access rate of the Israeli population in general and the 50% access rate of the Arab population in particular, it is possible that not all of the population was exposed to these materials. However, it is reasonable to assume that if the public needs information about the HPV vaccine or any other health information, it prioritizes the official health authorities’ websites.

Moreover, this study was based mainly on qualitative and quantitative content analysis of explanatory materials targeting different population groups in Israel. The research did not consider how these materials were received and accepted by the public, that is, it did not assess their effectiveness among the public. Future research on audience analysis is recommended to examine how the public receives the explanatory materials. In addition, for the deep dimensions, the number of indicators was quite small (5 indicators). Therefore, drawing statistical conclusions was problematic in the comparison between materials written in Hebrew and those written in Arabic.

The research focused on official health authorities’ informational materials and did not generalize to the materials of other countries. Other studies might be comparative and study the transfer of information between different countries.

Conclusions

The public health authorities have 2 main objectives in the context of vaccinations. One is to raise the vaccination rates and the other is to provide full and culturally sensitive information to give the public the tools to make intelligent decisions. The findings of this study indicated that despite the high uptake rate for HPV vaccination in the Arab population, the health authorities did not exercise full transparency and CS in transmitting the association between engaging in sexual relations and necessity of the vaccination. Thus, the major challenge for the health authorities is to find ways to implement the objective of communicating information about the vaccination in a way that is transparent and culturally sensitive, even if this raises questions and fears among the public deriving from their culture.

Specifically, explanatory health materials in general and those promoting the HPV vaccine in particular must provide all the relevant information available in the literature today in a manner that is comprehensive, detailed, culturally sensitive, and based on scientific evidence to enable each individual to make an independent and intelligent decision based on informed consent.

The main recommendation of this study is to make CS the first priority in designing explanatory materials targeting minority population groups. Referring to the cultural attributes of ethnic minority groups in such explanatory materials is important for empowering the population and stimulating open and appropriate discourse.

In addition, because the media have the power to influence a person’s health views and behaviors, the research recommends using media strategies that implement transparency in providing complete information.

Acknowledgments

The authors would like to thank the reviewers for their valuable comments and suggestions to improve the quality of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Coding of categorical indicators of surface dimension and coding of categorical indicators of deep dimension.

[DOCX File, 22KB - jmir_v21i5e13373_app1.docx]

Multimedia Appendix 2

Human papillomavirus vaccine materials published in Arabic and Hebrew: details and index scores.

[DOCX File, 21KB - jmir_v21i5e13373_app2.docx]

References

1. Kreuter MW, McClure SM. The role of culture in health communication. Annu Rev Public Health 2004;25:439-455. [doi: 10.1146/annurev.publhealth.25.101802.123000] [Medline: 15015929]
2. Craig Lefebvre R, Flora JA. Social marketing and public health intervention. Health Educ Q 2016 Sep 04;15(3):299-315. [doi: 10.1177/109019818801500305]
3. Gächter S, Herrmann B. Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. Philos Trans R Soc Lond B Biol Sci 2009 Mar 27;364(1518):791-806 [FREE Full text] [doi: 10.1098/rstb.2008.0275] [Medline: 19073476]
4. Würz A, Nurm Ü, Ekdahl K. Enhancing the role of health communication in the prevention of infectious diseases. J Health Commun 2013;18(12):1566-1571. [doi: 10.1080/10810730.2013.840698] [Medline: 24298888]
5. Betsch C, Böhm R, Airhihenbuwa CO, Butler R, Chapman GB, Haase N, et al. Improving medical decision making and health promotion through culture-sensitive health communication: an agenda for science and practice. Med Decis Making 2016 Dec;36(7):811-833. [doi: 10.1177/0272989X15600434] [Medline: 26296619]
6. Alden DL, Friend J, Schapira M, Stiggelbout A. Cultural targeting and tailoring of shared decision making technology: a theoretical framework for improving the effectiveness of patient decision aids in culturally diverse groups. Soc Sci Med 2014 Mar;105:1-8. [doi: 10.1016/j.socscimed.2014.01.002] [Medline: 24606791]
7. Dutta MJ. Communicating Health: A Culture-centered Approach. Cambridge: Polity Press; 2008.
8. Dutta MJ. Communicating about culture and health: theorizing culture-centered and cultural sensitivity approaches. Commun Theory 2007 Aug;17(3):304-328. [doi: 10.1111/j.1468-2885.2007.00297.x]
9. Guidry JJ, Fagan P, Walker V. Cultural sensitivity and readability of breast and prostate printed cancer education materials targeting African Americans. J Natl Med Assoc 1998 Mar;90(3):165-169. [Medline: 9549980]
10. Kreps GL, Kunimoto EN. Effective Communication in Multicultural Health Care Settings. Thousand Oaks, CA: Sage Publications; 1994.

11. Soskolne V. Dynamics of culture and health: perceived behavioral control and differences in smoking behavior between Arab and Jewish cardiac patients in Israel. J Socialomics 2015;05(01):- [FREE Full text] [doi: 10.4172/2167-0358.1000146]
12. Resnicow K, Baranowski T, Ahluwalia JS, Braithwaite RL. Cultural sensitivity in public health: defined and demystified. Ethn Dis 1999;9(1):10-21. [Medline: 10355471]
13. Hall S. Encoding/decoding. In: Durham MG, Kellner DM, editors. Media and Studies: Keyworks. Oxford: Blackwell; 2001:166-176.
14. Reese LE, Vera EM. Culturally relevant prevention:the scientific and practical considerations of community-based programs. Couns Psychol 2016 Jun 30;35(6):763-778. [doi: 10.1177/0011000007304588]
15. Resnicow K, Soler R, Braithwaite R, Ahluwalia J, Butler J. Cultural sensitivity in substance use prevention. J Community Psychol 2000;28(3):271-290. [doi: 10.1002/(SICI)1520-6629(200005)28:3<271::AID-JCOP4>3.0.CO;2-I]
16. Dutta MJ. Decolonizing communication for social change: a culture-centered approach. Commun Theor 2015 Apr 30;25(2):123-143. [doi: 10.1111/comt.12067]
17. Doak LG, Doak CC, Meade CD. Strategies to improve cancer education materials. Oncol Nurs Forum 1996 Sep;23(8):1305-1312. [Medline: 8883075]
18. Centers for Disease ControlPrevention (CDC). Recommendations on the use of quadrivalent human papillomavirus vaccine in males--Advisory Committee on Immunization Practices (ACIP), 2011. MMWR Morb Mortal Wkly Rep 2011 Dec 23;60(50):1705-1708 [FREE Full text] [Medline: 22189893]
19. Dunne EF, Unger ER, Sternberg M, McQuillan G, Swan DC, Patel SS, et al. Prevalence of HPV infection among females in the United States. J Am Med Assoc 2007 Feb 28;297(8):813. [doi: 10.1001/jama.297.8.813]
20. Ministry of Health. 2016. Cervical cancer in Israel-data update URL: https://www.health.gov.il/PublicationsFiles/cervical_cancer_2016.pdf [accessed 2019-04-30] [WebCite Cache ID 781cFQOC1]
21. Bosch FX, Lorincz A, Muñoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol 2002 Apr;55(4):244-265 [FREE Full text] [Medline: 11919208]
22. Muñoz N, Bosch FX, Castellsagué X, Díaz M, de Sanjose S, Hammouda D, et al. Against which human papillomavirus types shall we vaccinate and screen? the international perspective. Int. J. Cancer 2004 Apr 12;111(2):278-285. [doi: 10.1002/ijc.20244]
23. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999 Sep;189(1):12-19. [doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F] [Medline: 10451482]
24. Marlow LA, Wardle J, Forster AS, Waller J. Ethnic differences in human papillomavirus awareness and vaccine acceptability. J Epidemiol Commun Health 2009 Sep 17;63(12):1010-1015. [doi: 10.1136/jech.2008.085886]
25. Calloway C, Jorgensen CM, Saraiya M, Tsui J. A content analysis of news coverage of the HPV vaccine by US newspapers, January 2002–June 2005. J Womens Health 2006 Sep;15(7):803-809. [doi: 10.1089/jwh.2006.15.803]
26. Brabin L, Roberts SA, Stretch R, Baxter D, Elton P, Kitchener H, et al. A survey of adolescent experiences of human papillomavirus vaccination in the Manchester study. Br J Cancer 2009 Oct 6;101(9):1502-1504. [doi: 10.1038/sj.bjc.6605362]
27. Markowitz LE, Tsu V, Deeks SL, Cubie H, Wang SA, Vicari AS, et al. Human papillomavirus vaccine introduction--the first five years. Vaccine 2012 Nov;30:F139-F148. [doi: 10.1016/j.vaccine.2012.05.039]
28. Drolet M, Bénard E, Boily M, Ali H, Baandrup L, Bauer H, et al. Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect Dis 2015 May;15(5):565-580. [doi: 10.1016/S1473-3099(14)71073-4]
29. Ladner J, Besson M, Rodrigues M, Audureau E, Saba J. Performance of 21 HPV vaccination programs implemented in low and middle-income countries, 2009–2013. BMC Public Health 2014 Jun 30;14(1):670. [doi: 10.1186/1471-2458-14-670]
30. World Health Organization. Countries using HPV vaccine in national immunization schedule and planned introductions. Geneva: World Health Organization; 2013.
31. Ministry of Health Israel. 2013. The papilloma virus vaccine URL: https://www.health.gov.il/NewsAndEvents/SpokemanMesseges/Pages/09102013_1.aspx [accessed 2019-04-30] [WebCite Cache ID 781chPnpx]
32. Gesser-Edelsburg A, Walter N, Shir-Raz Y, Green MS. Voluntary or mandatory? The valence framing effect of attitudes regarding HPV vaccination. J Health Commun 2015 Jul 06;20(11):1287-1293. [doi: 10.1080/10810730.2015.1018642]
33. Gesser-Edelsburg A, Shir-Raz Y. Risk Communication and Infectious Diseases in an Age of Digital Media. London and New York: Routledge; 2016.
34. Garland SM, Hernandez-Avila M, Wheeler CM, Perez G, Harper DM, Leodolter S, et al. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. N Engl J Med 2007 May 10;356(19):1928-1943. [doi: 10.1056/NEJMoa061760]
35. Garland S, Steben M, Sings H, James M, Lu S, Railkar R, et al. Natural history of genital warts: analysis of the placebo arm of 2 randomized phase iii trials of a quadrivalent human papillomavirus (types 6, 11, 16, and 18) vaccine. J Infect Dis 2009 Mar 15;199(6):805-814. [doi: 10.1086/597071]
36. Georgousakis M, Jayasinghe S, Brotherton J, Gilroy N, Chiu C, Macartney K. Population-wide vaccination against human papillomavirus in adolescent boys: Australia as a case study. Lancet Infect Dis 2012 Aug;12(8):627-634. [doi: 10.1016/S1473-3099(12)70031-2]

37. Wheeler CM, Castellsagué X, Garland SM, Szarewski A, Paavonen J, Naud P, et al. Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. Lancet Oncol 2012 Jan;13(1):100-110. [doi: 10.1016/S1470-2045(11)70287-X]
38. Sankaranarayanan R, Prabhu PR, Pawlita M, Gheit T, Bhatla N, Muwonge R, Indian HPV Vaccine Study Group. Immunogenicity and HPV infection after one, two, and three doses of quadrivalent HPV vaccine in girls in India: a multicentre prospective cohort study. Lancet Oncol 2016 Jan;17(1):67-77 [FREE Full text] [doi: 10.1016/S1470-2045(15)00414-3] [Medline: 26652797]
39. Palmer T, Wallace L, Pollock KG, Cuschieri K, Robertson C, Kavanagh K, et al. Prevalence of cervical disease at age 20 after immunisation with bivalent HPV vaccine at age 12-13 in Scotland: retrospective population study. Br Med J 2019 Apr 03;365:l1161. [doi: 10.1136/bmj.l1161]
40. Harper DM, Williams KB. Prophylactic HPV vaccines: current knowledge of impact on gynecologic premalignancies. Discov Med 2010 Jul;10(50):7-17 [FREE Full text] [Medline: 20670593]
41. Joura EA, Kjaer SK, Wheeler CM, Sigurdsson K, Iversen O, Hernandez-Avila M, et al. HPV antibody levels and clinical efficacy following administration of a prophylactic quadrivalent HPV vaccine. Vaccine 2008 Dec;26(52):6844-6851. [doi: 10.1016/j.vaccine.2008.09.073]
42. Lu B, Kumar A, Castellsagué X, Giuliano AR. Efficacy and safety of prophylactic vaccines against cervical HPV infection and diseases among women: a systematic review & meta-analysis. BMC Infect Dis 2011 Jan 12;11(1):-. [doi: 10.1186/1471-2334-11-13]
43. Sawaya GF, Smith-McCune K. HPV vaccination—more answers, more questions. N Engl J Med 2007 May 10;356(19):1991-1993. [doi: 10.1056/NEJMe078060]
44. Villa LL, Costa RL, Petta CA, Andrade RP, Ault KA, Giuliano AR, et al. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. Lancet Oncol 2005 May;6(5):271-278. [doi: 10.1016/S1470-2045(05)70101-7]
45. Hildesheim A, Herrero R, Wacholder S, Rodriguez AC, Solomon D, Bratti MC, et al. Effect of human papillomavirus 16/18 L1 viruslike particle vaccine among young women with preexisting infection. J Am Med Assoc 2007 Aug 15;298(7):743. [doi: 10.1001/jama.298.7.743]
46. Reiter PL, Brewer NT, Gottlieb SL, McRee A, Smith JS. Parents' health beliefs and HPV vaccination of their adolescent daughters. Soc Sci Med 2009 Aug;69(3):475-480. [doi: 10.1016/j.socscimed.2009.05.024] [Medline: 19540642]
47. Siegrist C, Lewis EM, Eskola J, Evans SJ, Black SB. Human papilloma virus immunization in adolescent and young adults. The Pediatric Infectious Disease Journal 2007;26(11):979-984. [doi: 10.1097/INF.0b013e318149dfea]
48. Blitshteyn S. Postural tachycardia syndrome after vaccination with Gardasil. Eur J Neurol 2010 Jul;17(7):e52. [doi: 10.1111/j.1468-1331.2010.03021.x] [Medline: 20402758]
49. Blitshteyn S. Postural tachycardia syndrome following human papillomavirus vaccination. Eur J Neurol 2014;21(1):135-139. [doi: 10.1111/ene.12272] [Medline: 24102827]
50. Brinth L, Theibel AC, Pors K, Mehlsen J. Suspected side effects to the quadrivalent human papilloma vaccine. Dan Med J 2015 Apr;62(4):A5064. [Medline: 25872549]
51. Barloese MC, Mehlsen J, Brinth L, Lundberg HI, Jennum PJ, Jensen RH. Reduced baroreflex sensitivity in cluster headache patients. Headache 2015 Jun;55(6):815-824. [doi: 10.1111/head.12586] [Medline: 26084238]
52. Kinoshita T, Abe R, Hineno A, Tsunekawa K, Nakane S, Ikeda S. Peripheral sympathetic nerve dysfunction in adolescent Japanese girls following immunization with the human papillomavirus vaccine. Intern Med 2014;53(19):2185-2200 [FREE Full text] [Medline: 25274229]
53. Martínez-Lavín M, Martínez-Martínez LA, Reyes-Loyola P. HPV vaccination syndrome. A questionnaire-based study. Clin Rheumatol 2015 Nov;34(11):1981-1983. [doi: 10.1007/s10067-015-3070-3] [Medline: 26354426]
54. Nishioka K, Yokota Y. The 2nd BIO Rheumatology International Congress (BRIC)/ The 9th Global Arthritis Research Network (GARN) 2014 meeting in Moscow. Int J Rheum Dis 2014 Oct 08;17(S2):6-29. [doi: 10.1111/1756-185X.12504]
55. Souayah N, Michas-Martin P, Nasar A, Krivitskaya N, Yacoub HA, Khan H, et al. Guillain-Barré syndrome after Gardasil vaccination: data from Vaccine Adverse Event Reporting System 2006-2009. Vaccine 2011 Jan 29;29(5):886-889. [doi: 10.1016/j.vaccine.2010.09.020] [Medline: 20869467]
56. Colafrancesco S, Perricone C, Tomljenovic L, Shoenfeld Y. Human papilloma virus vaccine and primary ovarian failure: another facet of the autoimmune/inflammatory syndrome induced by adjuvants. Am J Reprod Immunol 2013 Oct;70(4):309-316. [doi: 10.1111/aji.12151] [Medline: 23902317]
57. Little DT, Ward HR. Premature ovarian failure 3 years after menarche in a 16-year-old girl following human papillomavirus vaccination. Case Reports 2012 Oct 03;2012:bcr2012006879. [doi: 10.1136/bcr-2012-006879]
58. Little DT, Ward HR. Adolescent premature ovarian insufficiency following human papillomavirus vaccination. J Investig Med High Impact Case Rep 2014 Oct 28;2(4):232470961455612. [doi: 10.1177/2324709614556129]
59. Tomljenovic L, Colafrancesco S, Perricone C, Shoenfeld Y. Postural orthostatic tachycardia with chronic fatigue after HPV vaccination as part of the “Autoimmune/Auto-inflammatory Syndrome Induced by Adjuvants”. Journal of Investigative Medicine High Impact Case Reports 2014 Mar 18;2(1):23247096145278l. [doi: 10.1177/2324709614527812]

60. Bodemer N, Müller SM, Okan Y, Garcia-Retamero R, Neumeyer-Gromen A. Do the media provide transparent health information? A cross-cultural comparison of public information about the HPV vaccine. Vaccine 2012 May;30(25):3747-3756. [doi: 10.1016/j.vaccine.2012.03.005]
61. Bennett P, Calman K, Curtis S, Smith D. Risk Communication and Public Health. 2nd edition. Oxford: Oxford University Press; 2010.
62. Corcoran N. Communicating Health: Strategies for Health Promotion. 2nd edtion. Thousand Oaks, CA: Sage; 2013.
63. Du Pré A. Communicating About Health: Current Issues and Perspectives. Third Edition. New York: Oxford University Press; 2010.
64. Walaski PM. Risk and Crisis Communications: Methods and Messages. 1st Edition. Hoboken, NJ: John Wiley & Sons Inc; 2011.
65. Neumeyer-Gromen A, Bodemer N, Müller S, Gigerenzer G. [Do media reports and brochures provide informed decisions about cervical cancer prevention?]. Bundesgesundheitsbl 2011 Oct 22;54(11):1197-1210. [doi: 10.1007/s00103-011-1347-5]
66. Attena F, Valdes Abuadili A, Marino S. The informed consent in Southern Italy does not adequately inform parents about infant vaccination. BMC Public Health 2014 Feb 28;14(1):-. [doi: 10.1186/1471-2458-14-211]
67. Grandahl M, Oscarsson M, Stenhammar C, Nevéus T, Westerling R, Tydén T. Not the right time: why parents refuse to let their daughters have the human papillomavirus vaccination. Acta Paediatr 2014 Jan 08;103(4):436-441. [doi: 10.1111/apa.12545]
68. Reiss DR, Karako-Eyal N. Informed consent to vaccination: theoretical, legal, and empirical insights. SSRN Journal 2019 Feb 05:- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3319395. [doi: 10.2139/ssrn.3319395]
69. Nabulsi M, Khalil Y, Makhoul J. Parental attitudes towards and perceptions of their children's participation in clinical research: a developing-country perspective. J Med Ethics 2010 Aug 16;37(7):420-423. [doi: 10.1136/jme.2010.035899]
70. Cooper Robbins SC, Bernard D, McCaffery K, Brotherton JM, Skinner SR. “I just signed”: factors influencing decision-making for school-based HPV vaccination of adolescent girls. Health Psychology 2010;29(6):618-625. [doi: 10.1037/a0021449]
71. Morales-Campos DY, Markham CM, Peskin MF, Fernandez ME. Hispanic mothers' and high school girls' perceptions of cervical cancer, human papilloma virus, and the human papilloma virus vaccine. J Adolesc Health 2013 May;52(5):S69-S75. [doi: 10.1016/j.jadohealth.2012.09.020]
72. Mupandawana ET, Cross R. Attitudes towards human papillomavirus vaccination among African parents in a city in the north of England: a qualitative study. Reprod Health 2016 Aug 22;13(1):97 [FREE Full text] [doi: 10.1186/s12978-016-0209-x] [Medline: 27549328]
73. Niccolai LM, Hansen CE, Credle M, Ryan SA, Shapiro ED. Parents' views on human papillomavirus vaccination for sexually transmissible infection prevention: a qualitative study. Sex Health 2014;11(3):274. [doi: 10.1071/SH14047] [Medline: 24990400]
74. Stephens DP, Thomas TL. Cultural values influencing immigrant Haitian mothers' attitudes toward human papillomavirus vaccination for daughters. J Black Psychol 2013 Apr;39(2):156-168 [FREE Full text] [doi: 10.1177/0095798412461807] [Medline: 25342865]
75. Perkins RB, Clark JA, Apte G, Vercruysse JL, Sumner JJ, Wall-Haas CL, et al. Missed opportunities for HPV vaccination in adolescent girls: a qualitative study. Pediatrics 2014 Aug 18;134(3):e666-e674. [doi: 10.1542/peds.2014-0442]
76. Schmidt-Grimminger D, Frerichs L, Black Bird AE, Workman K, Dobberpuhl M, Watanabe-Galloway S. HPV knowledge, attitudes, and beliefs among Northern Plains American Indian adolescents, parents, young adults, and health professionals. J Canc Educ 2013 Apr 6;28(2):357-366. [doi: 10.1007/s13187-013-0468-y]
77. Siu JY. Perceptions of and barriers to vaccinating daughters against Human Papillomavirus (HPV) among mothers in Hong Kong. BMC Women's Health 2014 Jun 2;14(1):73. [doi: 10.1186/1472-6874-14-73]
78. Cover JK, Nghi NQ, LaMontagne DS, Huyen DT, Hien NT, Nga LT. Acceptance patterns and decision-making for human papillomavirus vaccination among parents in Vietnam: an in-depth qualitative study post-vaccination. BMC Public Health 2012 Aug 09;12:629 [FREE Full text] [doi: 10.1186/1471-2458-12-629] [Medline: 22877158]
79. Gordon D, Waller J, Marlow LA. Attitudes to HPV vaccination among mothers in the British Jewish community: reasons for accepting or declining the vaccine. Vaccine 2011 Oct;29(43):7350-7356. [doi: 10.1016/j.vaccine.2011.07.083]
80. Hughes CC, Jones AL, Feemster KA, Fiks AG. HPV vaccine decision making in pediatric primary care: a semi-structured interview study. BMC Pediatr 2011 Aug 30;11:74 [FREE Full text] [doi: 10.1186/1471-2431-11-74] [Medline: 21878128]
81. Marlow LA, Wardle J, Waller J. Attitudes to HPV vaccination among ethnic minority mothers in the UK: an exploratory qualitative study. Human Vaccines 2014 Oct 27;5(2):105-110. [doi: 10.4161/hv.5.2.7368] [Medline: 19164939]
82. Perkins RB, Pierre-Joseph N, Marquez C, Iloka S, Clark JA. Why do low-income minority parents choose human papillomavirus vaccination for their daughters? J Pediatr 2010;157(4):617-622 [FREE Full text] [doi: 10.1016/j.jpeds.2010.04.013] [Medline: 20472250]
83. Sanders Thompson VL, Arnold LD, Notaro SR. African American parents' HPV vaccination intent and concerns. J Health Care Poor Underserved 2012 Feb;23(1):290-301 [FREE Full text] [doi: 10.1353/hpu.2012.0007] [Medline: 22643477]

84. Centers for Disease ControlPrevention (CDC). FDA licensure of quadrivalent human papillomavirus vaccine (HPV4, Gardasil) for use in males and guidance from the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 2010 May 28;59(20):630-632 [FREE Full text] [Medline: 20508594]
85. Rosenthal SL, Rupp R, Zimet GD, Meza HM, Loza ML, Short MB, et al. Uptake of HPV vaccine: demographics, sexual history and values, parenting style, and vaccine attitudes. J Adolesc Health 2008 Sep;43(3):239-245. [doi: 10.1016/j.jadohealth.2008.06.009]
86. Schuler CL, Reiter PL, Smith JS, Brewer NT. Human papillomavirus vaccine and behavioural disinhibition. Sex Transm Infect 2011 Jun;87(4):349-353 [FREE Full text] [doi: 10.1136/sti.2010.048017] [Medline: 21357601]
87. Stretch R, Roberts SA, McCann R, Baxter D, Chambers G, Kitchener H, et al. Parental attitudes and information needs in an adolescent HPV vaccination programme. Br J Cancer 2008 Nov 4;99(11):1908-1911. [doi: 10.1038/sj.bjc.6604766]
88. Forster A, Wardle J, Stephenson J, Waller J. Passport to promiscuity or lifesaver: press coverage of HPV vaccination and risky sexual behavior. J Health Commun 2010 Mar;15(2):205-217. [doi: 10.1080/10810730903528066] [Medline: 20390987]
89. Hilton S, Hunt K, Langan M, Bedford H, Petticrew M. Newsprint media representations of the introduction of the HPV vaccination programme for cervical cancer prevention in the UK (2005-2008). Soc Sci Med 2010 Mar;70(6):942-950 [FREE Full text] [doi: 10.1016/j.socscimed.2009.11.027] [Medline: 20064682]
90. Stadlin S, Bednarczyk RA, Omer SB. Medical exemptions to school immunization requirements in the United States--association of state policies with medical exemption rates (2004-2011). J Infect Dis 2012 Oct 01;206(7):989-992. [doi: 10.1093/infdis/jis436] [Medline: 22936834]
91. Liddon NC, Leichliter JS, Markowitz LE. Human papillomavirus vaccine and sexual behavior among adolescent and young women. Am J Prev Med 2012 Jan;42(1):44-52. [doi: 10.1016/j.amepre.2011.09.024] [Medline: 22176845]
92. Marchand E, Glenn BA, Bastani R. HPV vaccination and sexual behavior in a community college sample. J Community Health 2013 Dec;38(6):1010-1014 [FREE Full text] [doi: 10.1007/s10900-013-9710-0] [Medline: 23728823]
93. Mullins TL, Zimet GD, Rosenthal SL, Morrow C, Ding L, Shew M, et al. Adolescent perceptions of risk and need for safer sexual behaviors after first human papillomavirus vaccination. Arch Pediatr Adolesc Med 2012 Jan;166(1):82-88 [FREE Full text] [doi: 10.1001/archpediatrics.2011.186] [Medline: 22213755]
94. Casciotti DM, Smith KC, Tsui A, Klassen AC. Discussions of adolescent sexuality in news media coverage of the HPV vaccine. J Adolesc 2014 Feb;37(2):133-143. [doi: 10.1016/j.adolescence.2013.11.004] [Medline: 24439619]
95. Sinai M, Shehade FM. Let's (not) talk about sex: challenges in integrating sex education in traditional Arabic Society in Israel. Int J Adv Counselling 2018 Jul 18:1-15. [doi: 10.1007/s10447-018-9355-x]
96. Gesser-Edelsburg A, Abed Elhadi Arabia M. Discourse on exposure to pornography content online between Arab adolescents and parents: qualitative study on its impact on sexual education and behavior. J Med Internet Res 2018 Oct 09;20(10):e11667 [FREE Full text] [doi: 10.2196/11667] [Medline: 30305264]
97. Smohaa S. Jewish-Arab relations in Israel as a Jewish democratic state. In: Yaar E, Shavit Z, editors. Trends in Israeli Society. Tel Aviv: The Open University; 2011:231-263.
98. Azaiza F. The perception and utilisation of social support in times of cultural change: the case of Arabs in Israel. Int J Soc Welf 2008;17(3):198-203. [doi: 10.1111/j.1468-2397.2008.00542.x]
99. Casper MJ, Carpenter LM. Sex, drugs, and politics: the HPV vaccine for cervical cancer. Sociol Health Illn 2008 Sep;30(6):886-899 [FREE Full text] [doi: 10.1111/j.1467-9566.2008.01100.x] [Medline: 18761509]
100. Gańczak M, Barss P, Alfaresi F, Almazrouei S, Muraddad A, Al-Maskari F. Break the silence: HIV/AIDS knowledge, attitudes, and educational needs among Arab university students in United Arab Emirates. J Adolesc Health 2007 Jun;40(6):572.e1-572.e8. [doi: 10.1016/j.jadohealth.2007.01.011] [Medline: 17531765]
101. Wight D, Plummer ML, Mshana G, Wamoyi J, Shigongo ZS, Ross DA. Contradictory sexual norms and expectations for young people in rural Northern Tanzania. Soc Sci Med 2006 Feb;62(4):987-997. [doi: 10.1016/j.socscimed.2005.06.052] [Medline: 16139937]
102. Natan M, Aharon O, Palickshvili S, Gurman V. Attitude of Israeli mothers with vaccination of their daughters against human papilloma virus. J Pediatr Nurs 2011 Feb;26(1):70-77. [doi: 10.1016/j.pedn.2009.07.006] [Medline: 21256414]
103. Daley EM, Vamos CA, Buhi ER, Kolar SK, McDermott RJ, Hernandez N, et al. Influences on human papillomavirus vaccination status among female college students. J Womens Health (Larchmt) 2010 Oct;19(10):1885-1891. [doi: 10.1089/jwh.2009.1861] [Medline: 20815737]
104. El-Sayed AM, Galea S. The health of Arab-Americans living in the United States: a systematic review of the literature. BMC Public Health 2009 Jul 30;9:272 [FREE Full text] [doi: 10.1186/1471-2458-9-272] [Medline: 19643005]
105. Schwartz M, Ibrahim R, Kahn-Horwitz J. Multi-literate experience as a treasure chest for young learners of English as a foreign language. Read Writ 2016 May 19;29(7):1293-1315. [doi: 10.1007/s11145-016-9633-0]
106. Okun BS. An investigation of the unexpectedly high fertility of secular, native-born Jews in Israel. Popul Stud (Camb) 2016 Dec;70(2):239-257. [doi: 10.1080/00324728.2016.1195913] [Medline: 27414186]
107. Landau JM. The Arabs in Israel: A Political Study. London: Routledge; 2015.
108. Rose S, Spinks N, Canhoto AI. Management Research: Applying the Principles. London: Routledge; 2014.
109. 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]

110. 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]
111. Segars AH, Kohut GF. Strategic communication through the world wide web: an empirical model of effectiveness in the ceo's letter to shareholders. J Management Studs 2001 Jun;38(4):535-556. [doi: 10.1111/1467-6486.00248]
112. Curtin M, Fossey E. Appraising the trustworthiness of qualitative studies: guidelines for occupational therapists. Aust Occ Ther J 2007 Jun;54(2):88-94. [doi: 10.1111/j.1440-1630.2007.00661.x]
113. Denzin NK. Triangulation 2.0. J Mix Methods Res 2012 Feb 23;6(2):80-88. [doi: 10.1177/1558689812437186]
114. Flick U. Mixing methods, triangulation,integrated research: challenges for qualitative research in a world of crisis. In: Denzin NK, Giardina MD, editors. Qualitative Inquiry and Global Crises. Walnut Creek, CA: Left Coast Press; 2011:132-135.
115. Lewis-Beck MS, Bryman A, Futing Liao T. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: Sage Publications; 2004.
116. Gray JH, Densten IL. Integrating quantitative and qualitative analysis using latent and manifest variables. Qual Quant 1998;32(4):419-431. [doi: 10.1023/A:1004357719066]
117. Neuendorf KA. The Content Analysis Guidebook Second Edition. Thousand Oaks, CA: Sage Publications; 2017.
118. Mobaraki AE, Söderfeldt B. Gender inequity in Saudi Arabia and its role in public health. East Mediterr Health J 2010 Jan;16(1):113-118 [FREE Full text] [Medline: 20214168]
119. Brotherton JM, Giuliano AR, Markowitz LE, Dunne EF, Ogilvie GS. Monitoring the impact of HPV vaccine in males-considerations and challenges. Papillomavirus Res 2016 Dec;2:106-111 [FREE Full text] [doi: 10.1016/j.pvr.2016.05.001] [Medline: 29074169]
120. Grulich AE, Hillman R, Brotherton JM, Fairley CK. Time for a strategic research response to anal cancer. Sex Health 2012 Dec;9(6):628-631. [doi: 10.1071/SH12116] [Medline: 23380237]
121. Palefsky JM. Human papillomavirus-related disease in men: not just a women's issue. J Adolesc Health 2010 Apr;46(4 Suppl):S12-S19 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.01.010] [Medline: 20307839]
122. Denny LA, Franceschi S, de Sanjosé S, Heard I, Moscicki AB, Palefsky J. Human papillomavirus, human immunodeficiency virus and immunosuppression. Vaccine 2012 Nov 20;30 Suppl 5:F168-F174. [doi: 10.1016/j.vaccine.2012.06.045] [Medline: 23199960]
123. Fraser C, Tomassini JE, Xi L, Golm G, Watson M, Giuliano AR, et al. Modeling the long-term antibody response of a human papillomavirus (HPV) virus-like particle (VLP) type 16 prophylactic vaccine. Vaccine 2007 May 22;25(21):4324-4333. [doi: 10.1016/j.vaccine.2007.02.069] [Medline: 17445955]
124. Tomljenovic L, Shaw CA. Human papillomavirus (HPV) vaccine policy and evidence-based medicine: are they at odds? Ann Med 2013 Mar;45(2):182-193. [doi: 10.3109/07853890.2011.645353] [Medline: 22188159]
125. Manhart LE, Koutsky LA. Do condoms prevent genital HPV infection, external genital warts, or cervical neoplasia? A meta-analysis. Sex Transm Dis 2002 Nov;29(11):725-735. [Medline: 12438912]
126. Anhang R, Goodman A, Goldie SJ. HPV communication: review of existing research and recommendations for patient education. CA Cancer J Clin 2004;54(5):248-259 [FREE Full text] [doi: 10.3322/canjclin.54.5.248] [Medline: 15371283]
127. Kelly BJ, Leader AE, Mittermaier DJ, Hornik RC, Cappella JN. The HPV vaccine and the media: how has the topic been covered and what are the effects on knowledge about the virus and cervical cancer? Patient Educ Couns 2009 Nov;77(2):308-313 [FREE Full text] [doi: 10.1016/j.pec.2009.03.018] [Medline: 19395221]
128. Nygård J, Skare G, Thoresen S. The cervical cancer screening programme in Norway, 1992–2000: changes in Pap smear coverage and incidence of cervical cancer. J Med Screen 2016 Nov 7;9(2):86-91. [doi: 10.1136/jms.9.2.86] [Medline: 12133929]
129. Wong LP, Wong YL, Low WY, Khoo EM, Shuib R. Knowledge and awareness of cervical cancer and screening among Malaysian women who have never had a Pap smear: a qualitative study. Singapore Med J 2009 Jan;50(1):49-53 [FREE Full text] [Medline: 19224084]
130. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010;60(5):277-300 [FREE Full text] [doi: 10.3322/caac.20073] [Medline: 20610543]
131. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer 2006 Jun 15;118(12):3030-3044 [FREE Full text] [doi: 10.1002/ijc.21731] [Medline: 16404738]
132. Kines RC, Cerio RJ, Roberts JN, Thompson CD, de Los Pinos E, Lowy DR, et al. Human papillomavirus capsids preferentially bind and infect tumor cells. Int J Cancer 2016 Feb 15;138(4):901-911 [FREE Full text] [doi: 10.1002/ijc.29823] [Medline: 26317490]
133. Hakim M, Peled-Raz M. Human papilloma virus vaccine: the appropriate policy for Israel. Medicine and Law 2007;37:55-65 https://www.health.gov.il/English/Topics/Vaccination/HPV/Pages/default.aspx.
134. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55(2):74-108 [FREE Full text] [doi: 10.3322/canjclin.55.2.74] [Medline: 15761078]
135. Goldstein MA, Goodman A, del Carmen MG, Wilbur DC. Case records of the Massachusetts General Hospital. Case 10-2009. A 23-year-old woman with an abnormal Papanicolaou smear. N Engl J Med 2009 Mar 26;360(13):1337-1344. [doi: 10.1056/NEJMcpc0810837] [Medline: 19321871]

136. Hathaway JK. HPV: diagnosis, prevention, and treatment. Clin Obstet Gynecol 2012 Sep;55(3):671-680. [doi: 10.1097/GRF.0b013e31825caa36] [Medline: 22828099]
137. Schiffman M, Kjaer SK. Chapter 2: natural history of anogenital human papillomavirus infection and neoplasia. J Natl Cancer Inst Monogr 2003(31):14-19. [Medline: 12807940]
138. Mesch GS. Ethnic origin and access to electronic health services. Health Informatics J 2016 Dec;22(4):791-803. [doi: 10.1177/1460458215590863] [Medline: 26261219]
139. Mesch GS, Talmud I. Ethnic differences in Internet access: the role of occupation and exposure. Inform Commun Soc 2011 Jun;14(4):445-471. [doi: 10.1080/1369118X.2011.562218]
140. Scollon R, Scollon SW, Jones RH. Intercultural Communication: A Discourse Approach. Third Edition. Malden, MA: Wiley-Blackwell; 2012.
141. Vigoda-Gadot E, Miller-Mor Atias R, Cohen N, Mizrahi S. Public Trust, Civil Participation and Performance Evaluation – Comparative Examination of Health Services and Medical Funds in Israel: Civilian Positions Analysis. Haifa: The Division of Public Administration and Policy, School of Political Science, University of Haifa; 2015.
142. Friedman DB, Hoffman-Goetz L. Assessment of cultural sensitivity of cancer information in ethnic print media. J Health Commun 2006 Jun;11(4):425-447. [doi: 10.1080/10810730600671920] [Medline: 16720539]
143. Ho EY, Tran H, Chesla CA. Assessing the cultural in culturally sensitive printed patient-education materials for Chinese Americans with type 2 diabetes. Health Commun 2015;30(1):39-49 [FREE Full text] [doi: 10.1080/10410236.2013.835216] [Medline: 24446839]
144. Nutbeam D. Health promotion glossary. Health Promot 1986 May;1(1):113-127. [doi: 10.1093/heapro/13.4.349] [Medline: 10318625]
145. Wilson FL, Baker LM, Nordstrom CK, Legwand C. Using the teach-back and Orem's Self-care Deficit Nursing theory to increase childhood immunization communication among low-income mothers. Issues Compr Pediatr Nurs 2008;31(1):7-22. [doi: 10.1080/01460860701877142] [Medline: 18300059]
146. Mesch G, Mano R, Tsamir J. Minority status and health information search: a test of the social diversification hypothesis. Soc Sci Med 2012 Sep;75(5):854-858. [doi: 10.1016/j.socscimed.2012.03.024] [Medline: 22633160]
147. Haber G, Malow RM, Zimet GD. The HPV vaccine mandate controversy. J Pediatr Adolesc Gynecol 2007 Dec;20(6):325-331. [doi: 10.1016/j.jpag.2007.03.101] [Medline: 18082853]
148. Javitt G, Berkowitz D, Gostin LO. Assessing mandatory HPV vaccination: who should call the shots? J Law Med Ethics 2008;36(2):384-95, 214. [doi: 10.1111/j.1748-720X.2008.00282.x] [Medline: 18547207]
149. Abu-Baker K. A Rocky Road: Arab Women as Political Leaders in Israel. Jerusalem: Center for the Study of Arab Society, Van Leer Jerusalem Institute; 1998.
150. Alhadiri I. The Patriarchal System and the Issue of Gender Among the Arabs. Beirut: Al-Saqi Publishers; 2003.
151. Almarnisi F. The Virgin and Patriarchialism: Women and Sexuality in Muslim Society. Damascus: Al Mada Publishing; 2004.
152. Wong LP. HPV information needs, educational messages and channel of delivery preferences: views from developing country with multiethnic populations. Vaccine 2009 Feb;27(9):1410-1415. [doi: 10.1016/j.vaccine.2008.12.045] [Medline: 19150379]
153. Waller J, McCaffery KJ, Forrest S, Wardle J. Human papillomavirus and cervical cancer: issues for biobehavioral and psychosocial research. Ann Behav Med 2004 Feb;27(1):68-79. [doi: 10.1207/s15324796abm2701_9] [Medline: 14979865]
154. Krieger JL, Katz ML, Eisenberg D, Heaner S, Sarge M, Jain P. Media coverage of cervical cancer and the HPV vaccine: implications for geographic health inequities. Health Expect 2013 Sep;16(3):e1-12 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00721.x] [Medline: 21895901]
155. Daley EM, Perrin KM, McDermott RJ, Vamos CA, Rayko HL, Packing-Ebuen JL, et al. The psychosocial burden of HPV: a mixed-method study of knowledge, attitudes and behaviors among HPV+ women. J Health Psychol 2010 Mar;15(2):279-290. [doi: 10.1177/1359105309351249] [Medline: 20207671]
156. Guimond ME, Salman K. Modesty matters: cultural sensitivity and cervical cancer prevention in muslim women in the United States. Nurs Womens Health 2013;17(3):210-6; quiz 217. [doi: 10.1111/1751-486X.12034] [Medline: 23773193]
157. Central Bureau of Statistics. Selected Data on Health from the 2017 Social Survey: Preventive Medicine - Immunizations and Medical Examinations for the Early Detection of Malignant Diseases. Jerusalem: Central Bureau of Statistics; 2018.
158. Abdelmutti N, Hoffman-Goetz L. Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil: a content analysis of Canadian and U.S. national newspaper articles. Women Health 2009;49(5):422-440. [doi: 10.1080/03630240903238776] [Medline: 19851946]
159. Habel MA, Liddon N, Stryker JE. The HPV vaccine: a content analysis of online news stories. J Womens Health (Larchmt) 2009 Mar;18(3):401-407. [doi: 10.1089/jwh.2008.0920] [Medline: 19281323]
160. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. J Health Psychol 2010 Apr;15(3):446-455. [doi: 10.1177/1359105309353647] [Medline: 20348365]
161. Gesser-Edelsburg A, Shir-Raz Y, Green MS. Why do parents who usually vaccinate their children hesitate or refuse? General good vs individual risk. J Risk Res 2014 Dec 12;19(4):405-424. [doi: 10.1080/13669877.2014.983947]

162. Bennett P, Calman K, Curtis S, Smith D. Risk Communication and Public Health Second Edition. Oxford: Oxford University Press; 2010.
163. Maar M, Lightfoot N, Sutherland M, Strasser R, Wilson K, Lidstone-Jones C, et al. Thinking outside the box: Aboriginal people's suggestions for conducting health studies with Aboriginal communities. Public Health 2011 Nov;125(11):747-753. [doi: 10.1016/j.puhe.2011.08.006] [Medline: 22093920]
164. Kass NE, Faden RR. Ethics and learning health care: the essential roles of engagement, transparency, and accountability. Learn Health Sys 2018 Sep 18;2(4):e10066. [doi: 10.1002/lrh2.10066]

Abbreviations

CS: cultural sensitivity
HMO: health maintenance organization
HPV: human papillomavirus
Pap: Papanicolaou
WHO: World Health Organization

Edited by G Eysenbach; submitted 11.01.19; peer-reviewed by B Torr, C Rossmann, M Fadda, M Das; comments to author 07.02.19; revised version received 17.03.19; accepted 23.04.19; published 17.05.19.

Please cite as:
Abed Elhadi Shahbari N, Gesser-Edelsburg A, Mesch GS
Case of Paradoxical Cultural Sensitivity: Mixed Method Study of Web-Based Health Informational Materials About the Human Papillomavirus Vaccine in Israel
J Med Internet Res 2019;21(5):e13373
URL: http://www.jmir.org/2019/5/e13373/
doi:10.2196/13373
PMID:31102371

Review

Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review

Raphaël Chevrier[1,2], MD; Vasiliki Foufi[1,2], PhD; Christophe Gaudet-Blavignac[1,2], BSc, MD; Arnaud Robert[1,2], MSc; Christian Lovis[1,2], MD, MPH, FACMI

[1]Division of Medical Information Sciences, University Hospitals of Geneva, Geneva, Switzerland

[2]Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:
Raphaël Chevrier, MD
Division of Medical Information Sciences
University Hospitals of Geneva
Rue Gabrielle-Perret-Gentil, 4
Geneva, 1205
Switzerland
Phone: 41 223790847
Email: raphael.chevrier@hcuge.ch

Abstract

Background: The secondary use of health data is central to biomedical research in the era of data science and precision medicine. National and international initiatives, such as the Global Open Findable, Accessible, Interoperable, and Reusable (GO FAIR) initiative, are supporting this approach in different ways (eg, making the sharing of research data mandatory or improving the legal and ethical frameworks). Preserving patients' privacy is crucial in this context. De-identification and anonymization are the two most common terms used to refer to the technical approaches that protect privacy and facilitate the secondary use of health data. However, it is difficult to find a consensus on the definitions of the concepts or on the reliability of the techniques used to apply them. A comprehensive review is needed to better understand the domain, its capabilities, its challenges, and the ratio of risk between the data subjects' privacy on one side, and the benefit of scientific advances on the other.

Objective: This work aims at better understanding how the research community comprehends and defines the concepts of de-identification and anonymization. A rich overview should also provide insights into the use and reliability of the methods. Six aspects will be studied: (1) terminology and definitions, (2) backgrounds and places of work of the researchers, (3) reasons for anonymizing or de-identifying health data, (4) limitations of the techniques, (5) legal and ethical aspects, and (6) recommendations of the researchers.

Methods: Based on a scoping review protocol designed a priori, MEDLINE was searched for publications discussing de-identification or anonymization and published between 2007 and 2017. The search was restricted to MEDLINE to focus on the life sciences community. The screening process was performed by two reviewers independently.

Results: After searching 7972 records that matched at least one search term, 135 publications were screened and 60 full-text articles were included. (1) Terminology: Definitions of the terms de-identification and anonymization were provided in less than half of the articles (29/60, 48%). When both terms were used (41/60, 68%), their meanings divided the authors into two equal groups (19/60, 32%, each) with opposed views. The remaining articles (3/60, 5%) were equivocal. (2) Backgrounds and locations: Research groups were based predominantly in North America (31/60, 52%) and in the European Union (22/60, 37%). The authors came from 19 different domains; computer science (91/248, 36.7%), biomedical informatics (47/248, 19.0%), and medicine (38/248, 15.3%) were the most prevalent ones. (3) Purpose: The main reason declared for applying these techniques is to facilitate biomedical research. (4) Limitations: Progress is made on specific techniques but, overall, limitations remain numerous. (5) Legal and ethical aspects: Differences exist between nations in the definitions, approaches, and legal practices. (6) Recommendations: The combination of organizational, legal, ethical, and technical approaches is necessary to protect health data.

Conclusions: Interest is growing for privacy-enhancing techniques in the life sciences community. This interest crosses scientific boundaries, involving primarily computer science, biomedical informatics, and medicine. The variability observed in the use of the terms de-identification and anonymization emphasizes the need for clearer definitions as well as for better education and dissemination of information on the subject. The same observation applies to the methods. Several legislations, such as the American Health Insurance Portability and Accountability Act (HIPAA) and the European General Data Protection Regulation

(GDPR), regulate the domain. Using the definitions they provide could help address the variable use of these two concepts in the research community.

(J Med Internet Res 2019;21(5):e13484) doi:10.2196/13484

KEYWORDS

anonymization; anonymisation; de-identification; deidentification; pseudonymization; privacy; confidentiality; secondary use; data protection; scoping review

Introduction

Background

In 2003, the National Institutes of Health (NIH) released its final statement on sharing research data. The NIH made the provision of a data-sharing plan mandatory for any funding starting at US $500,000 per year [1]. This statement, among other published work [2-5], accelerated the sharing of research data worldwide in parallel to the growing availability of data and information technologies. In this context, the research community gained an unprecedented capacity to access and analyze large amounts of health data, originating partly from nonresearch sources. The use of medical data for a different purpose than the one it was initially collected for is commonly called “secondary use of medical data” [3]. This particular use of health data is subject to technical and semantic problems as well as legal, ethical, and societal concerns. To comply with the legal and ethical principles, researchers have two main options to access and use medical data for a secondary purpose [6]. One option is to gain patients’ consent specifically for the new purpose of their research. This is generally complicated and costly [7]. Alternatively, they can de-identify the data, since the law permits the disclosure of clinical information if it has been correctly de-identified [8]. Institutional review boards (IRBs) generally waive the need for consent in this situation [9]. The existence of the second option gives de-identification and anonymization a pivotal role in biomedical research. Consequently, the availability of reliable techniques to protect privacy becomes essential for the research community to leverage the secondary use of medical data [10].

Despite all efforts, an important gap still exists between the needs and the access to massive data in science. Large collaborative data-sharing projects are somehow below expectations and the research community is calling for improved open data and open science [11]. Some authors have proposed explanations as to why data sharing is more complicated in practice than in theory [3]. An article has considered the influence of policies and of our capacity to protect the data on our ability to share it [12]. Reviews have been published on the techniques and systems aiming at protecting health data privacy [13,14]; one has collected and studied the known re-identification attacks on health data [15], and another has looked specifically into the security and privacy issues related to electronic health records [16]. Various techniques aim at protecting the medical data subjects’ privacy. Those that do not strictly represent an anonymization or de-identification process are not part of the scope of this review. Cryptography, privacy-preserving record linkage [17], and differential privacy [18] are among these techniques.

Although advanced probabilistic privacy-enhancing methods have been studied and applied for over three decades in other areas [19], their application to medical data is a fairly recent interest for the biomedical research community. A striking example is the late introduction of *data anonymization* (2016) and other central concepts of health data privacy (eg, *personally identifiable information*) in the Medical Subject Headings (MeSH) thesaurus of the US National Library of Medicine. Over the last few years, a great amount of expert literature was produced on anonymization and de-identification techniques for medical data. However, publications providing the readers with a broad understanding of these techniques, and addressing their application in life sciences and clinical research in a comprehensive way, are lacking. As a consequence, the fundamental concepts remain either unknown to the research community or difficult to comprehend. Adding to the confusion, a well-documented and long-standing ambiguity exists in the vocabulary used by those who contribute to the practice [20,21]. In particular, the terms *de-identification* and *anonymization* have been used with different meanings by researchers. De-identification is frequently, but not exclusively, used in the biomedical literature to refer to rule-based techniques. These techniques often apply the rules provided in the *Safe Harbor* method of the American legislation (ie, the Health Insurance Portability and Accountability Act [HIPAA]). On the other hand, anonymization is commonly, but not exclusively, used in the biomedical literature to refer to statistical or probabilistic techniques. Turning to the legislations to clarify the meaning of these terms can bring further confusion. Although researchers tend to use two different terms—de-identification and anonymization—to refer to one approach or the other, the American law itself regards both approaches (ie, rule-based or probabilistic) as ways to achieve de-identification. The first follows the *Safe Harbor* method—§164.514(b)(2)—and the second follows the *Expert Determination* method—§164.514(b)(1). The European legislation (ie, the General Data Protection Regulation [GDPR]), on the other hand, does not use either of the terms.

A growing number of health care data breaches are being reported [22], some resulting directly from a failure to anonymize or de-identify the data properly [23]. In this context, it seems essential to review the literature published on this rapidly evolving domain to inform researchers, doctors, lawmakers, and the public about instruments that are becoming indispensable to researchers. This is true, especially since these instruments have bearing on subjects of paramount importance for the future of medical research, namely, data sharing, data privacy, and public trust in health care research and institutions.

Objectives

The aim of this work is to better understand how the life sciences research community defines, comprehends, and uses the concepts of de-identification and anonymization. Providing a broad perspective on the field, this review should also contextualize these concepts and their application in today's biomedical research domain. To attain these goals, the reviewers identified six key aspects to study, which are presented in Textbox 1. These aspects are central to this work; they guided the data collection and they structured the Results section.

Textbox 1. Objectives: subjects of focus for this scoping review.

1. Vocabulary, definitions, and understandings of the terms *anonymization* and *de-identification*.
2. Authors' backgrounds and places of work.
3. Reasons for anonymizing or de-identifying health data.
4. Limitations of anonymization and de-identification techniques.
5. Legal and ethical implications of the practice.
6. Experts' recommendations.

Methods

Overview

Scoping reviews represent an increasingly popular type of review [24], which allows for the mapping of concepts in a field of interest. They are intended to study complex and overlapping domains, particularly when they have not been reviewed comprehensively before [25]. To conduct this work, the authors used the guidance proposed by the Joanna Briggs Institute for the conduct of scoping reviews [26].

The first step of the scoping review process was to perform a preliminary, nonsystematic survey of the literature regarding de-identification and anonymization. This survey identified the key concepts, the concerns, the challenges, and the gaps in the domain. This information was used to define the study's objectives and to design the study protocol.

Article Identification and Selection

Search Strategy

Aiming to focus this work on the life sciences researchers' community, the articles were sourced selectively from one database: MEDLINE [27]. To maximize the sensitivity and specificity of the search query, several strategies were tested and implemented. The terms "de-identification" and "anonymization"; their lexical variants (eg, "de-identif*," and "anonymi*"); and their spelling variants (eg, "deidentified" without hyphen and "anonymisation" in British English) were used. Alternative spellings were proven effective in a previous literature review on the same topic [28]. Numerous candidate terms were tested, here are some examples: "privacy protection," "data protection," "confidentiality," "personal data," "medical data," ”re-identification," and "breaches". None of these terms increased the sensitivity of the search compared to the terms "de-identification," "anonymization," and their variants. The same conclusion was reached regarding the use of the MeSH-controlled vocabulary. Finally, search-field descriptors—[ti] (Title) and [tiab] (Title/Abstract)—were used, and the terms were combined between themselves using Boolean operators. A full description of the search query is provided in the Results section.

Inclusion Criteria

This work analyses the literature published between November 1, 2007, and November 1, 2017. Only original research articles and review articles available in full text through the University of Geneva's library network were considered. Additionally, publications had to meet at least one of the following three criteria to be included:

1. The subject of the article is the process of rendering medical data as less identifiable using computer techniques (ie, de-identification or anonymization).
2. The article focuses on sharing medical data; however, protecting the patients' privacy using computer techniques is also discussed.
3. The article presents legal and ethical aspects of sharing medical data, and the concept of de-identification or anonymization is discussed.

Exclusion Criteria

The literature addressing certain data types was excluded: video recordings, photographic images, radiological images, and geolocation data. This decision was made on the basis of the information found during the preliminary literature survey [28,29] and was confirmed after discussions with experts. Short reports, posters, and editorials were also excluded.

Data Collection

Based on the list of six objectives (see Textbox 1), information categories were defined (see Table 1). Quantitative and qualitative data were collected from the articles. Quantitative information was extracted for certain categories and statistical analysis was performed on this data. Qualitative information was collected for the categories not suited to quantitative analysis. This second approach was nonetheless important, as it allowed us to bring together the views of some experts and to identify consensus or disagreements.

Table 1. Categories of information used to collect quantitative and qualitative data from the reviewed articles.

Type of data	Categories of information
Quantitative	• Journal • Year of publication • Author(s) • Authors' backgrounds • Authors' places of work • Presence of the terms "de-identification" and "anonymization" • Definitions of the terms "de-identification" and "anonymization" • Meanings given to the terms "de-identification" and "anonymization"
Qualitative	• Purposes of de-identification and anonymization • Limitations of the privacy-enhancing techniques • Ethical or legal considerations • Suggestions and recommendations • Data utility and information loss • Data sharing in biomedical research • Types of data subjected to anonymization or de-identification • Public opinion on privacy-enhancing techniques and health data sharing

To determine the backgrounds of the authors, points were attributed to domains (medicine, computer science, law, etc) according to each author's professional affiliation and academic qualifications. Up to three authors were included per publication (ie, first, second, and last author), based on a previous research study, which showed that the most significant contributions were made by these authors [30]. All publications included in the review were considered. The information about the authors was collected manually from the articles, from the authors' or organizations' websites, and from other sources, such as Google Scholar, Open Researcher and Contributor ID (ORCID), ResearchGate, etc.

Results

Study Selection and Characteristics

The literature search retrieved 135 articles from the sizeable number of existing records—7972 after the removal of duplicates—containing at least one of the search terms used. The breakdown of the search query shows the number of records at each level (see Figure 1).

The search query identified 135 records in MEDLINE corresponding to the keyword search; the records were then manually screened according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology (see Figure 2). Among them, 103 records were in the considered time frame. Three records were excluded because the full text could not be retrieved. An additional 40 articles were excluded based on the focus of the paper, the data type considered, or the publication type. During this process, five records raised questions about their potential eligibility. A third reviewer was involved to reach consensus.

Figure 1. Architecture and breakdown of the search query with the number of records at each level. [ti]: Title; [tiab]: Title/Abstract.

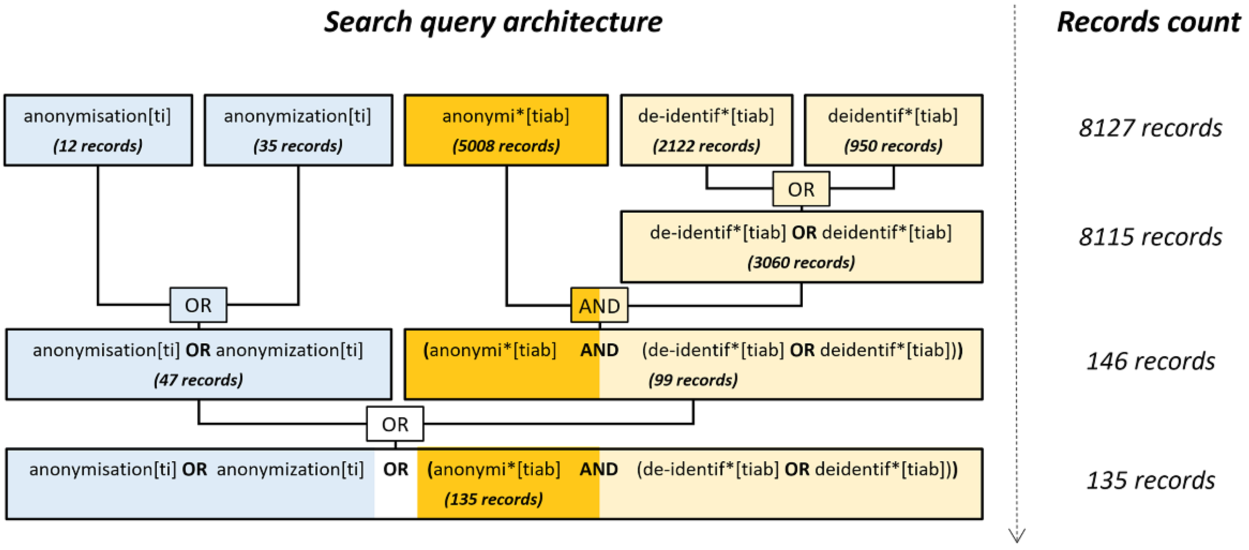
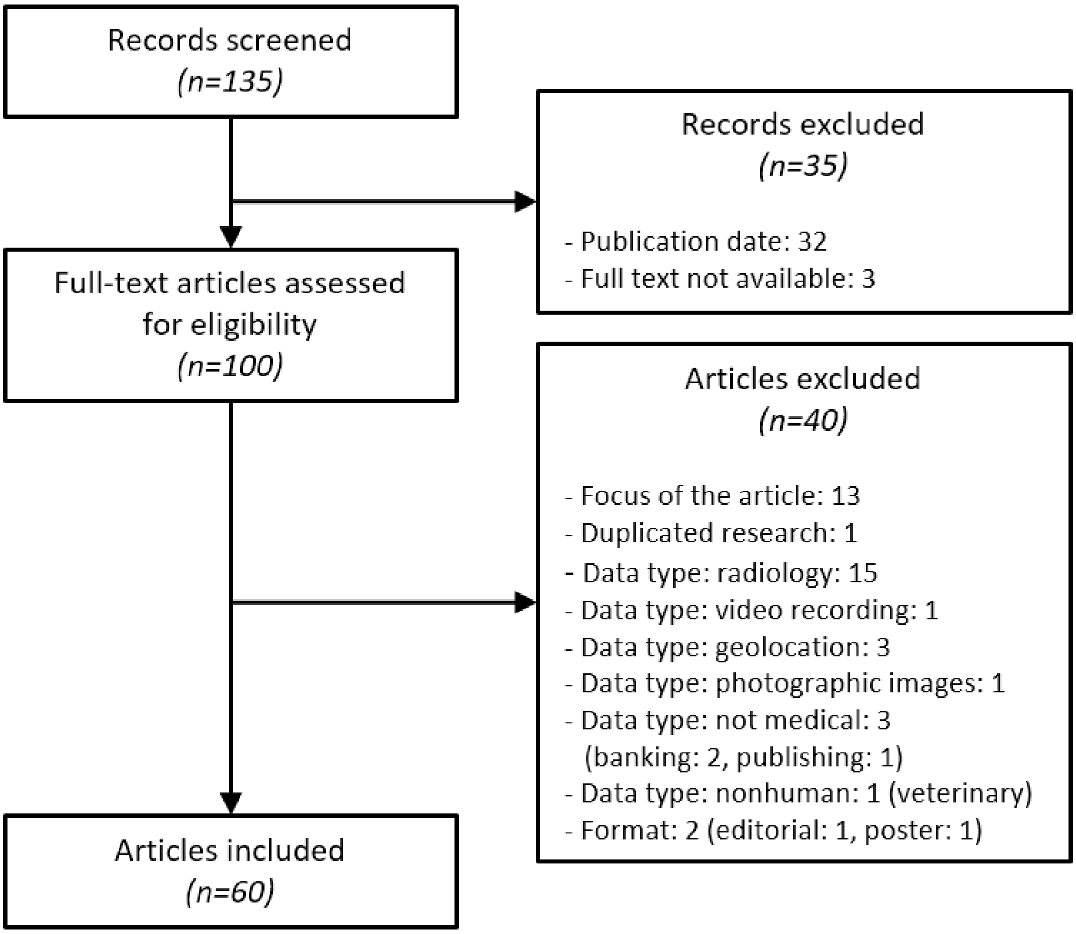


Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for the scoping review process (ie, screening, eligibility, and inclusion).



The process resulted in the inclusion of 60 articles; the list is available in Multimedia Appendix 1. The 60 articles came from 32 different scientific journals. Corrected for the five journals that did not have a registered impact factor in 2017, the average impact factor of the journals included in this review was 2.859, ranging from 9.504 for the highest to 0.304 for the lowest, with a median of 2.766. More than a third of the articles (23/60, 38%) were published after 2015 (see Table 2).

Table 2. Characteristics of the 60 articles included in the review and of the journals where they were published.

Characteristics	Count (N=60), n (%)
Year of publication	
2008-2009	5 (8)
2010-2011	11 (18)
2012-2013	10 (17)
2014-2015	11 (18)
2016-2017	23 (38)
Scientific domains of the journals	
Biomedical informatics	32 (53)
Engineering	8 (13)
Public health, methodology, and epidemiology	6 (10)
Bioethics and law & health policies	5 (8)
Medicine: biomedical sciences	5 (8)
Medicine: clinical	4 (7)

Vocabulary and Definitions

Half of the articles (29/60, 48%) provided a definition of de-identification or anonymization (see Table 3).

The attempts at defining the terms were rare, and the definitions often vague, inconsistent, or even contradictory (see Table 4). Referring to the HIPAA *Safe Harbor* method of de-identification, one article correctly recommended the removal of 18 types of protected health information (PHI) [13]. Another suggested the removal of 17 types of PHI [31]. Regarding the processing of the types of PHI, one article proposed to "hide" or "remove" them [13], while another suggested to "extract" or "replace" them with pseudonyms [32]. Concerning anonymization, the variability was similar. One article presented the process as the removal of the patients' names [33]. Another considered it a much more radical alteration of the data, which would be virtually impossible to reverse [28].

Conflicting representations of de-identification and anonymization were uncovered (see Textbox 2). In some articles, the terms are used interchangeably to refer to the same concept [34-37], while in others they outline strictly different processes [13,19,28,38].

The researchers' representations of de-identification and anonymization, as similar or different concepts, were counted from the reviewed articles to determine whether or not there was a consensus among the experts. The results are presented in Table 5. The 38 authors who used both terms were evenly split between those who considered the two notions to be identical (19/60, 32%) and those who considered them to be different (19/60, 32%).

The 19 researchers who only used or discussed one concept in the core of their articles mentioned the second one in the keywords or title. From the reviewers' perspective, this finding reinforces the idea that de-identification and anonymization are synonyms in many people's minds.

Table 3. Presence of definitions for the terms *de-identification* or *anonymization* in the reviewed articles.

Terms with definitions	Count (N=60), n (%)
De-identification	26 (43)
Anonymization	12 (20)
Both	9 (15)
None	31 (52)

Table 4. Examples of attempts to define the terms *de-identification* or *anonymization*.

Terms	Definitions
De-identification	"For clinical data to be considered de-identified, the HIPAA 'Safe Harbor' technique requires 18 data elements (called PHI: Protected Health Information) to be removed...de-identification only means that explicit identifiers are hidden or removed." [13]
	"Under Safe Harbor, data are considered de-identified if 17 listed types of identifiers are removed." [31]
	"de-identification where explicit identifiers (e.g., Protected Health Information [PHI] elements) are extracted or replaced with 'pseudonyms'" [32]
	"De-identification of medical record data refers to the removal or replacement of personal identifiers so that it would be difficult to reestablish a link between the individual and his or her data. Although a de-identified dataset may contain an encrypted patient identifier with which authorized individuals could relink a patient with his or her dataset, this dataset must not contain data that will allow an unauthorized individual to infer a patient's identity from the existing data elements." [28]
Anonymization	"The anonymization consists in removing the patients' names from the records: unfortunately, other pieces of information enable to identify the patients." [33]
	"anonymization implies that the data cannot be linked to identify the patient" [13]
	"the process of rendering data into a form which does not identify individuals and where identification is not likely to take place" [10]
	"Data anonymization is the process of conditioning a dataset such that no sensitive information can be learned about any specific individual." [19]
	"Anonymization refers to the irreversible removal of the link between the individual and his or her medical record data to the degree that it would be virtually impossible to reestablish the link." [28]

Textbox 2. Discrepancies in understanding and using de-identification and anonymization in relation to each other.

Anonymization = de-identification:

- “Access to de-identified (anonymized) health records would in many cases be sufficient.” [34]
- “Anonymization: Redaction, perturbation, or generalization of those attributes that could be used, alone or in combination, to associate a given record with a specific person. Also called ‘de-identification.’” [35]
- “Recent renewed interest in de-identification (also known as ‘anonymisation’) has led to the development of a series of systems in the United States with very good performance on challenge test sets.” [36]
- “As has been seen, the European regime for privacy does not require the de-identification (anonymization) of personal data used in genomic databases or biobanks.” [37]

Anonymization ≠ de-identification:

- “we note that a recent analysis of matching attacks against a large, public, de-identified (although not anonymized) dataset independently came up” [19]
- “Anonymization and de-identification are often used interchangeably, but de-identification only means that explicit identifiers are hidden or removed, while anonymization implies that the data cannot be linked to identify the patient (i.e. de-identified is often far from anonymous).” [13]
- “De-identification of medical record data refers to the removal or replacement of personal identifiers so that it would be difficult to reestablish a link between the individual and his or her data...Anonymization refers to the irreversible removal of the link between the individual and his or her medical record data to the degree that it would be virtually impossible to reestablish the link.” [28]
- “The term ‘anonymization’ is not identical to ‘de-identification.’ De-identification is the removal of attributes known to increase the risk of identification, and this can be seen as a preliminary step for producing anonymous data. It requires, however, a further assessment as to whether the de-identification process achieves anonymization.” [38]

Table 5. Researchers’ understanding of de-identification and anonymization as similar or different concepts.

Use of the terms in the articles	Count (N=60), n (%)
Only use or discuss one concept	19 (32)
De-identification and anonymization are two different concepts	19 (32)
De-identification and anonymization are used interchangeably	19 (32)
Ambiguous with regard to the meaning of both terms	3 (5)

Authors’ Backgrounds and Places of Work

Applying the scoring system presented in the Methods section, we counted 163 authors for the 60 publications. A total of 248 background points were attributed to 19 different domains (see Table 6).

The first seven fields represent 90% of the researchers’ backgrounds. On average, one researcher was awarded 1.52 research field points. A total of 14 researchers published more than one article (ie, 2-8 articles). Out of 14 prolific authors, 13 (93%) had a background in the three leading domains. Removing the duplicates revealed 121 unique authors. The number of domains and their ranking remained unchanged with and without duplicates, with a slightly smaller gap between the first three domains and the others when duplicates were removed. The background of 7 authors could not be found; this represents a margin of error of 4.3%.

Regarding the place of work, the United States was the largest contributor with 25 articles (25/60, 42%), followed by Germany, the United Kingdom, and Canada combined (23/60, 38%). The predominance of publications from the US-based research groups is noticeable particularly between 2010 and 2012. After this period, their contribution decreases in absolute number and, more importantly, in relation to other groups, due to the arrival of new groups from 2014 and the rapid growth of publications on the topic of anonymization and de-identification. As a result, the leading position that American researchers (ie, Canada and the United States) held until 2013 was caught up to by researchers from other countries in 2014. Since 2015, European groups have been publishing an equal or greater number of articles than the Americans on this topic (see Figure 3).

Purpose of Anonymization and De-Identification

Most often, the authors mentioned the secondary use of medical data without specification as to the purpose of their research. When specified, their objective was to enable and support biomedical research [7,32,39-41]. Regarding the research domains, *genetics and genomics* [42-49] were the most frequently cited, followed by *personalized health and precision medicine* [48,50-52]. Improvement in the domains of *epidemiology* and *public health surveillance and reporting* were among other anticipated benefits of developing privacy protection techniques [8,53]. The protection of privacy was implicit in most projects but was also explicitly cited as a standalone objective in some publications [50,54]. Complying with regulations and policies was a motivation expressed by certain authors [46,55]. Several other reasons were found, as shown in Textbox 3.

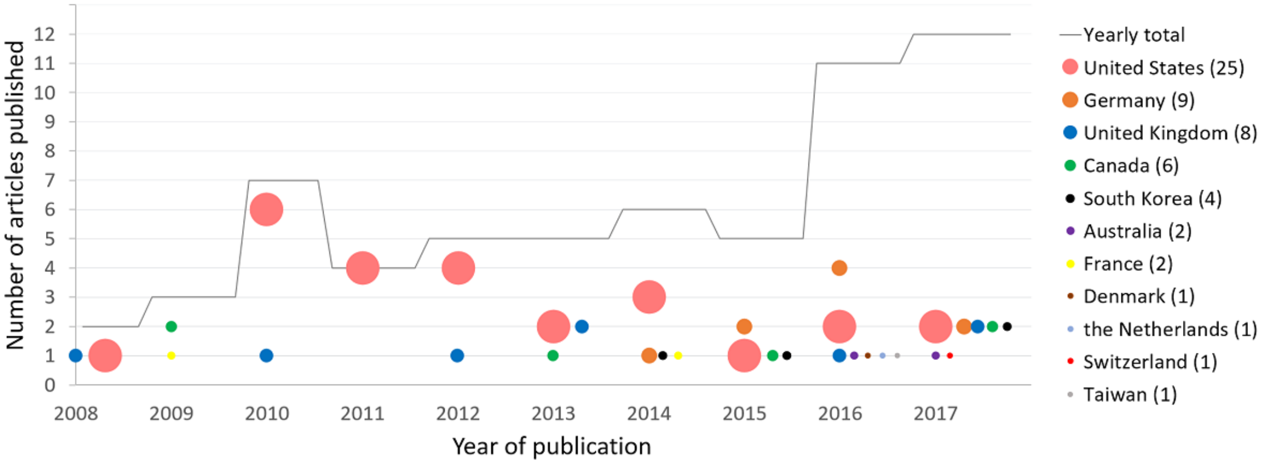
Table 6. Background points awarded to the authors of the reviewed articles. The authors are separated by authorship position: first, second, and last.

Research field	First author (N=92), n (%)	Second author (N=72), n (%)	Last author (N=84), n (%)	Total count (N=248), n (%)
Computer science	36 (14)	26 (10)	29 (12)	91 (36.7)
Biomedical informatics	16 (6)	15 (6)	16 (6)	47 (19.0)
Medicine (MD[a])	13 (5)	9 (4)	16 (6)	38 (15.3)
Epidemiology and statistics	6 (2)	3 (1)	7 (3)	16 (6.5)
Mathematics and biomathematics	6 (2)	5 (2)	5 (2)	16 (6.5)
Law	3 (1)	3 (1)	2 (1)	8 (3.2)
Psychology	2 (1)	3 (1)	2 (1)	7 (2.8)
Linguistics	2 (1)	0 (0)	2 (1)	4 (1.6)
Project management	1 (0)	1 (0)	1 (0)	3 (1.2)
Bioethics and humanities	1 (0)	2 (1)	0 (0)	3 (1.2)
Public health	1 (0)	0 (0)	1 (0)	2 (0.8)
Neuroscience	2 (1)	0 (0)	0 (0)	2 (0.8)
Behavioral economy	0 (0)	2 (1)	0 (0)	2 (0.8)
Journalism	1 (0)	1 (0)	0 (0)	2 (0.8)
Biology and microbiology	1 (0)	0 (0)	1 (0)	2 (0.8)
Physics	1 (0)	1 (0)	0 (0)	2 (0.8)
Health care administration	0 (0)	0 (0)	1 (0)	1 (0.4)
Ecology and evolution	0 (0)	1 (0)	0 (0)	1 (0.4)
Business (MBA[b])	0 (0)	0 (0)	1 (0)	1 (0.4)

[a] MD: Doctor of Medicine.

[b] MBA: Master of Business Administration.

Figure 3. Representation of the 60 publications according to the date of publication, the number of articles per year, and the authors' locations. The size of the discs used on the graph represents each country's contribution in number of articles over the studied period (10 years). The exact count is shown between brackets next to each country's name.



Textbox 3. Additional reasons expressed by experts for de-identifying or anonymizing health data.

1. Publication in biomedical journals [56].
2. Teaching [34].
3. Spontaneous reporting systems to collect adverse drug events [57].
4. Limiting the administrative burden of consent in research [7,38].
5. Facilitating clinical trial data publication [35].
6. Facilitating population screening programs [58].
7. Enabling the creation of medical text corpora for natural language processing (NLP) research and development [37].
8. Protecting particularly sensitive information (eg, mental health data) [59].
9. Producing reports on prescribing patterns and drug utilization and to perform economic studies [60].
10. Performing comparative effectiveness studies [45].

Limitations of Anonymization and De-Identification Techniques

Technical and Operational Limitations

Anonymization and de-identification are time-consuming tasks, particularly when textual data is concerned [61]. The necessity for manual intervention is seen as a weakness that leaves room for human error and contributes to lengthening the procedure [62]. The difficulty in generalizing and scaling the de-identification and anonymization procedures, as well as the absence of broadly accepted metrics to judge their results, are recurrent concerns raised in publications [8,13,38,39,60]. The complexity of these procedures depends on the type of information involved. Structured information (eg, tabular data) is generally easier to process than unstructured information (eg, textual data) [29]. Specific types of information (eg, diagnoses of rare diseases [60]) are more identifying than others. Some types are even considered identifying by nature (eg, large genome sequencing) and presumably impossible to render anonymous [38,63]. More generally, balancing the probability of re-identification with the amount of distortion applied to the data is seen as a challenge [7,59]. Unable to overcome the interdependence between data quality and data identifiability, one has to be compromised for the other: “no existing anonymization algorithm provides both perfect privacy protection and perfect analytic utility” [19]. The re-identification risk depends on the availability of additional information. Using data linkage techniques, the presence of individuals in the protected dataset can be revealed and their personal information re-identified [51,64,65]. Because the amount of information available for comparison can only be estimated, the re-identification risk will always remain an estimate [66]. Additionally, this risk will increase over time [13,48]. These inherent weaknesses have led some researchers to express doubts about the reliability of anonymization or de-identification techniques [35,66,67].

Limitations in Accessibility and Governance

The substantial cost and the limited access to trained professionals are seen as hindrances for institutions wanting to share their data [29,33,66]. Disparities in the availability of anonymization and de-identification systems between English-speaking countries and the rest of the world is expressed by certain authors [33,40]. Textual data is primarily concerned by this problem with a critical need for natural language processing (NLP) systems in varied languages [36,68]. Authors report the lack of practical guidelines and training to assist the researchers [31,69]. They also report an absence of a consensus “regarding the effective governance of secondary research uses, beyond adherence to the terms of informed consent” [70]. Finally, several researchers point out the confusion affecting the terminology as a flaw in itself, increasing the risk of re-identification through misconceptions and misunderstandings [38,71].

Legal and Ethical Implications

General

Privacy laws and regulations provide the legal framework for the collection, processing, and sharing of personal data [71]. Differences exist between nations in the definitions, approaches, and legal practices [66]. Commonly, legal experts agree that relying on legislation alone to protect privacy would be an error [71]. Legislations are effective when used in conjunction with ethical principles, commitments in data use agreements (DUAs), and technical safeguards provided by the de-identification and anonymization process [52,72]. Stricter DUAs can be used to mitigate the loss of data quality that would otherwise be required if the technical process alone had to guarantee the privacy [48,73]. Current rules and regulations are seen by some authors as too soft to discourage attempts at re-identifying data, however, the same authors recommend consistency over severity in prosecuting the misuse of health data [71].

Accountability

The legal responsibilities and the ethical obligations are shared by all those involved in the collection or in the use of the data (ie, institutions or individuals) [10]. Research participants generally believe that anyone who uses their information, regardless of when and under which circumstances, share these responsibilities [50].

Institutional Review Boards

Review boards play an important role in the secondary use of health data. Although de-identified or anonymized data, in some cases, are not considered individually identifiable health information, research projects involving such data generally require IRB submission and approval. In these situations, the IRB assesses the information the subjects received, what they consented to, and whether the proposed research could be conflicting with their interests [31,72]. Eventually, if the IRB approves the project in question, it waives the obligation for informed consent [7,13,31].

Experts' Recommendations

The highest level of protection can only be provided by multidisciplinary approaches combining organizational, legal, ethical, and technical safeguards [10,59,72-74]. Relying exclusively on one of these aspects would be a mistake [75]. More information and training should be provided to researchers about privacy protection and about the risks associated with data sharing [60,69,74]. Numerous researchers express the necessity to review and update the current legal framework [10,31,56,59,71]. Many authors consider the ambiguity of the vocabulary and the misuse of terms as a problem that urgently requires a cooperative effort from the expert community [19,38,56,74]. When applying anonymization techniques, researchers generally recommend favoring privacy over data quality in the process of de-identification and anonymization [75,76].

Discussion

Principal Findings

The development and the application of privacy-enhancing techniques to health data has come to represent a research domain in its own right. This domain is growing rapidly, as demonstrated by the increasing number of publications and the arrival and geographical spread of new research groups. Researchers come from different disciplines and often have qualifications in several fields themselves. Computer science, biomedical informatics, and medicine are the most prevalent backgrounds overall; the main purpose driving the development and application of privacy-enhancing techniques to medical data is to facilitate biomedical research.

At the beginning of the 2000s, great hopes leaned on our abilities to develop technical safeguards that would unleash the potential of the secondary use of medical data. Almost 20 years later, our knowledge and competences have significantly improved, although every advance has come with new interrogations and challenges. Methods are still difficult to generalize or scale and inevitably alter the data quality, which can notably hinder its use for research. A successful exercise lessens the risk of re-identification while maintaining a sufficient level of data quality for research to be performed. In this aim, legal and contractual safeguards are essential and their use can be tailored (ie, made stricter or more lenient) to each situation to mitigate the technical limitations. The research community emphasizes that the different approaches (ie, organizational, legal, ethical, and technical) are complementary and necessary to provide an acceptable level of protection. What is an acceptable level of protection, however, is not easily defined. It varies both in the views of different experts and in the legislations of the different countries.

This work confirms and further illustrates the existence of a disconcerting confusion in the domain's vocabulary affecting the understanding of the concepts at multiple levels. The vagueness and lack of consensus among the experts is worrying and requires actions. The life sciences research community is aware of this situation and is calling for clear and standardized definitions and for cross-border regulatory frameworks.

Propositions

Clear Definitions

Appropriate use of the terms *de-identification* and *anonymization* should be promoted and incentivized. As a first step in this direction, the authors of this work suggest that future publications on the subject include definitions or state which definitions are referred to. Although not universal, clear definitions are provided in two major legislations on personal data protection (ie, the GDPR and the HIPAA) and should be used where appropriate.

The GDPR defines anonymous information as "an information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable" [77]. It is an irreversible state. Accordingly, the term *anonymous* should not be used to describe the process of rendering data less identifiable, which is the prevailing representation of de-identification and anonymization in the biomedical literature. To refer to the concept of rendering data less identifiable, or to the techniques that are used in this aim, the GDPR defines the term *pseudonymization*: "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

Finally, the term *de-identification* comes from the American legislation where definitions are provided: HIPAA §164.514(a) and (b). Authors using this term in their publications should refer to these definitions.

Development of Clear Guidelines

In a manner that already exists in most clinical disciplines, international guidelines regarding privacy protection should be developed, agreed upon, and made widely available to the stakeholders in the field of biomedical research. These guidelines should clarify the concepts, the definitions, and the techniques, as well as their results and risks.

Improved Dissemination and Education

A striking result of this work is the lack of information dissemination and education at all levels. It is critical that the research community gains access to the appropriate information, definitions, and guidelines on the subjects of data privacy and data protection. The public and the media should benefit from this improved access and understanding. Building trust is essential for life sciences research to leverage today's technological capabilities in accessing, sharing, and analyzing data. With this aim, information dissemination is key (see Textbox 4).

Limitations of This Work

There are several limitations to this work. As for any scoping review based on free-text searches, contributions may have been missed despite having maximized the search sensitivity. Privacy protection of health data is a rapidly evolving domain. Between the end of the scoping review and January 2019, 14 additional publications would have to be assessed for eligibility (ie, 114 vs 100).

The fact that the literature search was limited to MEDLINE introduces a strong but deliberate selection bias toward the domain of life sciences. Within life sciences, *genomics*, *personalized health*, and *precision medicine* may be overrepresented due to their growing popularity in recent years and their characteristic need for large amounts of sensitive data. During the data collection, it was not possible to find the background of 7 authors. This represents 4.3% of the total author count (N=163), which should not impact the validity of the results.

Conclusions

Health data is increasingly produced and used. This wealth of information should not be left dormant as it represents a real potential to fuel research and improve medicine. Multidisciplinary safeguards (ie, ethical, organizational, legal, and technical) are required to guarantee the privacy of health data subjected to secondary use. Creating an overall trusted environment to leverage scientific research in life sciences is essential. It requires building on safe and strong foundations, to have processes and structures in place to enforce these foundations, and to communicate widely with the public and the media.

Textbox 4. Recommendations for future work.

- Future publications should include definitions or state which definitions are referred to.
- Existing definitions proposed by major legislations (ie, the Health Insurance Portability and Accountability Act [HIPAA] and the General Data Protection Regulation [GDPR]) should be used where applicable.
- Global and specific guidelines should be developed to define the field of application, the process, the expected results, and the risk of the different technical approaches to privacy protection.
- Information dissemination and education should be improved across the research community for all stakeholders.

Acknowledgments

We would like to thank W Martin, LLM, for his precious help in better understanding some legal concepts.

Authors' Contributions

RC performed the preliminary literature survey; codesigned the study protocol; performed the screening and inclusion process, data collection, and data analysis; and wrote the manuscript. VF performed the screening and inclusion process and data collection and helped in writing the manuscript. CGB participated in the discussions concerning the eligibility of some articles and helped with the interpretation of the results. AR participated in the data analysis and created the figures. CL codesigned the study protocol and participated in the data analysis and the redaction of the manuscript. The manuscript has been reviewed and approved by all authors.

Conflicts of Interest

CL is Editor-in-Chief for JMIR Medical Informatics.

Multimedia Appendix 1

List of the 60 articles reviewed in this work.

[PDF File (Adobe PDF File), 500KB - jmir_v21i5e13484_app1.pdf]

References

1. Final NIH Statement on Sharing Research Data. Bethesda, MD: National Institutes of Health; 2003 Feb 26. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html [accessed 2019-01-24] [WebCite Cache ID 75fHDUrPp]
2. Institute of Medicine (IOM). Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: The National Academies Press; 2015.
3. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Expert Panel. Toward a national framework for the secondary use of health data: An American Medical Informatics Association White Paper. J Am Med Inform Assoc 2007;14(1):1-9 [FREE Full text] [doi: 10.1197/jamia.M2273] [Medline: 17077452]

4. Pisani E, Whitworth J, Zaba B, Abou-Zahr C. Time for fair trade in research data. Lancet 2010 Feb 27;375(9716):703-705. [doi: 10.1016/S0140-6736(09)61486-0] [Medline: 19913902]
5. Dukes P, Clement-Stoneham G. Data Sharing Policy. Version 2.2. London, UK: Medical Research Council; 2016 Sep. URL: https://www.mrc.ac.uk/documents/pdf/mrc-data-sharing-policy/ [accessed 2018-02-13] [WebCite Cache ID 6xCclmaMF]
6. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. BMJ 2015 Mar 20;350:h1139 [FREE Full text] [doi: 10.1136/bmj.h1139] [Medline: 25794882]
7. El Emam K, Dankar FK, Issa R, Jonker E, Amyot D, Cogo E, et al. A globally optimal k-anonymity method for the de-identification of health data. J Am Med Inform Assoc 2009;16(5):670-682 [FREE Full text] [doi: 10.1197/jamia.M3144] [Medline: 19567795]
8. Sengupta S, Calman NS, Hripcsak G. A model for expanded public health reporting in the context of HIPAA. J Am Med Inform Assoc 2008;15(5):569-574 [FREE Full text] [doi: 10.1197/jamia.M2207] [Medline: 18579843]
9. Willison DJ, Emerson C, Szala-Meneok KV, Gibson E, Schwartz L, Weisbaum KM, et al. Access to medical records for research purposes: Varying perceptions across research ethics boards. J Med Ethics 2008 Apr;34(4):308-314. [doi: 10.1136/jme.2006.020032] [Medline: 18375687]
10. Tucker K, Branson J, Dilleen M, Hollis S, Loughlin P, Nixon MJ, et al. Protecting patient privacy when sharing patient-level data from clinical trials. BMC Med Res Methodol 2016 Jul 08;16 Suppl 1:77 [FREE Full text] [doi: 10.1186/s12874-016-0169-4] [Medline: 27410040]
11. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. N Engl J Med 2016 Jun 23;374(25):2411-2413. [doi: 10.1056/NEJMp1605348] [Medline: 27168194]
12. Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J Investig Med 2010 Jan;58(1):11-18 [FREE Full text] [doi: 10.2310/JIM.0b013e3181c9b2ea] [Medline: 20051768]
13. Meystre SM, Friedlin FJ, South BR, Shen S, Samore MH. Automatic de-identification of textual documents in the electronic health record: A review of recent research. BMC Med Res Methodol 2010 Aug 02;10:70 [FREE Full text] [doi: 10.1186/1471-2288-10-70] [Medline: 20678228]
14. Gkoulalas-Divanis A, Loukides G, Sun J. Publishing data from electronic health records while preserving privacy: A survey of algorithms. J Biomed Inform 2014 Aug;50:4-19 [FREE Full text] [doi: 10.1016/j.jbi.2014.06.002] [Medline: 24936746]
15. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011;6(12):e28071 [FREE Full text] [doi: 10.1371/journal.pone.0028071] [Medline: 22164229]
16. Fernández-Alemán JL, Señor IC, Lozoya P, Toval A. Security and privacy in electronic health records: A systematic literature review. J Biomed Inform 2013 Jun;46(3):541-562 [FREE Full text] [doi: 10.1016/j.jbi.2012.12.003] [Medline: 23305810]
17. Vatsalan D, Christen P, Verykios VS. A taxonomy of privacy-preserving record linkage techniques. Inf Syst 2013 Sep;38(6):946-969. [doi: 10.1016/j.is.2012.11.005]
18. Dankar FK, El Emam K. Transactions on Data Privacy. Catalonia, Spain: IIIA-CSIC; 2013 Apr. Practicing differential privacy in health care: A review URL: https://pdfs.semanticscholar.org/65a5/37c9cd327c2925676f59ddffa01cf4afbe51.pdf [accessed 2019-05-23] [WebCite Cache ID 78aeXVavZ]
19. Lasko TA, Vinterbo SA. Spectral anonymization of data. IEEE Trans Knowl Data Eng 2010 Mar 01;22(3):437-446 [FREE Full text] [doi: 10.1109/TKDE.2009.88] [Medline: 21373375]
20. Knoppers BM, Saginur M. The Babel of genetic data terminology. Nat Biotechnol 2005 Aug;23(8):925-927. [doi: 10.1038/nbt0805-925] [Medline: 16082354]
21. Phillips M, Knoppers BM. The discombobulation of de-identification. Nat Biotechnol 2016 Dec 08;34(11):1102-1103. [doi: 10.1038/nbt.3696] [Medline: 27824850]
22. Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information. Washington, DC: US Department of Health and Human Services; 2019. Cases currently under investigation URL: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf [accessed 2019-01-24] [WebCite Cache ID 75fKWiXfo]
23. Culnane C, Rubinstein B, Teague V. Health Data in an Open World. Melbourne, Australia: The University of Melbourne; 2017 Dec 18. URL: https://arxiv.org/ftp/arxiv/papers/1712/1712.05627.pdf [accessed 2019-05-12] [WebCite Cache ID 78byKIiYt]
24. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 09;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
25. Levac D, Colquhoun H, O'Brien KK. Scoping studies: Advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
26. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
27. US National Library of Medicine. 2019. Fact Sheet: MEDLINE® Journal Selection URL: https://www.nlm.nih.gov/lstrc/jsel.html [accessed 2018-05-24] [WebCite Cache ID 78byuh1EF]

28. Kushida CA, Nichols DA, Jadrnicek R, Miller R, Walsh JK, Griffin K. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. Med Care 2012 Jul;50 Suppl:S82-S101 [FREE Full text] [doi: 10.1097/MLR.0b013e3182585355] [Medline: 22692265]
29. Kayaalp M. Patient privacy in the era of big data. Balkan Med J 2018 Dec 20;35(1):8-17 [FREE Full text] [doi: 10.4274/balkanmedj.2017.0966] [Medline: 28903886]
30. Baerlocher MO, Newton M, Gautam T, Tomlinson G, Detsky AS. The meaning of author order in medical research. J Investig Med 2007 May;55(4):174-180. [doi: 10.2310/6650.2007.06044] [Medline: 17651671]
31. Choi HJ, Lee MJ, Choi C, Lee J, Shin S, Lyu Y, et al. Establishing the role of honest broker: Bridging the gap between protecting personal health data and clinical research efficiency. PeerJ 2015;3:e1506 [FREE Full text] [doi: 10.7717/peerj.1506] [Medline: 26713253]
32. Ye H, Chen ES. Attribute Utility Motivated k-anonymization of datasets to support the heterogeneous needs of biomedical researchers. AMIA Annu Symp Proc 2011;2011:1573-1582 [FREE Full text] [Medline: 22195223]
33. Chazard E, Mouret C, Ficheur G, Schaffar A, Beuscart J, Beuscart R. Proposal and evaluation of FASDIM, a Fast And Simple De-Identification Method for unstructured free-text clinical records. Int J Med Inform 2014 Apr;83(4):303-312. [doi: 10.1016/j.ijmedinf.2013.11.005] [Medline: 24370391]
34. Pantazos K, Lauesen S, Lippert S. Preserving medical correctness, readability and consistency in de-identified health records. Health Informatics J 2017 Dec;23(4):291-303. [doi: 10.1177/1460458216647760] [Medline: 27199298]
35. O'Neill L, Dexter F, Zhang N. The risks to patient privacy from publishing data from clinical anesthesia studies. Anesth Analg 2016 Dec;122(6):2017-2027. [doi: 10.1213/ANE.0000000000001331] [Medline: 27172145]
36. Grouin C, Rosier A, Dameron O, Zweigenbaum P. Testing tactics to localize de-identification. Stud Health Technol Inform 2009;150:735-739. [Medline: 19745408]
37. Townend D. EU laws on privacy in genomic databases and biobanking. J Law Med Ethics 2016 Dec;44(1):128-142. [doi: 10.1177/1073110516644204] [Medline: 27256129]
38. Sariyar M, Schlünder I. Reconsidering anonymization-related concepts and the term "identification" against the backdrop of the European legal framework. Biopreserv Biobank 2016 Oct;14(5):367-374 [FREE Full text] [doi: 10.1089/bio.2015.0100] [Medline: 27104620]
39. Prasser F, Kohlmayer F, Kuhn KA. Efficient and effective pruning strategies for health data de-identification. BMC Med Inform Decis Mak 2016 Apr 30;16:49 [FREE Full text] [doi: 10.1186/s12911-016-0287-2] [Medline: 27130179]
40. Shin S, Park YR, Shin Y, Choi HJ, Park J, Lyu Y, et al. A de-identification method for bilingual clinical texts of various note types. J Korean Med Sci 2015 Jan;30(1):7-15 [FREE Full text] [doi: 10.3346/jkms.2015.30.1.7] [Medline: 25552878]
41. Kohlmayer F, Prasser F, Eckert C, Kuhn KA. A flexible approach to distributed data anonymization. J Biomed Inform 2014 Aug;50:62-76 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.002] [Medline: 24333850]
42. Heatherly RD, Loukides G, Denny JC, Haines JL, Roden DM, Malin BA. Enabling genomic-phenomic association discovery without sacrificing anonymity. PLoS One 2013;8(2):e53875 [FREE Full text] [doi: 10.1371/journal.pone.0053875] [Medline: 23405076]
43. Loukides G, Gkoulalas-Divanis A, Malin B. Anonymization of electronic medical records for validating genome-wide association studies. Proc Natl Acad Sci U S A 2010 Apr 27;107(17):7898-7903 [FREE Full text] [doi: 10.1073/pnas.0911686107] [Medline: 20385806]
44. Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B. Anonymization of longitudinal electronic medical records. IEEE Trans Inf Technol Biomed 2012 May;16(3):413-423 [FREE Full text] [doi: 10.1109/TITB.2012.2185850] [Medline: 22287248]
45. Malin B, Benitez K, Masys D. Never too old for anonymity: A statistical standard for demographic data sharing via the HIPAA Privacy Rule. J Am Med Inform Assoc 2011;18(1):3-10 [FREE Full text] [doi: 10.1136/jamia.2010.004622] [Medline: 21169618]
46. Loukides G, Denny JC, Malin B. The disclosure of diagnosis codes can breach research participants' privacy. J Am Med Inform Assoc 2010;17(3):322-327 [FREE Full text] [doi: 10.1136/jamia.2009.002725] [Medline: 20442151]
47. Loukides G, Gkoulalas-Divanis A. Utility-aware anonymization of diagnosis codes. IEEE J Biomed Health Inform 2013 Jan;17(1):60-70. [doi: 10.1109/TITB.2012.2212281] [Medline: 22893444]
48. Prasser F, Kohlmayer F, Kuhn KA. The importance of context: Risk-based de-identification of biomedical data. Methods Inf Med 2016 Aug 05;55(4):347-355. [doi: 10.3414/ME16-01-0012] [Medline: 27322502]
49. Heatherly R, Rasmussen LV, Peissig PL, Pacheco JA, Harris P, Denny JC, et al. A multi-institution evaluation of clinical profile anonymization. J Am Med Inform Assoc 2016 Apr;23(e1):e131-e137 [FREE Full text] [doi: 10.1093/jamia/ocv154] [Medline: 26567325]
50. Goodman D, Johnson CO, Bowen D, Smith M, Wenzel L, Edwards K. De-identified genomic data sharing: The research participant perspective. J Community Genet 2017 Jul;8(3):173-181 [FREE Full text] [doi: 10.1007/s12687-017-0300-1] [Medline: 28382417]
51. Prasser F, Kohlmayer F, Spengler H, Kuhn KA. A scalable and pragmatic method for the safe sharing of high-quality health data. IEEE J Biomed Health Inform 2018 Dec;22(2):611-622. [doi: 10.1109/JBHI.2017.2676880] [Medline: 28358693]

52. Prasser F, Bild R, Kuhn KA. A generic method for assessing the quality of de-identified health data. Stud Health Technol Inform 2016;228:312-316. [Medline: 27577394]
53. El Emam K, Moher E. Privacy and anonymity challenges when collecting data for public health purposes. J Law Med Ethics 2013 Mar;41 Suppl 1:37-41. [doi: 10.1111/jlme.12036] [Medline: 23590738]
54. Majeed A, Ullah F, Lee S. Vulnerability- and diversity-aware anonymization of personally identifiable information for improving user privacy and utility of publishing data. Sensors (Basel) 2017 May 08;17(5):1-23 [FREE Full text] [doi: 10.3390/s17051059] [Medline: 28481298]
55. Xia W, Heatherly R, Ding X, Li J, Malin BA. R-U policy frontiers for health data de-identification. J Am Med Inform Assoc 2015 Sep;22(5):1029-1041 [FREE Full text] [doi: 10.1093/jamia/ocv004] [Medline: 25911674]
56. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: Guidance for journal editors, authors, and peer reviewers. Trials 2010 Jan 29;11:9 [FREE Full text] [doi: 10.1186/1745-6215-11-9] [Medline: 20113465]
57. Lin W, Yang D, Wang J. Privacy preserving data anonymization of spontaneous ADE reporting system dataset. BMC Med Inform Decis Mak 2016 Dec 18;16 Suppl 1:58 [FREE Full text] [doi: 10.1186/s12911-016-0293-4] [Medline: 27454754]
58. Bartholomäus S, Hense HW, Heidinger O. Blinded Anonymization: A method for evaluating cancer prevention programs under restrictive data protection regulations. Stud Health Technol Inform 2015;210:424-428. [Medline: 25991179]
59. Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang C, Jackson RG, et al. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC Med Inform Decis Mak 2013 Jul 11;13:71 [FREE Full text] [doi: 10.1186/1472-6947-13-71] [Medline: 23842533]
60. El Emam K, Dankar FK, Vaillancourt R, Roffey T, Lysyk M. Evaluating the risk of re-identification of patients from hospital prescription records. Can J Hosp Pharm 2009 Jul;62(4):307-319 [FREE Full text] [Medline: 22478909]
61. Meystre S, Heider P, Kim Y, Trice A, Underwood G. Clinical text automatic de-identification to support large scale data reuse and sharing: Pilot results. In: Proceedings of the AMIA 2018 Annual Symposium. 2018 Nov 05 Presented at: AMIA 2018 Annual Symposium; November 3-7, 2018; San Francisco, CA URL: https://symposium2018.zerista.com/event/member/509567
62. Shaw DM. Blinded by the light: Anonymization should be used in peer review to prevent bias, not protect referees. EMBO Rep 2015 Aug;16(8):894-897 [FREE Full text] [doi: 10.15252/embr.201540943] [Medline: 26174615]
63. Sinnott R, Ajayi O, Stell A, Young A. Towards a virtual anonymisation grid for unified access to remote clinical data. Stud Health Technol Inform 2008;138:90-101. [Medline: 18560111]
64. Lu Y, Sinnott RO, Verspoor K. A semantic-based k-anonymity scheme for health record linkage. Stud Health Technol Inform 2017;239:84-90. [Medline: 28756441]
65. Tamersoy A, Loukides G, Denny JC, Malin B. Anonymization of administrative billing codes with repeated diagnoses through censoring. AMIA Annu Symp Proc 2010 Nov 13;2010:782-786 [FREE Full text] [Medline: 21347085]
66. Thorogood A, Zawati MH. International guidelines for privacy in genomic biobanking (or the unexpected virtue of pluralism). J Law Med Ethics 2015;43(4):690-702. [doi: 10.1111/jlme.12312] [Medline: 26711410]
67. Narayanan A, Felten EW. randomwalker.info. 2014 Jul 09. No silver bullet: De-identification still doesn't work URL: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf [accessed 2019-01-24] [WebCite Cache ID 75fQcTcoA]
68. Foufi V, Gaudet-Blavignac C, Chevrier R, Lovis C. De-identification of medical narrative data. Stud Health Technol Inform 2017;244:23-27. [Medline: 29039370]
69. O'Keefe CM, Westcott M, O'Sullivan M, Ickowicz A, Churches T. Anonymization for outputs of population health and health services research conducted via an online data center. J Am Med Inform Assoc 2017 May 01;24(3):544-549. [doi: 10.1093/jamia/ocw152] [Medline: 28011594]
70. Fullerton SM, Lee SS. Secondary uses and the governance of de-identified data: Lessons from the human genome diversity panel. BMC Med Ethics 2011 Sep 26;12:16 [FREE Full text] [doi: 10.1186/1472-6939-12-16] [Medline: 21943371]
71. Phillips M, Dove ES, Knoppers BM. Criminal prohibition of wrongful re-identification: Legal solution or minefield for big data? J Bioeth Inq 2017 Dec;14(4):527-539 [FREE Full text] [doi: 10.1007/s11673-017-9806-9] [Medline: 28913771]
72. Kohlmayer F, Prasser F, Kuhn KA. The cost of quality: Implementing generalization and suppression for anonymizing biomedical data with minimal information loss. J Biomed Inform 2015 Dec;58:37-48 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.007] [Medline: 26385376]
73. Arbuckle L, Moher E, Bartlett SJ, Ahmed S, El Emam K. Montreal Accord on Patient-Reported Outcomes (PROs) use series - Paper 9: Anonymization and ethics considerations for capturing and sharing patient-reported outcomes. J Clin Epidemiol 2017 Sep;89:168-172. [doi: 10.1016/j.jclinepi.2017.04.016] [Medline: 28433677]
74. Smith C. Preventing unintended disclosure of personally identifiable data following anonymisation. Stud Health Technol Inform 2017;235:313-317. [Medline: 28423805]
75. Cimino JJ. The false security of blind dates: Chrononymization's lack of impact on data privacy of laboratory data. Appl Clin Inform 2012;3(4):392-403 [FREE Full text] [doi: 10.4338/ACI-2012-07-RA-0028] [Medline: 23646086]
76. South BR, Mowery D, Suo Y, Leng J, Ferrández Ó, Meystre SM, et al. Evaluating the effects of machine pre-annotation and an interactive annotation interface on manual de-identification of clinical text. J Biomed Inform 2014 Aug;50:162-172 [FREE Full text] [doi: 10.1016/j.jbi.2014.05.002] [Medline: 24859155]

77. The European Parliament and the Council of the European Union. Official Journal of the European Union. Volume 59. Luxembourg: Publications Office of the European Union; 2016 May 04. Legislation URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN [accessed 2019-05-27] [WebCite Cache ID 78gSMqiNn]

Abbreviations

DUA: data use agreement
GDPR: General Data Protection Regulation
GO FAIR: Global Open Findable, Accessible, Interoperable, and Reusable
HIPAA: Health Insurance Portability and Accountability Act
IRB: institutional review board
MBA: Master of Business Administration
MD: Doctor of Medicine
MeSH: Medical Subject Headings
NIH: National Institutes of Health
NLP: natural language processing
ORCID: Open Researcher and Contributor ID
PHI: protected health information
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
[ti]: Title
[tiab]: Title/Abstract

Edited by M Focsa; submitted 24.01.19; peer-reviewed by B Knoppers, AJ Greenberg, J Goris; comments to author 13.02.19; revised version received 29.03.19; accepted 26.04.19; published 31.05.19.

Please cite as:
Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C
Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review
J Med Internet Res 2019;21(5):e13484
URL: http://www.jmir.org/2019/5/e13484/
doi:10.2196/13484
PMID:31152528

Original Paper

Estimating Determinants of Attrition in Eating Disorder Communities on Twitter: An Instrumental Variables Approach

Tao Wang[1,2], MSc; Emmanouil Mentzakis[1], PhD; Markus Brede[3], PhD; Antonella Ianni[1,4], PhD

[1]Department of Economics, University of Southampton, Southampton, United Kingdom
[2]The Alan Turing Institute, London, United Kingdom
[3]Department of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom
[4]Department of Economics, Università Cà Foscari, Venice, Italy

Corresponding Author:
Tao Wang, MSc
Department of Economics
University of Southampton
Building 58/1010
University Road
Southampton, SO17 1BJ
United Kingdom
Phone: 44 7864373080
Email: t.wang@soton.ac.uk

Abstract

Background: The use of social media as a key health information source has increased steadily among people affected by eating disorders (EDs). Research has examined characteristics of individuals engaging in online communities, whereas little is known about discontinuation of engagement and the phenomenon of participants dropping out of these communities.

Objective: This study aimed to investigate the characteristics of dropout behaviors among eating disordered individuals on Twitter and to estimate the causal effects of personal emotions and social networks on dropout behaviors.

Methods: Using a snowball sampling method, we collected a set of individuals who self-identified with EDs in their Twitter profile descriptions, as well as their tweets and social networks, leading to 241,243,043 tweets from 208,063 users. Individuals' emotions are measured from their language use in tweets using an automatic sentiment analysis tool, and network centralities are measured from users' following networks. Dropout statuses of users are observed in a follow-up period 1.5 years later (from February 11, 2016 to August 17, 2017). Linear and survival regression instrumental variables models are used to estimate the effects of emotions and network centrality on dropout behaviors. The average levels of attributes among an individual's followees (ie, people who are followed by the individual) are used as instruments for the individual's attributes.

Results: Eating disordered users have relatively short periods of activity on Twitter with one half of our sample dropping out at 6 months after account creation. Active users show more negative emotions and higher network centralities than dropped-out users. Active users tend to connect to other active users, whereas dropped-out users tend to cluster together. Estimation results suggest that users' emotions and network centralities have causal effects on their dropout behaviors on Twitter. More specifically, users with positive emotions are more likely to drop out and have shorter lasting periods of activity online than users with negative emotions, whereas central users in a social network have longer lasting participation than peripheral users. Findings on users' tweeting interests further show that users who attempt to recover from EDs are more likely to drop out than those who promote EDs as a lifestyle choice.

Conclusions: Presence in online communities is strongly determined by the individual's emotions and social networks, suggesting that studies analyzing and trying to draw condition and population characteristics through online health communities are likely to be biased. Future research needs to examine in more detail the links between individual characteristics and participation patterns if better understanding of the entire population is to be achieved. At the same time, such attrition dynamics need to be acknowledged and controlled when designing online interventions so as to accurately capture their intended populations.

(J Med Internet Res 2019;21(5):e10942) doi:10.2196/10942

KEYWORDS

medical informatics; eating disorders; social media; attrition; emotions; social network; causality; instrumental variables

Introduction

Background

Eating disorders (EDs), such as anorexia and bulimia, are complex mental disorders defined by extreme obsessions with body weight or shape and unusual eating behaviors [1]. These diseases have the highest mortality rate of any mental illness [2], intractable comorbidities, [3] and worldwide prevalence [4], having become a major public health concern. Although a variety of treatment options have emerged over recent years [5], populations affected by EDs are often hard to reach through traditional health care services. This is mainly because of fear of stigma or a feeling of shame; many sufferers conceal their ED symptoms and never seek professional treatment or support [6,7]. To keep struggles with illnesses private, people often seek health-related information and support through online peer-to-peer communities, particularly via social media sites such as Twitter and Facebook. Participation in online communities is common in ED populations [8] and has been suggested as a screening factor for EDs [3]. This provides an opportunity for health care professionals to deliver health support to these hard-to-reach populations through online communities [9-13]. Moreover, as online communities present a relatively anonymous environment for individuals to naturally self-disclose and socialize [14], user-generated data online provide a large amount of records about individuals' concerns, thoughts, emotions, and social interactions [15-17], which can complement traditional data sources (eg, surveys and interviews) in understanding risk factors of EDs. Hence, growing research has focused on characterizing individuals' behavioral patterns in online communities [15-19] so as to better understand EDs and promote population-level well-being.

One notable characteristic of online ED communities is their participants have widely different stances on EDs [8,20,21]. Some communities encourage members to discuss their struggles with EDs, share treatment options, and offer support toward recovery from EDs, so called *prorecovery* communities [20-22]. There are also many *anti-recovery* or *pro-ED* communities in which members often deny an ED to be a disorder and instead promote EDs as a healthy lifestyle choice [8,23]. These pro-ED communities can negatively affect health and quality of life among people with and without EDs, through reinforcing an individual's identity around EDs [24], promoting thin ideals [25], and disseminating harmful practices for weight loss [8]. Recent studies have shown that individuals' language use online strongly indicates their pro-ED or prorecovery stances [15,17,20], as well as emotions of depression, helplessness, and anxiety that reflect their mental disorders [16]. Other studies have also examined interactions between pro-ED and prorecovery communities on Flickr [21], anorexia-related misinformation [18], sentiments in comments on ED-related videos on YouTube [26], characteristics of removed pro-ED content, [27] and lexical variation of pro-ED tags on Instagram [19,28]. Yet, prior studies have largely focused on examining how people engage in and maintain an online ED community, whereas little is known about how people drop out of such a community. As a dynamic process, people who join and actively engage in a community at earlier stages can have less participation and leave the community at later stages. Understanding the attrition processes of online communities can enhance our knowledge of the dynamics in these communities.

Studying the attrition process of an online community can also have practical implications for disease prevention and health interventions. Given the ease of accessibility of social media for many individuals (eg, via mobile devices), increasing attention has focused on using online communities to deliver health interventions [9-13,29,30]. One of the most popular approaches is to deliver health lessons and behavior-changing instructions via online communities [9-13,29]. Although pilot studies based on small samples have demonstrated the effectiveness of these approaches in reducing body dissatisfaction and disordered eating [12,13], evidence from interventions for a variety of health behaviors (eg, smoking, diet, exercise, and sexual health) suggests that attrition (ie, participant loss) is one of the most common challenges in online interventions [10,29]. This is known as the *law of attrition* of online interventions [31]. A recent study has shown a high attrition rate in an online intervention for EDs [32], though this intervention is delivered via a purposely designed website rather than a general social media site. Thus, an important goal in conducting successful interventions via online communities is to improve members' retention, as members who remain longer are more likely to receive these interventions and have more opportunities to promote a target behavior change. To achieve this goal, a critical first step is to understand what factors influence members' retention in an online community.

Previous studies have shown that people's decisions of retention or dropout in online communities are associated with a variety of factors [33,34], including personality traits (eg, shyness and the Big Five traits) [35,36], interests [37], recognition in a community [38-41], and support from others [42,43]. However, such an association is not adequate to conclude the presence of a causal relationship [44,45] between an individual's attributes and her or his online participation. This is because an association can arise from non-causal relationships. For example, most previous studies focus on the use of self-reported surveys and rely on participants' reports of their own personality, concerns, and behaviors [35,39,46]. This can introduce considerable retrospective bias and measurement errors, leading to a coincidental association between 2 unrelated variables, particularly in small samples. Even if variables are measured rather than self-reported [43,47], participation in an online community is inherently self-selected (eg, sharing common interests) and members can drop out for many different reasons (eg, effect of an online or offline event). Thus, unobservable factors (ie, confounding variables) may affect both a main predictor and participation outcomes, causing a spurious association. Moreover, in some cases reverse causality can lead to an association. For example, previous studies suggest that feelings of social isolation are linked to frequent social media use [35,48] whereas recent studies indicate that social media use is linked to increased feelings of social isolation [49]. Technically speaking, the issues of measurement errors, confounding variables, and reverse causality can cause endogeneity, which refers to an explanatory variable of interest being correlated with the error term in a regression model [44]. In these cases, traditional methods such as ordinary least squares (OLS) give biased and inconsistent estimates of the effect of interest. It is therefore not surprising that mixed results exist in previous studies. For example, a positive association between individuals' expertise and online participation was found in a study by Tausczik and Pennebaker [46] whereas a negative association was found in another study by Cook et al [38].

Objectives

This study aimed to estimate determinants of dropout in an online ED community, while addressing the endogeneity issues by using an instrumental variable (IV) approach [44]. Specifically, we analyzed tweeting activities for a large set of individuals who self-identified with EDs on Twitter for over 1.5 years and identified the presence of dropout if a user ceased to post tweets in the observation period. We explored determinants of a user's dropout based on the incentive theory [34,50], which argues that people's engagement in an activity can be driven by (a) intrinsic motivation which refers to doing something because it is interesting or enjoyable and (b) extrinsic motivation which refers to doing something because it earns an external reward. We focused on intrinsic motivation captured by personal emotions and extrinsic motivation captured by sociometric status in an online peer-to-peer community. Rather than using self-reports [35,39,46], we measured users' emotions based on their emotional expressions in tweets using sentiment analysis techniques [51] and quantified users' sociometric statuses by network centrality [52] in the social network of an ED community on Twitter. On the basis of these measured variables, IV estimators both for the decision to drop out and for the time to drop out were implemented to achieve consistent estimates of the effects of personal emotions and network centrality on dropout in an online ED community. To better understand the estimation results, we further examined heterogeneity in tweeting interests (ie, topics discussed in tweets) of users with differing levels of characteristics (eg, emotions) and dropout outcomes. To our knowledge, this study is the first to systematically characterize the determinants of dropout behaviors in online ED communities. A total of 3 research questions were examined: (a) what are the general characteristics of the attrition process in an online ED community? (b) how do intrinsic and extrinsic factors affect the decision of an individual to drop out of the community? and (c) how do these factors affect the duration of time until the occurrence of dropout?

Methods

Data Collection

Our data are collected from Twitter, a microblogging platform that allows millions of users to self-disclose and socialize. As many social media platforms such as Facebook and Instagram have taken moderation actions to counteract pro-ED content and user accounts [28], Twitter has not yet enforced actions to limit such content [53]. This makes Twitter a unique platform to study the attrition process naturally happening in an online ED community and allows us to examine individuals' behaviors in a nonreactive way. Our study protocol was approved by the Ethics Committee at the University of Southampton. All data used in our study are *public* information on Twitter and available through the Twitter application programming interfaces (APIs). No personally identifiable information was used in this study. Our data collection process included 3 phases:

1. First, we collected a set of individuals who self-identified with EDs on Twitter using a snowball sampling approach. Specifically, we tracked the public tweet stream using "eating disorder," "anorexia," "bulimia," and "EDNOS" (ie, eating disorder not otherwise specified) from January 8 to 15, 2016. This resulted in 1169 tweets that mentioned EDs. From the authors of these tweets, we identified 33 users who self-reported both ED-related keywords (eg, "eating disorder," "anorexia," and "bulimia") and personal bio-information (eg, body weight and height) in their profile descriptions (ie, a sequence of user-generated text describing their accounts below profile images). Starting from these seed users, we expanded the user set using snowball sampling through their social networks of followees and followers. At each sampling stage, we filtered out non-English speaking accounts and finally obtained 3380 unique ED users who self-reported ED-related keywords and bio-information in their profile descriptions. Note that our focus in this work is studying individuals who are affected by EDs rather than those who are related to EDs. The inclusion of bio-information in user sampling allowed us to filter out ED-related therapists, institutes, or organizations, as these users often displayed ED-related keywords but did not show bio-information in their Twitter profile descriptions. Details about the data collection of ED users can be found in our previous study [54].
2. Then, we collected all friends (including followees and followers) of each ED user, leading to a large social network consisting of 208,063 users. For each user, we retrieved up to 3200 (the limit returned from Twitter APIs) of their most recent tweets and obtained 241,243,043 tweets in total. The data collection process finished on February 11, 2016.
3. Finally, we opened a follow-up observation period for all users on August 17, 2017, to obtain measurements on users' activities online. In the second observation, we only collected users' profile information which includes users' last posted statuses.

To verify the quality of our collected sample, 2 members of the research team classified a random sample of 1000 users on whether they were likely to be a true ED user based on their posted tweets, images, and friends' profiles. Users were classified as "disordered" if they frequently and intensively posted their body weights, details of their dietary regimen (eg, calories), struggles with eating (eg, "I want to eat but cannot"), pictures of themselves, and self-reports of being disordered or in recovery in tweets and followed ED-related friends (eg, user profiles with ED-related keywords). The process revealed a

95.2% match between the identified ED individuals in the data collection stage and those classified as ED during inspection. Although it is impossible to diagnose individuals' disorders based on their online behaviors, this inspection provides a strong indication that the collected users are likely to be affected by EDs rather than those who merely talk about EDs online. See [54] for details of data validation.

Estimation Framework

There are 2 different models specified to estimate the effects of emotions and network centrality on dropout. First, we specified a linear probability model on the whole sample to estimate the effects of individuals' characteristics observed in the first observation period on the probability of dropping out in the second observation period. Finally, we estimated survival models to explore the effects of individuals' characteristics observed in the first observation on the time to dropout in the second observation (ie, the duration from our first observation to the dropout in our second observation). However, similar to all social media studies, only a limited number of individuals' characteristics are available for our estimations and these are mostly observed through user-generated data online. This leads to confounding variable bias, as unobservable factors can be correlated with both the main explanatory variables (ie, emotions and network centrality) and dropout outcomes. For example, undergoing hospital treatment can simultaneously affect a person's emotional state and the use of social media. Furthermore, previous studies have shown that social media use is associated with increased depression [55], social anxiety, [49] and body dissatisfaction [56,57], implying an effect of online participation on individuals' emotions (ie, reverse causality). Both confounding variables and reverse causality result in biased estimates of the effects of emotions and network centrality on dropout. This problem can be addressed by using a randomized controlled trial in which emotions or network centralities are randomly assigned to users by researchers [58]. Such a trial, however, is not feasible, because of ethical and practical limitations [59].

Here, we utilized an alternative approach for estimating the effects of interest that is based on IV regression, an econometric technique to infer causal relations from observational data [44]. This technique has been applied to a variety of contexts, from identifying the causal effect of education on earning [60], the effect of a health treatment [61], to estimating social contagion effects on both online [59] and offline behaviors [62]. Formally, consider a model $Y=Beta1*X1 + Beta2*X2 + u$, where $X1$ is endogenous, $X2$ is exogenous, u is a random error term, and *Beta*s are effects to be estimated. IV methodology uses an instrument Z (which is [i] not contained in the explanatory equation, [ii] correlated with $X1$, namely $cov(Z,X1)$ is not equal to 0, and [iii] uncorrelated with u, ie, $cov(Z,u)=0$, conditional on the other covariates such as $X2$) and runs a first stage reduced-form regression $X1 = B1*Z + B2*X2 + v$, where v is a random error and B s are coefficients. The causal effect of $X1$ on Y is then given in a second stage regression $Y=Beta1*X3 + Beta2*X2 + u$, where $X3$ is the predicted value of $X1$ from the first stage. See [44] for more details.

Measures

A number of variables are needed for estimations. All independent variables and IVs are measured in the first observation period (unless otherwise stated), whereas dependent variables are measured in the second observation period.

Dropout Outcomes as Dependent Variables

Following previous studies [42,43], we identified the presence of dropout if a user ceased to post tweets. Specifically, in the linear probability models, we encoded the dropout status of a user as 0 (denoting *nondropout*) if the user had updated posts in our second observation and 1 (denoting *dropout*) otherwise.

In the survival models, each user has a 2-variable outcome: (a) a censoring variable denoting whether the event of dropout occurs and (b) a variable of survival time denoting the duration of time until the occurrence of dropout. We censored the occurrence of a "dropout event" in 2 ways. First, users are said to drop out if they have not posted tweets for more than a fixed threshold interval π before our second observation (so called *identical interval censoring*). As people use social media platforms with different activity levels (eg, some users post every several hours whereas other users only post once every couple of days), our second censoring method further accounts for personalized posting activities of individuals (called *personalized interval censoring*). In this method, users are said to drop out if they have not posted tweets for more than a variate threshold interval $\lambda\pi + (1–\lambda) Ii$ before our second observation, where π is a fixed threshold, Ii is the average posting interval of individual i in our first observation period, and λ is a tunable parameter to control the effects of individual activities. We tuned the parameters by maximizing the agreement between the estimated dropout states based on users' activities in our first observation and the observed states in our second observation. See Multimedia Appendix 1 for details. For users who were censored as dropped-out, we set their survival times as the durations from our first observation to their last postings in our second observation. For those who were censored as non-dropped-out, we set their survival times as the whole time period between our 2 observations.

Emotions and Network Centrality as Main Explanatory Variables

Individuals' emotions were measured through their language used in tweets. There is a variety of sentiment analysis algorithms to measure emotional expressions in texts [51,63]. In this study, we used SentiStrength [51] as (a) it has been used to measure the emotional content in online ED communities and has shown good interrater reliability [26] and (b) it is designed for short informal texts with abbreviations and slang and is thus suitable to process tweets [51]. After removing mention marks, hashtags, and URLs, each tweet was assigned a scaled value ranging from −4 to 4 by SentiStrength, where negative and positive scores indicate the strength of negative and positive emotions, respectively, and 0 denotes neutral emotions. We quantified a user's emotional state by the average score of all tweets posted by the user. All retweets were excluded, as retweets reflect the emotions of their original authors more than those of their retweeters. To obtain robust results from the language processing algorithms, we only considered users who had more than 10 tweets and posted more than 50 words.

Network centrality measures the importance of a person in a social network; people well-recognized by their peers often have high centralities in a group [52]. To measure a user's centrality in the ED community, we built a who-follows-whom network among ED users and their friends, where a directed edge runs from node A representing user A to node B representing user B if A follows B on Twitter. Although there are various measures of network centrality, we focused on coreness centrality [64] as it has been shown to outperform other measures such as degree and betweenness centrality [52] in detecting influential nodes in complex networks [65] and cascades of users leaving an online community [66,67]. We measured the sociometric status of a user in the ED community by the in-coreness centrality [68] of a node in the generated network using the package igraph version 0.7.0 [69].

Aggregated Emotions and Network Centrality of Friends as Instrumental Variables

As IVs for a user's attributes, we used average emotions and network centrality over all followees of the user, namely people who are followed by the user. The choice of these IVs is based on the following considerations. First, we considered the relevance assumption of our instruments, requiring that the characteristics of followees be correlated to the user's characteristics, namely $cov(Z,X1)$ is not equal to 0. We expected the followees' updates to act as information sources for a user and followees' behaviors as well as emotions manifested in their tweets can influence the user. Previous work [54] has shown the presence of homophily among ED users on Twitter suggesting that users who share similar emotional and network attributes tend to follow one another. Furthermore, the empirical existence and strength of the relevance property are tested in a first-stage regression and presented along with the structural estimates of the models.

Finally, we examined the exogeneity requirement (ie, $cov(Z,u)=0$), where followees' emotions and centrality must not have a direct effect on the dropout decision of the user other than through their effect on the user's emotions. Although we have taken such assumptions to be reasonable, we identified a pathway through which direct links could arise. Followees' attributes (eg, emotions) could affect a user's dropout through their effects on followees' own dropouts. For example, followees' emotional states may affect their own dropouts, and a feeling of loneliness because of friends' leaving may then drive the target user to drop out. To control for this channel, we measured the proportion and durations of followees that remained active in our second observation (regardless of whether the target user dropped out or not). Furthermore, we changed the definition of followees (that are used to create the instruments) to those who are followed by a user but do not follow the user back (called *single way followees*). In this setting, the reverse causality of a user's dropout on followees' attributes was nullified, which strengthened the exogeneity assumption on IVs and controls.

Estimation Covariates

Our estimates control for several covariates that may affect users' tweeting activities, as listed in Table 1. First, we measured users' social capital on Twitter (eg, the numbers of social connections and the levels of engagement in sharing content) to capture the fact that people with different levels of popularity may have different tendencies to share content online [70]. Note that, although the numbers of followees and followers can be regarded as the in- and out-degree centralities of a user in the whole social network on Twitter (ie, the "global" social capital), we were interested in the "local" network centrality in the ED-specific communities. Second, as previous studies show an association between social media use and depression [55], we measured historical activity levels of users (ie, active days) to capture effects that previous engagement may relate to both users' emotions and their future engagement. We also measured users' activity frequencies (eg, posting frequency) to capture their patterns of Twitter usage. Third, the covariates on observational bias were used to control for effects caused by incomplete observations. For example, a limited number of tweets were retrieved and used to measure emotions for a user. All variables on social capital, activity level, and observational bias were measured from users' profile information and tweets collected in our first observation. Finally, as discussed above, we included the proportion and average durations of followees that were active in our second observation to capture the channel that followees' emotions affect a user's dropout through their effects on followees' own dropouts.

Model Estimations

Instrumental Variables Estimation in Linear Regression Model

We use standard 2-stage least squares estimators for linear probability models. In the first stage, we ran an auxiliary regression and predicted the endogenous variables (ie, an individual's emotional state and network centrality) based on IVs and exogenous covariates. In the second stage regression, we substituted the endogenous variables of interest with their predicted values from the first stage. Estimation was conducted through the package of applied econometrics with R [71], and robust standard errors were computed.

Instrumental Variables Estimation in Survival Model

We used a Kaplan-Meier estimator [72] to estimate the survival function from data. Aalen's additive hazards model [73] was used to estimate the effects of users' attributes on the time to dropout. Compared with the proportional hazards models in which the ratios of hazard functions (ie, hazard ratios) for different strata were assumed to be constant over time [74], the additive model was more flexible and applied under less restrictive assumptions. To compute an IV estimator in an additive hazards model, we used a control function based approach which was proposed by Tchetgen et al [61]. The timereg package [75] was used for the implementation of the estimation algorithm. Standard errors were obtained through nonparametric bootstrap.

Table 1. Covariates used in estimations.

Control effect and covariate	Description
Social capital	
#Followees	Number of total followees
#Posts	Number of total posts, including tweets and retweets
#Followers	Number of total followers
Activity level	
Active days	Number of days from account creation to last posting
#Followee/day	Average number of followees per day
#Posts/day	Average number of posts per day
#Followers/day	Average number of followers per day
Observational bias	
#Tweets in use	Number of tweets in use to measure emotions
#Followees in use	Number of followees whose attributes are used as instruments
Alternative causal channel	
%Active followees	Proportion of followees being active between 2 observations
Avg. duration of followees	Average days of followees being active between 2 observations

Results

Descriptive Statistics

We obtained 2906 users who posted more than 10 tweets (excluding retweets) and 50 words in our data, where 84.61% (*n*=2459) of users had no posting activities during our 2 observation periods. Among the 357 users who self-reported gender information in their Twitter profile descriptions, 84.0% of them (300) were female. The mean age was 17.3 years among ED users who self-reported age (*n*=1015). On the basis of the timestamps of account creation and last posting, we used the Kaplan-Meier estimator to estimate the “lifetime” of a user on Twitter, namely the duration from account creation to the last posting. The estimated median lifetime of these users on Twitter was 6 months. That is, one half of the entire cohort drops out at 6 months after creating an account. Figure 1 visualizes the social network between dropouts and nondropouts among ED users, laid out by the Fruchterman-Reingold algorithm [76]. We noted that users with the same dropout states tended to cluster together. Computing Newman’s homophily coefficient *r* [77] of this network by users’ dropout states, we found *r*=0.09 (*z*=16.84 and *P*<.001 compared with a null model, see Multimedia Appendix 1), suggesting that users with the same dropout states tended to befriend one another. See Multimedia Appendix 1 for details of data statistics.

Estimation Results of Linear Probability Models

Table 2 shows estimated results in the linear models with 2 different IV specifications. In the first specification, we used all followees of a user to create IVs for the user’s attributes. The results are given in columns 2 to 3, in which both OLS and IV estimators show that positive emotions are associated with a higher probability of dropout (beta=.044; *P*=.007 and beta=.29; *P*<.001, respectively), with largely comparable coefficients for covariates. Compared with the OLS estimator, the IV estimator of the effect of emotions on dropout was remarkably stronger. The Wu-Hausman test further showed a significant difference between the OLS and IV estimators (*P*<.01), suggesting the presence of endogeneity. These results indicated that ignoring endogeneity in the OLS estimation leads to an underestimation of the effect of interest. Moreover, the *F* statistics in the first stage regressions showed that the relevance of IVs exceeds the conventional standard of *F*=10 [78], indicating the validity of our IVs.

Figure 1. The who-follows-whom network among eating disorder users on Twitter. Node colors represent dropout statues, where the red color denotes dropout and the green color denotes nondropout. Node size is proportional to the in-coreness centrality.

Table 2. Estimated effects of emotions on dropout using ordinary least squares (OLS) and instrumental variables (IVs) models.

Variable	All followees (n=2906[a])				Single way followees (n=2898[a])			
	Ordinary least squares		Instrumental variables		Ordinary least squares		Instrumental variables	
	Beta	P[b] value	Beta	P[b] value	Beta	P[b] value	Beta	P[b] value
Emotions[c,d]	.044	.005	0.29	<.001	.064	<.001	.304	<.001
#Followees	−.0004	.01	−.0002	.18	−.0001	.11	−.0001	.22
#Posts	−.00000	.18	−0.00000	.35	−.00001	.06	−.00001	.14
#Followers	.00001	.53	.00001	.58	.00001	.7	0	.82
Active days	−.0003	<.001	−.0003	<.001	−.0003	<.001	−.0003	<.001
#Followee/day	.001	.01	.002	<.001	.001	.03	.002	.003
#Posts/day	.0002	.72	−.001	.38	.0001	.85	−.001	.32
#Followers/day	−.003	.01	−.005	<.001	−.003	.02	−.004	.002
#Tweets in use	−.00004	.002	−.00004	.003	−.00003	.03	−.00003	.03
#Followees in use	.0004	.03	.0002	.38	0	.96	−.0001	.41
%Active followees	−1.159	<.001	−.812	<.001	−.939	<.001	−.655	<.001
Avg. duration of followees	.001	.004	.001	.16	.001	.005	.0005	.19
Constant	1.27	<.001	1.273	<.001	1.246	<.001	1.251	<.001

[a] The numbers of samples that are used in estimations.

[b] P values are computed based on heteroscedasticity-consistent standard errors.

[c] F statistic tests the significance of the instrument from a first-stage regression of a user's emotions on followees' emotions (ie, the instrument) and the rest of the covariates, where $F=440.26$ ($P<.001$) when all followees are used to build an instrument and $F=158.21$ ($P<.001$) when only single way followees are used.

[d] Wu-Hausman (H) statistic tests the difference in estimates between OLS and IV, where $H=42.24$ ($P<.001$) when all followees are used to build an instrument and $H=14.54$ ($P<.001$) when only single way followees are used. Rejecting the null hypothesis suggests the presence of endogeneity.

Columns 4 to 5 show results of the second IV specification in which only single way followees are used to create IVs. Users who have no single way followees are excluded as instruments for these users' attributes were not available. Thus, the number of observations decreased as compared with that in the first IV specification. Moreover, as data on a smaller number of friends were used in the second IV specification, the relevance of IV became weaker but still passed the conventional test in the first stage regression. Despite such changes, the 2 specifications produced largely similar results. Computing Wald tests of equality of coefficients between the 2 IV models, we found that the estimated effects of emotions on dropout were statistically the same across different IV specifications (P=.8), potentially suggesting robustness of the results.

Note that network centrality was excluded from the linear models. This is because many users had dropped out long before our first observation (see Figure S2 in Multimedia Appendix 1), and the social networks of such users might largely change from the dates of their dropouts to our first observation. For example, a user might be followed by new followers when these followers were unaware of the dropout of this user. In these cases, network centralities in the future are used to explain dropouts in the past, which can produce misleading results in the linear models. In fact, including network centrality in the above linear models produces statistically insignificant effect of centrality on the dropout decision, confirming our argument that network centrality is irrelevant to the binary decision to drop out or not.

Estimation Results of Survival Models

In the survival models, we only considered users who were active past our first observation period, so as to examine the effect of network centralities in our first observation on users' activities in the second observation period. Table 3 shows mean coefficients of emotions and network centrality in the survival models. All models are estimated controlling for the full list of covariates but are omitted from the tables because of space concerns. The complete results are available from the authors. Following [61], the effects of all covariates are assumed to be time dependent in estimations. Both the standard and IV models on the identical interval censored data show that (a) positive emotions lead to a shorter survival time (P<.05 in the IV model) and (b) a core position in social networks is associated with a longer survival time (P<.05 in both models). Estimations on the personalized interval censored data and using different IV specifications gave similar results. The strong relevance of IVs in the first stage regressions confirms the validity of IVs across different models. A comparison of results between the linear and survival models further shows that these models have consistent estimators for the effect of emotions on dropout, namely positive emotions increase the likelihood to drop out.

Table 3. Estimated effects of emotions and centrality on survival time using Aalen’s additive hazards models.

Variable[a,b]	All followees (n=447[c])		Single way followees (n=445[c])	
	Standard (95% CI[d])	IV[e] (95% CI)	Standard (95% CI)	IV (95% CI)
Identical interval censoring				
Emotions	−0.018 (−0.037 to 0.0002)	−0.043 (−0.083 to −0.004)	−0.018 (−0.036 to 0.0006)	−0.061 (−0.116 to −0.011)
Centrality	0.001 (0.0008 to 0.0011)	0.001 (0.0007 to 0.0011)	0.001 (0.0008 to 0.0011)	0.001 (0.0006 to 0.0011)
Personalized interval censoring				
Emotions	−0.016 (−0.034 to 0.0031)	−0.038 (−0.08 to 0.002)	−0.015(−0.034 to 0.0026)	−0.056 (−0.115 to −0.007)
Centrality	0.001 (0.0008 to 0.0011)	0.001 (0.0008 to 0.0012)	0.001 (0.0008 to 0.0011)	0.001 (0.0007 to 0.0011)

[a] F-statistic tests the joint significance of the 2 instruments from a first-stage regression of a user’s emotions on followees’ emotions and followees’ centralities (ie, the instruments) plus the rest of the covariates, where $F=66.11$ ($P<.001$) when all followees are used to build instruments and $F=34.99$ ($P<.001$) when only single way followees are used.

[b] F-statistic tests the joint significance of the 2 excluded instruments from a first-stage regression of a user’s centrality on followees’ emotions and followees’centralities (ie, the instruments) plus the rest of the covariates, where $F=27.85$ ($P<.001$) when all followees are used to build instruments and $F=12.62$ ($P<.001$) when only single way followees are used.

[c] The numbers of samples that are used in estimations.

[d] CIs for coefficients are obtained from 1000 bootstrap replicates. A coefficient is significant at $P<.05$ if 0 is not in 95% CIs.

[e] IVs stands for instrumental variables.

Underlying Connection Between Emotions and Dropout

To better understand the relationships between emotions and dropout, we examined posting interests among users with different dropout statuses and emotional states based on hashtags used in users’ tweets (see Multimedia Appendix 1 for details). We found nondropouts were interested in advocating a thin ideal (eg, using hashtags “mythinspo” and “skinny4xmas”) and promoting a pro-ED identity (eg, “edlogic” and “beautiful”). In contrast, dropouts engage in discussing their health problems (eg, “selfharmprobz,” “bulimicprobz,” and “anorexicprobz”) and offering emotional support for others (eg, “anasisters” and “stayingstrong”), which implies a tendency of these users to recover from disorders [20-22]. Similarly, we split all ED users into 3 equal-size sets based on their emotional scores and examined hashtags used by each set of users. We found that users with negative emotions often engage in promoting thin ideals (eg, “bonespo” and “mythinspo”), showing largely overlapping interests with the nondropouts. In contrast, users with neutral and positive emotions were more interested in discussing their health problems (eg, “anorexicprobz” and “bulimicprobz”), opposing pro-ED promotions (eg, “reversethinspo”), and encouraging healthier body image and behaviors (eg, “fitfam” and “fitness”), showing similar interests with the dropouts. See Multimedia Appendix 1 for more detailed lists of hashtags.

Measuring the Spearman rank correlation ρ between pairwise lists of hashtags posted by users with a given state (eg, dropped out or not, and positive or negative), we found a positive correlation between negative users and nondropouts in hashtag usage (ρ=0 . 36; P=.003 in Table 4), indicating similar posting interests among these users. A similar pattern occurred between positive users and dropouts. In contrast, users with other pairs of states showed a negative correlation or noncorrelation in hashtag usage, indicating their discrepancies in posting interests. Note that all tags in 2 lists *Di* and *Dj* are considered in computing the correlation ρ (*Di*, *Dj*); tags in each list are ranked by TF-IDF scores [79] and the TF-IDF score of tag *t* in list *Di* is 0 if *Di* does not contain *t*. These results revealed a possible underlying connection between positive emotions and dropout. Compared with users with positive emotions, those with negative emotions had more similar interests to active members (ie, nondropouts) in the ED community. Finding similarities with other members in a community can enhance a sense of belonging to the community and positively increase intention to engage in community activities [33,37]. Therefore, it is not surprising that negative users are less likely to drop out than positive users in our estimations.

Table 4. Spearman rank correlations between pairwise lists of hashtags posted by users with a given dropout state and by users with a given emotional state, respectively.

Group	Negative (n=61[a])	Neutral (n=108)	Positive (n=110)
Nondropout (n=54[a])	0.36 (P=.003)[b]	−0.21 (P=.03)	−0.66 (P<.001)
Dropout (n=227)	−0.33 (P<.001)	−0.04 (P=.57)	0.12 (P=.07)

[a] The number of hashtags posted by users with a given state.

[b] The Spearman correlations ρ of hashtags posted by users with different dropout and emotional states, where ρ ranges from -1 to 1 and 0 indicates no correlation. P values testing for noncorrelation are reported in parentheses.

Discussion

Principal Findings

This study provided the first estimates of the effects of personal emotions and interpersonal social networks on dropout in online ED communities. This study has several strengths. First, we based our analysis on the incentive theory to explore determinants of users' online behaviors (ie, dropout), allowing us to study users' behaviors in a more systematic way than most previous studies that often focus on a single type of determinant (eg, individual attributes [35,36] or social attributes [42,43,67]). Second, we used automated sentiment analysis techniques to measure users' emotions and network analysis methods to quantify users' sociometric statuses in an online community, leading to higher efficiency than traditional research methods such as surveys [35,37,39,41,48]. Finally, we applied an IV approach to both linear probability and survival models, which enabled us to achieve a more consistent estimate of human behavior in online settings than traditional methods (eg, OLS) used in previous studies [39,41,47]. Overall, we found that positive emotions increased the likelihood of dropout in ED individuals and accelerated the dropout process on Twitter. In contrast, a central position in the social network of ED individuals at an earlier stage was associated with prolonged participation of an individual at a later stage. These findings were verified across a variety of robustness checks.

Despite differences in methodology, our findings aligned with previous studies in psychological and social media research [5,33,35]. Our results suggested that ED users with negative emotions had high levels of participation on Twitter. This aligned with previous survey studies on social media use (eg, Facebook use), where people with social anxiety and shyness (ie, personality traits that are often correlated with multiple negative emotions such as feeling lonely, isolated, and unhappy [80]) were found to spend more time online [35,48,81]. An explanation for this is the online disinhibition effect [82]. Specifically, because of anonymity in online interactions, people with social inhibitions (eg, those who are socially anxious or shy and those with a stigmatized health problem [83]) might be more willing to share personal feelings and reveal themselves in online interactions than offline interactions to meet their social and intimacy needs [48]. Additional analyses on users' posting interests revealed that users with negative emotions share similar interests with active users. This allowed us to confirm the validity of our results via the social capital theory [39,40], namely sharing common attributes (eg, interests and vision) with other members can enhance a sense of belonging and positive feeling toward a community, which drives people to actively engage in the community.

Consistent with positive associations between network centrality and active participation in other online communities [67,70], we find that central users in the social network of an ED community tend to have a longer-lasting participation in the community. This result is expected for several reasons. First, users who are centrally embedded in a group have a relatively high number of social ties with other members, which can lead these users to feel being socially accepted and approved, as well as a strong sense of belonging to the group. Previous studies have consistently shown that recognition from other members and identification within an online community increase an individual's commitment to the community [34,39-41]. Finally, information shared by central users is likely to spread to the majority of a community through social ties, and their central positions in the community may promote other members to trust such information [70]. This implies that central users have a greater potential than peripheral users in influencing members' opinions, emotions, and behaviors in online communities [84]. Thus, compared with peripheral users, feeling influential may provide an additional incentive for central users to continue participating.

In line with previous studies on online ED communities [15,17,21], we found that ED users on Twitter have different stances on EDs, where users with negative emotions often share pro-ED content and those with positive emotions often share prorecovery content. As pro-ED content often contains thin-ideal images and harmful tips for weight loss and control [8,24,25], this result aligned with clinical evidence on ED treatment showing that more emotional distress is associated with a higher risk to learn and develop dysfunctional coping behaviors among ED sufferers [5]. Thus, as suggested by previous studies [85], engaging in pro-ED content may serve as a coping mechanism to deal with emotional pressures and stress of EDs. A possible explanation for the association between engaging in harmful online content and coping with stress is sensation seeking [86], a basic personality trait defined as the seeking of varied, novel, complex and intense sensations and experiences, and the willingness to take risks. Several studies have shown that sensation seeking is prominent in adolescence (ie, the age at which disordered eating often develops [1]) and closely related to pathological internet use, such as use of violent sites [87] and internet dependence [88].

Our study also offered new insights into online ED communities. First, ED users have a high dropout rate (85% in our sample) and a short lifespan between an account creation to lost posting on Twitter (with 6 months of median time to drop out). This aligns with views of online ED communities as hidden, secretive groups [30], but also indicates the dynamic characteristics of these communities. Second, users who discuss their health problems and share prorecovery content (ie, prorecovery users) have lower levels of posting activities (ie, a higher dropout rate) than those who share pro-ED content (ie, pro-ED users) on Twitter. This can be explained as follows. Owing to common interests in EDs, prorecovery and pro-ED groups are likely to be connected in the same social networks, and content shared within a group is hence likely to be visible to the other group. However, exposure to content from the antagonist group can have distinct effects in pro-ED and prorecovery groups. Exposure to pro-ED content is harmful for prorecovery users and can impede their recovery process [3,24], whereas exposure to prorecovery content can instead stimulate harmful behaviors in pro-ED users (eg, actively sharing pro-ED content) [21]. Thus, prorecovery users might tend to leave such an online community to avoid a risk of further deterioration or relapse. Our finding may also explain why pro-ED content is found being more pervasive than prorecovery content across social media sites [15,17,21], for example, almost 5 times in terms of unique publishers on Tumblr [15]. Finally, ED users tend to connect with others with the same dropout states on Twitter. This implies that whether an individual drops out from online communities depends on whether others in the individual's social networks drop out. In other words, dropout in online ED communities is not only a function of individual experience or individual choice but also a property of group interactions, such as homophily [89] and social contagion effects [59].

Implications

Our findings are of practical relevance to the promotion of public health over social media. First, the decision to maintain active participation in an online community can be caused by intrinsic and extrinsic characteristics or traits of the participants, such as personal emotions, interests, and social networks. Such self-selection bias can lead to the sample not being representative of the whole population, and hence researchers need to consider both active and dropped-out users for a well-rounded picture of online health communities. This is particularly important for public health officials to make special efforts to reach these dropouts and offer more intensive support when they are trying to recover. Second, high attrition rates are often regarded as negative outcomes in online interventions, particularly in those delivered over a purposely designed website [11,31,32]. However, this may or may not be the case in interventions over general social media sites (eg, Twitter) depending on how targeted populations use these sites. For example, when an intervention is delivered in an online community in which members often shared harmful content, a high attrition rate (ie, members dropping out of the harmful community) may not be a negative outcome. Using automated data-mining techniques to track users' behaviors (eg, emotions and posting interests), as used in this study, can provide more detailed information about people's use of online health communities and improve our understanding of attrition in online interventions. Third, interventions that recommend content containing positive emotions to ED users (not limited to ED-related content but more general content containing happiness and inspiration) may reduce their engagement in a harmful online community. This aligns with Fredrickson's broaden-and-build model which argues that cultivating positive emotions is useful to prevent and treat mental health problems [90]. Finally, intervention strategies could be tailored for different individuals depending on their positions in the social network of an online community. For example, identifying central individuals as change agents might enhance the efficacy and cost-effectiveness of an intervention because of their greater influence potential through larger numbers of social ties [91] and also their longer-lasting effects through longer-term participation in the community.

Limitations

First, we recognized that self-diagnosis information on Twitter might be itself self-censored by users to align with their personality traits and perceptions of their audience on the platform. People may not use tags such as "eating disorder" to self-report their experience of illness and would be excluded by our collection methods. Also, although over 208,000 users and over 241,000,000 tweets are studied in this study, a small sample of rich social media data is used to explore the attrition of ED communities on Twitter. Thus, generalization of our results to all ED-related online communities should be cautious. Second, our measures of dropout are based on posting activity, whereas some people primarily use Twitter to receive outside information but rarely post their own information. We have little activity data on these users and hence less understanding of the characteristics of their dropout. This thus raises important issues that need further research to enhance our understanding of attrition in online health communities, such as consensus and clarity about the definition of dropout. Third, our study focused on the Twitter platform, without validation on other platforms. However, stopping using a platform can be related to the attractiveness of the platform. Hence, future research is also in need to examine many other factors that we did not explore but can affect dropout on social media, such as individual personality, physical health states, perceptions, and purposes of using a particular social media platform. Fourth, user accounts on a social media site are often not unique—an individual may have multiple accounts on the same site. Thus, we cannot be certain whether individuals who stopped using an account will engage in the same or similar online communities through other accounts. In other words, the dropout of a user account may not necessarily imply that an individual abandons a specific identity (eg, pro-ED) shared within a community. Finally, although our methodology allows us to establish a causal link from emotions to dropout behavior, it offers limited insights into the pathways through which this link exactly operates, and future work is needed to explore such issues in detail.

Conclusions

This study presented a systematic characterization of attrition in an ED community on Twitter. Our analysis offered the first attempt toward the estimation of the effects of personal emotions and network centrality on dropout behaviors in individuals affected by ED on Twitter. Our results provided new insights into the trajectories that ED communities develop online which can help public health officials to better understand individual needs in using online ED communities and provided tailored support for individuals with different needs.

Acknowledgments

This work is supported by ESRC Doctoral Training Centre (#ES/J500161/1), Institute for Life Sciences, WSI-RCSF and SocSCI-SIRF, University of Southampton, and the Alan Turing Institute, UK.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary information for estimating determinants of attrition in eating disorder communities on Twitter: an instrumental variables approach.

[PDF File (Adobe PDF File), 700KB - jmir_v21i5e10942_app1.pdf]

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). Washington, DC: The American Psychiatric Association; 2013.
2. Arcelus J, Mitchell AJ, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. Arch Gen Psychiatry 2011 Jul;68(7):724-731. [doi: 10.1001/archgenpsychiatry.2011.74] [Medline: 21727255]
3. Campbell K, Peebles R. Eating disorders in children and adolescents: state of the art review. Pediatrics 2014 Sep;134(3):582-592 [FREE Full text] [doi: 10.1542/peds.2014-0194] [Medline: 25157017]
4. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
5. Corstorphine E. Cognitive–emotional–behavioural therapy for the eating disorders: working with beliefs about emotions. Eur Eat Disorders Rev 2006 Nov;14(6):448-461. [doi: 10.1002/erv.747]
6. Swanson SA, Crow SJ, Le Grange D, Swendsen J, Merikangas KR. Prevalence and correlates of eating disorders in adolescents. Results from the national comorbidity survey replication adolescent supplement. Arch Gen Psychiatry 2011 Jul;68(7):714-723. [doi: 10.1001/archgenpsychiatry.2011.22] [Medline: 21383252]
7. Rich E. Anorexic dis(connection): managing anorexia as an illness and an identity. Sociol Health Illn 2006 Apr;28(3):284-305 [FREE Full text] [doi: 10.1111/j.1467-9566.2006.00493.x] [Medline: 16573717]
8. Wilson JL, Peebles R, Hardy KK, Litt IF. Surfing for thinness: a pilot study of pro-eating disorder web site usage in adolescents with eating disorders. Pediatrics 2006 Dec;118(6):e1635-e1643. [doi: 10.1542/peds.2006-1133] [Medline: 17142493]
9. Latkin CA, Knowlton AR. Social network assessments and interventions for health behavior change: a critical review. Behav Med 2015;41(3):90-97 [FREE Full text] [doi: 10.1080/08964289.2015.1034645] [Medline: 26332926]
10. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
11. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
12. McLean SA, Wertheim EH, Masters J, Paxton SJ. A pilot evaluation of a social media literacy intervention to reduce risk factors for eating disorders. Int J Eat Disord 2017 Jul;50(7):847-851. [doi: 10.1002/eat.22708] [Medline: 28370321]
13. Stapleton JL, Manne SL, Day AK, Levonyan-Radloff K, Pagoto SL. Healthy body image intervention delivered to young women via Facebook groups: formative study of engagement and acceptability. JMIR Res Protoc 2018 Feb 20;7(2):e54 [FREE Full text] [doi: 10.2196/resprot.9429] [Medline: 29463495]
14. Bazarova NN, Choi YH. Self-disclosure in social media: extending the functional approach to disclosure motivations and characteristics on social network sites. J Commun 2014 Jun 19;64(4):635-657. [doi: 10.1111/jcom.12106]
15. De Choudhury M. Anorexia on Tumblr: A Characterization Study. 2015 Presented at: Proceedings of the 5th International Conference on Digital Health 2015; May 18-20, 2015; Florence, Italy. [doi: 10.1145/2750511.2750515]
16. Chancellor S, Lin Z, Goodman E, Zerwas S, De Choudhury M. Quantifying and predicting mental illness severity in online pro-eating disorder communities. 2016 Presented at: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; February 27-March 02, 2016; San Francisco, California, USA p. 1171. [doi: 10.1145/2818048.2819973]
17. Chancellor S, Mitra T, De Choudhury M. Recovery Amid Pro-Anorexia: Analysis of Recovery in Social Media. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 May Presented at: CHI '16; May 07-12, 2016; San Jose, California, USA p. 2111-2123 URL: http://europepmc.org/abstract/MED/28840201 [doi: 10.1145/2858036.2858246]
18. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]

19. Chancellor S, Kalantidis Y, Pater J, De Choudhury M, Shamma D. Multimodal Classification of Moderated Online Pro-Eating Disorder Content. 2017 Presented at: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 06-11, 2017; Denver, Colorado, USA p. 3213. [doi: 10.1145/3025453.3025985]
20. Lyons EJ, Mehl MR, Pennebaker JW. Pro-anorexics and recovering anorexics differ in their linguistic Internet self-presentation. J Psychosom Res 2006 Mar;60(3):253-256. [doi: 10.1016/j.jpsychores.2005.07.017] [Medline: 16516656]
21. Yom-Tov E, Fernandez-Luque L, Weber I, Crain SP. Pro-anorexia and pro-recovery photo sharing: a tale of two warring tribes. J Med Internet Res 2012;14(6):e151 [FREE Full text] [doi: 10.2196/jmir.2239] [Medline: 23134671]
22. Wolf M, Theis F, Kordy H. Language use in eating disorder blogs: psychological implications of social online activity. J Lang Soc Psychol 2013 Feb 07;32(2):212-226. [doi: 10.1177/0261927X12474278]
23. Borzekowski DL, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: a content analysis of pro-eating disorder web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]
24. Maloney P. Online networks and emotional energy: how pro-anorexic websites use interaction ritual chains to (re) form identity. Inf Commun Soc 2013 Feb;16(1):105-124. [doi: 10.1080/1369118X.2012.659197]
25. Bardone-Cone AM, Cass KM. Investigating the impact of pro-anorexia websites: a pilot study. Eur Eat Disorders Rev 2006 Jul;14(4):256-262. [doi: 10.1002/erv.714]
26. Oksanen A, Garcia D, Sirola A, Näsi M, Kaakinen M, Keipi T, et al. Pro-anorexia and anti-pro-anorexia videos on YouTube: sentiment analysis of user responses. J Med Internet Res 2015 Nov 12;17(11):e256 [FREE Full text] [doi: 10.2196/jmir.5007] [Medline: 26563678]
27. Chancellor S, Lin Z, De Choudhury M. "This Post Will Just Get Taken Down": Characterizing Removed Pro-Eating Disorder Social Media Content. 2016 Presented at: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 07-12, 2016; San Jose, California, USA p. 1157-1162. [doi: 10.1145/2858036.2858248]
28. Chancellor S, Pater J, Clear T, Gilbert E, De Choudhury M. #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. 2016 Presented at: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; February 27-March 02, 2016; San Francisco, California, USA p. 1201. [doi: 10.1145/2818048.2819963]
29. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2014;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
30. Casilli AA, Pailler F, Tubaro P. Online networks of eating-disorder websites: why censoring pro-ana might be a bad idea. Perspect Public Health 2013 Mar;133(2):94-95. [doi: 10.1177/1757913913475756] [Medline: 23467530]
31. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
32. ter Huurne ED, Postel MG, de Haan HA, van der Palen J, DeJong CA. Treatment dropout in web-based cognitive behavioral therapy for patients with eating disorders. Psychiatry Res 2017 Dec;247:182-193. [doi: 10.1016/j.psychres.2016.11.034] [Medline: 27918968]
33. Malinen S. Understanding user participation in online communities: a systematic literature review of empirical studies. Comput Human Behav 2015 May;46:228-238. [doi: 10.1016/j.chb.2015.01.004]
34. Kollock P. The economies of online cooperation. In: Communities in cyberspace. London: Taylor & Francis Group; 1999.
35. Orr ES, Sisic M, Ross C, Simmering MG, Arseneault JM, Orr RR. The influence of shyness on the use of Facebook in an undergraduate sample. Cyberpsychol Behav 2009 Jun;12(3):337-340. [doi: 10.1089/cpb.2008.0214] [Medline: 19250019]
36. Hughes DJ, Rowe M, Batey M, Lee A. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. Comput Human Behav 2012 Mar;28(2):561-569. [doi: 10.1016/j.chb.2011.11.001]
37. Casaló LV, Flavián C, Guinalíu M. New members' integration: key factor of success in online travel communities. J Bus Res 2013 Jun;66(6):706-710. [doi: 10.1016/j.jbusres.2011.09.007]
38. Cook E, Teasley S, Ackerman M. Contribution, commercialization & audience: understanding participation in an online creative community. 2009 Presented at: Proceedings of the ACM 2009 international conference on Supporting group work; May 10-13, 2009; Sanibel Island, Florida, USA p. 41-50. [doi: 10.1145/1531674.1531681]
39. Chiu C, Hsu M, Wang ET. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. Decis Support Syst 2006 Dec;42(3):1872-1888. [doi: 10.1016/j.dss.2006.04.001]
40. Nahapiet J, Ghoshal S. Social capital, intellectual capital, and the organizational advantage. Acad Manag Rev 1998 Apr;23(2):242. [doi: 10.2307/259373]
41. Lai H, Chen TT. Knowledge sharing in interest online communities: a comparison of posters and lurkers. Comput Human Behav 2014 Jun;35:295-306. [doi: 10.1016/j.chb.2014.02.004]
42. Wang Y, Kraut R, Levine J. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. 2012 Presented at: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work; February 11-15, 2012; Seattle, Washington, USA p. 833. [doi: 10.1145/2145204.2145329]
43. Xing W, Goggins S, Introne J. Quantifying the effect of informational support on membership retention in online communities through large-scale data analytics. Comput Human Behav 2018 Sep;86:227-234. [doi: 10.1016/j.chb.2018.04.042]

44. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996 Jun;91(434):444-455. [doi: 10.2307/2291629]
45. Stovitz SD, Verhagen E, Shrier I. Distinguishing between causal and non-causal associations: implications for sports medicine clinicians. Br J Sports Med 2017 Nov 21 Epub ahead of print(forthcoming). [doi: 10.1136/bjsports-2017-098520] [Medline: 29162620]
46. Tausczik Y, Pennebaker J. Participation in an online mathematics community: differentiating motivations to add. 2012 Presented at: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work; February 11-15, 2012; Seattle, Washington, USA p. 207. [doi: 10.1145/2145204.2145237]
47. Wu B. Patient continued use of online health care communities: web mining of patient-doctor communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
48. Sheeks MS, Birchmeier ZP. Shyness, sociability, and the use of computer-mediated communication in relationship development. Cyberpsychol Behav 2007 Feb;10(1):64-70. [doi: 10.1089/cpb.2006.9991] [Medline: 17305450]
49. Primack BA, Shensa A, Sidani JE, Whaite EO, Lin LY, Rosen D, et al. Social media use and perceived social isolation among young adults in the US. Am J Prev Med 2017 Jul;53(1):1-8 [FREE Full text] [doi: 10.1016/j.amepre.2017.01.010] [Medline: 28279545]
50. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
51. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. J Am Soc Inf Sci 2010 Dec 15;61(12):2544-2558. [doi: 10.1002/asi.21416]
52. Friedkin NE. Theoretical foundations for centrality measures. Am J Sociol 1991 May;96(6):1478-1504. [doi: 10.1086/229694]
53. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. #Proana: pro-eating disorder socialization on Twitter. J Adolesc Health 2016 Jun;58(6):659-664. [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]
54. Wang T, Brede M, Ianni A, Mentzakis E. Detecting and Characterizing Eating-Disorder Communities on Social Media. 2017 Presented at: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining; February 06-10, 2017; Cambridge, United Kingdom p. 91-100. [doi: 10.1145/3018661.3018706]
55. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among US young adults. Depress Anxiety 2016 Apr;33(4):323-331. [doi: 10.1002/da.22466] [Medline: 26783723]
56. Mabe AG, Forney KJ, Keel PK. Do you "like" my photo? Facebook use maintains eating disorder risk. Int J Eat Disord 2014 Jul;47(5):516-523. [Medline: 25035882]
57. de Vries DA, Peter J, de Graaf H, Nikken P. Adolescents' social network site use, peer appearance-related feedback, and body dissatisfaction: testing a mediation model. J Youth Adolesc 2016 Jan;45(1):211-224 [FREE Full text] [doi: 10.1007/s10964-015-0266-4] [Medline: 25788122]
58. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
59. Coviello L, Sohn Y, Kramer AD, Marlow C, Franceschetti M, Christakis NA, et al. Detecting emotional contagion in massive social networks. PLoS One 2014;9(3):e90315 [FREE Full text] [doi: 10.1371/journal.pone.0090315] [Medline: 24621792]
60. Card D. The Causal Effect of Education on Earnings. In: Handbook of labor economics. vol. 3. Netherlands: Elsevier; 1999:1801.
61. Tchetgen TE, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. Epidemiology 2015 May;26(3):402-410 [FREE Full text] [doi: 10.1097/EDE.0000000000000262] [Medline: 25692223]
62. Aral S, Nicolaides C. Exercise contagion in a global social network. Nat Commun 2017 Dec 18;8:14753 [FREE Full text] [doi: 10.1038/ncomms14753] [Medline: 28418379]
63. Gonçalves P, Araújo M, Benevenuto F, Cha M. Comparing and combining sentiment analysis methods. 2013 Presented at: Proceedings of the first ACM conference on Online social networks; October 07-08, 2013; Boston, Massachusetts, USA p. 27-38. [doi: 10.1145/2512938.2512951]
64. Seidman S. Network structure and minimum degree. Soc Networks 1983 Sep;5(3):269-287. [doi: 10.1016/0378-8733(83)90028-x]
65. Kitsak M, Gallos L, Havlin S, Liljeros F, Muchnik L, Stanley H, et al. Identification of influential spreaders in complex networks. Nature Phys 2010 Aug 29;6(11):888-893. [doi: 10.1038/nphys1746]
66. Garcia D, Mavrodiev P, Schweitzer F. Social resilience in online communities: the autopsy of friendster. 2013 Presented at: Proceedings of the first ACM conference on Online social networks; October 07-08, 2013; Boston, Massachusetts, USA p. 39. [doi: 10.1145/2512938.2512946]
67. Garcia D, Mavrodiev P, Casati D, Schweitzer F. Understanding popularity, reputation, and social influence in the Twitter society. Policy & Internet 2017;9(3):343-364. [doi: 10.1002/poi3.151]
68. Giatsidis C, Thilikos D, Vazirgiannis M. D-cores: Measuring Collaboration of Directed Graphs Based on Degeneracy. 2013 Presented at: 2011 IEEE 11th International Conference on Data Mining; December 11-14, 2011; Vancouver, BC, Canada p. 343. [doi: 10.1109/icdm.2011.46]

69. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Systems 2006;1695(5):1-9 [FREE Full text]
70. Wasko MM, Faraj S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS Q 2005;29(1):35-57. [doi: 10.2307/25148667]
71. Kleiber C, Zeileis A. Applied Econometrics With R. New York: Springer; 2008.
72. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53(282):457-481. [doi: 10.1007/978-1-4612-4380-9_25]
73. Aalen O, Borgan O, Gjessing H. Survival And Event History Analysis: A Process Point Of View (statistics For Biology And Health). New York: Springer; 2008.
74. Cox D. Regression Models and Life-Tables. In: Breakthroughs in Statistics. New York: Springer; 1992:1-4612.
75. Scheike TH, Zhang M. Analyzing competing risk data using the R timereg package. J Stat Soft 2011;38(2). [doi: 10.18637/jss.v038.i02]
76. Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. Software Pract Exper 1991 Nov;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
77. Newman ME. Mixing patterns in networks. Phys Rev E 2003 Feb 27;67(2). [doi: 10.1103/physreve.67.026126]
78. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. J Bus Econ Stat 2002 Oct;20(4):518-529. [doi: 10.1198/073500102288618658]
79. Sparck JK. A statistical interpretation of term specificity and its application in retrieval. J Doc 1972 Jan;28(1):11-21. [doi: 10.1108/eb026526]
80. Farmer AS, Kashdan TB. Social anxiety and emotion regulation in daily life: spillover effects on positive and negative social events. Cogn Behav Ther 2012;41(2):152-162 [FREE Full text] [doi: 10.1080/16506073.2012.666561] [Medline: 22428662]
81. Amichai-Hamburger Y, Gazit T, Bar-Ilan J, Perez O, Aharony N, Bronstein J, et al. Psychological factors behind the lack of participation in online discussions. Comput Human Behav 2016 Feb;55:268-277. [doi: 10.1016/j.chb.2015.09.009]
82. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]
83. Lapidot-Lefler N, Barak A. The benign online disinhibition effect: could situational factors induce self-disclosure and prosocial behaviors? Cyberpsychology 2015 Aug;9(2):3. [doi: 10.5817/CP2015-2-3]
84. Tang J, Li J. Semantic mining of social networks. In: Synthesis Lectures on the Semantic Web: Theory and Technology. Santa Clara, California, United States: Atypon; Apr 30, 2015:1-205.
85. Mulveen R, Hepworth J. An interpretative phenomenological analysis of participation in a pro-anorexia internet site and its relationship with disordered eating. J Health Psychol 2006 Mar;11(2):283-296. [doi: 10.1177/1359105306061187] [Medline: 16464925]
86. Zuckerman M. Sensation seeking. In: The International Encyclopedia of Communication. Hoboken, New Jersey: Wiley; 2008:1-4.
87. Slater M. Alienation, aggression, and sensation seeking as predictors of adolescent use of violent film, computer, and website content. J Commun 2003;53(1):105-121. [doi: 10.1111/j.1460-2466.2003.tb03008.x]
88. Lin SS, Tsai C. Sensation seeking and internet dependence of Taiwanese high school adolescents. Comput Human Behav 2002 Jul;18(4):411-426. [doi: 10.1016/S0747-5632(01)00056-5]
89. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Sociol 2001 Aug;27(1):415-444. [doi: 10.1146/annurev.soc.27.1.415]
90. Fredrickson B. Cultivating positive emotions to optimize health and well-being. Prevention & Treatment 2000;3(1). [doi: 10.1037/1522-3736.3.1.31a]
91. Valente T. Network interventions. Science 2012 Jul 06;337(6090):49-53. [doi: 10.1126/science.1217330] [Medline: 22767921]

Abbreviations

API: application programming interface
ED: eating disorder
IV: instrumental variable
OLS: ordinary least squares

Edited by G Eysenbach; submitted 06.05.18; peer-reviewed by CA Almenara, Z Cooper, S Boucher; comments to author 30.08.18; revised version received 22.11.18; accepted 31.12.18; published 03.05.19.

Please cite as:
Wang T, Mentzakis E, Brede M, Ianni A
Estimating Determinants of Attrition in Eating Disorder Communities on Twitter: An Instrumental Variables Approach
J Med Internet Res 2019;21(5):e10942
URL: https://www.jmir.org/2019/5/e10942/
doi:10.2196/10942
PMID:31066718

Original Paper

Assessing Electronic Health Literacy in the State of Kuwait: Survey of Internet Users From an Arab State

Dari Alhuwail[1,2], PhD; Yousef Abdulsalam[3], PhD

[1]Department of Information Science, College of Computing Sciences and Engineering, Kuwait University, Al-Adailiya, Kuwait

[2]Health Informatics Unit, Dasman Diabetes Institute, Kuwait, Kuwait

[3]Department of Quantitative Methods and Information Systems, College of Business Administration, Kuwait University, Shuwaikh, Kuwait

Corresponding Author:
Dari Alhuwail, PhD
Department of Information Science
College of Computing Sciences and Engineering
Kuwait University
PO Box 5969
Al-Adailiya, Safat 13060
Kuwait
Phone: 965 24633274
Email: dari.alhuwail@ku.edu.kw

Abstract

Background: The internet and social media have become an important source for health information. In 2017, the State of Kuwait ranked first in mobile subscription penetration in the Arab world; nearly 90% of its population uses the internet. Electronic health (eHealth) literacy is important in populations that have easy and affordable access to internet resources to more effectively manage health conditions as well as improve general population health.

Objective: The aim of this study was to assess eHealth literacy levels across internet users in Kuwait and identify demographic characteristics that influence eHealth literacy. Furthermore, the study aimed to identify the reasons and type of information that people seek online. Finally, this study examined the utilization of various social media channels for accessing online health information. The social media platforms considered were as follows: WhatsApp, Twitter, Instagram, YouTube, Facebook, and Snapchat.

Methods: A cross-sectional anonymous Web-based survey was used to collect data about eHealth literacy and related information. The eHealth literacy scale (eHEALS), originally developed by Norman and Skinner, is measured using 8 Likert-type scales. A linear regression model estimates the effect of demographic variables such as age, gender, and education on eHealth literacy while controlling for participants' perceived usefulness and importance of the internet. Participants were also surveyed about their frequency in using social media platforms for seeking health information.

Results: Kuwait's composite eHEALS, based on a sample of 386 participants, was 28.63, which is very similar to eHEALS observed among adult populations in other developed countries. Females in Kuwait demonstrated a higher average eHEALS compared with males. Among the social media platforms, the survey results indicated that YouTube is the most frequently used to seek health information, with Facebook being the least frequently used.

Conclusions: Internet users in Kuwait appear confident in their ability to search for health-related information online compared with other populations, as indicated by aggregate eHEALS scores. Considering this finding, government and health care organizations should shift more efforts from traditional media toward online health information, focusing on the social media outlets that people in Kuwait find more useful for seeking health information.

(J Med Internet Res 2019;21(5):e11174) doi:10.2196/11174

KEYWORDS

eHEALS; literacy; health information; information-seeking; informatics; Arab; Kuwait

Introduction

Nearly 4.1 billion individuals have access to the internet worldwide [1]. Among its many purposes, the internet has become a popular destination for individuals seeking information about health, diet, and lifestyle [2-4]. Accordingly, there has been an uptick in the propagation of electronic health (eHealth) apps that support health care delivery [5], and the topic has received increased attention from health care institutions and health informatics researchers [6].

Online Health Information as a Health Care Resource for Patients

Internet searches for health-related information online have been increasing exponentially, catalyzed by affordable access to the World Wide Web and the proliferation of mobile phones, including smartphones [7]. However, despite this large supply of health-related information online, many individuals are not able to use this information to make informed health decisions [8]. Seeking health information online requires eHealth literacy, which is defined as the ability to read, use computers, search for information, understand health information, and put it into context [9]. A prerequisite for one’s health management is an adequate level of knowledge about how to manage his or her diseases, conditions, and lifestyle[10-12]. Therefore, eHealth literacy becomes important for patients to be in charge of their health. Norman and Skinner [13] developed the eHealth Literacy Scale (eHEALS) to measure an individual’s skills in finding, evaluating, and applying online health information. This scale has since been widely adopted and tested in numerous contexts [14-23].

Combating Noncommunicable Diseases With Information

Noncommunicable diseases (NCDs), such as diabetes and hypertension, have a great toll on health care systems [24,25]. Patients with NCDs must manage their conditions and adapt their lifestyles to gradually reduce the need for constant supervision of a health care professional [26]. Online health information is especially valuable for patients living with NCDs, who must eventually transition to self-management of their condition by seeking health information independently [27]. Providing patient-centered information resources to people with NCDs can save lives, improve quality of life, and ultimately reduce the burden on the health care system [28,29]. Therefore, it is essential that patients are able to independently identify, locate, process, and use the necessary information that can help them manage these diseases [14,30].

Examining Electronic Health Literacy in an Internet-Savvy and Disease-Prone Population

The accessibility and widespread use of the internet and social media in the Middle East region can be a cost-effective mechanism for delivering health information to masses and creating support communities [31]. In this study, we consider the case of the State of Kuwait and the prevalence of eHealth literacy in Kuwait. This context is interesting because some of the population’s characteristics create great potential for eHealth literacy to augment traditional health care delivery channels. Kuwait is an oil-rich country with a high per capita gross domestic product and an affluent population that ranked first in mobile subscription penetration in the Arab world with nearly 90% of its population having access to the internet [32].

The World Health Organization reports that 72% of all deaths in Kuwait are attributed to NCDs [33]. The population in Kuwait struggles with numerous health issues such as high rates of obesity and other NCDs across most demographics such as diabetes (14.6% of the population), hypertension (15.7% of the population), and osteoarthritis (16%) [34]. Recent research reports that almost 40% of Kuwaiti citizens had prediabetes or diabetes [35].

The combination of internet, affluence, and high rate of NCDs provides a lucrative opportunity to empower patients and reduce health care system costs through the dissemination of online health information. For example, through the use of social media, patients, as well as their caregivers, can share personal information relevant to the disease, ask for help, and seek disease-specific emotional support [36,37].

This study targets the internet users in Kuwait and aims to (1) assess eHealth literacy rates, (2) gauge perceptions about the utility and importance of the internet as a source of health information, (3) determine demographic characteristics (such as age, internet usage, gender, and education) that influence eHealth literacy rates, (4) identify reasons for seeking health information online, and (5) examine the frequency of using social media channels to seek health information.

Methods

Study Design

This cross-sectional study [38] was initiated after obtaining the necessary ethical approvals from the Research Ethics Committee at Kuwait University, and it was conducted in full accordance with the World Medical Association Declaration of Helsinki. A convenience sampling approach was followed to recruit participants from the general population in Kuwait between September and October 2017 based on 3 inclusion criteria: respondents were aged 18 years or older, resided in Kuwait, and agreed to voluntarily participate in this anonymous study.

Survey Instrument

The Web-based, self-administered, voluntary, and anonymous questionnaire consisted of the following 5 parts:

First, it involved collecting demographic information including age, gender, education, and internet use.

Second, respondents were asked about the reasons for seeking online health information and the types of health information sought online. Under each of these categories, participants provided binary, yes or no, answers to choices related to that category.

Third, the 8 items of the eHEALS were included [13]. Respondents rated their level of agreement with each of the 8 items on a scale of 1 to 5, with 1 being the lowest and 5 being the highest. Therefore, the theoretical range of the composite score is between 8 and 40. Furthermore, 2 supplemental eHEALS questions measuring perceived usefulness and importance of accessing health information online, used in similar studies [39], were also included. These questions were also measured on a scale of 1 to 5.

Fourth, following previous studies, 2 additional questions related to the reasons for and types of information sought online were included [15,40].

Finally, participants rated, on a scale of 1 to 5 (1=*Never*, 5=*Always*), their frequency in using social media platforms as a venue for seeking health information. Participants rated the following 6 social media platforms: WhatsApp, Twitter, YouTube, Snapchat, Instagram, and Facebook.

Bilingual faculty members in the Information Science and the Health Informatics disciplines translated the survey instrument from English to Arabic. Initially, the Information Science faculty member translated the survey from English to Arabic and then a Health Informatics faculty member checked this translation. A graduate student back-translated the survey from Arabic to English. The back-translated survey was checked for accuracy and clarity and approved by the researchers. Moreover, 10 individuals, who were colleagues of 1 of the researchers, were contacted to pilot the survey with the researcher face to face. In addition, 5 individuals tested the English version of the survey, and the other 5 individuals tested the translated Arabic version. Feedback from the pilot survey provided valuable insights to clarify concepts, questions, design, and the structure of the final survey instrument.

Data Collection

The survey was published online using LimeSurvey hosted on the researchers’ server. The online survey was made available in both the Arabic and English languages.

A broad combination of strategies similar to those used by other studies was used to identify the convenience sample [41]. In addition to disseminating the survey to the researchers’ contacts and to capitalize on the heavy traffic around universities and shopping centers, open invitations to participate were circulated via printed posters at these locations. The posters were strategically displayed at the front entrance of 8 shopping centers distributed across Kuwait’s 6 governorates. The posters were also displayed on student boards across 11 colleges in Kuwait’s state university. The recruitment posters at universities allowed inviting a mixed group of students, faculty, and staff who were likely to use the internet given the nature of higher education. The recruitment posters at shopping centers allowed to expand the sampling pool to include a diverse set of the general population in Kuwait who regularly visit the cooperative store but may not currently attend a university or have graduated from a university.

The researchers also circulated electronic invitations through various mediums including emails, short message service (SMS) text messages, and social media. Initially, the researchers sent email invitations to their network of family, friends, and colleagues, inviting them to participate and help pass on the invitation to others. SMS text messages were sent via WhatsApp to an expanded network of the researchers’ contacts, especially as some potential participants may not use or check emails. To recruit participants via social media, the researchers mainly used Twitter to announce the study. The researchers approached major Twitter accounts followed by people in Kuwait to help voluntarily tweet or retweet about the study and how to participate.

Data Analysis

The survey data were analyzed using the R software (version 3.5) developed by R Core Team [42]. The eHEALS score was analyzed, and its distribution was examined. Ordinary least squares regression was utilized to assess how demographics and perceptions can influence the eHEALS. Factors of interest were collected via the survey to examine their correlation with eHEALS. Factors included in the analysis were age, gender, education, and internet usage. We controlled for the perceived usefulness and perceived importance of the internet because these 2 measures were significant correlates to the eHEALS.

P values for all statistical tests were reported, and we considered an alpha of .05 when reporting that a test statistic is significant. In other words, we consider statistical tests to be significant when they show a *P* value of less than .05.

Results

Respondents' Characteristics

Participant demographic information is shown in Table 1. In total, 615 people attempted the survey, and 386 fully completed it (completion rate of 62.7%, 386/615). Of the respondents, 63% (243/386) were females, and 37% (143/386) were males. Furthermore, 63% of participants (244/386) accessed the internet for 3 hours or more per day. Compared with the general population in Kuwait, more females participated in the survey than males. In terms of age, our sample was more skewed toward older participants compared with the general population in Kuwait.

Electronic Health Literacy

The eHealth literacy score, which was calculated as the composite of the 8 items, had a mean of 28.63, median of 29, and SD of 5.6. The dispersion of the observations appears normally distributed with a range between 8 and 40.

Instrument Dimensionality

We ran a confirmatory factor analysis for a single factor model where all 8 items load onto 1-factor and a 3-factor model. The 3-factor model is based on recent studies that parse the eHEALS items into subgroups [43,44]. The fit statistics of the 2 models are presented in Table 2. We found that the 3-factor model demonstrated a better fit based on the global fit indices. However, the correlations between the 3 factors were statistically significant, indicating that they are subscales to an overarching unidimensional structure.

Table 1. Respondents' demographics (n=386).

Demographic	Statistics, n (%)	Kuwait population, %[a]
Gender		
Male	143 (37.0)	61.2
Female	243 (63.0)	38.7
Age (years)		
0-19	22[b] (5.7)	26.7
20-29	113 (29.3)	10.7
30-39	121 (31.3)	22.8
40-49	56 (14.5)	22.4
50-59	57 (14.8)	11.4
60-69	16 (4.1)	4.4
70-79	1 (0.3)	1.2
Education level		
Primary school or lower	8 (2.1)	—[c]
High school	69 (17.9)	—
Diploma	43 (11.1)	—
Bachelor’s degree	185 (47.9)	—
Master's degree	48 (12.4)	—
Doctorate degree	33 (8.5)	—
Occupation		
Student	82 (21.2)	—
Employed	232 (60.1)	—
Unemployed	26 (6.7)	—
Retired	46 (11.9)	—
Internet use		
Less than 1 hour per day	32 (8.3)	—
1-3 hours per day	110 (28.5)	—
3-5 hours per day	99 (25.6)	—
5+ hours per day	145 (37.6)	—

[a] Kuwait population statistics are from the Kuwait Central Bureau of Statistics estimate for January 1, 2018.

[b] Respondents in this group were either 18 or 19 years old.

[c] Data not available.

Table 2. Confirmatory factor analysis.

Goodness of fit statistics	1-Factor model	3-Factor model
Comparative fit index	.899	.931
Tucker-Lewis index	.858	.886
Root mean square error of approximation	.142	.128
Standardized root mean square residual	.053	.045

In response to the item *How useful do you feel the internet (including social media) is in helping you in making decisions about your health?*, 81% (312/386) of the participants considered the internet useful or very useful. In response to *How important is it for you to be able to access health resources on the internet (including social media)?*, 77% (296/386) of the participants expressed that it was important (or very important) to be able to access health resources on the internet.

The 2 items are correlated with one another with a correlation coefficient of .52. In addition, both perceived usefulness and perceived importance of the internet correlate highly with the eHealth literacy score, at .426 and .431, respectively.

Factors Related to Electronic Health Literacy

Demographic characteristics were considered when investigating the study's third aim of determining any factors that contribute to eHealth literacy rates. The results of the regression model are presented in Table 3. The model was statistically significant, explaining 28% of the variance in eHEALS observations ($F_{11,374}$=12.29, $P<.001$).

Age (as measured by year of birth) was not a significant predictor of eHEALS after controlling for other factors. With regard to *gender*, the model predicts that males have a lower eHEALS score than females (β=−.23, $P<.05$) after controlling for other factors.

In terms of *education*, individuals with a Doctorate degree are expected to have a higher eHEALS score compared with bachelor's degree holders by an average of 3.5 points, controlling for all other factors. Finally, individuals who used the internet more than 5 hours a day on average are expected to have a higher eHEALS compared with individuals who used the internet an average of 1 to 3 hours daily.

Types and Reasons for Seeking Health Information Online

In addition to measuring the eHEALS score, the study aimed to identify the type of health information sought via the internet and social media, as well as the reasons that the participants sought them. The top health information sought online by participants included information about a *disease or medical problem* and *medical treatment or procedure*. The least common types were *online support group* and *search for a particular physician or hospital* (refer to Table 4).

In terms of reasons for seeking health information online, most participants cited reasons related to having a general curiosity about a topic (ie, *to be more informed* and *just out of interest*). On the other hand, the least common reasons for seeking health information online were *limited time with health professional* and *disagree with health professional's opinion* (refer to Table 5).

Table 3. Regression model predicting electronic health literacy (n=386).

Variable[a]	B (SE)	β	*t* (df)	*P* value
Intercept	25.93 (46.97)	−.17	−1.72 (374)	.087
Usefulness	2.06 (0.39)	.28	5.34 (374)	<.001
Importance	1.87 (0.36)	.28	5.18 (374)	<.001
Gender (Male=1, Female=0)	−1.29 (0.53)	−.23	−2.42 (374)	.016
Age	−0.01 (0.02)	−.02	−0.30 (374)	.766
Education[b]				
Primary school	0.41 (1.78)	.07	0.23 (374)	.816
High school	0.73 (0.75)	.13	0.98 (374)	.329
Diploma	0.40 (0.84)	.07	0.47 (374)	.639
Master's degree	1.24 (0.80)	.22	1.55 (374)	.122
Doctorate degree	3.52 (0.97)	.62	3.64 (374)	<.001
Internet usage[c]				
Less than 1 hour	0.25 (1.01)	.04	0.25 (374)	.801
3-5 hours	0.87 (0.70)	.15	1.24 (374)	.217
More than 5 hours	1.42 (0.67)	.25	2.13 (374)	.034

[a] R^2=0.28. *F*=12.29 on 12 and 374 degrees of freedom ($P<.001$).

[b] Education variables are binary variables that compare with bachelor's degree holders.

[c] Internet usage variables are binary variables that compare with the 1- to 3-hour usage group.

Table 4. Types of health information sought online (n=386).

Types of health information sought online[a]	n (%)[b]
A disease or medical problem	284 (73.6)
Medical treatment or procedure	237 (61.4)
Diet, nutrition, and vitamins	219 (52.3)
Medication	202 (56.7)
Sports and exercise	196 (50.8)
A particular physician or hospital	137 (35.5)
Online support groups	35 (9.1)

[a] For each item, participants were requested to answer *yes* or *no*.

[b] Percentages represent the proportion of participants (out of the 386) who answered *yes*.

Table 5. Reasons for seeking health information online (n=386).

Reasons for seeking health information online[a]	n (%)[b]
To be more informed	224 (58.0)
Just out of interest	203 (47.2)
Help manage my own condition	182 (36.0)
Look for alternative or additional treatment options	146 (52.6)
Clarify information that has been given to me by a health professional	139 (31.6)
Check information discussed during a consultation with a health professional	122 (37.8)
Have information to read	107 (27.7)
Insufficient information from a health professional	79 (14.8)
Limited time with a health professional	57 (20.5)
Disagree with a health professional's opinion	52 (13.5)

[a] For each item, participants were requested to answer *yes* or *no*.

[b] Percentages represent the proportion of participants (out of the 386) who answered *yes*.

Social Media Channels for Health Information

Participants indicated their frequency of using various social media platforms for seeking health information online. Table 6 provides the means and SDs of participant responses, as well as the prevalence of these platforms among the general population in Kuwait. All these platforms demonstrate a relatively high level of market penetration in Kuwait led by WhatsApp and Facebook with 84% and 75% penetration, respectively [45].

Participants generally conveyed that they *Never* or *Rarely* use these social media platforms for seeking health information. The exception was YouTube, where about 50% of the participants indicated that they use the platform *Always* or *Sometimes* compared with 28% who indicated *Never* or *Rarely*. Facebook was the least desirable social media platform for seeking health information, with 71% of the respondents indicating that they never use this platform for seeking health information. This was followed by Snapchat with 45% unfavorable opinions (*Never* or *Rarely*). Twitter was also not deemed a reliable source of health information. WhatsApp and Instagram showed a relatively uniform number of responses across the 1 to 5 scale.

Table 6. Use frequency of social media platforms for seeking health information (n=377).

Social media platform	Mean (SD)[a]	Platform penetration in Kuwait, 2015, %[b]
YouTube	3.34 (1.31)	43
Instagram	2.96 (1.42)	43
WhatsApp	2.92 (1.53)	84
Twitter	2.72 (1.53)	41
Snapchat	2.34 (1.49)	—[c]
Facebook	1.60 (1.11)	75

[a] The mean is based on a 1 to 5 scale (1=Never, 5=Always).

[b] Source: Arab Social Media Report [45]

[c] No data.

Discussion

Principal Findings

To date, a limited number of studies have investigated eHealth literacy in the Middle East [17]. To our knowledge, no previous research has studied eHealth literacy rates in Kuwait, as well as the types of health information people seek and reasons why people seek health information online. The importance of eHealth literacy will continue to grow as more people gain access to the World Wide Web and as patients increasingly expect to be active consumers of health care services [16].

When the eHEALS scale was first developed in 2006, it was initially validated as a unidimensional scale [13]. This scale has since been measured and revalidated across many populations and demographics. More recent studies have shown that the 8 items reflect a 3-factor scale [43,44]. The 3 factor–scale measures *awareness* (2 items), *skill* (3 items), and *evaluation* (3 items). Nonetheless, researchers who have observed the multidimensional scale note a significant correlation among all 3 factors, which is indicative of an overarching unidimensional structure [44].

Despite some evidence that several subscales preside within eHEALS, we follow Hyde's [43] recommendation to analyze the eHEALS as a unidimensional factor. Therefore, our analysis considers the full eHEALS, which also allows us to compare our results with that of other studies. The reported eHEALS score in this study is similar to the eHEALS scores observed in other developed countries (see Table 7). The majority of the participants indicated that the internet is a valuable source for health-related information.

Furthermore, 2 of the most important factors in predicting a participant's eHealth literacy were the perceived *importance of accessing health information through the internet* and the *perceived usefulness of the internet* in guiding health-related decisions. This finding is consistent with previous studies and commonly associated with eHEALS [47].

The results revealed that females, on average, demonstrate higher eHEALS than males. After controlling for factors such as education, internet usage, and age, gender's effect on eHEALS was statistically significant in our regression analysis (Table 3). In terms of using health services, previous studies observe that females visited physicians more often than males, consumed more diagnostic services, and had more hospitalizations even after excluding pregnancy-related visits [49,50]. When comparing differences in internet use, research has shown that the main drivers for internet usage among females were interpersonal communication and educational use, whereas entertainment and leisure were main drivers for males [51]. An educational motive combined with higher consumption of health services provides some logic as to why females may demonstrate higher eHealth literacy over males.

Our results indicate limited perceived value from accessing health information through social media, as demonstrated by low reliance on social media platforms (with YouTube being a slight exception). This finding corroborates previous studies, which find that *evidence related to the efficacy and effectiveness of social media is currently limited* [52]. Although the use of social media may have a promising future, there is a present need to engage and educate consumers about accessing health information via social media [53,54].

Many of the participants in this study indicated that they searched YouTube the most for health information but did not place high value on other social media platforms such as Twitter, Snapchat, and Facebook. In a classroom setting, online videos *[are] by far the most common type of social media used…with 80% of faculty reporting some form of class use of online video* [55]. This in-classroom finding may be extended to a general context as well, making online videos an effective medium for consuming information. Alternatively, it may be the ease of searchability of YouTube, which is a content-centric platform as compared with the other user-centric platforms.

Table 7. Similar studies examining Electronic Health Literacy Scale around the world.

Study (N)	Country	Year[a]	Electronic Health Literacy Scale	SD	Group
Britt et al [46] (422)	United States	2017	31.92[b]	5.68	College students
Sudbury-Riley et al [16] (313)	United States	2017	30.48[b]	6.40	Baby boomers
Sudbury-Riley et al [16] (407)	United Kingdom	2017	29.28[b]	6.32	Baby boomers
Sudbury-Riley et al [16] (276)	New Zealand	2017	28.72[b]	6.72	Baby boomers
Richtering et al [47] (453)	Australia	2017	27.2	4.91	Moderate-to-high cardiovascular risk
Giudice et al [48] (868)	Italy	2016	28.20	6.20	Health-literate group versus general public
Chung et al [39] (500)	Korea	2016	28.08[b]	6.43	Young adults
Tubaishat et al [17] (541)	Jordan	2016	28.96[b]	4.64	Undergraduate nursing students
Tennant et al [19] (283)	United States	2015	29.05	5.75	Baby boomers
Lee et al [15] (400)	Australia	2015	29.50	4.30	Patients with chronic health conditions
Suri et al [20] (1,062)	Singapore	2015	23.44[b]	—[c]	College students
James et al [41] (881)	Florida, United States	2014	30.40	7.80	African American adults
Choi et al [21]; Study 1 (73)	Texas, United States	2013	28.24	6.08	Low-income adults, under 60 years
Choi et al [21]; Study 2 (218)	Texas, United States	2013	25.76[b]	6.80	Low-income adults over 60 years
Mitsutake et al [18] (2,115)	Japan	2012	23.40	6.40	General population
Ghaddar et al [22] (261)	Texas, United States	2012	30.60	5.90	High school students
Van der Vaart et al [9] Study 1 (189)	Netherlands	2011	28.20	5.90	Patients with rheumatic diseases
Van der Vaart et al [9] Study 2 (88)	Netherlands	2011	27.60	5.90	Stratified sample of the Dutch population
Mitsutake et al [23] (2,970)	Japan	2009	23.50	6.50	Japanese internet users

[a] This refers to the year the study was conducted and not necessarily the year it was published.

[b] The indicated studies reported eHEALS as an average of the 8 items. To allow for a direct comparison with the other studies and our own (sum of the 8 items), we converted the average scale into a sum scale by multiplying the mean and SD by 8.

[c] Not applicable.

Facebook appears to have the least credibility as a source of health information even though it is the most widely recognized platform in Kuwait [45] and the most widely used platform for news [56]. Furthermore, an explanation for this may be a general lack of reliability of information that spreads on that platform. In fact, the spreading of fake news and rumors on Facebook has become so widespread that the company has attempted to take direct action [57]. Rumors and fake information that circulate via Twitter have also been criticized and researched in recent years [58,59]. This drives us to hypothesize that it is perhaps the perceived quality of information on a platform that drives preference for health information sources or perhaps it is the type of content; quality online videos generally take more effort to produce than news articles or blog posts and are, therefore, less likely to be authored by rumor peddlers. More research is needed to confirm or refute these observations.

In this study, most participants believed that online health information helped them make decisions about their health. This presents an opportunity for health care organizations, professionals, and government agencies providing health care services to play a more active role in monitoring, evaluating, and curating health information online [60]. It can be useful to establish policies and guidelines that ensure the credibility and quality of information similar to the HealthOnNet or DISCERN certification efforts of online health-related resources [61-63].

Comparison With Other Electronic Health Literacy Scale Studies

In this study, we report a mean for eHEALS of 28.63 and an SD of 5.69. These findings are similar to other studies that evaluate eHEALS (refer to Table 7). Although our results are comparable with other countries, it is notable that studies conducted in some Asian countries such as Japan and Singapore report lower eHEALS scores. We acknowledge that it can be difficult to compare eHEALS because of the heterogeneity of the populations being studied. We hypothesize that the geographic location, cultural and language barriers could affect eHEALS [64] because of the lesser availability of health-related information in languages other than English. This may not be the case for the State of Kuwait as English is a secondary official language and is taught at all levels of its educational system.

Study Limitations

Several limitations of this study need to be considered. Given the lack of previous research about eHEALS in the region, we started with the survey approach. Although the eHEALS scale is widely used to assess eHealth literacy, we recognize this as a potential limitation as the scale only captures participants' perceptions and not their actual performance [9]. In the future, more experimental studies are required to measure actual eHealth literacy performance and develop a more reliable self-assessment instrument. In addition, the results may have been influenced by the voluntary bias of the participants who were already interested in using the internet to search for information online. It may also be informative in future efforts to ask how often participants sought the health information for themselves versus for others such as a child or an elderly person.

Given our convenience sampling approach, we are not able to accurately report the survey response rates. However, this approach was useful because it provided us with a sizable sample in a reasonable amount of time. Another limitation of the study is that our sample may not accurately represent the population of Kuwait. Although our sampling approach allows for the collection of a large sample, it introduces the potential for sample selection bias. As shown in Table 1, the sample's age distribution does not accurately reflect Kuwait's population. Furthermore, university graduates are overrepresented in the sample. The 20- to 29-year old demographic and those with a university degree are also overrepresented in our sample, likely because 1 of the survey's distribution channels was university colleges. Given that younger people are generally more tech-savvy and computer-literate, this sample bias may have inflated the eHEALS results. Therefore, we caution readers about the generalizability of the study and call for future research in the region to validate or update our study's results.

Conclusions

This study reports on eHealth literacy rates among internet users in Kuwait, as well as their perceptions about the utility and importance of the internet as a source of health information. The findings reveal high eHEALS score among the participants, suggesting that many internet users in Kuwait are confident in their ability to search for health-related information online. This high confidence and the high frequency of accessing the internet reported by the participants presents a myriad of opportunities to better engage patients digitally and conveniently. Our exploration of social media platforms as outlets for eHealth may provide guidance about how to best reach the intended audiences and stimulate further research. YouTube appears to be the most effective platform for delivering health information. Health care organizations, professionals, and government agencies providing health care services need to play a more active role in monitoring, evaluating, and curating online health information. There is a need to establish policies and guidelines that ensure the credibility and quality of information.

Acknowledgments

The authors wish to acknowledge Dr Sana BuHamra for her support in reviewing the manuscript and verifying the results.

Conflicts of Interest

None declared.

References

1. Internet World Stats. 2018. Internet Usage Statistics URL: https://www.internetworldstats.com/stats.htm [accessed 2019-05-10] [WebCite Cache ID 78GDn1Xj0]
2. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
3. Reid P, Borycki EM. Emergence of a new consumer health informatics framework: introducing the healthcare organization. Stud Health Technol Inform 2011;164:353-357. [Medline: 21335736]
4. Lee C, Gray SW, Lewis N. Internet use leads cancer patients to be active health care consumers. Patient Educ Couns 2010 Dec;81(Suppl):S63-S69 [FREE Full text] [doi: 10.1016/j.pec.2010.09.004] [Medline: 20889279]
5. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
6. Alqudsi-ghabra T, Al-Bannai T, Al-Bahrani M. The Internet in the Arab Gulf Cooperation Council (AGCC): Vehicle of Change. Int J Internet Sci 2011;6(1):44-67 [FREE Full text]
7. Ranallo PA, Kilbourne AM, Whatley AS, Pincus HA. Behavioral health information technology: from chaos to clarity. Health Aff (Millwood) 2016 Jun 1;35(6):1106-1113. [doi: 10.1377/hlthaff.2016.0013] [Medline: 27269029]
8. 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]
9. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 9;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
10. Bo A, Friis K, Osborne RH, Maindal HT. National indicators of health literacy: ability to understand health information and to engage actively with healthcare providers - a population-based survey among Danish adults. BMC Public Health 2014 Oct 22;14:1095 [FREE Full text] [doi: 10.1186/1471-2458-14-1095] [Medline: 25339154]

11. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
12. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
13. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
14. Schulz PJ, Nakamoto K. Health literacy and patient empowerment in health communication: the importance of separating conjoined twins. Patient Educ Couns 2013 Jan;90(1):4-11. [doi: 10.1016/j.pec.2012.09.006] [Medline: 23063359]
15. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer Use of. J Med Internet Res 2015 Dec 29;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
16. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the measurement properties of the eHealth literacy scale (eHEALS) among baby boomers: a multinational test of measurement invariance. J Med Internet Res 2017 Dec 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
17. Tubaishat A, Habiballah L. eHealth literacy among undergraduate nursing students. Nurse Educ Today 2016 Jul;42:47-52. [doi: 10.1016/j.nedt.2016.04.003] [Medline: 27237352]
18. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Dec 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
19. 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]
20. Suri VR, Majid S, Chang Y, Foo S. Assessing the influence of health literacy on health information behaviors: a multi-domain skills-based approach. Patient Educ Couns 2016 Dec;99(6):1038-1045. [doi: 10.1016/j.pec.2015.12.017] [Medline: 26794667]
21. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May 2;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
22. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health 2012 Jan;82(1):28-36. [doi: 10.1111/j.1746-1561.2011.00664.x] [Medline: 22142172]
23. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012 Nov 13;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
24. Chaker L, Falla A, van der Lee SJ, Muka T, Imo D, Jaspers L, et al. The global impact of non-communicable diseases on macro-economic productivity: a systematic review. Eur J Epidemiol 2015 May;30(5):357-395 [FREE Full text] [doi: 10.1007/s10654-015-0026-5] [Medline: 25837965]
25. Gostin LO. Non-communicable diseases: healthy living needs global governance. Nature 2014 Jul 10;511(7508):147-149. [Medline: 25013878]
26. Xiao Y. The. Chronic Dis Transl Med 2015 Sep;1(3):145-151 [FREE Full text] [doi: 10.1016/j.cdtm.2015.08.001] [Medline: 29063000]
27. Zakaria N, AlFakhry O, Matbuli A, Alzahrani A, Arab NS, Madani A, et al. Development of Saudi e-health literacy scale for chronic diseases in Saudi Arabia: using integrated health literacy dimensions. Int J Qual Health Care 2018 May 1;30(4):321-328. [doi: 10.1093/intqhc/mzy033] [Medline: 29584913]
28. Hussein SH, Almajran A, Albatineh AN. Prevalence of health literacy and its correlates among patients with type II diabetes in Kuwait: a population based study. Diabetes Res Clin Pract 2018 Jul;141:118-125. [doi: 10.1016/j.diabres.2018.04.033] [Medline: 29729374]
29. Ahn S, Basu R, Smith ML, Jiang L, Lorig K, Whitelaw N, et al. The impact of chronic disease self-management programs: healthcare savings through a community-based intervention. BMC Public Health 2013 Dec 6;13:1141 [FREE Full text] [doi: 10.1186/1471-2458-13-1141] [Medline: 24314032]
30. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
31. Alanzi T. Role of social media in diabetes management in the Middle East region: systematic review. J Med Internet Res 2018 Dec 13;20(2):e58 [FREE Full text] [doi: 10.2196/jmir.9190] [Medline: 29439941]
32. Salem F. The Arab World Online 2017: Digital Transformations and Societal Trends in the Age of the 4th Industrial Revolution (Vol. 3). Dubai: MBR School of Government; 2017.
33. World Health Organization. Noncommunicable Diseases (NCD) Country Profiles 2018, Kuwait. Geneva, Switzerland; 2018.
34. World Health Survey in Kuwait: Summary Report 2013. Sulaibikhat, Kuwait: Ministry of Health, Kuwait; 2015.

35. Alkandari A, Longenecker JC, Barengo NC, Alkhatib A, Weiderpass E, Al-Wotayan R, et al. The prevalence of pre-diabetes and diabetes in the Kuwaiti adult population in 2014. Diabetes Res Clin Pract 2018 Oct;144:213-223. [doi: 10.1016/j.diabres.2018.08.016] [Medline: 30179683]
36. Patel R, Chang T, Greysen SR, Chopra V. Social media use in chronic disease: a systematic review and novel taxonomy. Am J Med 2015 Dec;128(12):1335-1350. [doi: 10.1016/j.amjmed.2015.06.015] [Medline: 26159633]
37. AlQarni ZA, Yunus F, Househ MS. Health information sharing on Facebook: an exploratory study on diabetes mellitus. J Infect Public Health 2016;9(6):708-712 [FREE Full text] [doi: 10.1016/j.jiph.2016.08.015] [Medline: 27618634]
38. Levin KA. Study design III: cross-sectional studies. Evid Based Dent 2006;7(1):24-25. [doi: 10.1038/sj.ebd.6400375] [Medline: 16557257]
39. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): reliability and validity testing in younger adults recruited online. J Med Internet Res 2018 Dec 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
40. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin Shabr F, et al. Association of online health information–seeking behavior and self-care activities among type 2 diabetic patients in Saudi Arabia. J Med Internet Res 2015 Aug 12;17(8):e196 [FREE Full text] [doi: 10.2196/jmir.4312] [Medline: 26268425]
41. James DC, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research among African Americans, Florida, 2014–2015. Prev Chronic Dis 2016 Dec 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
42. R Core Team. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2019-05-13] [WebCite Cache ID 78KdHKxdL]
43. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-factor structure of the ehealth literacy scale among magnetic resonance imaging and computed tomography outpatients: a confirmatory factor analysis. JMIR Hum Factors 2018 Feb 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
44. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and validity of the telephone-based eHealth literacy scale among older adults: cross-sectional survey. J Med Internet Res 2017 Dec 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
45. TNS. Arab Social Media Report. Dubai, UAE: Arab Social Media Influencers Summit; 2015:1-64.
46. Britt RK, Collins WB, Wilson K, Linnemeier G, Englebert AM. eHealth literacy and health behaviors affecting modern college students: a pilot study of issues identified by the American College Health Association. J Med Internet Res 2017 Dec 19;19(12):e392 [FREE Full text] [doi: 10.2196/jmir.3100] [Medline: 29258979]
47. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
48. Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, et al. Correlation between eHealth literacy and health literacy using the eHealth literacy scale and real-life experiences in the health sector as a proxy measure of functional health literacy: cross-sectional web-based survey. J Med Internet Res 2018 Oct 31;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
49. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract 2000 Feb;49(2):147-152. [Medline: 10718692]
50. Rosenstock IM. Why people use health services. Milbank Mem Fund Q 1966 Jul;44(3):Suppl:94-Suppl127. [Medline: 5967464]
51. Weiser EB. Gender differences in internet use patterns and internet application preferences: a two-sample comparison. Cyberpsychol Behav 2000 Apr;3(2):167-178. [doi: 10.1089/109493100316012]
52. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J 2014 Mar;20(1):50-58. [doi: 10.1177/1460458213476969] [Medline: 24550564]
53. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
54. 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]
55. Moran M, Seaman J, Tinti-Kane H. ERIC - Education Resources Information Center. 2011. Teaching, Learning, and Sharing: How Today's Higher Education Faculty Use Social Media Internet URL: https://eric.ed.gov/?id=ED535130 [accessed 2019-01-02] [WebCite Cache ID 757hfDfoe]
56. Moon A. Reuters. 2017 Sep 08. Two-thirds of American adults get news from social media: survey Internet URL: https://www.reuters.com/article/us-usa-internet-socialmedia/two-thirds-of-american-adults-get-news-from-social-media-survey-idUSKCN1BJ2A8 [accessed 2019-05-10] [WebCite Cache ID 78GFQyYTi]

57. Wingfield N, Isaac M, Benner K. The New York Times. 2016 Nov 14. Google and Facebook Take Aim at Fake News Sites URL: https://www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html [WebCite Cache ID 78GFgRdCT]
58. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Dec 9;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
59. Oh O, Agrawal M, Rao H. Community intelligence and social media services: a rumor theoretic analysis of tweets during social crises. MIS Q 2013;37(2):407-426 [FREE Full text]
60. Bianco A, Zucco R, Nobile CG, Pileggi C, Pavia M. Parents seeking health-related information on the internet: cross-sectional study. J Med Internet Res 2013 Sep 18;15(9):e204 [FREE Full text] [doi: 10.2196/jmir.2752] [Medline: 24047937]
61. Boyer C, Dolamic L. Automated detection of HONcode website conformity compared to manual detection: an evaluation. J Med Internet Res 2015 Jun 2;17(6):e135 [FREE Full text] [doi: 10.2196/jmir.3831] [Medline: 26036669]
62. 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]
63. Khazaal Y, Chatton A, Zullino D, Khan R. HON label and DISCERN as content quality indicators of health-related websites. Psychiatr Q 2012 Mar;83(1):15-27. [doi: 10.1007/s11126-011-9179-x] [Medline: 21547515]
64. Schyve PM. Language differences as a barrier to quality and safety in health care: the Joint Commission perspective. J Gen Intern Med 2007 Nov;22(Suppl 2):360-361 [FREE Full text] [doi: 10.1007/s11606-007-0365-3] [Medline: 17957426]

Abbreviations

eHEALS: Electronic Health Literacy Scale
eHealth: electronic health
NCDs: noncommunicable diseases
SMS: short message service

Edited by G Eysenbach; submitted 30.05.18; peer-reviewed by M Stellefson, K Schroeder, JR Bautista, C Bakker; comments to author 30.09.18; revised version received 14.01.19; accepted 10.04.19; published 24.05.19.

Please cite as:
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;21(5):e11174
URL: http://www.jmir.org/2019/5/e11174/
doi:10.2196/11174
PMID:31127723

Original Paper

Discovering Clinical Information Models Online to Promote Interoperability of Electronic Health Records: A Feasibility Study of OpenEHR

Lin Yang[1], PhD; Xiaoshuo Huang[1], BEng; Jiao Li[1], PhD

[1] Institute of Medical Information / Medical Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

Corresponding Author:
Jiao Li, PhD
Institute of Medical Information / Medical Library
Chinese Academy of Medical Sciences & Peking Union Medical College
No 3 Yabao Road, Chaoyang District
Beijing, 100020
China
Phone: 86 18618461596
Email: li.jiao@imicams.ac.cn

Abstract

Background: Clinical information models (CIMs) enabling semantic interoperability are crucial for electronic health record (EHR) data use and reuse. Dual model methodology, which distinguishes the CIMs from the technical domain, could help enable the interoperability of EHRs at the knowledge level. How to help clinicians and domain experts discover CIMs from an open repository online to represent EHR data in a standard manner becomes important.

Objective: This study aimed to develop a retrieval method to identify CIMs online to represent EHR data.

Methods: We proposed a graphical retrieval method and validated its feasibility using an online CIM repository: openEHR Clinical Knowledge Manager (CKM). First, we represented CIMs (archetypes) using an extended Bayesian network. Then, an inference process was run in the network to discover relevant archetypes. In the evaluation, we defined three retrieval tasks (medication, laboratory test, and diagnosis) and compared our method with three typical retrieval methods (BM25F, simple Bayesian network, and CKM), using mean average precision (MAP), average precision (AP), and precision at 10 (P@10) as evaluation metrics.

Results: We downloaded all available archetypes from the CKM. Then, the graphical model was applied to represent the archetypes as a four-level clinical resources network. The network consisted of 5513 nodes, including 3982 data element nodes, 504 concept nodes, 504 duplicated concept nodes, and 523 archetype nodes, as well as 9867 edges. The results showed that our method achieved the best MAP (MAP=0.32), and the AP was almost equal across different retrieval tasks (AP=0.35, 0.31, and 0.30, respectively). In the diagnosis retrieval task, our method could successfully identify the models covering “diagnostic reports,” “problem list,” “patients background,” “clinical decision,” etc, as well as models that other retrieval methods could not find, such as “problems and diagnoses.”

Conclusions: The graphical retrieval method we propose is an effective approach to meet the uncertainty of finding CIMs. Our method can help clinicians and domain experts identify CIMs to represent EHR data in a standard manner, enabling EHR data to be exchangeable and interoperable.

(J Med Internet Res 2019;21(5):e13504) doi:10.2196/13504

KEYWORDS

openEHR; clinical information model; health information interoperability; information retrieval; probabilistic graphical model

Introduction

Electronic health record (EHR) data can be used and reused for many purposes, including managing an individual patient’s care, medical and health services research, and management of health care facilities. More recently, EHR data has been defined as a part of real-world data [1] and is increasingly seen as a viable source of data for regulatory decisions [2]. However, bias can occur in different steps of the data chain, which might lead to incomparable or invalid analysis results [3].

Semantic interoperability is essential for accurate and advanced health-related computing, shared EHRs, and coordination of clinical care across clinical systems [4,5]. According to ISO/TS 18308 (a standard published by the International Organization for Standardization defining the set of requirements for EHR architecture), it is the ability for data shared by systems to be understood at the level of fully defined domain concepts [6]. To achieve this, a two-level clinical modeling methodology is proposed to separate clinical knowledge from information models [7]. It distinguishes two models: the reference model (RM), which contains the basic and stable properties of health record information, and the clinical information model (CIM), which formally defines clinical concepts (or domain content models) in a standardized and reusable manner, such as blood pressure [8,9]. In this scenario, CIMs in agreement at an organizational, regional, national, or international level will provide a firm basis for establishing semantic interoperability [9].

This two-level modeling approach is used in the ISO/CEN EN13606 (a standard designed to achieve semantic interoperability in EHR communication) [10] and openEHR (described subsequently) [11], as well as Health Level Seven (HL7) version 3 Clinical Document Architecture (HL7's primary standard for representing structured clinical documentation on patients) and Care Provision messages (information structures used to communicate information between providers of care) [12]. For openEHR and ISO/CEN EN13606, CIMs are defined in the form of archetypes, whereas those of HL7 are in the form of HL7 templates. According to the systematic review done by Moreno-Conde et al [13], archetypes are the preferred type of technical artifacts, and openEHR is most frequently mentioned. Therefore, CIMs in our study specifically refer to openEHR archetypes.

OpenEHR is an open-source EHR standard ensuring universal interoperability among all forms of electronic data [14-21]. It is well known for its two-level design paradigm, consisting of an RM, archetypes, and templates. *Archetypes* are computable clinical content specifications that formalize the patterns and requirements for the representation of health information content [9]. To achieve common, coherent, and clinician-approved archetypes, the openEHR community provides a Web-based controlled authoring environment for a wide range of domain experts, especially clinicians, to participate in the creation of archetypes. All contributions are open access and freely available under a Creative Commons license. Archetypes are general purpose, reusable, and composable; therefore, searching for reusable archetypes from archetype repositories is essential throughout the development process [22,23]. Documents with complete archetype design specifications are the input; lists of existing reusable archetypes, either complete or needing modifications, and new archetypes to be developed from scratch are the output [23]. The crucial problem is how to find the relevant ones from open repositories to help identify reusable archetypes.

The openEHR community provides the Clinical Knowledge Manager (CKM) [24] to be a library of openEHR archetypes. It supports their retrieval based on clinical concepts in different sections of archetypes. When the end user enters a term, the CKM will return the archetype that contains the word in metadata, definition, or ontology section. It could help find reusable archetypes [25]. However, domain experts are mainly concerned about whether the concept name and core data items are covered [17,26,27], and they may be not familiar with openEHR archetypes, especially clinicians. For better results, end users usually need to do a large amount of preparatory work, which may include classifying and rearranging data [27], abstracting clinical concepts from data schemas [17], and identifying archetype-friendly concepts from clinical statements [26]. It is an iterative and time-consuming process.

We aimed to develop a retrieval method to identify archetypes online to represent EHR data and optimize existing retrieval results of the CKM. Archetypes usually have their own hierarchical structures, and semantic relationships occur between different archetypes; therefore, we considered that the graphical representation of this potential knowledge might support the retrieval of CIMs. Previous studies show that graphs could efficiently represent clinical knowledge [28-30], and the Bayesian network, as a probabilistic graphical model, is an effective methodology to meet the uncertainty of information needs. Rotmensch et al [30] used a naive Bayes classifier and a Bayesian network to automatically construct a health knowledge graph from electronic medical records. However, in retrieval tasks, differences between Bayesian network-based information retrieval methods mainly lie in the structure of the network, and this structure depends on dependencies between the variables involved in the problem. The basic Bayesian network consists of two different sets of variables, a set of indexing terms and a set of documents in the collection, and the relationships between them [31]. Related research has been conducted to extend a simple Bayesian network for better results. Some methods focus on the structure of the term subnetwork using a polytree [32,33] or two term layers [34,35] to represent term relationships. Some focus on the structure of the document subnetwork using two document layers [36] to represent document relationships. Compared with the previous studies, we focused on the probabilistic graphical representation of openEHR archetype sets, which depends on relationships between the variables involved in finding relevant archetypes, and how the inference process is carried out, aiming for better retrieval performance.

Methods

Information Need Analysis

To find relevant archetypes from the open repository, we first had to understand which kinds of terms end users tended to enter. As archetype modeling methodology [23] shows, domain experts identify core clinical concepts and related data elements involved in a particular scenario and organize them into mind maps or design tables. These archetype design specifications are the main source of search keywords. We considered that the input of end users was mainly the names of clinical concepts or related data elements.

Ideally, the user enters the clinical concept and the system feeds back the archetype defining the concept, or the user enters data elements related to a concept and the system feeds back the archetype that covers all the data elements. However, it is difficult to distinguish clinical concepts and data elements from the end user's input, unless it forces users to input separately. More importantly, data elements defined by end users may be the concept in an archetype repository, or the defined concept is the data element of an archetype. If we match concepts and data elements separately, users may miss some important relevant archetypes.

Based on these considerations, we tried to translate the problem into identifying potentially relevant clinical concepts from the input. We proposed to reorganize the archetype collection with the dependencies between clinical concepts, data elements, and archetypes and used a probabilistic approach to meet the uncertainty of user information needs.

Graphical Retrieval Method Based on an Extended Bayesian Network

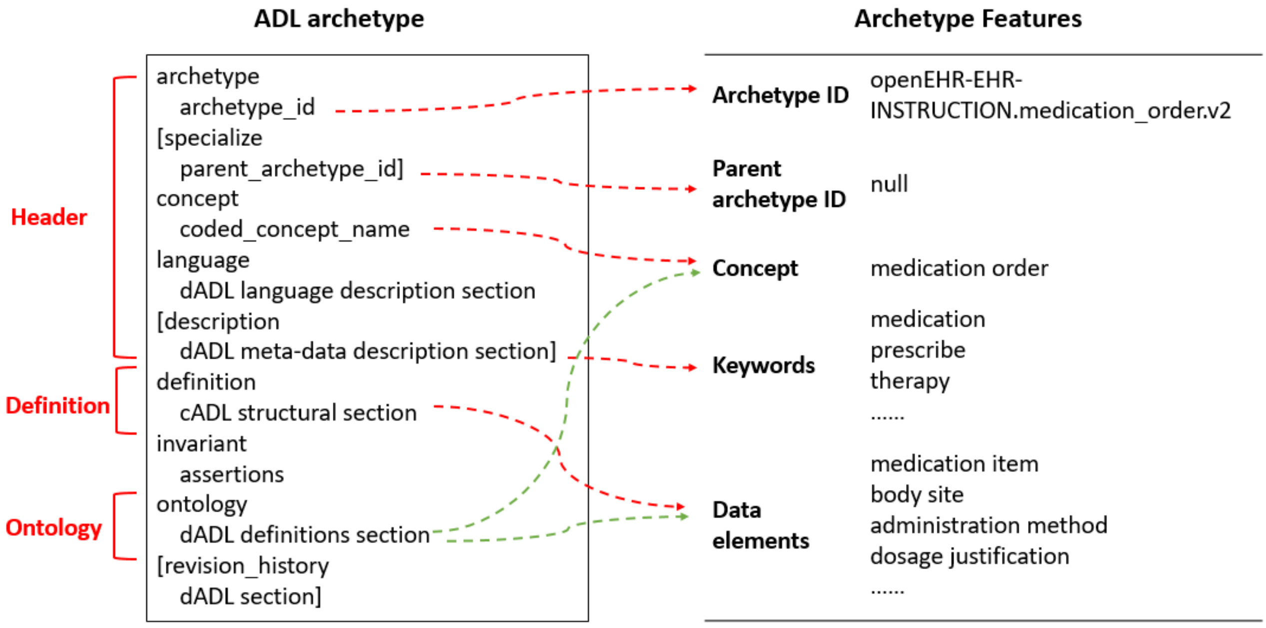
Archetype Feature Identification and Extraction

Based on information need analysis, we attempted to use clinical concepts and data elements to represent each archetype. An archetype is expressed in Archetype Definition Language (ADL) and mainly consists of three sections (Figure 1). The header contains a unique identifier for the archetype and includes some descriptive information, such as concept name and keywords; the definition contains the main formal definition of the archetype, including all possible data elements that could be relevant for the clinical concept; and the ontology contains the code that represents the meaning of nodes. We considered that clinical concepts were the topics of archetypes, whereas keywords and data elements explained the meaning of topics from different perspectives. Thus, we extracted archetype ID, concepts, keywords, and data elements based on ADL files parsing as features (Figure 1).

There are also relationships between archetypes, including specialization and aggregation. An archetype is a specialization of another if it mentions that archetype as its parent and only makes changes to its definition. Aggregation enables any subset of archetypes to be stated as the allowed set for use in a compositional parent archetype. In general, archetypes tend to provide highly reusable models of real-world content with local constraining left to templates, which may result in matching as many archetypes as possible when defining archetype slots. For example, "openEHR-EHR-CLUSTER.device_details.v1" allows the inclusion of 199 archetypes. We thought that such cases might blur the semantic relationship between archetypes. In addition, version control is an integral part of the openEHR architecture. When an archetype updates, the old version could not be found in the archetype library. Therefore, we only added the parent archetype ID as the feature (Figure 1).

Furthermore, there are four main categories of archetypes, including COMPOSITION, SECTION, ENTRY, and CLUSTER, each defined as part of the openEHR RM. A COMPOSITION is a container class, whereas a SECTION is an organizing class, each containing ENTRY objects [16]. The ENTRY class is further specialized into ADMIN_ENTRY, OBSERVATION, EVALUATION, INSTRUCTION, and ACTION subclasses, of which the latter four are kinds of CARE_ENTRY. CLUSTERS are reusable archetypes for use within any ENTRY or other CLUSTER. In addition, the openEHR designs Demographic archetypes for demographic information. Thereby, archetypes could be mainly divided into COMPOSITION, SECTION, ENTRY, CLUSTER, and DEMOGRAPHIC. However, these archetype categories will not obscure the clinical content, and we did not use these as the feature.

Figure 1. An example of archetype feature identification and extraction.



Clinical Resources Network Modeling

We attempted to use a three-level Bayesian network to represent the dependencies among data elements, concepts, and archetypes (Figure 2). The first is the data element layer. It contains the set of indexing data elements T={T_i, i=1...M}, M being the number of data elements from a given archetype collection. Each data element node is linked to its corresponding concept node in the clinical concept layer. The second is the clinical concept layer. It contains the set of indexing concepts C={C_j, j=1...N}, N being the number of concepts. The third layer contains the set of archetypes A={A_k, k=1...K}, K being the total number of archetypes in the collection. If A_k is a specialization of another archetype A_p which defines C_j, there is a link joining any concept node C_j and any archetype node A_k.

However, data elements are unevenly distributed across different types of archetypes, especially for container classes. When two archetypes have few data elements and terms used are totally different, such as "openEHR-EHR-COMPOSITION .medication_list.v0" and "openEHR-EHR-SECTION.medication _order_list.v0," it is difficult to find correlation between them.

Therefore, we tried to include relationships between concepts in the model to extend the similarity between archetypes. Relationships between concepts were measured by estimating conditional probabilities of relevance of every concept given that another concept was considered relevant [36]. Let $e(C_i)$ be an event representing some type of evidence about the relevance of a concept C_i. In openEHR, the evidence could be "keywords," "purpose," "use," or other semantic information. In this case, we considered that $e(C_i)$ as the event [$KW_l = kw_l, \forall KW_l \in C_i$], KW being the keywords used to describe the concept. Given a concept C_j, we calculated the probabilities $p(c_j| e(C_i)) \ \forall C_i \in C$ using equation (a) in Figure 3, where the weight was computed by equation (d) in Figure 3 and M_k was the number of keywords. After decreasing the ordering of $p(c_j|e(C_i))$, the top n concepts $R_n(C_j)$ were the ones that were more related to C_j. Then, we included in the network-explicit dependence relationships between C_j and each concept $C_i \in R_n(C_j)$.

To determine the topology of the Bayesian network, we used a concept subnetwork with two layers instead of the original concept layer. We duplicated each concept node C_j to obtain another concept node C_j, thus forming a new concept layer, and the arcs connecting the two layers went from $C_i \in R_n(C_j)$ to C_j. Thus, this directed acyclic graph had the set of variables $V=T \cup C \cup C \cup A$. The new topology avoids connections between nodes in the same layer and facilitates the inference process.

The overall modeling procedure is summarized in Figure 4. First, we extracted archetype ID, clinical concept, and data elements from the ADL files (detailed in section "archetype feature identification and extraction"). Second, we learned the dependencies between concepts (detailed previously). Third, we graphically represented the dependencies between the variables.

Parameters Estimation in the Clinical Resources Network

In this section, we will discuss how to estimate the probability distributions of each node in the network.

Data Element Nodes

A data element node has no parents; therefore, we had to store the probability of relevance $p(t_i)$ and the probability of being nonrelevant. We used the estimator (Figure 3, equation b), where M is the number of terms used to index the concept collection.

Figure 2. Topology of three-level clinical resources network. A: archetype; C: clinical concept; T: data element.

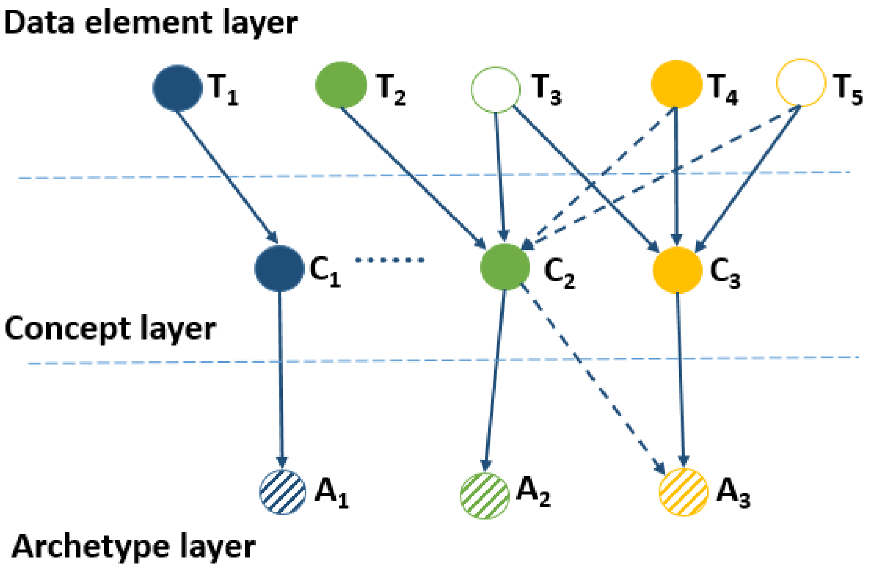
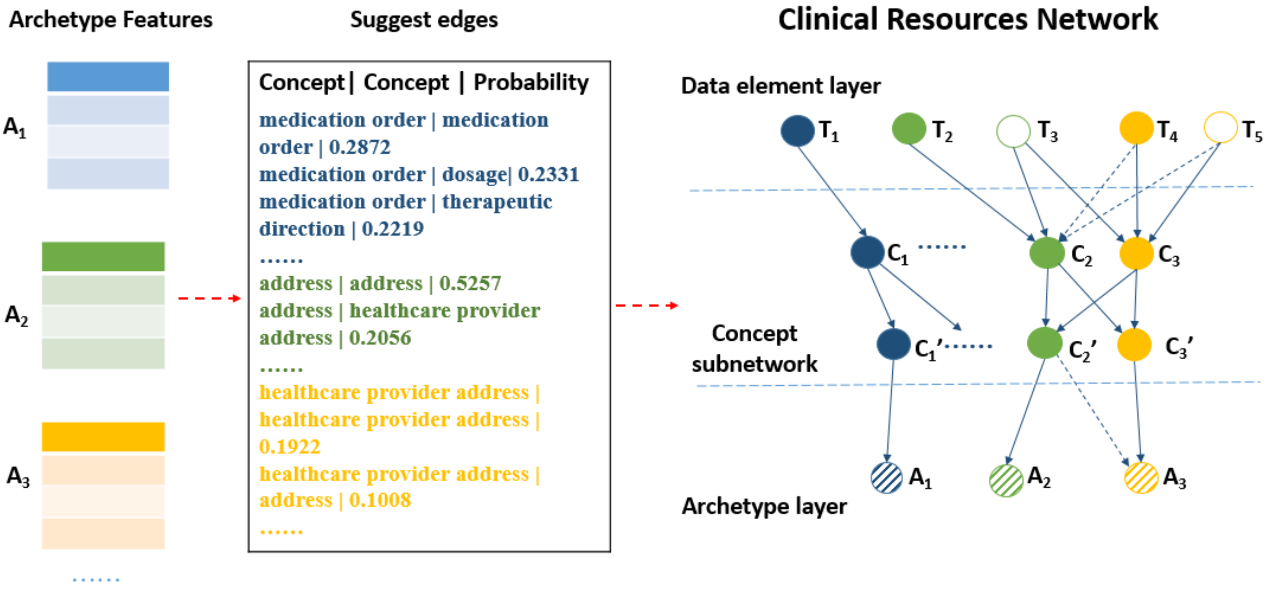


Figure 3. Equations used in our method.

(a)	$\mathrm{p}\left(c_j \middle| e(C_i)\right) = \frac{1}{M_k}\sum_{KW_l \in C_j} w_{lj} + \frac{M_k - 1}{M_k}\sum_{KW_l \in C_j \cap C_i} w_{lj}$
(b)	$\mathrm{p}(t_i) = \frac{1}{M}$ and $\mathrm{p}(\bar{t}_i) = 1 - \mathrm{p}(t_i)$
(c)	$\mathrm{p}(c_j \| \mathrm{pa}(C_j)) = \sum_{T_i \in Pa(C_j), t_i \in \mathrm{pa}(C_j)} w_{ij}$
(d)	$w_{ij} = \alpha^{-1} \frac{tf_{ij} \times idf_i^2}{\sqrt{\sum_{T_k \in Pa(C_j)} tf_{kj} \times idf_k^2}}$
(e)	$\mathrm{p}(c'_j \| \mathrm{pa}(C'_j)) = \frac{1}{S_j}\sum_{C_i \in Pa(C'_j), c_i \in \mathrm{pa}(C'_j)} \mathrm{p}\left(c_j \middle| e(C_i)\right)$
(f)	$\mathrm{p}(a_k \| \mathrm{pa}(A_k)) = \sum_{C'_j \in Pa(A_k), C'_j \in \mathrm{pa}(A_k)} v_{jk}$
(g)	$v_{jk} = \begin{cases} \frac{\alpha}{n_1}, & R(Pa(A_k), A_k) = \text{'nonspecialized'} \\ \frac{\beta}{n_2}, & R(Pa(A_k), A_k) = \text{'specialized'} \end{cases}$
(h)	$\mathrm{p}(t_i \| \mathrm{Q}) = \begin{cases} 1.0, & T_i \in Q \\ \frac{1}{M}, & T_i \notin Q \end{cases}$
(i)	$\mathrm{p}(c_j \| \mathrm{Q}) = \sum_{T_i \in Pa(C_j)} w_{ij} p(t_i \| Q)$
(g)	$\mathrm{p}(c'_j \| \mathrm{Q}) = \frac{1}{S_j}\sum_{C_i \in Pa(C'_j)} \mathrm{p}\left(c_j \middle| e(C_i)\right) p(c_i \| Q)$
(k)	$\mathrm{P}(a_k \| \mathrm{Q}) = \sum_{C'_j \in Pa(A_k)} v_{jk} p(c'_j \| Q)$
(l)	$\mathrm{BM25F}(a_k) = \sum_{t_i \in q \cap a_k} \frac{x_{a_k, t_i}}{k_1 + x_{a_k, t_i}} w_{t_i}$
(m)	$x_{a_k, t_i} = \sum_f \frac{x_{a_k, f, t_i}}{1 + B_f(\frac{l_{a_k, f}}{l_f} - 1)} w_f$

Figure 4. Clinical resources network modeling pipeline. A: archetype; C: clinical concept; C : duplicated clinical concept; T: data element.



Concept	Concept	Probability
medication order	medication order	0.2872
medication order	dosage	0.2331
medication order	therapeutic direction	0.2219
……		
address	address	0.5257
address	healthcare provider address	0.2056
……		
healthcare provider address	healthcare provider address	0.1922
healthcare provider address	address	0.1008
……		

Concept Nodes

For each concept node C_j in the concept subnetwork, we needed to estimate a set of conditional probability distributions $p(c_j|pa(C_j))$. $Pa(C_j)$ represents the parent nodes set of concept C_j, containing all the data elements belonging to concept C_j, and $pa(C_j)$ is a possible configuration of value associated with the parent set $Pa(C_j)$. We used the estimator (Figure 3, equations c and d) proposed by De Campos et al [33], where α is a normalizing constant (assure $\sum_{T_i \in Pa(C_j)} w_{ij} \leq 1 \ \forall \ C_j \in C$), tf_{ij} is the term frequency of data element T_i in concept C_j, and idf_i is the inverse concept frequency of T_i in the whole concept collection; $idf_i = 1 + log\ (N / n_i)$, N being the total number of concepts, and n_i being the total number of concepts containing T_i.

For each concept node C_j, we need to estimate a set of conditional probability distributions $p(c_j|pa(C'_j))$. We used the estimator (Figure 3, equation e) proposed by Acid et al [36], where $S_j = \sum_{C_k \in Pa(C'_j)} p(c_j|e(C_k))$ and the values $p(c_j|e(C_k))$ are obtained when modeling the network.

Archetype Nodes

For each archetype node A_k, we needed to estimate a set of conditional probability distributions $p(a_k|\ pa(A_k))$. $Pa(A_k)$ represents the parent node sets of archetype A_k, containing all the concepts belonging to archetype A_k, and $pa(A_k)$ is a possible configuration of values associated with the parent set $Pa\ (A_k)$. v_{jk} is a constant to represent the weight of a concept for an archetype. The estimator is shown in Figure 3, equations (f) and (g), where $R(Pa(A_k), A_k)$ represents two different relationships between the concept and archetype, n_1 is the number of "nonspecialized" archetypes of one concept, and n_2 is the number of "specialized" archetypes, whereas α and β are coefficients for the weight.

Relevant Archetype Discovering: Inference in the Clinical Resources Network

To find relevant archetypes is to estimate the probability of relevance $p\ (a_k|Q)$ for each archetype, Q being an end user query.

Given a query Q, the set of terms used to formulate the query will be a new piece of evidence. The retrieval process starts by placing the evidence in the data element subnetwork. Then, the inference process is run in the clinical resources network. This allows us to obtain the probability of relevance of each archetype, given that the terms in the query are relevant, $p\ (a_k|Q)$. Finally, the archetypes will be sorted in decreasing order of probability to carry out the evaluation process. The inference process is composed of four stages.

1. Terms in the data element layer are marginally independent; therefore, the probability of relevance $p(t_i|Q)$ is calculated by equation (h) in Figure 3.
2. Based on the propagation process, the conditional probability of concept C_j in the concept subnetwork for the query Q could be calculated by equation (i) in Figure 3.
3. The conditional probability of concept C_j in the concept subnetwork for the query Q could be computed using equation (g) in Figure 3.
4. The conditional probability of archetype A_k for the query Q, $p(a_k|Q)$ could be carried out using information obtained in the previous step by the equation (k) in Figure 3.

Therefore, the propagation with this topology is to evaluate equations (h), (i), (g), and (k) in Figure 3.

Experiment Setup

Test Queries

We defined test queries with the following considerations: first, clinical concepts to be retrieved should be essential components of the EHR; second, there should be needs to reuse these clinical contents [37], such as medical events prediction [38], clinical research [39], and disease research [40]; third, queries should allow us to test the performance of retrieval methods in related archetypes identification, including specialized archetypes and compositional parent archetypes. Based on these criteria, we selected medication, laboratory test, and diagnosis as retrieval tasks and formulated three queries (Table 1).

Data Source

We downloaded all available archetypes from the CKM [24] for a total of 526 on August 30, 2018. All files were in ADL format. We used the ADL parser [41] to extract features. Among these CIMs, three archetypes did not use English as the description language, so the total number changed to 523.

Relevance Assessment

To evaluate retrieval results, we first had to identify relevant archetypes in three retrieval tasks as the gold standard. We manually annotated all 523 archetypes, according to their relevance to each query, to formulate three benchmark datasets. Given a query and an archetype, three annotators were asked to judge if the archetype was relevant. The labeling instructions were as follows: a label was relevant when the archetype could cover the potential clinical concept inferred from the given query; a label was nonrelevant otherwise. We took the majority vote to decide the relevance of an archetype. These three benchmark datasets were used as ground truth for the medication, laboratory test, and diagnosis retrieval tasks.

Baseline Methods

To validate the performance of our method, three typical retrieval methods were selected as baselines: CKM, BM25F, and simple Bayesian network.

Table 1. Test queries.

Query	Retrieval task	Input terms
1	Medication	Medicine name, total daily amount, allowed period, and order start date/time
2	Laboratory test	Report, test name, and test results
3	Diagnosis	Problem/diagnosis, test diagnosis, date/time of onset, and body site

BM25F is an extension of the BM25 ranking function, which is applicable to structured documents consisting of multiple fields. It combines the term frequencies (weighted accordingly to their field importance) and uses the resulting pseudofrequency in the BM25 ranking function. In this study, we supposed that an archetype was decomposed into two fields, concept and data elements, and used the function (Figure 3, equations l and m) proposed by Zaragoza et al [42], where w_{ti} is the RSJ relevance weight for term t_i, $x_{ak, f, ti}$ is the term frequency of term t_i in the field type f of archetype a_k, $l_{ak, f}$ is the length of that field, l_f is the average field length for that field type, and B_f is a field-dependent parameter.

For the Bayesian network, the structure is illustrated in Figure 2. The propagation with this topology is to evaluate equations (h), (i), and (k) in Figure 3.

Results

Overview of Clinical Resources Network

Table 2 shows the distribution of archetypes across different clinical domains.*Clinical domain classification* refers to the concept schema proposed by Hruby et al [39].

Table 3 shows the distribution of archetypes, concepts, and data elements across different types of archetypes in the collection. In addition, there were 31 specialized archetypes, 11 of whose parent archetypes are no longer in the CKM.

Then, we learned the dependencies between concepts. Table 4 shows the top relevant concepts suggested by four different percentages of values of $p(c_j|e(C_i))$ for “dosage” and “examination of a lung,” respectively.

After that, we constructed four clinical resource networks, G_1, G_2, G_3, and G_4, according to the top 3%, 5%, 8%, and 10% of values, respectively. Each graph consisted of 5513 nodes, which were 3982 data element nodes, 504 concept nodes, 504 duplicated concept nodes, and 523 archetype nodes, with 6366 edges from T to C and 543 edges from C to A. For edges C to C , G_1 had 1590 arcs, G_2 had 2485 arcs, G_3 had 2958 arcs, and G_4 had 3263 arcs.

Evaluation of the Performance

To compare the performance of different graphs in supporting retrieval, we calculated the average precision (AP) values for the 11 standard recall points of each graph for the test queries and then computed the mean average precision (MAP) values. The results (Table 5) showed that the retrieval method based on G_3 achieved the best MAP (MAP=0.32), with an AP of 0.35, 0.31, and 0.3 for each query, respectively.

Table 2. Distribution of archetypes across different clinical domains.

Clinical domain and subdomains	Archetypes, n
Patient	
Demographic	42
Health characteristic	32
Patient	6
Pretreatment diagnosis	
Clinical assessment	73
Pretreatment diagnosis	26
Procedure	6
Intent	1
Treatment	
Treatment	39
Prescribed	12
Surgery	9
Detection/Treatment results	184
Organizational/Provider characteristics	26
Outcomes	24
Patient environment factors	6
Other	37
Total	523

Table 3. Distribution of archetypes, concepts, and data elements.

Archetype type subtypes	Archetypes, n	Concepts, n	Elements, n	Data elements per concept, mean
Cluster	198	198	1567	7.9
Composition	25	25	45	1.8
Entry				
Action	15	15	252	16.8
Evaluation	51	51	432	8.5
Observation	164	163	1511	9.3
Instruction	8	8	124	15.5
Admin	4	4	69	17.3
Section	26	26	88	3.4
Demographic	32	29	169	5.8
Total	523	504	3982	7.9

Table 4. Top edge suggestions for “dosage” and “examination of lung.”

Clinical concept	Different threshold of $p(c_j\|e(C_i))$[a]			
	Top 3%	Top 5%	Top 8%	Top 10%
Dosage	Dosage	Dosage	Dosage	Dosage
	Medication order	Medication order	Medication order	Medication order
		Therapeutic direction	Therapeutic direction	Therapeutic direction
			Medication	Medication
			Medication authorization	Medication authorization
Examination of lung	Examination of a lung	Examination of a lung	Examination of a lung	Examination of a lung
	Auscultation of lung	Auscultation of lung	Auscultation of lung	Auscultation of lung
	Pulmonary function test	Pulmonary function test	Pulmonary function test	Pulmonary function test
	Macroscopic findings-lung cancer	Macroscopic findings-lung cancer	Macroscopic findings-lung cancer	Macroscopic findings-lung cancer
				Examination findings-posterior chamber of eye
				Examination of a breast
				Examination of a burn

[a] c_j=”dosage” and “examination of lung,” respectively.

Next, we compared the results of our method based on G_3 with baseline methods. To comprehensively validate the performance, we selected the MAP, AP, and precision at 10 (P@10) as evaluation metrics. Archetypes in the CKM are updated regularly, so it is difficult for us to compare the result on the same collection. We searched relevant archetypes in the CKM for the three queries given on December 12, 2018, and evaluated its performance against the ground truth. The result (Table 6) shows that our method outperforms all the baseline methods, achieving the best AP and P@10 across different test queries, as well as the best MAP. For instance, for query 1, our method, CKM, Bayesian network, and BM25F achieved a P@10 of 0.50, 0.40, 0.20, and 0.20, respectively. Furthermore, we can observe that the MAP of BM25F (MAP=0.177) and Bayesian network (MAP=0.127) was lower than that of CKM (MAP=0.227), which means that there are limitations in using clinical concepts and data elements to represent each archetype. Our approach takes into account the semantic associations between concepts and effectively compensates for this deficiency.

The same trend is observed when evaluating precision-recall graphs across all test queries. Figure 5 shows the precision-recall curves evaluated against the ground truth. Here, BM25F falls short in performance. For instance, for a recall of 0.3, our method, CKM, Bayesian network, and BM25F achieved a precision of 0.38, 0.30, 0.05, and 0, respectively. Additionally, the 11-point MAP curve of the Bayesian network is similar to that of our approach, but the performance is much worse than ours. Meanwhile, compared with the curve of the CKM, our curve is smoother and has higher precision when the recall is below 0.6. These results may be explained by the fact that dependencies between concepts could help identify relevant archetypes.

Table 5. Average precision performance of graphs with different similarity thresholds.

Graphs with different similarity thresholds[a]	Mean average precision	Average precision		
		Query 1 (medication)	Query 2 (laboratory test)	Query 3 (diagnosis)
G_1 (top 3%)	0.253	0.36	0.10	0.30
G_2 (top 5%)	0.277	0.27	0.26	0.30
G_3 (top 8%)	0.320	0.35	0.31	0.30
G_4 (top 10%)	0.313	0.33	0.31	0.30

[a] Graphs with percentages of values of $p(c_j|e(C_i))$.

Table 6. Retrieval performance comparison.

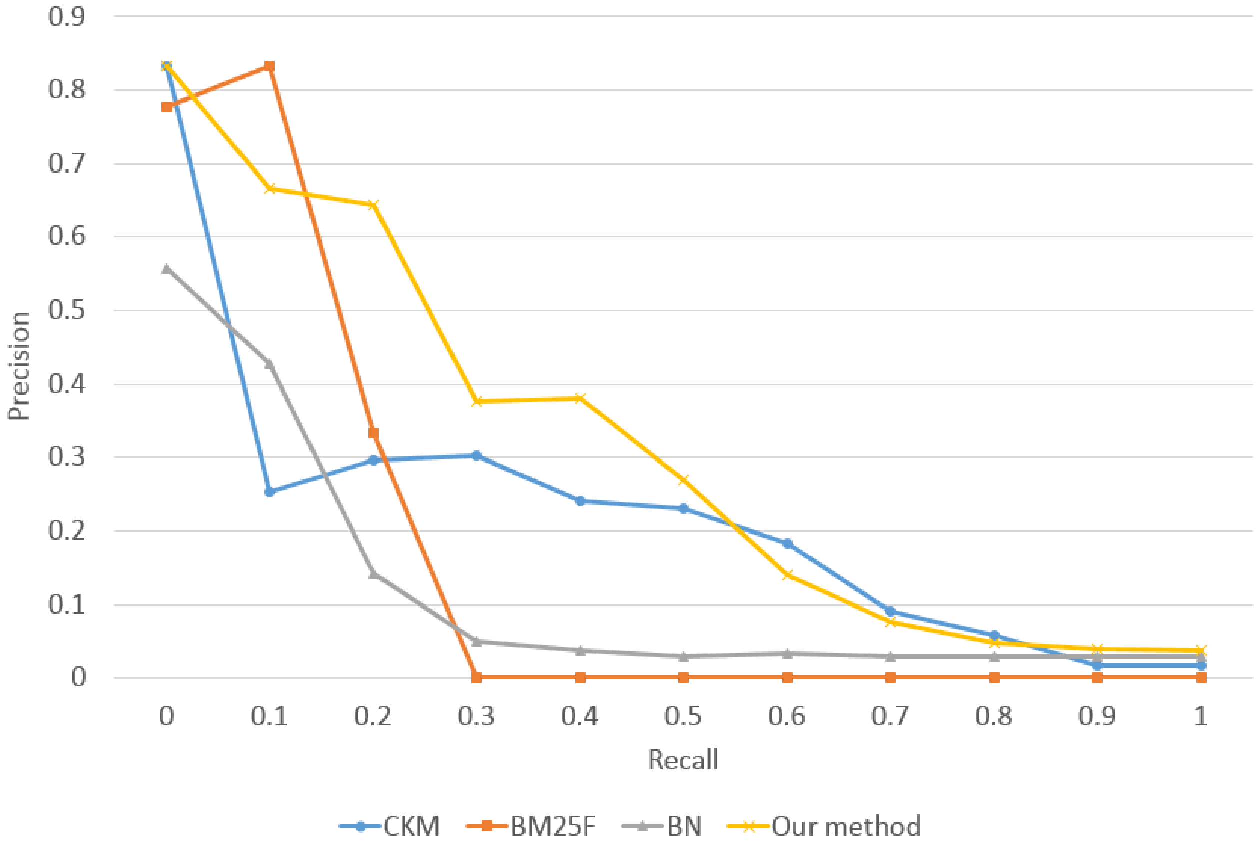
Method	MAP[a]	Query 1 (medication)		Query 2 (laboratory test)		Query 3 (diagnosis)	
		AP[b]	P@10[c]	AP	P@10	AP	P@10
CKM	0.227	0.26	0.40	0.31	0.30	0.11	0.10
BM25F	0.177	0.08	0.20	0.18	0.30	0.27	0.30
Bayesian network	0.127	0.11	0.20	0.22	0.30	0.05	0.10
Our method	0.320	0.35	0.50	0.31	0.50	0.30	0.30

[a] MAP: mean average precision.

[b] AP: average precision.

[c] P@10: precision at 10.

Figure 5. Precision-recall curves of the four retrieval methods. BM25F: an extension of the BM25 ranking function; BN: Bayesian network; CKM: Clinical Knowledge Manager.



Discussion

Principal Findings

The dual model methodology used by openEHR distinguished the clinical content domain from the technical domain, which enabled reusable CIMs (archetypes) [9]. We were interested in identifying relevant CIMs online to standardize clinical concept representation within EHRs, so we developed a graphical retrieval method based on an extended Bayesian network and validated its feasibility using an online clinical information knowledge source: OpenEHR CKM. We combined a qualitative representation of the retrieval task, by using a graphical representation of relationships among data elements, concepts, and archetypes, with quantitative representation of the uncertainty of information needs, using a probabilistic approach. Compared with three typical retrieval methods (BM25F, Bayesian network, and CKM) in the medication, laboratory test, and diagnosis retrieval tasks, our method achieved the best MAP (MAP=0.32). In the diagnosis retrieval task, CKM and BM25F could not find the relevant archetype "openEHR-EHR-SECTION.problems_and_diagnoses.v1." Our method could successfully identify the models covering "diagnostic reports," "problem list," "patients background," "clinical decision," etc, as well as "problems and diagnoses."

Although end users were mainly concerned about whether an archetype covered the concept name and core data items, we could not obtain satisfied performances without considering any potential knowledge that might be mined from the collection. Here, BM25F and Bayesian network just used clinical concepts and data elements as main features to represent each archetype and performed worse compared with the other models. In the laboratory test retrieval task, the recall of BM25F was 0.158, whereas ours was 1.0 and CKM was 0.895. In the diagnosis retrieval task, the value of precision at 3 of Bayesian network was 0, whereas ours was 1.0 and CKM was 0.333. A possible reason was that we used exact matching instead of fuzzy matching. The most important reason was that they only encoded the dependence relationships between variables and did not take into account the semantic associations between them. Previous studies showed that using the structure of existing knowledge resources and distributional statistics drawn from text corpora could help estimate semantic similarity and relatedness between medical concepts [43]. In the openEHR framework, archetypes should map to clinical terminologies (such as SNOMED CT). However, most archetypes currently in the CKM lacked this kind of mapping, which could have limited the calculation of semantic relatedness. In this study, we learned relationships between concepts by a probabilistic approach and constructed a concept subnetwork with two layers. The results showed that the performance significantly improved, which explained the effectiveness of using prior knowledge to improve retrieval results.

Accordingly, how to find the top *n* concepts relevant with each concept became crucial. We used e(C_i) as an event representing some type of evidence about the relevance of a concept C_i, and keywords were used as evidence in the experiment. With their help, we could find that the concepts "medication list" and "medication order list" were related, even though their concept name and data elements were totally different. There was also other semantic information that could be used as evidence, such as "purpose" and "use." How to use them to better support retrieval might need to be further clarified. However, this method could also include in the network some lower relevant concepts, as shown in the column "Top 10%" in Table 4. For better results, we used AP and MAP as evaluation metrics to help select relevant concepts; meanwhile, we noticed that many concepts had the same values of conditional probabilities. This was because of the probabilistic approach we applied, which reminded us that we could not simply select the top *n* concepts as the relevant ones. Here, we adopted concepts with top *n* percentages of values of conditional probabilities.

When modeling clinical resources network, we took the relationship of specialization between archetypes into consideration. It helped us find "openEHR-EHR-COMPOSITION.report-result.v1," a specialized archetype of "openEHR-EHR- COMPOSITION.report.v1," which BM25F could not find. In addition, we could also find relevant compositional parent archetypes successfully, even though we did not use the relationship of aggregation. For example, in the diagnosis retrieval task, our method could find "openEHR-EHR-SECTION.clinical_decision.v0," which defined an archetype slot to allow "openEHR-EHR-EVALUATION.problem_diagnosis.v1." It was because the compositional archetype used the clinical concept of the allowed archetype as its data element. When we linked the data element node to its corresponding concept node, we in fact modeled the relationship of aggregation.

The key idea of our approach lay in identifying potentially relevant clinical concepts from the input. In a two-level model methodology, clinicians were usually the end users. In most scenarios, they were not familiar with openEHR archetypes and did not know what archetype-friendly concepts were. This requires the retrieval method to be as insensitive to the input as possible. For example, take the medication retrieval task. If the user inputs "medication item, order start date/time, dosage, dose unit, comment," using some frequent words in the archetype library, the CKM performed better than ours. The AP value of CKM was 0.82 (P@10=0.7, recall=1) whereas ours was 0.45 (P@10=0.6, recall=1). However, when the user used uncommon words, such as "medicine name" (Table 1), our method, CKM, Bayesian network, and BM25F achieved an AP of 0.35, 0.26, 0.11, and 0.08, respectively. In addition, as Table 6 shows, our AP was almost equal across different retrieval tasks (0.35, 0.31, and 0.30, respectively), whereas the other retrieval methods were not. From the clinical domain, queries 2 and 3 mainly belonged to the topic of detection/treatment results, whereas query 1 belonged to treatment, which indicated that our performance was relatively stable across different clinical domains. All these showed that our method was more robust than the others.

Additionally, better retrieval results could help users to identify reusable archetypes quickly, promote reuse of archetypes, and improve standardization of CIMs, thereby enhancing interoperability of EHRs. Archetype modeling methodology [15,23] showed that clinicians and domain experts should compare archetype design specifications with retrieved archetypes to decide whether new archetypes need to be developed or whether an existing one could be adapted. Our method could successfully identify relevant archetypes that the CKM could not find, such as "openEHR-EHR-SECTION.problems_and_diagnoses.v1" in the diagnosis retrieval task. If this archetype was the case need, domain experts might create a new one as they thought it did not exist in the CKM. Our method achieved the best recall (recall=1) in different retrieval tasks, which could help reuse archetypes and promote the semantic interoperability of EHRs.

Limitations

Our study has important limitations. First, it is a feasibility study based on openEHR archetypes. Whether our method can be applied to other CIMs, such as HL7 templates, and to what extent it needs to be localized still need to be clarified and validated. In fact, the key features used in our method are data elements, clinical concepts, CIMs (archetypes), and their relationships. It indicates that our method has potential feasibility if these features are available for other CIMs. Which results are potentially possible will be discussed in future work.

Second, our method presented in this study lacks the calculation of the semantic relevance of synonyms or homonyms, both for queries and network modeling. However, relevant semantic computing methods [43] can be applied to our retrieval method. With their help, we may be able to identify that "medication item" and "medicine item" referred to the same term, and the results would be improved. In the future, we will validate its feasibility and effectiveness.

Third, we did not validate the impact of our method on interoperability. In fact, the basic problem of semantic interoperability in EHRs must be solved from the perspective of the business domains the concepts originally belong to. Our approach only addresses specific technical issues in the CIM modeling process.

Furthermore, there are other limitations. First, the relevant archetypes in the three retrieval tasks that we manually annotated may be controversial, according to different experts. Second, we compared our performances with the CKM on different archetype collections, which may lead to inaccurate results.

Conclusions

In this paper, we proposed an extended Bayesian network retrieval method for finding relevant CIMs. We graphically represented openEHR archetypes using an extended Bayesian network with two concept layers. The results show that it is an effective approach to meet the uncertainty of retrieval tasks, and the key step in modeling this network is to learn the dependencies between concepts. Our better retrieval results could encourage clinicians and domain experts to reuse existing CIMs to represent EHR data in a standard manner, thereby enhancing the interoperability of EHRs. Furthermore, our study provided how the inference process was carried out. Comparing the results of our method with baseline methods, we had the best performance. To optimize the method, further research should focus on the potential feasibility for other CIMs and the calculation of semantic relevance of synonyms or homonyms.

Acknowledgments

This research is supported by the Chinese Academy of Medical Sciences (grant #2017PT63010, 2018PT33024) and the National Key R&D Program of China (grant #2016YFC0901901, 2017YFC0907503).

Conflicts of Interest

None declared.

References

1. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence-what is it and what can it tell us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
2. Framework for FDA's Real-World Evidence Program. Silver Spring, MD: US Food and Drug Administration; 2018 Dec. URL: https://www.fda.gov/media/120060/download [accessed 2019-05-15] [WebCite Cache ID 78NqFnwnD]
3. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible sources of bias in primary care electronic health record data use and reuse. J Med Internet Res 2018 May 29;20(5):e185 [FREE Full text] [doi: 10.2196/jmir.9134] [Medline: 29844010]
4. European Commission. 2012 Dec 07. eHealth Action Plan 2012-2020: innovative healthcare for the 21st century URL: https://ec.europa.eu/digital-single-market/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century [accessed 2019-05-18] [WebCite Cache ID 78SkxTsZO]
5. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
6. International Organization for Standardization. 2011 Apr. ISO 18308:2011 health informatics-requirements for an electronic health record architecture URL: https://www.iso.org/standard/52823.html [accessed 2019-05-15] [WebCite Cache ID 78NrkRGeJ]
7. Rector AL, Nowlan WA, Kay S, Goble CA, Howkins TJ. A framework for modelling the electronic medical record. Methods Inf Med 1993 Apr;32(2):109-119. [Medline: 8321129]

8. Goossen W, Goossen-Baremans A, van der Zel M. Detailed clinical models: a review. Healthc Inform Res 2010 Dec;16(4):201-214 [FREE Full text] [doi: 10.4258/hir.2010.16.4.201] [Medline: 21818440]
9. Leslie H. ResearchGate. 2014 Jul. The openEHR approach URL: https://www.researchgate.net/publication/277667443_The_openEHR_approach [accessed 2019-05-15] [WebCite Cache ID 78NsTRNDB]
10. International Organization for Standardization. ISO 13606 Standard URL: https://www.iso.org/home.html [accessed 2019-05-15] [WebCite Cache ID 78Nsir48b]
11. OpenEHR. URL: https://www.openehr.org/ [accessed 2019-05-15] [WebCite Cache ID 78Nt2sQ3h]
12. Health Level Seven. URL: http://www.hl7.org/ [accessed 2019-05-15] [WebCite Cache ID 78NtFA8bQ]
13. Moreno-Conde A, Moner D, Cruz WD, Santos MR, Maldonado JA, Robles M, et al. Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. J Am Med Inform Assoc 2015 Jul;22(4):925-934. [doi: 10.1093/jamia/ocv008] [Medline: 25796595]
14. Wang L, Min L, Wang R, Lu X, Duan H. Archetype relational mapping-a practical openEHR persistence solution. BMC Med Inform Decis Mak 2015 Nov 05;15:88 [FREE Full text] [doi: 10.1186/s12911-015-0212-0] [Medline: 26541142]
15. Min L, Tian Q, Lu X, An J, Duan H. An openEHR based approach to improve the semantic interoperability of clinical data registry. BMC Med Inform Decis Mak 2018 Mar 22;18(Suppl 1):15 [FREE Full text] [doi: 10.1186/s12911-018-0596-8] [Medline: 29589572]
16. Cardoso de Moraes JL, de Souza WL, Pires LF, do Prado AF. A methodology based on openEHR archetypes and software agents for developing e-health applications reusing legacy systems. Comput Methods Programs Biomed 2016 Oct;134:267-287 [FREE Full text] [doi: 10.1016/j.cmpb.2016.07.013] [Medline: 27480749]
17. Min L, Liu J, Lu X, Duan H, Qiao Q. An implementation of clinical data repository with openehr approach: from data modeling to architecture. Stud Health Technol Inform 2016;227:100-105. [Medline: 27440296]
18. Marco-Ruiz L, Moner D, Maldonado JA, Kolstrup N, Bellika JG. Archetype-based data warehouse environment to enable the reuse of electronic health record data. Int J Med Inform 2015 Sep;84(9):702-714. [doi: 10.1016/j.ijmedinf.2015.05.016] [Medline: 26094821]
19. Wulff A, Haarbrandt B, Tute E, Marschollek M, Beerbaum P, Jack T. An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR. Artif Intell Med 2018 Jul;89:10-23 [FREE Full text] [doi: 10.1016/j.artmed.2018.04.012] [Medline: 29753616]
20. Chen R, Klein GO, Sundvall E, Karlsson D, Ahlfeldt H. Archetype-based conversion of EHR content models: pilot experience with a regional EHR system. BMC Med Inform Decis Mak 2009 Jul 01;9:33 [FREE Full text] [doi: 10.1186/1472-6947-9-33] [Medline: 19570196]
21. Saalfeld B, Tute E, Wolf K, Marschollek M. Introducing a method for transformation of paper-based research data into concept-based representation with openEHR. Stud Health Technol Inform 2017;235:151-155. [Medline: 28423773]
22. Mar M, Begoña M. Towards the interoperability of computerised guidelines and electronic health records: an experiment with openEHR archetypes and a chronic heart failure guideline. In: Riaño D, ten Teije A, Miksch S, Peleg M, editors. Knowledge Representation for Health-Care. KR4HC 2010. Lecture Notes in Computer Science. Berlin: Springer; 2011:101-113.
23. Moner D, Maldonado JA, Robles M. Archetype modeling methodology. J Biomed Inform 2018 Dec;79:71-81 [FREE Full text] [doi: 10.1016/j.jbi.2018.02.003] [Medline: 29454107]
24. openEHR. Clinical Knowledge Manager URL: https://www.openehr.org/ckm/ [accessed 2019-05-18] [WebCite Cache ID 78SlOqhgw]
25. Teodoro D, Sundvall E, João Junior M, Ruch P, Miranda Freire S. ORBDA: an openEHR benchmark dataset for performance assessment of electronic health record servers. PLoS One 2018;13(1):e0190028 [FREE Full text] [doi: 10.1371/journal.pone.0190028] [Medline: 29293556]
26. Maranhão PA, Bacelar-Silva GM, Ferreira DN, Calhau C, Vieira-Marques P, Cruz-Correia RJ. Nutrigenomic information in the openEHR data set. Appl Clin Inform 2018 Jan;9(1):221-231 [FREE Full text] [doi: 10.1055/s-0038-1635115] [Medline: 29590680]
27. Pahl C, Zare M, Nilashi M, de Faria Borges MA, Weingaertner D, Detschew V, et al. Role of OpenEHR as an open source solution for the regional modelling of patient data in obstetrics. J Biomed Inform 2015 Jun;55:174-187 [FREE Full text] [doi: 10.1016/j.jbi.2015.04.004] [Medline: 25900270]
28. Finlayson SG, LePendu P, Shah NH. Building the graph of medicine from millions of clinical narratives. Sci Data 2014;1:140032 [FREE Full text] [doi: 10.1038/sdata.2014.32] [Medline: 25977789]
29. Goodwin T, Harabagiu SM. Automatic generation of a qualified medical knowledge graphits usage for retrieving patient cohorts from electronic medical records. 2013 Sep 16 Presented at: IEEE Seventh International Conference on Semantic Computing; Sep 16-18, 2013; Irvine, CA p. 978 URL: http://www.hlt.utdallas.edu/~travis/papers/icsc_2013.pdf
30. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. Sci Rep 2017 Jul 20;7(1):5994 [FREE Full text] [doi: 10.1038/s41598-017-05778-z] [Medline: 28729710]
31. Turtle HR, Croft WB. Efficient probabilistic inference for text retrieval. In: Proceedings RIAO '91 Intelligent Text and Image Handling. 1991 Presented at: RIAO '91 Intelligent Text and Image Handling; Apr 2-5, 1991; Barcelona, Spain p. 644-661 URL: https://dl.acm.org/citation.cfm?id=3171012

32. de Campos LM, Fernandez-Luna J, Huete J. The BNR model: foundations and performance of a bayesian network-based retrieval model. Int J Approx Reason 2003 Nov;34(2-3):265-285 [FREE Full text]
33. de Campos LM, Fernandez-Luna J, Huete J. Clustering terms in the bayesian network retrieval model: a new approach with two term-layers. Appl Soft Comput 2004 May;4(2):149-158 [FREE Full text]
34. Garrouch K, Omri M. Bayesian network based information retrieval model. 2017 Presented at: International Conference on High Performance Computing & Simulation; July 17, 2017; Genoa, Italy URL: https://www.researchgate.net/publication/317185437_Bayesian_Network_Based_Information_Retrieval_Model
35. Xu JM, Tang WS. A word similarity based belief network IR model with two term layers. 2009 Presented at: WRI Global Congress on Intelligent Systems; May 19-21, 2009; Xiamen, China p. 19-21 URL: https://ieeexplore.ieee.org/document/5209386
36. Acid S, de Campos LM, Fernandez-Luna J. An information retrieval model based on simple Bayesian networks. Int J Intell Syst 2003;18(2):251-265 [FREE Full text]
37. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical data reuse or secondary use: current status and potential future progress. Yearb Med Inform 2017 Aug;26(1):38-52 [FREE Full text] [doi: 10.15265/IY-2017-007] [Medline: 28480475]
38. Rajkomar A, Oren E, Chen K. Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine 2018 May;1(1):18 [FREE Full text]
39. Hruby GW, Hoxha J, Ravichandran PC, Mendonça EA, Hanauer DA, Weng C. A data-driven concept schema for defining clinical research data needs. Int J Med Inform 2016 Jul;91:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.03.008] [Medline: 27185504]
40. Denaxas SC, Morley KI. Big biomedical data and cardiovascular disease research: opportunities and challenges. Eur Heart J Qual Care Clin Outcomes 2015 Jul 01;1(1):9-16. [doi: 10.1093/ehjqcco/qcv005] [Medline: 29474568]
41. GitHub. Adl-parser URL: https://github.com/openEHR/java-libs/tree/master/adl-parser [accessed 2019-05-15] [WebCite Cache ID 78NxYGEFV]
42. Zaragoza H, Craswell N, Taylor M. Microsoft Cambridge at TREC 2004: Web and HARD track. 2004 Presented at: TREC 2004; Nov 16, 2004; Gaithersburg, MD URL: https://trec.nist.gov/pubs/trec13/papers/microsoft-cambridge.web.hard.pdf
43. Yu Z, Wallace BC, Johnson T, Cohen T. Retrofitting concept vector representations of medical concepts to improve estimates of semantic similarity and relatedness. Stud Health Technol Inform 2017;245:657-661 [FREE Full text] [Medline: 29295178]

Abbreviations

ADL: Archetype Definition Language
AP: average precision
CIM: clinical information model
CKM: Clinical Knowledge Manager
EHR: electronic health record
HL7: Health Level Seven
MAP: mean average precision
P@10: precision at 10
RM: reference model

Edited by G Eysenbach; submitted 27.01.19; peer-reviewed by X Lu, D Moner, G Tognola, J Lee, R Correia, W Goossen; comments to author 21.02.19; revised version received 18.04.19; accepted 02.05.19; published 28.05.19.

Please cite as:
Yang L, Huang X, Li J
Discovering Clinical Information Models Online to Promote Interoperability of Electronic Health Records: A Feasibility Study of OpenEHR
J Med Internet Res 2019;21(5):e13504
URL: http://www.jmir.org/2019/5/e13504/
doi:10.2196/13504
PMID:31140433

Original Paper

A Digital Behavioral Weight Gain Prevention Intervention in Primary Care Practice: Cost and Cost-Effectiveness Analysis

Anirudh Krishnan[1], BS; Eric Andrew Finkelstein[1], MHA, PhD; Erica Levine[2], BA, MPH; Perry Foley[2], MPH, MSW; Sandy Askew[2], MPH; Dori Steinberg[2], MS, RD, PhD; Gary G Bennett[2], PhD

[1]Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

[2]Duke Global Digital Health Science Center, Duke University, Durham, NC, United States

Corresponding Author:
Anirudh Krishnan, BS
Program in Health Services and Systems Research
Duke-NUS Medical School
8 College Road
Singapore, 169857
Singapore
Phone: 65 66015470
Fax: 65 62217372
Email: anirudh.krishnan@duke-nus.edu.sg

Abstract

Background: Obesity is one of the largest drivers of health care spending but nearly half of the population with obesity demonstrate suboptimal readiness for weight loss treatment. Black women are disproportionately likely to have both obesity and limited weight loss readiness. However, they have been shown to be receptive to strategies that prevent weight gain.

Objective: The aim of this study was to evaluate the costs and cost-effectiveness of a digital weight gain prevention intervention (Shape) for black women. Shape consisted of adaptive telephone-based coaching by health system personnel, a tailored skills training curriculum, and patient self-monitoring delivered via a fully automated interactive voice response system.

Methods: A cost and cost-effectiveness analysis based on a randomized clinical trial of the Shape intervention was conducted from the payer perspective. Costs included those of delivering the program to 91 intervention participants in the trial and were summarized by program elements: self-monitoring, skills training, coaching, and administration. Effectiveness was measured in quality-adjusted life years (QALYs). The primary outcome was the incremental cost per QALY of Shape relative to usual care.

Results: Shape cost an average of US $758 per participant. The base-case model in which quality of life benefits decay linearly to zero 5 years post intervention cessation, generated an incremental cost-effectiveness ratio (ICER) of US $55,264 per QALY. Probabilistic sensitivity analyses suggest an ICER below US $50,000 per QALY and US $100,000 per QALY in 39% and 98% of simulations, respectively. Results are highly sensitive to durability of benefits, rising to US $165,730 if benefits end 6 months post intervention.

Conclusions: Results suggest that the Shape intervention is cost-effective based on established benchmarks, indicating that it can be a part of a successful strategy to address the nation's growing obesity epidemic in low-income at-risk communities.

(J Med Internet Res 2019;21(5):e12201) doi:10.2196/12201

KEYWORDS

cost-effectiveness; cost-benefit analysis; obesity; telemedicine; women's health; minority health; weight gain prevention; weight gain

Introduction

Excess weight is estimated to account for 9% of total annual health care costs, with roughly half paid by public sector health programs [1]. Nationally, almost 55% of black women have obesity compared with 38% of white women [2]. As a result, black women are at greater risk for obesity-related chronic diseases, including stroke, coronary heart disease, and depression, and attributable costs [3,4]. As a result, interventions that successfully address excess weight in this at-risk group may confer significant health and economic benefits to individuals and society. However, reducing risk factors in this group is a challenge because, relative to other populations, black women express less interest in or readiness for weight loss treatment [5,6]. Moreover, weight loss interventions have shown consistently smaller weight loss among black women relative to their white counterparts [7-10]. Therefore, delivering interventions that seek to prevent weight gain, as opposed to promoting weight loss, might be a more successful treatment strategy [11].

Weight gain prevention strategies align with sociocultural norms among black communities that are tolerant of higher body weights [12-14]. Previous digital weight gain prevention interventions have shown moderate success in reducing weight gain among children and young adults with overweight, but no studies had studied their effectiveness among black women [15-17]. Bennett et al developed *The Shape Program* to test a tailored digital health solution aimed at helping black women prevent weight gain [18,19]. Results reveal Shape's effectiveness in preventing weight gain among black women. However, whether Shape is cost-effective remains unknown; that is the focus of this analysis.

A cost-effectiveness analysis is one strategy for understanding whether the benefits of an intervention are worth the costs. Many public sector agencies, such as the National Institute for Health and Care Excellence in the United Kingdom and the Health Intervention and Technology Assessment Program in Thailand, require cost-effectiveness analyses before considering a subsidy decision for a health intervention [20]. Although the United States does not systematically require cost-effectiveness analyses, they have gained popularity as a tool to compare the value of diverse interventions. Guidelines recommend that cost-effectiveness analyses report benefits in terms of a common metric such as the quality-adjusted life year (QALY), which consolidates diverse health benefits to facilitate comparisons of value among interventions targeting diverse population health gaps [21,22]. In this study, we present the costs and cost-effectiveness of Shape relative to usual care (UC) in terms of cost per QALY gained and compare this value to established benchmarks for cost-effectiveness. Given the risk of steady weight gain in the target population, third-party payers may be interested in knowing whether a successful weight gain prevention program, such as Shape, represents good use of scarce health care resources.

Methods

The Shape Program

The Shape Program (Shape) was designed to prevent weight gain in black female primary care patients whose body mass index (BMI) placed them in either the overweight (25 to 29.9 kg/m^2) or class 1 obese (30 to 34.9 kg/m^2) categories. Shape sought to promote the modification of obesogenic lifestyle behaviors (diet, physical activity, and leisure time activities). It leveraged key technological innovations to support personnel within a private, nonprofit community health center network [19]. In doing so, the program was able to augment the capacity of existing health systems to reach patients who otherwise would receive little or no weight management counseling. The Shape program included adaptive telephone-based coaching by health system personnel, personalized obesogenic behavior change goals assigned every 2 months, a tailored skills training curriculum, patient self-monitoring delivered via a fully automated interactive voice response system, 12 counseling calls with a registered dietitian, and a 12-month gym membership [19].

Shape's effectiveness relative to a light touch UC intervention was tested among 194 overweight and class 1 obese black women aged 25 to 44 years in a 2-arm parallel-group randomized controlled trial over 12 months followed by a 6-month follow-up period (ClinicalTrials.gov reference: NCT00938535) [18]. Additional inclusion criteria were having visited a member in the health center in the past 24 months, being a state resident, and being fluent in English. Participants were excluded if they were pregnant, up to 12 months postpartum, had a myocardial infarction or stroke in the past 2 years, or had any history of cognitive, developmental, or psychiatric disorders. In the trial, UC consisted of the *Aim for a Healthy Weight* brochure and semiannual newsletters on health topics not related to weight. Intent-to-treat analyses included outcome measurements for 91 participants randomized to receive the intervention and 94 UC participants [18]. At 12 and 18 months, Shape participants had lower weight gain than UC participants (mean difference of −1.4 kg and −1.7 kg at 12 and 18 months, respectively).

Cost Analysis

To estimate the incremental costs of Shape, we employed an activity-based costing method that links program resource consumption to specific program components [23,24]. This approach allows evaluators to map the resource flow of the program. Electronic budgetary records, staff interviews, and engagement data were utilized to estimate program costs. All costs were inflated to 2018 US dollars using the medical portion of the seasonally adjusted US consumer price index [25].

Cost-Effectiveness Analysis

The cost-effectiveness analysis, which consisted of quantifying the incremental costs and QALYs of Shape relative to UC, was conducted from the third-party payer perspective.

Incremental Cost

As virtually no costs were incurred in the usual-care arm, the incremental cost is set equal to the cost of program delivery of the Shape intervention (including self-monitoring, skills training, coaching, and administration costs). This excluded program *development* costs, as these represent sunk costs that would not need to be repeated if the program were more broadly adopted. The average per capita cost of program delivery was assigned to the 91 participants who received the intervention and were included in the intent-to-treat analysis.

Incremental Effectiveness

The primary measure of effectiveness in the trial was weight change from baseline to 12 months. We converted the weight change into a health-related quality of life (QoL) change score over this time period. This imputation followed the regression approach described in Finkelstein and Kruger [26] using data from Finkelstein et al [27] and restricting it to a sample of women with a BMI between 25 and 35. This age and gender restriction allowed for obtaining estimates in a subsample that best approximates the characteristics of the Shape study population. Using this restricted sample, we estimated the association between QoL change and weight change (in kilograms) while controlling for baseline BMI and age via the following equation:

$$\Delta \text{QoL}_i = \beta_1 \times \Delta \text{weight}_i + \beta_2 \times (\Delta \text{weight}_i)^2 + \beta_3 \times \text{baseline BMI}_i + \beta_4 \times \text{baseline age}_i + \varepsilon_i$$

Using a process of step-wise regression, iteratively dropping variables found not to be statistically significant at the 5% significance level, we identified the following relationship:

$$\Delta \text{QoL} = -0.0029 \times \Delta \text{weight} + 0.0002 \times \text{baseline age}$$

We used this equation to impute a QoL change for each individual in the Shape trial.

Cost-Effectiveness Analysis

As with the primary analysis, the cost-effectiveness analysis was based on the intention-to-treat sample, with missing observations in both trial arms treated as missing at random [18]. The numerator of the incremental cost-effectiveness ratio (ICER) is the incremental cost to deliver Shape. The denominator is the mean discounted QALYs gained by intervention participants minus mean discounted QALYs gained by the UC group. QALY estimates for each arm were generated by plotting a curve of ΔQoL against time from baseline and taking the area under this curve. All post-trial QALY estimates were discounted at 3.5% per annum.

In the base case, we used estimates of QoL change from baseline to each of 6 months, 12 months, and 18 months, and then assumed QoL benefits decay linearly until the end of the fifth year postcessation of the intervention, at which time we assume no further benefits.

Sensitivity Analyses

We assessed the sensitivity of our ICER to changes in key inputs using 1-way sensitivity analyses. We estimated the effect of the following changes on the ICER: (1) halving the cost of the intervention; (2) doubling and halving the costs incurred in each cost category; (3) doubling or halving the incremental effectiveness of the intervention with regard to UC; and (4) varying the duration of residual benefits post cessation from 5 years in the base case to 0.5 and 3 years.

Probabilistic Sensitivity Analyses

In addition, we conducted 10,000 simulations of the model to quantify the probability that the intervention is cost-effective for a range of willingness-to-pay thresholds that decision makers might consider. Cost was assumed to follow a gamma distribution, with an SD of 25% of mean costs; effectiveness was assumed to follow a normal distribution, with SDs equal to the SEs of effectiveness estimates.

The methods described in this section and reporting of results throughout the paper are consistent with the Consolidated Health Economic Evaluation Reporting Standards [21].

Results

Program Costs

The total cost of Shape was US $758 per participant in 2018 US dollars (Table 1) for the 1-year intervention. Program costs were allocated to 4 areas, including administration, self-monitoring, skills training, and counseling (Table 1). Administration costs, including personnel, costs of support staff training, and space and other overheads were the greatest consumer of program resources at an average cost of US $387 per participant. Telephone counseling costs were the second highest cost driver, driven largely by registered dietitians' personnel costs and cell phone plan subscriptions, at an average of US $149 per participant. Interactive self-monitoring included server and interactive voice response system maintenance costs and purchasing of pedometers and scales and cost an average of US $126 per participant. Tailored skills training costs US $95 per participant, primarily driven by the cost of printing training materials and providing kit bags to participants. Training the coaches front-loaded many of the costs in the first 2 years of Shape. Specifically, the average program costs in years 1 and 2 (US $17,401) were 53% higher than the average costs in years 3 to 5 (US $11,380).

The variability in Shape coaching costs in years 1 and 2 was further explored. Shape coaches placed 3968 calls to participants (an average of 44 calls per participant during the yearlong program). The majority of these calls (3316/3968, 83.6%) were attempts to reach participants to deliver coaching content, while 16.4% (652/3968) were considered successful coaching calls in which the curriculum was delivered in full. These successful calls were on average 21.2 min long (SD 10.1 min). The average amount of time that coaches spent on unsuccessful calls per participant was 27.6 min (SD 19.2 min) for the whole program period.

Table 1. Program delivery costs for 91 participants by program area and year (all figures in 2018 US $).

Program area	Year 1 (US $)	Year 2 (US $)	Year 3 (US $)	Year 4 (US $)	Year 5 (US $)	Total (US $)	Cost per participant[a] (US $)	Cost as percentage of total[b]
Interactive self-monitoring	7123	1140	1102	1068	1047	11,481	126	17%
Tailored skills training	1785	4585	2161	132	0	8662	95	13%
Telephone counseling	1877	3417	3301	3199	1794	13,588	149	20%
Administration	7665	7210	6967	6750	6619	35,212	387	51%
Total	18,450	16,352	13,531	11,148	9461	68,942	758	100%

[a] Calculated for a total of 91 intervention participants.

[b] Total does not sum to 100% due to rounding.

Cost-Effectiveness Analysis

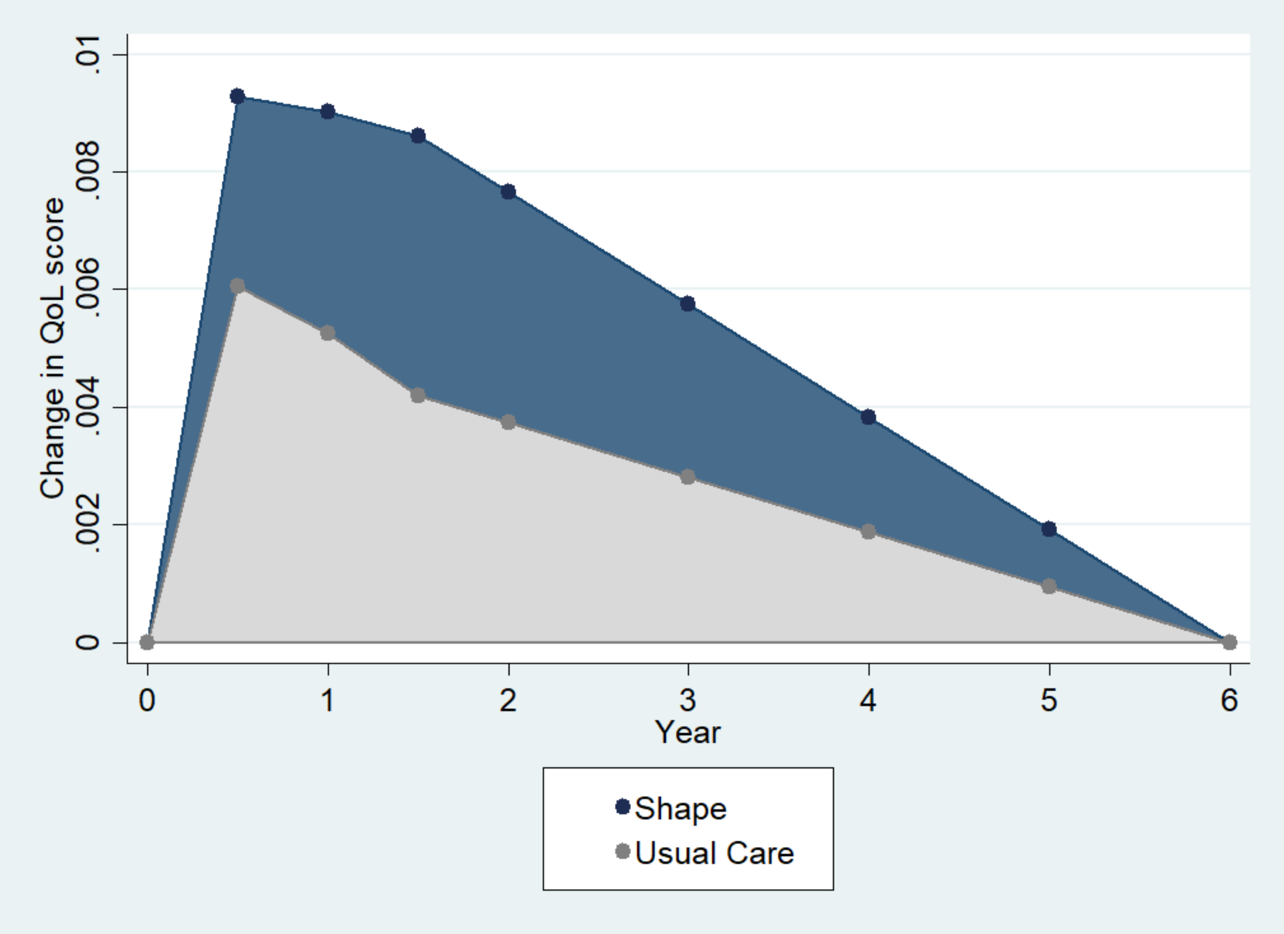
Incremental Cost

As virtually no cost was incurred in the UC arm, the incremental cost of Shape relative to UC was US $758 in the base case.

Effectiveness

As reported in Bennett et al, mean difference in weight change of the intervention and UC arms with regard to baseline approached statistical significance at 6 months (−1.1 kg [95% CI −2.3 to 0.04]), and was statistically significant at the 12-month (−1.4 kg [−2.8 to −0.1]) and 18-month (−1.7 kg [−3.3 to −0.2]) assessments [18]. The difference in weight change across arms was transformed to QoL change scores for Shape participants and UC participants at 6 months (+0.009 and +0.006, respectively), 12 months (+0.009 and +0.005, respectively), and 18 months (+0.009 and +0.004, respectively) from baseline. Figure 1 presents the graph of QoL change against time from baseline.

Figure 1. Estimated quality of life (QoL) change score plotted against time from baseline. The difference in area under the Shape curve and the Usual Care curve (ie, dark shaded region) represents the mean gain in quality-adjusted life years per participant in the base case analysis.



Base-Case Incremental Cost-Effectiveness Ratio

In the base case of a 1-year intervention followed by 5 years of linear decay of post-trial weight gain prevention benefits, we estimated an ICER of US $55,264 per QALY gained.

Sensitivity Analyses

One-way sensitivity analyses showed that halving incremental QALYs of the intervention arm with regard to UC raised the ICER to US $110,529 per QALY, whereas doubling incremental QALYs or halving incremental costs decreased the ICER to US $27,632 per QALY. When each Shape cost category was doubled or halved separately, the ICER ranged from US $41,130 to US $83,447 (administration costs); US $49,792 to US $66,122 (telephone counseling costs); US $51,766 to US $62,176 (tailored skills training costs); and US $50,637 to US $64,434 (interactive self-monitoring costs). With QoL benefits modeled to decay to zero within 3 years of cessation of the intervention, the ICER was US $77,644; with benefits ceasing 6 months after the intervention concluded, the ICER rose to US $165,730. Figure 2 presents a tornado diagram showing the results of the one-way sensitivity analyses.

Probabilistic Sensitivity Analyses

Figure 3 displays the cost-effectiveness acceptability curve. The figure reveals that 39.3% of simulations suggest that the incremental cost per QALY of Shape relative to UC is less than US $50,000, an oft-cited threshold for cost-effectiveness [28,29]. At a willingness to pay of US $100,000 per QALY (another commonly cited threshold), 98.3% of simulations suggested that Shape is cost-effective [30,31].

Figure 2. Results of one-way sensitivity analyses varying key parameters (incremental QALYs, incremental and category-specific costs, and duration of post-cessation benefits). ICER: incremental cost-effectiveness ratio; QALYs: quality-adjusted life years.

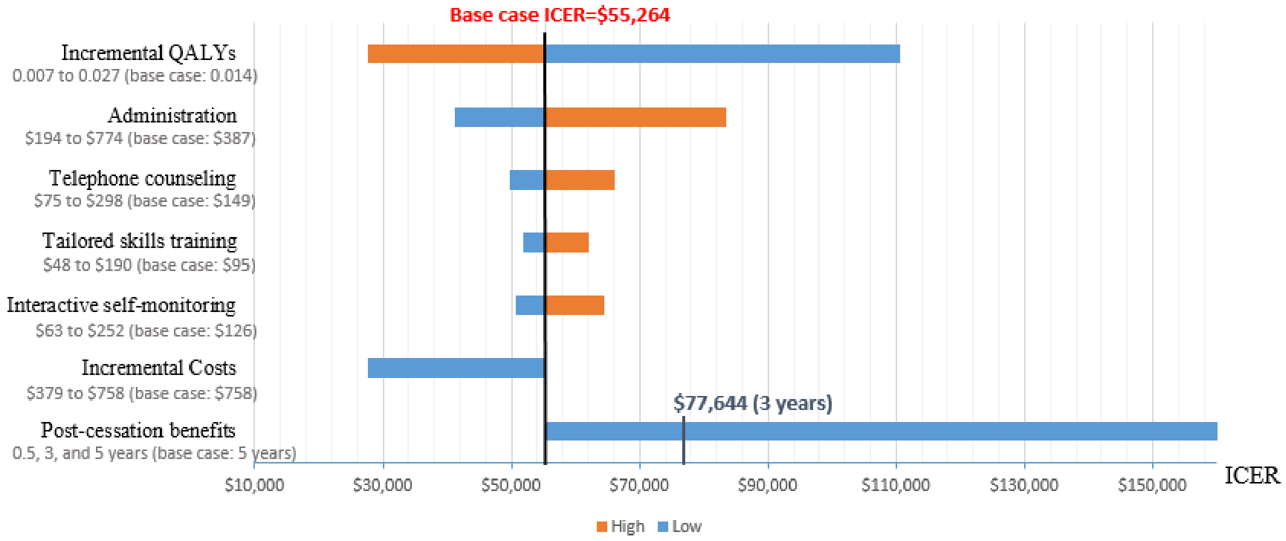
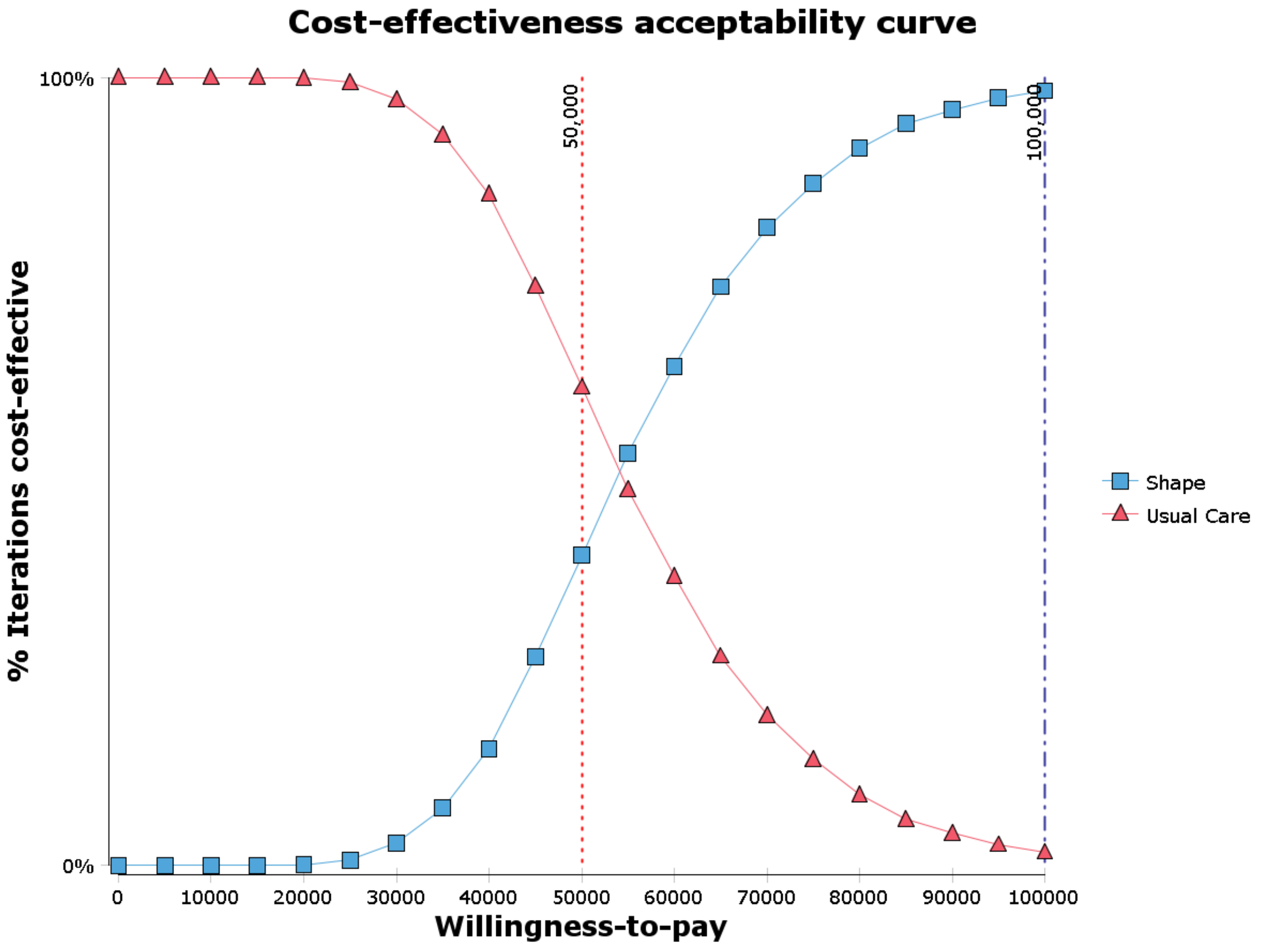


Figure 3. Cost-effectiveness acceptability curve compared against 2 potential cost-effectiveness thresholds.



Discussion

Principal Findings

This study presents the first evidence that a digital weight management program can be a cost-effective solution for preventing weight gain. In our base case, we estimated the incremental cost-effectiveness of Shape per QALY gained to be US $55,264, slightly higher than the often-quoted threshold of US $50,000 per QALY. The reality is that there is no established threshold for cost-effectiveness of health care interventions in the United States; indeed, the Patient Protection and Affordable Care Act specifically bars the use of a cost-effectiveness measure as a threshold [32]. Despite this, a threshold of US $50,000 per QALY has been widely used in the United States since 1992 [29]. In probabilistic sensitivity analyses, 39% of our simulations suggested that Shape has an ICER below this threshold. Some researchers even suggest a more appropriate threshold would be US $100,000 per QALY [30,31]. Compared against this threshold, 98% of our simulations suggest that Shape is cost-effective. Although there are no other weight gain prevention programs to compare against, Shape's ICER also compares favorably to the majority of the lifestyle and pharmacological interventions targeting weight loss [33]. Moreover, although this study did not quantify cost offsets from slower weight progression, Cawley et al suggest that annual savings from even moderate weight loss (or less weight gain relative to control) far exceed the US $758 annual cost of the program [34]. Moreover, there are several reasons to believe Shape's estimated cost of US $758 is likely to be an upper bound. Costs in years 3 to 5, when the program was recruiting at a much higher rate, were an average of 35% lower than that of the first 2 years (US $11,380 in years 3 to 5 compared with $17,401 in years 1 to 2; Table 1). If one considers only the per participant variable costs of the pedometer, scale, skills training kit bag, and coaching time, with the remaining costs averaged over a very large number of participants, per capita costs could be as low as US $243 (see Multimedia Appendix 1). This suggests that at full scale, Shape may be highly cost-effective. However, it may be that costs would need to be further reduced with no loss in outcomes for Shape to be highly scalable. For example, Weight Watchers OnlinePlus costs only US $160 per year [35]. Shape may need to demonstrate an average cost per participant in this range or better to increase the potential for scalability. This could be accomplished through better use of Shape's data to customize the intervention at the individual level and to intervene early for those most at risk of dropping out. This should be an area of future research.

Limitations

Although this study has many strengths, we identified 5 key limitations in this study. First, there is a lack of evidence on the persistence of weight gain prevention effects post intervention. The existing literature strongly suggests significant weight loss maintenance, and presumably QoL gains, up to 5 years after successful weight loss programs [36,37]. However, there are no corresponding data for weight gain prevention. We made the assumption that quality-of-life benefits would last 5 years beyond the intervention period and tested the sensitivity of the ICER to this assumption. Sensitivity analyses suggested that the ICER was highly sensitive to the duration over which benefits persist. Studies with longer-term follow-ups of both weight loss and QoL are required to validate our assumption in the context of weight gain prevention programs. Second, owing to data limitations, we were not able to directly assess QoL in participants in this study. Instead, we used an approach similar to that used in other cost-effectiveness studies [26,27] and assessed the sensitivity of our estimates to uncertainty in the QoL estimates using one-way and probabilistic sensitivity analyses. Analyses suggest that the ICER is moderately sensitive to the relationship between QoL and weight change. Third, estimates of QoL changes are sensitive to the method of elicitation. This is true whether one uses direct elicitation methods, such as standard gamble or time-tradeoff methods, or using patient-reported outcomes measures, such as the data from the SF-12 version 2 instrument used in this study [38]. This fact, combined with our imputation strategy which imputes QoL changes solely from weight change and age, ignoring other potentially important confounders, suggests that there is likely a high degree of error in our QoL estimates, as would be the case in most cost-effectiveness studies. We address this via sensitivity analyses using a wide range of QoL values that we believe capture reasonable lower and upper bounds for these estimates. Fourth, the trial did not measure potential cost offsets from reduced health care utilization that may result from improved participant health outcomes. As a result, cost-effectiveness results presented here may be conservative. Finally, this program was delivered to black women in a low-income rural community health center setting. Although the intervention could be fielded in any setting and to diverse populations, future studies would be needed to see if the results are generalizable. However, long-term studies that follow a cohort of participants over an extended period of time and link their weight loss to changes in health care utilization would be needed to truly confirm the long-term cost-effectiveness of the Shape intervention.

Conclusions

Although long-term studies are needed to confirm this result, this study suggests that Shape is likely to be a cost-effective intervention to prevent weight gain and reduce risks for chronic disease among high-risk black women in low-income rural communities, where obesity rates are also highest. It thus provides an additional strategy that these communities can rely on to effectively and efficiently respond to the nation’s growing obesity epidemic.

Acknowledgments

This study was funded by grants from the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK078798) and the National Institute on Drug Abuse (F32-DA034501 and R13-DA036339).

Conflicts of Interest

GGB holds equity in Coeus Health and serves on the scientific advisory boards of WW International (formerly Weight Watchers International) and Interactive Health. These organizations had no role in study design, data collection, data analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication. DS is a consultant with Omada Health. AK, EAF, EL, PF, and SA declare they have no conflicts of interest.

Multimedia Appendix 1

Costs of Shape Program by Program Activity (all costs in 2018 US $).

[PDF File (Adobe PDF File), 85KB - jmir_v21i5e12201_app1.pdf]

References

1. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer-and service-specific estimates. Health Aff (Millwood) 2009;28(5):w822-w831 [FREE Full text] [doi: 10.1377/hlthaff.28.5.w822] [Medline: 19635784]
2. Hales C, Carroll M, Fryar C, Ogden C. Prevalence of obesity among adults and youth: United States, 2015-2016. NCHS Data Brief 2017 Dec(288):1-8 [FREE Full text] [Medline: 29155689]
3. Hu FB, Stampfer MJ, Solomon C, Liu S, Colditz GA, Speizer FE, et al. Physical activity and risk for cardiovascular events in diabetic women. Ann Intern Med 2001 Jan 16;134(2):96-105. [Medline: 11177312]
4. Roberts RE, Deleger S, Strawbridge WJ, Kaplan GA. Prospective association between obesity and depression: evidence from the Alameda County Study. Int J Obes Relat Metab Disord 2003 Apr;27(4):514-521. [doi: 10.1038/sj.ijo.0802204] [Medline: 12664085]
5. Stevens J, Kumanyika SK, Keil JE. Attitudes toward body size and dieting: differences between elderly black and white women. Am J Public Health 1994 Aug;84(8):1322-1325. [Medline: 8059896]

6. Paeratakul S, White MA, Williamson DA, Ryan DH, Bray GA. Sex, race/ethnicity, socioeconomic status, and BMI in relation to self-perception of overweight. Obes Res 2002 May;10(5):345-350 [FREE Full text] [doi: 10.1038/oby.2002.48] [Medline: 12006633]
7. Kumanyika SK, Obarzanek E, Stevens VJ, Hebert PR, Whelton PK, Kumanyaka SK. Weight-loss experience of black and white participants in NHLBI-sponsored clinical trials. Am J Clin Nutr 1991 Dec;53(6 Suppl):1631S-1638S. [doi: 10.1093/ajcn/53.6.1631S] [Medline: 2031498]
8. Osei-Assibey G, Adi Y, Kyrou I, Kumar S, Matyka K. Pharmacotherapy for overweight/obesity in ethnic minorities and white Caucasians: a systematic review and meta-analysis. Diabetes Obes Metab 2011 May;13(5):385-393. [doi: 10.1111/j.1463-1326.2010.01346.x] [Medline: 21205118]
9. West DS, Elaine Prewitt T, Bursac Z, Felix HC. Weight loss of black, white, and Hispanic men and women in the Diabetes Prevention Program. Obesity (Silver Spring) 2008 Jun;16(6):1413-1420 [FREE Full text] [doi: 10.1038/oby.2008.224] [Medline: 18421273]
10. Wing RR, Anglin K. Effectiveness of a behavioral weight control program for blacks and whites with NIDDM. Diabetes Care 1996 May;19(5):409-413. [Medline: 8732700]
11. Benjamin RM, Yanovski SZ, Simons-Morton DG. Can primary care physician-driven community programs address the obesity epidemic among high-risk populations? JAMA Intern Med 2013 Oct 28;173(19):1778-1779. [doi: 10.1001/jamainternmed.2013.7776] [Medline: 23979565]
12. Flynn KJ, Fitzgibbon M. Body images and obesity risk among black females: a review of the literature. Ann Behav Med 1998;20(1):13-24. [Medline: 9755347]
13. Hebl MR, Heatherton TF. The stigma of obesity in women: The difference is black and white. Pers Soc Psychol Bull 2016 Jul 02;24(4):417-426. [doi: 10.1177/0146167298244008]
14. Lovejoy M. Disturbances in the social body: differences in body image and eating problems among African American and white women. Gend Soc 2016 Jun 30;15(2):239-261. [doi: 10.1177/089124301015002005]
15. Rodearmel SJ, Wyatt HR, Barry MJ, Dong F, Pan D, Israel RG, et al. A family-based approach to preventing excessive weight gain. Obesity (Silver Spring) 2006 Aug;14(8):1392-1401 [FREE Full text] [doi: 10.1038/oby.2006.158] [Medline: 16988082]
16. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mHealth lifestyle program with telephone support (TXT2BFiT) to prevent unhealthy weight gain in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2015;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
17. West DS, Monroe CM, Turner-McGrievy G, Sundstrom B, Larsen C, Magradey K, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. J Med Internet Res 2016 Jun 13;18(6):e133 [FREE Full text] [doi: 10.2196/jmir.5474] [Medline: 27296086]
18. Bennett GG, Foley P, Levine E, Whiteley J, Askew S, Steinberg DM, et al. Behavioral treatment for weight gain prevention among black women in primary care practice: a randomized clinical trial. JAMA Intern Med 2013 Oct 28;173(19):1770-1777 [FREE Full text] [doi: 10.1001/jamainternmed.2013.9263] [Medline: 23979005]
19. Foley P, Levine E, Askew S, Puleo E, Whiteley J, Batch B, et al. Weight gain prevention among black women in the rural community health center setting: the Shape Program. BMC Public Health 2012 Jun 15;12:305 [FREE Full text] [doi: 10.1186/1471-2458-12-305] [Medline: 22537222]
20. National Institute for HealthCare Excellence (NICE). National Institute for Health and Care Excellence (NICE). 2015 Oct. NICE Process and Methods Guides: Developing NICE Guidelines: The Manual URL: https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf [accessed 2019-05-02] [WebCite Cache ID 7858Rp6Rz]
21. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health 2013;16(2):231-250 [FREE Full text] [doi: 10.1016/j.jval.2013.02.002] [Medline: 23538175]
22. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. J Am Med Assoc 2016 Sep 13;316(10):1093-1103. [doi: 10.1001/jama.2016.12195] [Medline: 27623463]
23. Crowley DM, Jones DE, Greenberg MT, Feinberg ME, Spoth RL. Resource consumption of a diffusion model for prevention programs: the PROSPER delivery system. J Adolesc Health 2012 Mar;50(3):256-263 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.07.001] [Medline: 22325131]
24. Frick KD. Microcosting quantity data collection methods. Med Care 2009 Jul;47(7 Suppl 1):S76-S81 [FREE Full text] [doi: 10.1097/MLR.0b013e31819bc064] [Medline: 19536026]
25. Bureau of Labor Statistics. BLS Data Viewer. Washington, DC: United States Department of Labor; 2018. CPI-All Urban Consumers URL: https://beta.bls.gov/dataViewer/view/timeseries/CUSR0000SAM [accessed 2019-05-02] [WebCite Cache ID 78595lxjb]
26. Finkelstein EA, Kruger E. Meta- and cost-effectiveness analysis of commercial weight loss strategies. Obesity (Silver Spring) 2014 Sep;22(9):1942-1951 [FREE Full text] [doi: 10.1002/oby.20824] [Medline: 24962106]

27. Finkelstein EA, Kruger E, Karnawat S. Cost-effectiveness analysis of Qsymia for weight loss. Pharmacoeconomics 2015 Jul;33(7):699-706 [FREE Full text] [doi: 10.1007/s40273-014-0182-6] [Medline: 24986038]
28. Braithwaite RS, Meltzer DO, King JT, Leslie D, Roberts MS. What does the value of modern medicine say about the $50,000 per quality-adjusted life-year decision rule? Med Care 2008 Apr;46(4):349-356. [doi: 10.1097/MLR.0b013e31815c31a7] [Medline: 18362813]
29. Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Rev Pharmacoecon Outcomes Res 2008 Apr;8(2):165-178. [doi: 10.1586/14737167.8.2.165] [Medline: 20528406]
30. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness--the curious resilience of the $50,000-per-QALY threshold. N Engl J Med 2014 Aug 28;371(9):796-797. [doi: 10.1056/NEJMp1405158] [Medline: 25162885]
31. Dresser R, editor. Malignant: medical ethicists confront cancer. New York: Oxford University Press; Mar 2, 2012.
32. Neumann PJ, Weinstein MC. Legislating against use of cost-effectiveness information. N Engl J Med 2010 Oct 14;363(16):1495-1497. [doi: 10.1056/NEJMp1007168] [Medline: 20942664]
33. Finkelstein EA, Verghese NR. Incremental cost-effectiveness of evidence-based non-surgical weight loss strategies. Clin Obes 2019 Apr;9(2):e12294. [doi: 10.1111/cob.12294] [Medline: 30677252]
34. Cawley J, Meyerhoefer C, Biener A, Hammer M, Wintfeld N. Savings in medical expenditures associated with reductions in body mass index among US adults with obesity, by diabetes status. Pharmacoeconomics 2015 Jul;33(7):707-722 [FREE Full text] [doi: 10.1007/s40273-014-0230-2] [Medline: 25381647]
35. WeightWatchers. 2018. OnlinePlus plan: online health and weight loss solution URL: https://www.weightwatchers.com/us/plans/onlineplus [accessed 2018-09-11] [WebCite Cache ID 72N9Om1eg]
36. Anderson JW, Konz EC, Frederich RC, Wood CL. Long-term weight-loss maintenance: a meta-analysis of US studies. Am J Clin Nutr 2001 Nov;74(5):579-584 [FREE Full text] [Medline: 11684524]
37. Lowe MR, Miller-Kovach K, Phelan S. Weight-loss maintenance in overweight individuals one to five years following successful completion of a commercial weight loss program. Int J Obes Relat Metab Disord 2001 Mar;25(3):325-331. [doi: 10.1038/sj.ijo.0801521] [Medline: 11319628]
38. Ware J, Kosinski M, Dewey J, Gandek B. SF-36 health survey: manualinterpretation guide. Lincoln, RI: Quality Metric Inc; 2000.

Abbreviations

BMI: body mass index
ICER: incremental cost-effectiveness ratio
QALY: quality-adjusted life year
QoL: quality of life
UC: usual care

Edited by G Eysenbach; submitted 17.09.18; peer-reviewed by J Most, T Powell-Wiley, K Tamura; comments to author 19.02.19; revised version received 03.04.19; accepted 08.04.19; published 17.05.19.

Please cite as:
Krishnan A, Finkelstein EA, Levine E, Foley P, Askew S, Steinberg D, Bennett GG
A Digital Behavioral Weight Gain Prevention Intervention in Primary Care Practice: Cost and Cost-Effectiveness Analysis
J Med Internet Res 2019;21(5):e12201
URL: http://www.jmir.org/2019/5/e12201/
doi:10.2196/12201
PMID:31102373

Review

Theories Predicting End-User Acceptance of Telemedicine Use: Systematic Review

Lorenz Harst[1*], BA, MA; Hendrikje Lantzsch[2*], BA; Madlen Scheibe[3], Dipl Soc, Dr Rer Medic

[1]Research Association Public Health, Center of Evidence-based Healthcare, University Clinic Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[2]Master Program Health Sciences / Public Health, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[3]Center for Evidence-Based Healthcare, University Clinic Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[*]these authors contributed equally

Corresponding Author:
Lorenz Harst, BA, MA
Research Association Public Health, Center of Evidence-based Healthcare
University Clinic Carl Gustav Carus
Technische Universität Dresden
House 106
Löscherstraße 18
Dresden, 01307
Germany
Phone: 49 351 3177 ext 223
Fax: 49 351 3177170
Email: lorenz.harst@tu-dresden.de

Abstract

Background: Only a few telemedicine applications have made their way into regular care. One reason is the lack of acceptance of telemedicine by potential end users.

Objective: The aim of this systematic review was to identify theoretical predictors that influence the acceptance of telemedicine.

Methods: An electronic search was conducted in PubMed and PsycINFO in June 2018 and supplemented by a hand search. Articles were identified using predefined inclusion and exclusion criteria. In total, two reviewers independently assessed the title, abstract, and full-text screening and then individually performed a quality assessment of all included studies.

Results: Out of 5917 potentially relevant titles (duplicates excluded), 24 studies were included. The Axis Tool for quality assessment of cross-sectional studies revealed a high risk of bias for all studies except for one study. The most commonly used models were the Technology Acceptance Model (n=11) and the Unified Theory of Acceptance and Use of Technology (n=9). The main significant predictors of acceptance were perceived usefulness (n=11), social influences (n=6), and attitude (n=6). The results show a superiority of technology acceptance versus original behavioral models.

Conclusions: The main finding of this review is the applicability of technology acceptance models and theories on telemedicine adoption. Characteristics of the technology, such as its usefulness, as well as attributes of the individual, such as his or her need for social support, inform end-user acceptance. Therefore, in the future, requirements of the target group and the group's social environment should already be taken into account when planning telemedicine applications. The results support the importance of theory-guided user-centered design approaches to telemedicine development.

(J Med Internet Res 2019;21(5):e13117) doi:10.2196/13117

KEYWORDS

systematic review; telemedicine; technology; patient compliance

Introduction

Definition and Delimitation of Telemedicine and Related Terms

Telemedicine, as well as every other digital health technology, comes with the promise of changing care delivery for the better, be it by reaching traditionally underserved regions [1] or populations [2] or by enhancing patient-provider communication to facilitate shared decision making [3].

Digital health care is known under various terms, for example, telemedicine, electronic health (eHealth), telehealth, or, as new digital devices came to be used, mobile health (mHealth). This study focused on telemedicine, as telemedicine involves health care services being delivered by health care providers in a patient-centered manner, from a geographical distance, using ICT (Information and Communication Technology) [4]. The term is hereby clearly delimitated from other modes of digital care delivery. eHealth, for example, also encompasses electronic management of patient data, whereas telehealth covers the use of Internet of Things to enable self-management of health and the quantified self [5].

Acceptance and Diffusion of Telemedicine

The possibly high potential of telemedicine can only be fulfilled if telemedicine reaches a high diffusion throughout the health care system. Nevertheless, although being used for over 50 years now [6], telemedicine still does not mostly overcome the pilot project stage [7] and therefore never prevails in regular care [8,9]. Among the most commonly used applications are those enabling digital data storage and exchange or telecounseling for diagnostic purposes [10,11]. Those, however, are not covered by the definition of telemedicine, as care is not directly delivered to the patient. Therefore, patient acceptance is not a relevant factor for those applications.

Rogers argues that innovations that are compatible with their environment are more easily adopted than those that are not suitable to past experiences of the adoption units [12]. However, telemedicine entails changes in patient-provider communication, patient assessment, and engagement [13,14]. As Riley et al argue, many of the applications deploy so-called *cues to action*, demanding behavior change by the end user [15].

Therefore, it is necessary to study the factors influencing end-user acceptance of telemedicine, even more so as acceptance is a prerequisite for adoption of an innovation and therefore its diffusion [12]. Nevertheless, acceptance is often reduced to the study of the usability, that is, certain design features of technology [16]. For the implementation of Health Information Systems, Ifinedo finds factors influencing Canadian nurses' acceptance to go way beyond technical features [17]. Therefore, Hastall et al call for a holistic assessment of acceptance, incorporating not only features of the technology but also characteristics of the end user [18]. When structuring this holistic approach by studying individual, social, environmental, and technological factors of acceptance they recur on behavioral models of acceptance that have been proven to be effective in predicting behavior change in all sorts of health interventions—for example, by Sahay et al [19].

Theoretical Background of Technology Acceptance

The oldest yet still widely used model for health behavior change is the *Health Belief Model (HBM)* that focuses on the individual assessment of vulnerability, outcomes and costs of behavior change, and external cues toward behavior change [20]. It can be applied, in combination with other theories, to explain user's acceptance of wearables [21], which suggests transferability to telemedicine.

On the basis of the focus of the HBM on individual perceptions, the *Theory of Reasoned Action (TRA)* was formulated, also focusing on the attitude toward the intended outcome behavior but adding measures of subjective norm, that is, the perception of the behavior in question by those whose opinion is valued by the individual [22]. TRA is applicable for predicting attitude toward the use of a teleconsultation system in neurology [23]. Perceived control over one's own health was added to the theory later on, developing the *Theory of Planned Behavior (TPB)* [24]. The TPB is applicable to the use of fitness apps [25]. TRA and TPB constructs were used by Davis to explain acceptance of technology as a precondition for its use. *The Technology Acceptance Model (TAM)*, along with several additions, defines use as predicted by attitude toward the use, which again is a function of perceived usefulness and perceived ease of use, both being value judgements of the design features [26]. The TAM was further developed by Venkatesh, adding several components from previous behavioral theories, such as performance and outcome expectancy from the Social Cognitive Theory [27] and subjective norm from TRA and TPB [28]. His final model is called the *Unified Theory of Acceptance and Use of Technology (UTAUT)*. As relevant predictors of acceptance vary between these 2 models, so do the definitions of technology acceptance within the models. Although in the TAM, acceptance is defined as "actual system use" [26], in the UTAUT, it is defined as "use behavior" [28]. Nevertheless, both models aim to study acceptance. Both TAM and UTAUT have been used excessively to explain the use of several digital health applications, such as data sharing systems [29] and assessment tools for cognitive functions [30].

In contrast to the aforementioned theories, where acceptance or use is a dependent variable, Normalization Process Theory postulates that collective action—that is, for the purpose of this paper, technology use—is one of several highly interconnected variables. Among them are group processes and organizing structures [31].

In addition to the predictors derived from the theories and models, other variables may have the potential to determine individual peoples' willingness to use telemedicine. The nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework lists several possible domains challenging the implementation process of telemedicine. The individual end user, on whom this study focuses, according to the NASSS framework, interacts with the organizational and societal context and is constrained by his or her medical status, as well as technological features [32]. The authors of the NASSS framework conclude that there is a lack of theoretical foundation for individual adoption processes. Such, the UTAUT is especially suitable to close this research gap, as it also encompasses the organizational and technological infrastructure in which the individual acceptance unit lives [28].

There is scarce evidence on which theory or model of technology acceptance or health behavior change is best suited to explain the acceptance and therefore use of telemedicine, as defined above [33,34]. Lai gives a comprehensive review on existing technology acceptance models and theories [35]. Evidence synthesis of theoretical predictors can be found solely for the acceptance of health information systems, such as eHealth records [36], yet it is not found for telemedicine defined as narrow as it is defined by Sood et al [4].

Individual studies that focus on theoretical components as predictors of the acceptance of telemedicine exist. However, there is still no systematic overview of theoretical components that are able to empirically explain the acceptance of telemedicine. This study aimed to fill this void by answering the following research question:

> *Which theoretical components are empirically associated with end-user acceptance of telemedicine?*

Methods

Design

This systematic review was conducted according to the standardized strategy provided by the Cochrane Collaboration [37], and it also follows the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) checklist [38]. A review protocol was created a priori and published at Prospero (Number CRD42018098658).

Inclusion and Exclusion Criteria

The Population, Intervention, Comparison, Outcome, and Study Design (PICOS) criteria were used for deriving the inclusion and exclusion criteria for the review [37]. As this review aimed to explain the acceptance of telemedicine, and not the effectiveness, the comparison was omitted. The population studied included patients and health care providers as well as their respective direct social environment. As the term *telemedicine* is not consistently used, it cannot be clearly delimitated from related terms. Therefore, terms such as eHealth and mHealth were also part of the search strategy. Studies examining telemedicine that was used to deliver health care services in a patient-centered manner over a geographical distance were included. If this was not the case, those studies were excluded during the full-text assessment at the latest. Only studies aiming to explain end-user acceptance as a primary outcome using a theoretical underpinning were included. As the explanatory power of theoretical components was to be identified, only quantitative research designs were included. The same is true for studies published in English and German and for studies published in peer-reviewed journals. A further specification of the inclusion and exclusion criteria can be found in Table 1.

Literature Search

The search string was a combination of the building blocks of the PICOS. Electronic searches were conducted in PubMed, as it is the most important and conclusive database for medical research, and PsycINFO in June 2018. PsycINFO was chosen as it is a database for psychological research, and it was therefore deemed likely to list studies featuring theoretical foundations for technology acceptance. For PubMed, Medical Subject Headings terms were used, whereas for PsycINFO, the functional equivalent, that is, the Thesaurus, was used. The term telemedicine is used ambiguously and partly synonymously with others [5]. As there are applications that are not explicitly called telemedicine but meet the telemedicine definition [39], different digital health terms, such as eHealth, mHealth, and telehealth, were included in the search string. The database-specific search strings can be found in the appendix (see Multimedia Appendix 1).

A hand search was conducted. In addition to a forward search in Web of Science, major publications in the field of telemedicine were searched, such as *Telemedicine and eHealth* and the *Journal of Medical Internet Research*. In addition, a search took place in the Institute of Electrical and Electronics Engineers Xplore database, which appears not to be designed for comprehensive search strings. To cover the research fields of informatics and information systems, a hand search was also conducted in the journals *Management Information Systems Quarterly* and the *American Journal of Information Systems*. References of the included studies were assessed to identify landmark studies, that is, those cited by more than one of the included papers.

Table 1. Inclusion and exclusion criteria for the review according to the Population, Intervention, Outcome, and Study Design scheme.

Category	Inclusion criteria	Exclusion criteria
Population	Patients, social environment (relatives and peers or peer groups), and health care providers	Nonhuman populations, not patients, not health care providers, and veterinarians
Intervention	Telemedicine-delivered patient-centered health care services with involvement of health care providers	No telemedicine, that is, no patient-centered health care services delivered, no involvement of health care providers
Outcome	Acceptance of health technologies on the basis of theoretical components	No theory-based factors (derived from correlations, causal models, eg, multivariate regression analyses or Structural Equation Modeling or effect strengths calculated from group comparisons), no statements about acceptance, and theories
Study design	Intervention studies (randomized or nonrandomized controlled trials), observational studies (cohort studies, cross-sectional studies, and case-control studies), and studies published in English or German language	Qualitative studies (in-depth interviews, expert interviews, focus groups, and delphi), reviews, editorials, letters to the editor, studies not published in English or German, or not published in peer-reviewed journals

Identification and Selection of Studies

Predefined inclusion criteria were applied independently by 2 raters (LH and HL) to screen for potentially relevant titles and abstracts within all studies obtained from the database and the hand search. In a next step, all possibly eligible studies were subsequently screened as full texts, also by 2 independent reviewers (LH and HL). Articles that did not meet the aforementioned inclusion criteria were excluded. Each of the 2 reviewers documented the reasons for exclusion so that a direct comparison was possible and a transparent procedure was ensured. Any disagreement over the suitability of certain studies was discussed among the raters and resolved by consensus.

Data Extraction and Presentation

Study characteristics (bibliographical information, study design, study population, type of telemedicine application, theoretical model or theory, and statistical methods) were extracted independently by the 2 reviewers (LH and HL). *Statistical methods* encompassed the dependent variable, significant predictors (rooted within theory), and measures for internal consistency and reliability as reported by the authors of the included studies, as well as the statistical analysis conducted. Disagreements were discussed, and a consensus between both extractors was reached. The entire extraction table was discussed by all authors (LH, HL, and MS). Only those predictors rooted within acceptance theories or models were extracted, which eliminated those added by the authors of the included studies to increase the variance explained. Both restrictions are in line with the research question.

For all theories discovered, as well as all the significant predictors, first frequencies and then variances explained and effect strengths were presented in tables. Afterward, median variance that was explained by each study, and median effect strengths of the theoretical predictors used within each study were also calculated and presented in tables. The median is a proper measure, as it is “the middle score of a set of ordered observations” [40]. Therefore, all statistical values for variance explained (R^2) and each predictor (odds ratios [ORs] and betas) were listed and the middle value was either discernible (in an uneven list of values) or calculated as the arithmetic mean of the 2 middle values (in an even list of values) [40].

Assessment of Methodological Quality

Quality assessment was conducted using the Appraisal Tool for Cross-Sectional Studies (AXIS Tool) for quantitative studies [41]. Quality assessment was conducted by 2 authors (LH and HL) independently. The focus of the assessment procedure was on methodological issues, such as the selection of study participants, the reliability and validity of the outcome measurements, and the consideration of potential confounding factors and bias in the results. On the basis of these criteria, an evaluation was carried out on a 2-step scale from 0=not satisfyingly explained to 1=satisfyingly explained. If items concerning methods and results (as explained above) were all rated 1, the risk of bias within the results of the study was deemed *low*, otherwise, it was deemed *high*.

Results

Search Results

The electronic database search resulted in 6188 potentially eligible articles. A hand search resulted in a total of 13 additional studies. After removing 283 duplicates, 5821 articles were excluded by independent screening of titles and abstracts. Of the resulting 97 full texts, 73 were excluded, as they did not meet the inclusion criteria. The main reason for study exclusion was the use of an inadequate intervention, that is, the intervention studied did not fall within the definition of telemedicine (36 times). For studies where the full text was not available, the authors were contacted. Owing to nonresponse, 7 full texts could not be procured until the end of August 2018.

Finally, 24 papers met the predefined inclusion criteria and formed the basis for data extraction. A total of 20 of the papers resulted from the application of the search strings in PubMed and PsycINFO, whereas 4 additional ones were uncovered by hand search. A total of 3 of the 4 were found in relevant journals in the field of telemedicine and information systems research and by checking references of the included studies for landmark studies, the third relevant study was included. The PRISMA flow chart in Figure 1 shows the process of the study selection.

A list of studies excluded during full-text screening, complete with the reasons for exclusion, can be found in the appendix (see Multimedia Appendix 2). Table 2 shows the most important data extracted for each study. For a list of all extracted data, please see Multimedia Appendix 3.

Figure 1. Flow chart of studies included and excluded from the systematic review.

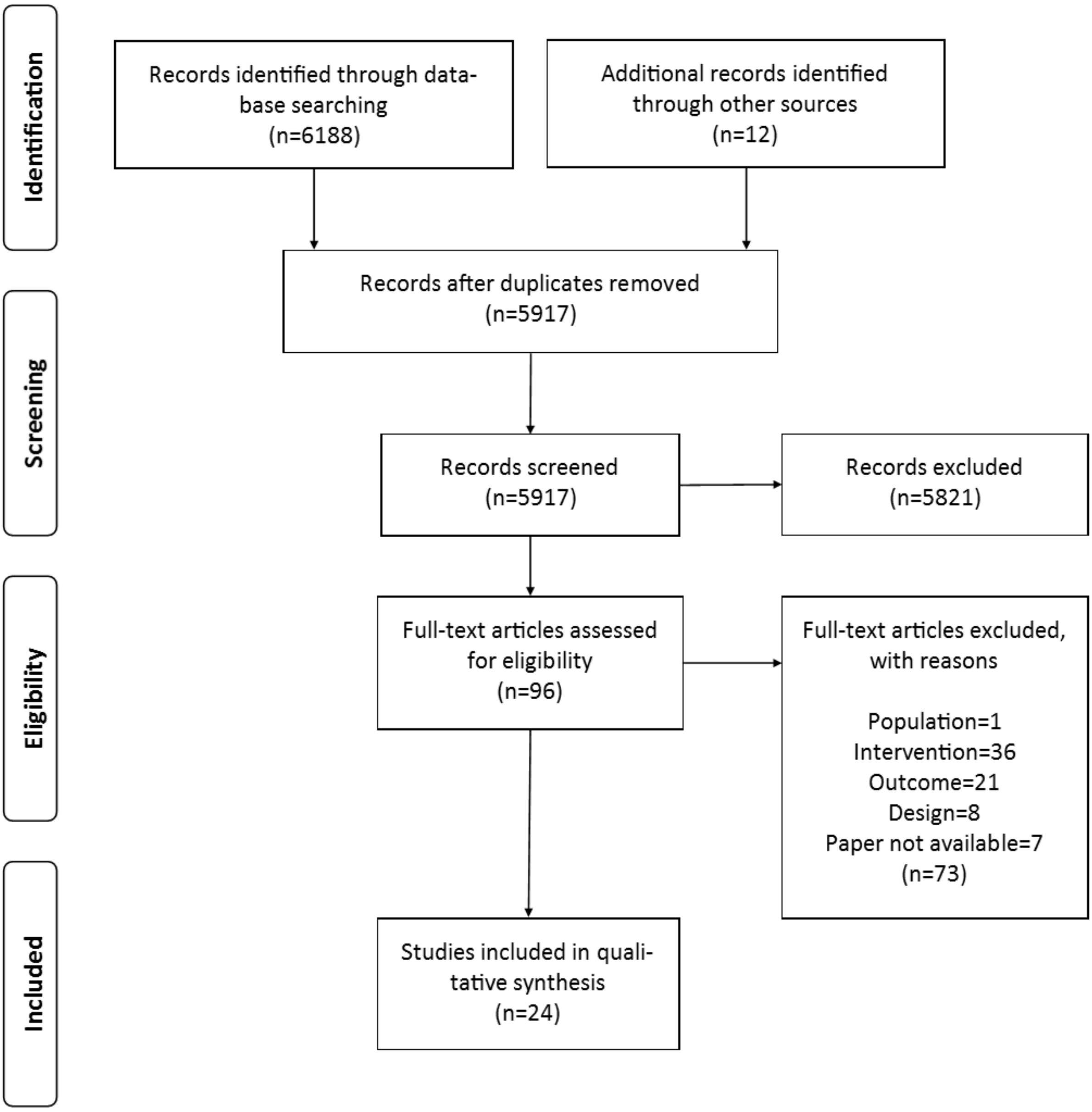


Table 2. Characteristics and outcomes of all included studies.

Author (year); Journal	Theoretical model or theory	Components of the model or theory with significant explanatory power	Effect strength and significance[a]
Health care provider			
Asua et al (2012); BMC Medical Informatics and Decision Making	TAM[b]; DOI[c]; TIB[d]	PU[e] (TAM); PEOU[f] (TAM); *Extended TAM:* Compatibility (DOI); Facilitators (TIB)	*Original Technology Acceptance Model:* PU: OR[g] 5.28[h] (95% CI 3.14 to 10.01); PEOU: OR 1.93[i] (95% CI 1.11 to 2.37); Final Model: Nagelkerke R^2=0.63; *Extended TAM:* PU: OR 2.65[i] (95% CI 1.15 to 6.12); Compatibility: OR 3.06[j] (95% CI 1.30 to 7.18); Facilitators: OR 4.90[h] (95% CI 2.38 to 10.09); Final model: Nagelkerke R^2=0.72
Gagnon et al (2012); Telemedicine and e-health	TAM	PU (TAM); *Modified model:* Facilitators (TIB)	*Original Technology Acceptance Model*:; PU: OR 5.28[h] (95% CI 2.12 to 13.11); Nagelkerke R^2=0.42; *Modified model:* Facilitators: OR 4.96[j] (95% CI 1.59 to 15.55); Final model: Nagelkerke R^2=0.54
Hennemann et al (2017); Journal of Health Communication	UTAUT[k]	SI[l]; PE[m]	SI: beta=.37[h] (95% CI 0.25 to 0.61); PE: beta=.28[h] (95% CI 0.12 to 0.44); Final model: R^2=0.63
James et al (2016); Journal of Diabetes Science and Technology	TAM	PEOU; SN[n]	Independent Predictors of Diabetes Educators' Intentions to Use; *Apps*: PEOU: OR 1.15[i] (95% CI 1.07 to 1.31); Final model: R^2=0.71; *Video conferencing*:; PEOU: OR 1.21[h] (95% CI 1.08 to 1.35); SN: OR 1.21[j] (95% CI 1.07 to 1.37); Final model: R^2=0.68
Kuhn et al (2015); Professional Psychology: Research and Practice	DOI	Complexity	Complexity: OR .35[h] (95% CI 0.23 to 0.55); Final model: Nagelkerke R^2=0.53
Orruño et al (2011); Journal of Telemedicine and Telecare	TAM; TIB; TRA[o]	PU (TAM); PEOU (TAM); Facilitators (TIB)	*Original Technology Acceptance Model*; PU on intention: OR 8.4,[h] (95% CI 3.4 to 21.0); PEOU on intention: OR 7.4,[h] (95% CI 2.9 to 19.0); Nagelkerke R^2=0.71; *Modified Technology Acceptance Mode:* Facilitators on intention: OR 9.9[h] (95% CI 2.80 to 34.94); Final model: Nagelkerke R^2=0.78
Saigi-Rubió et al (2014); Implementation Science	TAM; DOI; TRA; TPB; TR[p]	Level of ICT use (TR); Optimism (TR)	Technology Readiness Index: Level of ICT Use (Spain): b=2.661[j]; Level of ICT Use (Columbia): b=1.212[j]; Optimism (Bolivia): b=0.484[h]; Final model: Nagelkerke R^2 (Spain): 0.275; Nagelkerke R^2 (Columbia): 0.161; Nagelkerke R^2 (Bolivia): 0.197
Saigi-Rubió et al (2016); International Journal of Technology Assessment in Health Care	TAM; TPB; TRA	PU (cost reduction, quality of care; TAM); ATT[q] (confidentiality and security; TAM); SN[r] (patients, medical staff; TRA)	PU (cost reduction) on BI[s]: b=1.342[i]; ATT (security and confidentiality) on BI: b=0.798[i]; SN (patients) on BI: b=.583[j]; SN (medical staff) on BI: b=1.005[j]; Moderations: SN (patients)xPU (quality of care) on BI: b=.347[j]; SN (patients)xPU (cost reduction) on BI: b=.462[i]; SN (medical stuff)xPU (quality of care) on BI: b=.366[i]; SN (medical stuff)xPU (cost reduction) on BI: beta=.488[i]; SN (administration)xPU (cost reduction) on BI: beta=.571[i]; Final model: Nagelkerke R^2=0.481; CI NS[t]
Spaulding et al (2005); Journal of Telemedicine and Telecare	DOI	Relative advantage (provider); Relative advantage (patient); Observability; Trialability; Opinion leader present	Relative advantage (provider): r=0.42[i]; Relative advantage (patient): r=0.42[i]; Observability: r=0.57[i]; Trialability: r=0.44[i]; Opinion leader present: r=0.52[i]; CI NS
van Houwelingen et al (2015); Journal of Gerontological Nursing	UTAUT	PU; EE[u]; SI	PU: beta=.435[h]; EE: beta=.28[h]; SI: beta=.216[i]; Final model: R^2=0.54; CI NS

Author (year); Journal	Theoretical model or theory	Components of the model or theory with significant explanatory power	Effect strength and significance[a]
Vanneste et al (2013); BMC Medical Informatics and Decision Making	UTAUT; SCT[v]	FC[w] (UTAUT); SE (SCT)	FC: beta=.287[h]; SE: beta=.218[h]; Final model: R^2=0.308; CI NS
Zhang et al (2010); Computers, Informatics, Nursing	TAM 2	SN; IM[x]; PEOU; PU	SN: beta=.323[j]; IM: beta=.227[j]; PEOU: beta=.35[j]; PU: beta=.422[h]; R^2=0.375; CI NS
Patients			
Cajita et al (2017); Journal of Cardiovascular Nursing	TAM; TIB	PEOU (TAM); PU (TAM)	*Block 5:* change in R^2=0.095[h]; PEOU: beta=.16[h] (95% CI 0.07 to 0.24); *Block 6:* change in R^2=0.130[h]; PU: beta=.33[h] (95% CI 0.24 to 0.41), Final model: R^2=0.353
de Veer et al (2015); BMC Health Services Research	UTAUT	PE; EE; SE	*PE:* Block 2: beta=.52[h]; Block 3: beta=.24[h]; Block 4: beta=.24[h]; Final model: beta=.24[h]; *EE:* Block 3: beta=.42[h]; Block 4: beta=.42[h]; Final model: beta=.35[h]; *SE:* Final model beta=.01[j]; Final model: R^2=0.41; CI NS
Dockweiler et al (2017); Gesundheitswesen	UTAUT	PE; EE	*PE* (5 significant variables); average effect: OR 11.325[i] (95% CI 2.666 to 49.015); *EE* (2 significant variables); average effect: OR 0.121[i] (95% CI 0.022 to 0.685); Final model: R^2= 0.765
Dou et al (2017); JMIR Mhealth Uhealth	TAM; TAM 2; Dual-Factor model; HBM[y]	PU (TAM); PHT[z] (HBM); Resistance to change (Dual-Factor Model)	PU on intention to use: beta=.616[j]; PHT on intention to use: beta=.305[j]; resistance to change on intention to use: beta=−.149[i]; Final model: R^2=0.412; CI NS
Hennemann et al (2016); Journal of Medical Internet Research	UTAUT	SI; PE; EE	SI: beta=.39[j] (95% CI 0.3 to 0.54); PE: beta=.31[h] (95% CI 0.19 to 0.43); EE: beta=.22[h] (95% CI 0.09 to 0.31); Final model: R^2=0.78
Hossain et al (2018); Telemedicine and e-Health	UTAUT; TAM	Social reference (means SI; UTAUT); ATT (TAM); FC (UTAUT)	SR: OR 9.73[j] (95% CI 4.16 to 22.78); ATT: OR 4.56[j] (95% CI 2.71 to 7.66); FC: OR 3.92[i] (95% CI 1.29 to 11.95); Final model: R^2=0.55
Huygens et al (2015); Interactive Journal of Medical Research	UTAUT	EE; PE; FC; SI; ATT	Service to ask questions by internet via email or a website: EE: OR 5.46 (95% CI 3.27 to 9.13); PE: OR 5.47 (95% CI 3.44 to 8.70); ATT: OR 5.85 (95% CI 3.63 to 9.43); FC: OR 7.91 (95% CI 4.53 to 13.82); SI: OR 4.34 (95% CI 2.46 to 7.68); No levels of significance reported
Lin and Yang (2009); Telemedicine and e-Health	TAM	PU; ATT; SN; PEOUxATT; PUxATT; SNxATT	*Direct effects:* ATTon BI (lambda=.76[j]); SN on BI (lambda=.16[i]); *Total effects*; ATT on BI (lambda=.76[j]); PU on BI (lambda=.62[j]); SN on BI (lambda=.42[j]); PEOU on BI (lambda=.3[j]); Final Model: R^2=0.8; CI NS
Peeters et al (2012); Journal of Clinical Nursing	DOI	Relative advantage; Compatibility; Complexity; Observability	Relative advantage: beta=.17[i]; Compatibility: beta=.2[j]; Complexity: beta=.19[j]; Observability: beta=.34[h]; Final model: R^2=0.61; CI NS
Rho et al (2015); Cluster Computing	UTAUT	PE; EE; SI; FC on EE; FC on PE	PE: beta=.345[j]; EE: beta=.227[i]; SI: beta=.246[i]; FCxPE on BI: beta=.176[j]; FCxEE on BI : beta=.153[j]; Final model: R^2=0.44; CI NS
Zhang et al (2017); Informatics for Health and Social Care	TAM; SCT; PMT	PU (TAM); PEOU (TAM)*PU; SE (Protection Motivation Theory)*PEOU*PU* AI; RE (Response Efficacy)*PEOU*PU*AI; RE (Protection Motivation Theory)*PEOU*PU* AI	*Direct effect:*; PU: beta=.3[h]; *Moderator:*; PUxSE: beta=.145[j]; PUxRE: beta=.359[h]; Final model: R^2=0.501; CI NS

Author (year); Journal	Theoretical model or theory	Components of the model or theory with significant explanatory power	Effect strength and significance[a]
Social environment			
Jen and Hung (2010); Telemedicine and e-Health	TPB; TAM	ATT; PU; PEOU	BI of adopting Mobile Health Services is explained directly by ATT; ATT on BI (beta=.547[j]); Final model: R²=0.641 of the variance in BI; CI NS

[a] Italics serve as subheadings for stepwise models.

[b] TAM: Technology Acceptance Model.

[c] DOI: Diffusion of Innovations Theory.

[d] TIB: Theory of Interpersonal Behavior.

[e] PU: Perceived Usefulness.

[f] PEOU: Perceived Ease of Use.

[g] OR: odds ratio.

[h] $P \leq .001$.

[i] $P \leq .05$.

[j] $P \leq .01$.

[k] UTAUT: Unified Theory of Acceptance and Use of Technology.

[l] SI: Social Influence.

[m] PE: Performance Expectancy.

[n] SN: Social Norm.

[o] TRA: Theory of Reasoned Action.

[p] TR: Technology Readiness.

[q] ATT: attitude.

[r] SN: Social Norm.

[s] BI: Behavioral Intention.

[t] NS: not specified.

[u] EE: Effort Expectancy.

[v] SCT: Social Cognitive Theory.

[w] FC: Facilitating Conditions.

[x] IM: Image.

[y] HBM: Health Belief Model.

[z] PHT: perceived health threat.

Study Characteristics

The publishing years ranged from 2005 to 2017. Among the included studies, the Netherlands (4 times) and Spain (4 times) stand out. Only 1 study comes from a developing country, that is, Bangladesh. The number of journals from different research fields shows the conclusiveness of the conducted search.

All studies were cross-sectional studies. The number of participants ranged from 84 to 1014 (mean n=266.25 (SD 210.07), median n=228). Accordingly, acceptance was tested for using inferential statistics in all cases. Only 1 study did not apply causal statistics but merely a correlation analysis [42]. Of the remaining 22, 7 conducted complex models of causality, Structural Equation Modeling (SEM) or Path Modeling. For the statistical analysis conducted within each study, see Multimedia Appendix 3.

Only 1 study relied completely on an existing, previously tested questionnaire for the applied theory or model [29]. All other studies applied self-developed questionnaires. However, a total of 3 studies did not specify how their questionnaire was tested for validity and reliability [42-44]. Of the remaining 20, 6 [29,45-49]—those using SEM—applied confirmatory factor analysis, whereas the others relied on Cronbach alpha statistics [29,45,47-62].

Participant Characteristics

The population is balanced regarding health care providers and patients, as both were studied 11 times. Only 1 study focused on the social environment, namely relatives, as the key population.

Only 8 of the included studies reported a mean age, ranging from 43.53 to 77.8 years. This led to an overall mean age of 53.72 (n=8) years. The overall mean age for patients was 61.93 years, whereas the overall mean age for health care providers was 47.1 years. The remaining 15 studies reported age groups with different cohorts, which makes it impossible to include them into calculations of an overall mean age.

The overall percentage of female participants was higher among the 10 studies focusing on health care providers (1 did not report on gender) than among the 11 studying patients’ acceptance (77.3 vs 41.4).

Telemedicine Applications

A wide variety of telemedicine applications were analyzed within the included studies. First, it should be noted that of the 24 studies, in 8 cases, the authors did not specify the type of telemedicine application they were studying, that is, they used generic terms such as eHealth [50-52] or telemedicine [42,43,45]. A total of 6 studies focused on mobile apps [46-48,53,54,63], whereas 3 were concerned with applications based on internet devices, such as Web-based aftercare [29,44,55]. A total of 2 applications were used for monitoring of disease parameters [56,57]. A total of 4 more applications targeted certain diseases [49,58-60].

It is also noteworthy that even if a certain telemedicine application was focused on, sometimes the application did not (yet) exist. Instead, certain features were shown to the participants, who were then asked to imagine whether they would be willing to use the would-be application [50,60].

Although it was intended beforehand to study acceptance relative to broader categories of telemedicine applications [39], the sometimes quite generic terms used in the included studies do not allow for such a nuanced analysis. For an overview of telemedicine applications studied, see Multimedia Appendix 3.

Medical Conditions

Along with the concrete telemedicine application, a target disease or medical condition for which telemedicine was supposed to be used was also not stated in 8 of the included studies [42-44,50-52,57,61]. A total of 8 studies dealt with the acceptance of telemedicine applications for chronic diseases, such as diabetes [45,58]. Furthermore, in 3 cases, mental health conditions were studied [54,55,60]. Among the remaining medical conditions targeted were skin lesions [59] and heart failure [53]. For an overview of medical conditions studied, see Multimedia Appendix 3.

Relevant Models and Predictors

Frequency of and Variance Explained by Theories and Models

The results are presented as follows: First, frequencies of the acceptance theories and models used within the 24 included studies are reported, along with their median variance explained, as calculated by the authors. Calculations were thought to be justified, as the theories and models studied proved applicable across a wide variety of telemedicine and medical conditions. Afterward, frequencies and median effect strengths of the predictors found to be significant by the included studies are presented. Throughout the Results section, a distinction regarding relevance of the theories and predictors will be made according to health care providers, patients, and their social environment (see Table 2).

As depicted in Table 3, the TAM is used 11 times within the included studies of this review, and therefore most often. The UTAUT, however, is used 9 times. It should be mentioned that the UTAUT was used 7 times, without any additional predictors from other models, whereas the same is true for the TAM only in 4 cases. Instead, it is used with a variety of other models, among them are the TRA [43,59] and the TPB [46]. The so-called TAM 2, an extension of the original TAM, [64] was used 2 times within the included studies.

Contradictory to that, the TAM still has the highest amount of variance explained among all the models included in this review: A median R^2 of 0.68 is achieved by the TAM, compared with an R^2 of 0.59 for the UTAUT. The TAM was used more often than the UTAUT, whether in combination or not.

The Diffusion of Innovations Theory by Rogers was used to explain acceptance far less (3 times), yet it still reaches a median R^2 of 0.57, which is, in part, because of the fact it was once used in combination with the TAM [56].

There was no other theory or model used alone, except for the TAM or the UTAUT, the Theory of Interpersonal Behavior (TIB) being one of the remaining, which was used in combination with others most often—3 times [53,56,59].

The most powerful combination of models is based on the TAM, adding components of the TIB and the TRA, with a variance explained of $R^2=0.78$ [59]. This R^2, however, is still lower than the one achieved by Lin and Yang, using only the TAM, which was 0.8 [49]. The only significant predictor added by the TIB, according to Orruño et al, is the presence of facilitators. The TRA does not add significant predictors at all.

The TAM was used to explain acceptance of health care providers 7 times and 3 times for the acceptance of patients. The median proportion of variance explained by the TAM was higher for health care providers ($R^2=0.63$) than for patients ($R^2=0.501$).

The UTAUT was used more often to explain acceptance of patients than of health care providers (5 vs 3 times). For patients, its median explanatory power is also higher ($R^2=0.55$) than for health care providers ($R^2=0.54$). For frequencies of the models and theories, as well as their median variance explained, see Tables 3 and 4. A complete list of combined models and theories can be found in the appendix (Multimedia Appendix 4).

Table 3. Frequency of theories and models used to explain acceptance.

Model/Theory	Frequency of use
Dual factor model	1
Health Belief Model	1
Protection Motivation Theory	1
Technology Readiness	1
Social Cognitive Theory	2
Technology Acceptance Model 2	2
Theory of Interpersonal Behavior	2
Theory of Planned Behavior	2
Theory of Reasoned Action	2
Diffusion of Innovations Theory	3
Unified Theory of Acceptance and Use of Technology	9
Technology Acceptance Model	11

Table 4. Median variance explained by each model alone (if theory or model was used alone).

Model/Theory and variance explained (per author)	Median variance explained
Technology Acceptance Model	
0.35 (Cajita et al)	0.68
0.42 (Gagnon et al)	0.68
0.63 (Asua et al)	0.68
0.68 (James et al)	0.68
0.71 (James et al)	0.68
0.71 (Orruño et al)	0.68
0.80 (Lin and Yang)	0.68
Unified Theory of Acceptance and Use of Technology	
0.41 (de Veer et al)	0.59
0.44 (Rho et al)	0.59
0.54 (van Houwelingen et al)	0.59
0.63 (Hennemann et al)	0.59
0.77 (Dockweiler et al)	0.59
0.78 (Hennemann et al)	0.59
Diffusion of Innovations Theory	
0.53 (Kuhn et al)	0.57
0.61 (Peeters et al)	0.57
Technology Acceptance Model 2	
0.38 (Zhang et al)	0.38

Frequency and Effect Strength of the Significant Predictors

UTAUT adds, among others, the predictor *social influence* to the basic TAM predictors. As a result, it accounts for the perception of an item of technology by others, whose opinion is valued by the individual. The predictor was uncovered as a significant predictor by 6 of the included studies and comprises the attitudes of colleagues, patients, or the direct social environment, such as families and friends, toward telemedicine [55,58,60]. The predictors *performance expectancy*, *effort expectancy*, and *facilitating conditions* (sometime just called *facilitators*), all part of the UTAUT, were used 6 times as well.

The predictor mentioned as significant most often (11 times) was *perceived usefulness*, which is not part of the UTAUT but of the original TAM. It reaches both high ORs (when logistic regression was performed) and high betas (when multiple linear regression was performed). From a patient perspective, usefulness is achieved when, for example, telemedicine use improves quality of life or makes the care process more convenient for the patient [48]. For health care providers, according to the studies included, usefulness is mainly associated with streamlining care processes, such as diagnosis and monitoring of disease parameters [56,59]. The other TAM predictor, *perceived ease of use*, was discovered to be significant 6 times. It mostly covers the degree of training it would take, both patients and health care professionals, to understand and learn how to use the telemedicine application in question [55,56], and it is sometimes used synonymously with effort expectation [45].

Attitude was a significant predictor in 6 cases. However, in some of the proposed SEMs, it was circumvented in favor of direct effects of perceived usefulness [46] and perceived ease of use [49]. Yet, *attitude* is the predictor with the highest beta, with regard to the studies using multiple linear regression (median beta=.76).

When studying both median ORs and median betas, only the height of the numbers is reasonably interpretable, as calculation of both values differs greatly. Then, *Perceived usefulness* is the most important predictor for acceptance by health care providers, with a median OR of 5.28 and a median beta of .43 (as calculated by the authors).

Taking into account both OR and beta, there are 2 almost equally important predictors for patient acceptance of telemedicine: *performance expectancy*, with a median OR of 8.4 and a median beta of .3, and *social influence*, with a median OR of 7.04 and a median beta of .25. Patients expect telemedicine to help them cope with their health problems and thereby improve their health [45,55].

For relatives, *attitude* toward an mHealth care service used to connect their elderly family members with health care providers is the most important predictor (beta=.55). Lin and Yang operationalize attitude as the willingness to use a telemedicine application, as it is considered the ideal solution for a given health problem [49]. As this was the only study focusing on the social environment of patients, no further analysis was conducted.

The dependent variables used in the 24 studies do not always fit those intended by the authors of the original model. *Intention to use* telemedicine (or synonyms such as *behavioral intention*) was used 19 times, probably as only 2 studies reported actual use of their application [54,61], the actual dependent variable in the TAM. *Acceptance* and *adoption* were each used once. For a complete list of frequencies and effect strengths of all predictors, see Table 5. The following predictors are not listed as they were each mentioned only once in the included studies: Perceived Health Threat (beta=.305), Resistance to Change (beta=.149), Trialability (r=0.44), Opinion Leader present (r=0.52), Image (beta=.227), Optimism (beta=.484).

Table 5. List of predictors of acceptance according to frequencies of use, odds ratios, betas, b's, and r's.

Factors affecting telemedicine acceptance	*P* value, median	n	Odds ratio, median	Beta/lambda, median	b, median	r, median
Perceived usefulness	.001	11	5.28	.43	1.34	—[a]
Performance expectancy	.001	6	8.4	.3	—	—
Perceived ease of use	.01	6	1.57	.26	—	—
Effort expectancy	.001	6	2.79	.25	—	—
Facilitating conditions/faciliators	.001	6	4.96	.29	—	—
Social influence	.01	6	7.04	.25	—	—
Attitude to use	.01	6	5.21	.76	—	—
Subjective norms	.01	5	1.21	.16	.58	—
Relative advantage	.05	3	—	.17	—	0.42
Compatibility	.01	2	3.06	.2	—	—
Complexity	.006	2	0.35	.19	—	—
Self-efficacy	.051	2	—	.01	.22	—
Observability	.026	2	—	.34	—	0.57
Level of ICT use	.018	2	—	—	1.94	—

[a] No data provided.

Methodological Quality Assessment

All but 1 study have a considerable risk for bias, as there is only 1 study in which all AXIS items from the Methods and Results section could be rated 1, which is the study on mHealth use intention of heart failure patients by Cajita et al [53]. It achieves 19 from a total of 20 points.

The objective of each study (24 times rated with 1), study design (24), sample size (24), as well as sample frame (22), was well described in most studies. Moreover, questions dealing with the measurement of outcome variables and the determination of statistical significance could always be rated 1.

Only a few studies have taken measures to address and categorize nonresponders. The evaluation of question 11—whether there was sufficient description of the study design and statistical methods applied to derive the results—shows that only about half of the studies have sufficiently described those, which is because of the fact that these studies have not reported any CIs or nonsignificant results. This diminishes the possibility to repeat their results. Moreover, half of the studies do not score high on questions 13 and 14, as their response rate is low, and there is no information provided about nonresponders, raising the question of nonresponse bias. Overall, the 24 included studies reached an AXIS score of 15.67. An overview of the rating for quality assessment according to the AXIS Tool can be found in the appendix (Multimedia Appendix 5).

Discussion

Importance of Acceptance Theories and Models in General

The results of the 24 original studies included in this systematic review support Hastall et al’s demands for a holistic analysis of technology acceptance in health care by relying on a theoretical background [18].

The UTAUT was used more often to explain acceptance for patients than for health care providers, which is likely because of the fact that the UTAUT includes variables of the construct, *social influence*. Those are also more important for patients than for providers. For health care providers, the TAM has the highest variance explained, relying also on the predictor *perceived usefulness*, which reaches both high ORs and betas.

It is noteworthy that, although it is still the most commonly used model, the TAM is combined with further predictors 7 times. Those predictors borrow heavily from the TIB, TPB, and TRA, thereby enabling the original TAM to also incorporate factors of acceptance not only rooted within technology but also rooted within the individual as the end user. The high prevalence of the UTAUT, used 9 times and only 2 times in combination, also adds to the importance of such factors, being a far more holistic model than the TAM. Together, these findings support Karsh’s statement that acceptance is not solely achieved by improving usability [16]. Apart from that, they fulfill, in parts, the demand articulated by Riley et al for development and validation of novel health behavior theories for mobile interventions. Such theories should, according to the authors, include features and attributes of the technology as well as characteristics of the end users [15].

Differences Between Patients and Health Care Professionals

Among the predictors added to the TAM, which are already part of the UTAUT, the most prevalent are those covering the social and organizational environment of the individual, for example, *social influence* and *facilitating conditions*. The latter supports results from qualitative observational research on telehealth readiness by older *patients* done by van Houwelingen et al. The authors conclude the need for easily available sources for technological support in case of problems with technology use [65]. *Social influence*, when phrased positively, can be understood as *social support*, and it was uncovered to be an important factor in technology acceptance, for example, for Web-based interventions for pregnant women by Berg et al [66]. Venkatesh et al included *social influences* into the UTAUT to pay respect to the fact that individuals’ acceptance behavior is being influenced by what they assume others might think about them when using a certain technology [28]. The results presented here show this is also true for telemedicine. However, they contradict those presented by Boessen et al, who state that intrinsic motivation to use a self-management tool trumps the perception of others, which serves as extrinsic motivation [67]. However, Peeters et al show that for people living alone, the positive effects of telemedicine use are much more observable (as defined by Diffusion of Innovations Theory), and so they were also more willing to adopt home telecare than those living with a partner or relative [62]. Although these results seem contradictory to the importance of social influence, they give another meaning to the concept, showing that telemedicine can, when used properly, provide social contact where there is none.

Social influence appears to be more important for patients than for *health care providers*. For the latter, *perceived usefulness* is the most important predictor (studying median ORs and betas), which is in line with the results provided by Mothuy-Blanc et al. They show that whether psychotherapists are willing to use telepsychotherapy is predicted only by whether they find it useful [68]. As the provision of the best treatment to the patient is every health care provider’s primary concern, these results are not surprising. This can also explain why the TAM, focusing on perceived usefulness and perceived ease of use, is much more important for health care providers than for patients.

Performance expectancy is an important predictor for both patients’ and providers’ acceptance. This makes their individual expectations toward the outcome of telemedicine use an important focus for further research, even though it is contradictory to the results of Koivumäki et al. They found no significant influence of performance expectancy on the adoption of digital preventive services [69].

The data presented here show the importance of easy-to-use applications, as perceived ease of use was shown to be a significant predictor 6 times, as well as effort expectancy. Scheibe et al show that design features, such as simple, intuitive menus, large icons and high color contrasts, are especially important for older users [70]. As time is always scarce in health care provision, easy-to-use technology is also important for health care providers. De Angelis et al can even show that health care providers are willing to disseminate health information via Facebook, mainly as the Social Network is easy to use [71].

Practicability of Technology Acceptance Models and Theories

On a more general level, the results presented here show that technology acceptance theories, as well as their basic behavioristic underpinnings, are applicable to the study of acceptance of telemedicine, even though they are quite old (the TAM was formulated in 1989, and the TRA was formulated in 1975). This is true despite the fact that none of those models and theories were originally formulated to fit health care technologies.

Moreover, the theories and models analyzed here are applicable to health care providers and patients alike, not to mention special types of diseases or telemedicine applications. There seems to be no special diagnosis that impacts acceptance and its preconditions, although it is noteworthy that there were not any studies focusing on telemedicine for patients with a cognitive impairment applicable for analysis. A pooling of data to calculate medians therefore seemed feasible.

Although an analysis of acceptance regarding different types of telemedicine applications was intended, only a superficial count of applications studied was feasible because of a lack of specification in some studies. However, no matter what kind of application was studied, the theories and models are applicable. If any insights can be gained from the few studies analyzing acceptance of a certain type of application, it is that acceptance is less of an issue when the basic device used is already familiar to the end user from everyday life, such as mobile phones or Web portals [72]. The results provided by Saigí-Rubió et al in their study from 2014 further stress this point, as they find previous use of ICT in their everyday life to be a significant predictor for physicians’ telemedicine use [73].

The applicability of the UTAUT, according to the variance explained, stresses the importance of holistic models of acceptance, incorporating not only characteristics of the individual adopter but also of his or her direct or indirect social environment. Facilitating conditions, being mentioned 6 times as relevant predictors, refer to the technological as well as the organizational infrastructure fostering acceptance. In a qualitative study by Cimperman et al, cost of the technology in use was mentioned as a major concern of older adults [74], which may appeal to funding agencies, such as insurance companies, to provide financial support.

It should be noted in this context that, although standardized and well-tested questionnaires exist for the TAM [26], as well as for the UTAUT [28] and the Diffusion of Innovations Theory [75], only 1 author cited here fully relied on the UTAUT questions suggested by Venkatesh et al in the original publication [29]. The remaining authors either made additions or changed the wording of several items. As none of the above mentioned models—nor the questionnaires used to test them—were originally developed for health care technologies, additions and changes to the questionnaires seem logical.

From a purely scientific standpoint, this research is proof that testing of theories and models in a variety of settings can be done by applying a systematic review of empirical studies, that is, by solely relying on secondary data.

Limitations

This study has several limitations. First, papers not published in English or German were excluded, which may constitute a selection bias (language bias). Although a comprehensive search strategy has been used, and an additional hand search was conducted, it is possible that some relevant studies were missed if the specific keywords were used neither by the authors nor by the databases searched. Apart from that, a publication bias toward positive results cannot be precluded.

All included studies were cross-sectional studies. It should be noted that acceptance or the decision to adopt an innovation is a dynamic process, taking place over time [12], especially when health behavior change is going hand in hand with it [76]. Such processes cannot be captured with a cross-sectional study that, by nature, only covers 1 point in time [62].

Another limitation might be that acceptance cannot be fully evaluated regarding different medical conditions, as these were not often indicated in the included studies. The same is true for concrete applications. Acceptance is most likely rated differently when the subject is given the chance to use a real application instead of having to rate acceptance of hypothetical applications or generic terms such as eHealth or telemedicine.

Furthermore, age and gender could be confounders in the interpretation of the results, as they differ in the individual studies. However, because of questionnaires tailored specifically for each research interest in each included study, the theories, models, and predictors are applicable despite such heterogeneity in demographic variables.

In terms of quality assessment, the AXIS tool does not provide a numerical scale for assessing quality of the studies. Thus, a subjective rating, depending on the research interest, is required. Although the authors of the tool state that this subjectivity provides greater flexibility in assessing the quality of a study [41], this can still be a limitation.

Owing to the heterogeneity across the statistical methods within the included studies, a meta-analysis could not be performed. The medians, as well as the SDs calculated here, are an approximation, based on the overall applicability of the theories and models. It shows that the theories and models, as well as the predictors, can be compared.

The review strictly followed the PRISMA. We deployed a concise and literature-based search string, which was critically reviewed by several peers. The methodology applied for this review was checked for validity by the PROSPERO foundation, where a protocol was registered beforehand. A thorough quality assessment with a focus on methodological issues was conducted for each included study.

Further Research Needs

As demonstrated in the review, social influence, also called social support, represents an important factor in telemedicine acceptance. The acceptance of telemedicine may benefit from the support of others perceived as important by the unit of adoption. In addition, telemedicine should be perceived as useful by the users or those who recommend them. All in all, models that are explicitly suited for health care technologies, such as telemedicine, need to be developed and empirically tested. The Health Information Technology Acceptance Model proposed by Kim et al in 2012 can serve as suitable model [77]. Even though it is based on existing technology acceptance and health behavioral theories, it was used in none of the studies found in this review.

This review highlights a lack of methodologically adequate studies. Future studies should have a longitudinal design and should consider the dimension of time, to allow for measuring the influence of mid- and long-term use of an innovation [78]. An important issue for further research in theory-based approaches to the measurement of acceptance is that the studies should examine acceptance of real telemedicine applications.

Complex, interdependent interactions within an organizational setting should also be tested for. The Normalization Process Theory can be helpful in this endeavor, as it proposes a nonlinear understanding of acceptance [31]. Theoretical factors promoting the implementation of telemedicine within a whole health care organization are not covered by the primary aim of this review. Yet, information about them can be derived, insofar as they can be subsumed under facilitating conditions. Little information was also found on predictors of telemedicine acceptance for relatives and peers of the primary users. Applying the NASSS framework, Greenhalgh et al find the complex interactions among the 6 levels, especially the organizations and the wider policy system, to be severely hindering telemedicine implementation [79]. The role of a supportive policy system as a predictor for telemedicine acceptance in a health care organization has been shown to be especially important in developing countries by Zailani et al. In addition, the authors stress the importance of the existing health culture as a mediator between individual technology assessment and telemedicine acceptance [80]. A more thorough investigation of the role played by the social environment of the end user, especially relatives, should also be conducted. Finally, the research presented here stops with the acceptance of telemedicine. However, another prerequisite of sustained use is task performance, that is, the ability to successfully use a technology [12]. Little research has been done on task performance in telemedicine use. However, Serrano and Karahanna have shown, on the basis of the task-individual-technology-fit theory, that skills in acquiring knowledge, problem-solving, and presenting solutions influence the successful use of teleconsulting systems [81]. Taking these results as a starting point, a shift from acceptance to performance research is feasible.

Conclusions

The results of this systematic review indicate that acceptance of telemedicine can be examined by using technology acceptance theories and models. On the basis of the included studies, acceptance was most often predicted by perceived usefulness, social influences, and attitude. To examine how adoption processes evolve over time, longitudinal research on existing applications would be advisable in the future. A brief summary of the study results can be found in Textbox 1.

Textbox 1. Brief summary of the study results.

What this study adds:

- Theories of technology acceptance are superior to common behavioral theories in explaining telemedicine acceptance.
- Not only features of the technology but also individual characteristics of the end user have to be considered when designing user-centered telemedicine.
- For patients, telemedicine acceptance of their social environment is crucial, as friends and families can support uptake of telemedicine use.
- For health care providers, usefulness of telemedicine in their clinical practice is of vital importance.

Acknowledgments

The authors wish to thank Stefanie Deckert, Patrick Timpel, Brita Sedlmayr and Hannes Schlieter for their valuable input to both the design of the study and the manuscript. The work for this review was partly funded by the European Social Fund and the Free State of Saxony (Grant no. 100310385).

Authors' Contributions

LH and HL participated in the design of the study and also conducted the search. LH, HL, and MS were responsible for critical evaluation, analysis, and presentation of the results. LH, HL, and MS drafted the manuscript, critically evaluated the article, and gave their final approval before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database-specific search strings.
[PDF File (Adobe PDF File), 31 KB - jmir_v21i5e13117_app1.pdf]

Multimedia Appendix 2
Excluded full texts with reasons for exclusion.

[PDF File (Adobe PDF File), 85 KB - jmir_v21i5e13117_app2.pdf]

Multimedia Appendix 3
Complete list of data extracted from all 24 included studies.
[PDF File (Adobe PDF File), 105 KB - jmir_v21i5e13117_app3.pdf]

Multimedia Appendix 4
List of models and theories used in combination.
[PDF File (Adobe PDF File), 26 KB - jmir_v21i5e13117_app4.pdf]

Multimedia Appendix 5
Results of quality assessment according to the Appraisal Tool for Cross-Sectional Studies.
[PDF File (Adobe PDF File), 89 KB - jmir_v21i5e13117_app5.pdf]

References

1. Jang-Jaccard J, Nepal S, Alem L, Li J. Barriers for delivering telehealth in rural australia: a review based on Australian trials and studies. Telemed J E Health 2014 May;20(5):496-504. [doi: 10.1089/tmj.2013.0189] [Medline: 24801522]
2. Akinlotan M, Bolin JN, Helduser J, Ojinnaka C, Lichorad A, McClellan D. Cervical cancer screening barriers and risk factor knowledge among uninsured women. J Community Health 2017 Aug;42(4):770-778 [FREE Full text] [doi: 10.1007/s10900-017-0316-9] [Medline: 28155005]
3. 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]
4. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
5. Otto L, Harst L, Schlieter H, Wollschlaeger B, Richter P, Timpel P. Towards a Unified Understanding of eHealth and Related Terms – Proposal of a Consolidated Terminological Basis. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018). Setúbal: SCITEPRESS; 2018 Presented at: 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018); January 19-21 , 2018; Funchal, Portugal.
6. Singh G, O'Donoghue J, Soon CK. Telemedicine: issues and implications. Technol Health Care 2002;10(1):1-10. [Medline: 11847443]
7. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
8. Boonstra A, van Offenbeek M. Towards consistent modes of e-health implementationtructurational analysis of a telecare programme's limited success. Inform Syst J 2010 Nov;20(6):537-561. [doi: 10.1111/j.1365-2575.2010.00358.x]
9. van Dyk L. A review of telehealth service implementation frameworks. Int J Environ Res Public Health 2014 Jan 23;11(2):1279-1298 [FREE Full text] [doi: 10.3390/ijerph110201279] [Medline: 24464237]
10. Atlas of eHealth country profiles. The use of eHealth in support of universal health coverage 2015. In: Global Observatory for eHealth. Geneva, Switzerland: World Health Organisation (WHO); 2015.
11. Global diffusion of eHealth. Making universal health coverage achievable. In: Global Observatory for eHealth. Geneva, Switzerland: World Health Organisation (WHO); 2015.
12. Rogers EM. Diffusion of Innovations. New York: Free Press; 1995.
13. Aguilar KM, Campbell RS, Fiester A, Simpson RL, Hertel C. Bringing care home: how telemonitoring can expand population health management beyond the hospital. Nurs Adm Q 2014;38(2):166-172. [doi: 10.1097/NAQ.0000000000000029] [Medline: 24569764]
14. Esser PE, Goossens RH. A framework for the design of user-centred teleconsulting systems. J Telemed Telecare 2009;15(1):32-39. [doi: 10.1258/jtt.2008.080601] [Medline: 19139218]
15. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
16. Karsh B. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
17. Ifinedo P. Empirical study of Nova Scotia nurses' adoption of healthcare information systems: implications for management and policy-making. Int J Health Policy Manag 2017 Aug 13;7(4):317-327 [FREE Full text] [doi: 10.15171/ijhpm.2017.96] [Medline: 29626399]
18. Hastall M, Dockweiler C, Mühlhaus J. Achieving end user acceptance: Building blocks for an evidence-based user-centered framework for health technology development and assessment. Cham, Switzerland: Springer; 2017 Presented at: 11th

International Conference on Universal Access in Human-Computer Interaction UAHCI; July 9–14, 2017; Vancouver, BC, Canada p. 13-25.

19. Sahay TB, Ashbury FD, Roberts M, Rootman I. Effective components for nutrition interventions: a review and application of the literature. Health Promot Pract 2006 Oct;7(4):418-427. [doi: 10.1177/1524839905278626] [Medline: 16928989]
20. Rosenstock IM. Historical origins of the Health Belief Model. Health Educ Monogr 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
21. Zhang M, Luo M, Nie R, Zhang Y. Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. Int J Med Inform 2017 Dec;108:97-109. [doi: 10.1016/j.ijmedinf.2017.09.016] [Medline: 29132639]
22. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Boston, MA: Addison Wesley; 1975.
23. Araújo MT, Paiva T, Jesuino JC, Magalhães M. General practitioners and neurotelemedicine. Stud Health Technol Inform 2000;78:45-67. [Medline: 11151607]
24. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition To Behavior (Springer Series In Social Psychology). Berlin: Springer; 1985:11-39.
25. Herrmann LK, Kim J. The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. Mhealth 2017;3:2 [FREE Full text] [doi: 10.21037/mhealth.2017.01.03] [Medline: 28293619]
26. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
27. Bandura A. Social Foundations Of Thought And Action: A Social Cognitive Theory. Upper Saddle River, NJ: Prentice Hall; 1986.
28. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
29. Vanneste D, Vermeulen B, Declercq A. Healthcare professionals' acceptance of BelRAI, a web-based system enabling person-centred recording and data sharing across care settings with interRAI instruments: a UTAUT analysis. BMC Med Inform Decis Mak 2013;13:129 [FREE Full text] [doi: 10.1186/1472-6947-13-129] [Medline: 24279650]
30. Costa CR, Iglesias MJ, Rifón LE, Carballa MG, Rodríguez SV. The acceptability of TV-based game platforms as an instrument to support the cognitive evaluation of senior adults at home. PeerJ 2017;5:e2845 [FREE Full text] [doi: 10.7717/peerj.2845] [Medline: 28070464]
31. May C, Finch T. Implementing, embedding, and integrating practices: an outline of Normalization Process Theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
32. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
33. Wade V, Gray L, Carati C. Theoretical frameworks in telemedicine research. J Telemed Telecare 2017 Jan;23(1):181-187. [doi: 10.1177/1357633X15626650] [Medline: 26787643]
34. Whitten P, Johannessen LK, Soerensen T, Gammon D, Mackert M. A systematic review of research methodology in telemedicine studies. J Telemed Telecare 2007;13(5):230-235. [doi: 10.1258/135763307781458976] [Medline: 17697509]
35. Lai PC. the literature review of technology adoption models and theories for the novelty technology. JISTEM 2017 Apr 30;14(1):21-38. [doi: 10.4301/S1807-17752017000100002]
36. Garavand A, Mohseni M, Asadi H, Etemadi M, Moradi-Joo M, Moosavi A. Factors influencing the adoption of health information technologies: a systematic review. Electron Physician 2016 Aug;8(8):2713-2718 [FREE Full text] [doi: 10.19082/2713] [Medline: 27757179]
37. Higgins JPT. Cochrane Handbook for Systematic Reviews of Interventions. Chichester: John Wiley & Sons; 2010.
38. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009;339:b2535 [FREE Full text] [Medline: 19622551]
39. Bashshur R, Shannon G, Krupinski E, Grigsby J. The taxonomy of telemedicine. Telemed J E Health 2011;17(6):484-494. [doi: 10.1089/tmj.2011.0103] [Medline: 21718114]
40. Field A. Discovering Statistics Using IBM SPSS Statistics: North American Edition. Los Angeles: Sage Publications Ltd; 2018:1025.
41. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016 Dec 08;6(12):e011458 [FREE Full text] [doi: 10.1136/bmjopen-2016-011458] [Medline: 27932337]
42. Spaulding RJ, Russo T, Cook DJ, Doolittle GC. Diffusion theory and telemedicine adoption by Kansas health-care providers: critical factors in telemedicine adoption for improved patient access. J Telemed Telecare 2005;11(Suppl 1):107-109. [doi: 10.1258/1357633054461903] [Medline: 16036015]
43. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jan;32(1-2):29-36. [doi: 10.1017/S0266462316000015] [Medline: 27472158]

44. Huygens MW, Vermeulen J, Friele RD, van Schayck OC, de Jong JD, de Witte LP. Internet services for communicating with the general practice: barely noticed and used by patients. Interact J Med Res 2015;4(4):e21 [FREE Full text] [doi: 10.2196/ijmr.4245] [Medline: 26601596]
45. Rho M, Kim H, Chung K, Choi I. Factors influencing the acceptance of telemedicine for diabetes management. Cluster Comput 2015 Mar 12;18(1):321-331. [doi: 10.1007/s10586-014-0356-1]
46. Jen W, Hung M. An empirical study of adopting mobile healthcare service: the family's perspective on the healthcare needs of their elderly members. Telemed J E Health 2010;16(1):41-48. [doi: 10.1089/tmj.2009.0093] [Medline: 20070162]
47. Zhang H, Cocosila M, Archer N. Factors of adoption of mobile information technology by homecare nurses: a technology acceptance model 2 approach. Comput Inform Nurs 2010;28(1):49-56. [doi: 10.1097/NCN.0b013e3181c0474a] [Medline: 19940621]
48. Zhang X, Han X, Dang Y, Meng F, Guo X, Lin J. User acceptance of mobile health services from users' perspectives: the role of self-efficacy and response-efficacy in technology acceptance. Inform Health Soc Care 2017 Mar;42(2):194-206. [doi: 10.1080/17538157.2016.1200053] [Medline: 27564428]
49. Lin S, Yang H. Exploring key factors in the choice of e-health using an asthma care mobile service model. Telemed J E Health 2009 Nov;15(9):884-890. [doi: 10.1089/tmj.2009.0047] [Medline: 19919195]
50. de Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
51. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
52. Hossain N, Yokota F, Sultana N, Ahmed A. Factors influencing rural end-users' acceptance of e-health in developing countries: a study on portable health clinic in bangladesh. Telemed J E Health 2018 Apr 17:-. [doi: 10.1089/tmj.2018.0039] [Medline: 29664328]
53. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017 Feb 28:E7. [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
54. Kuhn E, Crowley J, Hoffman J, Eftekhari A, Ramsey K, Owen J, et al. Clinician characteristics and perceptions related to use of the PE (prolonged exposure) coach mobile app. Prof Psychol Res Pr 2015 Dec;46(6):437-443. [doi: 10.1037/pro0000051]
55. Hennemann S, Beutel ME, Zwerenz R. Drivers and barriers to acceptance of web-based aftercare of patients in inpatient routine care: a cross-sectional survey. J Med Internet Res 2016 Dec 23;18(12):e337 [FREE Full text] [doi: 10.2196/jmir.6003] [Medline: 28011445]
56. Asua J, Orruño E, Reviriego E, Gagnon MP. Healthcare professional acceptance of telemonitoring for chronic care patients in primary care. BMC Med Inform Decis Mak 2012 Nov 30;12:139 [FREE Full text] [doi: 10.1186/1472-6947-12-139] [Medline: 23194420]
57. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59 [FREE Full text] [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
58. James S, Perry L, Gallagher R, Lowe J. Diabetes educators' intended and reported use of common diabetes-related technologies: discrepancies and dissonance. J Diabetes Sci Technol 2016 Dec;10(6):1277-1286 [FREE Full text] [doi: 10.1177/1932296816646798] [Medline: 27179011]
59. Orruño E, Gagnon MP, Asua J, Ben Abdeljelil A. Evaluation of teledermatology adoption by health-care professionals using a modified Technology Acceptance Model. J Telemed Telecare 2011;17(6):303-307. [doi: 10.1258/jtt.2011.101101] [Medline: 21844171]
60. Dockweiler C, Kupitz A, Hornberg C. [Acceptance of online-based therapy by patients with light to moderate depressive disorders]. Gesundheitswesen 2017 Nov 10:1013-1022. [doi: 10.1055/s-0043-119287] [Medline: 29126341]
61. van Houwelingen CT, Barakat A, Best R, Boot WR, Charness N, Kort HS. Dutch nurses' willingness to use home telehealth: implications for practice and education. J Gerontol Nurs 2015 Apr;41(4):47-56. [doi: 10.3928/00989134-20141203-01] [Medline: 25486115]
62. Peeters JM, de Veer AJ, van der Hoek L, Francke AL. Factors influencing the adoption of home telecare by elderly or chronically ill people: a national survey. J Clin Nurs 2012 Nov;21(21-22):3183-3193. [doi: 10.1111/j.1365-2702.2012.04173.x] [Medline: 22827253]
63. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
64. Venkatesh V, Davis F. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]

65. van Houwelingen CT, Ettema RG, Antonietti MG, Kort HS. Understanding older people's readiness for receiving telehealth: mixed-method study. J Med Internet Res 2018 Apr 06;20(4):e123 [FREE Full text] [doi: 10.2196/jmir.8407] [Medline: 29625950]
66. Berg M, Linden K, Adolfsson A, Lundin CS, Ranerup A. Web-based intervention for women with type 1 diabetes in pregnancy and early motherhood: critical analysis of adherence to technological elements and study design. J Med Internet Res 2018 May 02;20(5):e160 [FREE Full text] [doi: 10.2196/jmir.9665] [Medline: 29720365]
67. Boessen AB, Vermeulen J, de Witte LP. Acceptance and usability of a home-based monitoring tool of health indicators in children of people with dementia: a Proof of Principle (POP) study. Patient Prefer Adherence 2017;11:1317-1324 [FREE Full text] [doi: 10.2147/PPA.S135022] [Medline: 28814839]
68. Monthuy-Blanc J, Bouchard S, Maïano C, Séguin M. Factors influencing mental health providers' intention to use telepsychotherapy in First Nations communities. Transcult Psychiatry 2013 Apr;50(2):323-343. [doi: 10.1177/1363461513487665] [Medline: 23666941]
69. Koivumäki T, Pekkarinen S, Lappi M, Väisänen J, Juntunen J, Pikkarainen M. consumer adoption of future MyData-based preventive eHealth services: an acceptance model and survey study. J Med Internet Res 2017 Dec 22;19(12):e429 [FREE Full text] [doi: 10.2196/jmir.7821] [Medline: 29273574]
70. Scheibe M, Reichelt J, Bellmann M, Kirch W. Acceptance factors of mobile apps for diabetes by patients aged 50 or older: a qualitative study. Med 2 0 2015 Mar 02;4(1):e1 [FREE Full text] [doi: 10.2196/med20.3912] [Medline: 25733033]
71. De Angelis G, Davies B, King J, Wells GA, Brosseau L. The use of social media by arthritis health professionals to disseminate a self-management program to patients: a feasibility study. Digit Health 2017;3:2055207617700520 [FREE Full text] [doi: 10.1177/2055207617700520] [Medline: 29942586]
72. Jun S, Plint AC, Campbell SM, Curtis S, Sabir K, Newton AS. Point-of-care cognitive support technology in emergency departments: a scoping review of technology acceptance by clinicians. Acad Emerg Med 2018 May;25(5):494-507. [doi: 10.1111/acem.13325] [Medline: 28960689]
73. Saigí-Rubió F, Torrent-Sellens J, Jiménez-Zarco A. Drivers of telemedicine use: comparative evidence from samples of Spanish, Colombian and Bolivian physicians. Implement Sci 2014 Oct 08;9:128. [doi: 10.1186/s13012-014-0128-6] [Medline: 25293651]
74. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790 [FREE Full text] [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]
75. Atkinson NL. Developing a questionnaire to measure perceived attributes of eHealth innovations. Am J Health Behav 2007;31(6):612-621. [doi: 10.5555/ajhb.2007.31.6.612] [Medline: 17691874]
76. 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]
77. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
78. Abramson EL, Silver M, Kaushal R. Meaningful use status and participation in health information exchange among New York State hospitals: a longitudinal assessment. Jt Comm J Qual Patient Saf 2014 Oct;40(10):452-453. [Medline: 26111305]
79. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018 May 14;16(1):66 [FREE Full text] [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
80. Zailani S, Gilani MS, Nikbin D, Iranmanesh M. Determinants of telemedicine acceptance in selected public hospitals in Malaysia: clinical perspective. J Med Syst 2014 Sep;38(9):111. [doi: 10.1007/s10916-014-0111-4] [Medline: 25038891]
81. Serrano C, Karahanna E. The compensatory interaction between user capabilities and technology capabilities in influencing task performance: an empirical assessment in telemedicine consultations. MIS Q 2016;40(3):597-621. [doi: 10.25300/MISQ/2016/40.3.04]

Abbreviations

eHealth: electronic health
HBM: Health Belief Model
ICT: Information and Communication Technology
mHealth: mobile health
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
OR: odds ratio
PICOS: Population, Intervention, Comparison, Outcome, and Study Design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SEM: Structural Equation Modeling
TAM: Technology Acceptance Model
TIB: Theory of Interpersonal Behavior
TPB: Theory of Planned Behavior

TRA: Theory of Reasoned Action
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by G Eysenbach; submitted 12.12.18; peer-reviewed by A Ekeland, T Rashid Soron, X Guo, F Saigí-Rubió, S Shah; comments to author 14.02.19; revised version received 08.03.19; accepted 29.03.19; published 21.05.19.

Please cite as:
Harst L, Lantzsch H, Scheibe M
Theories Predicting End-User Acceptance of Telemedicine Use: Systematic Review
J Med Internet Res 2019;21(5):e13117
URL: http://www.jmir.org/2019/5/e13117/
doi:10.2196/13117
PMID:31115340

Original Paper

Validity of Online Screening for Autism: Crowdsourcing Study Comparing Paid and Unpaid Diagnostic Tasks

Peter Washington[1], BA, MS; Haik Kalantarian[2], BS, MS, PhD; Qandeel Tariq[2], BS, MS; Jessey Schwartz[2], BA; Kaitlyn Dunlap[2], BSc, MRES; Brianna Chrisman[1], BS; Maya Varma[3]; Michael Ning[2], BS; Aaron Kline[2], BS; Nathaniel Stockham[4], BS, MS; Kelley Paskov[2], BS, MS; Catalin Voss[3], BS, MS; Nick Haber[2,5,6,7], BS, PhD; Dennis Paul Wall[5,7,8], PhD

[1]Department of Bioengineering, Stanford University, Stanford, CA, United States
[2]Department of Biomedical Data Science, Stanford University, Stanford, CA, United States
[3]Department of Computer Science, Stanford University, Stanford, CA, United States
[4]Department of Neuroscience, Stanford University, Stanford, CA, United States
[5]Department of Pediatrics, Stanford University, Stanford, CA, United States
[6]Department of Psychology, Stanford University, Stanford, CA, United States
[7]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States
[8]Division of Systems Medicine, Department of Biomedical Data Science, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Dennis Paul Wall, PhD
Division of Systems Medicine
Department of Biomedical Data Science
Stanford University
1265 Welch Rd
Palo Alto, CA,
United States
Phone: 1 617 304 6031
Email: dpwall@stanford.edu

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2019/6/e14950/

Abstract

Background: Obtaining a diagnosis of neuropsychiatric disorders such as autism requires long waiting times that can exceed a year and can be prohibitively expensive. Crowdsourcing approaches may provide a scalable alternative that can accelerate general access to care and permit underserved populations to obtain an accurate diagnosis.

Objective: We aimed to perform a series of studies to explore whether paid crowd workers on Amazon Mechanical Turk (AMT) and citizen crowd workers on a public website shared on social media can provide accurate online detection of autism, conducted via crowdsourced ratings of short home video clips.

Methods: Three online studies were performed: (1) a paid crowdsourcing task on AMT (N=54) where crowd workers were asked to classify 10 short video clips of children as “Autism” or “Not autism,” (2) a more complex paid crowdsourcing task (N=27) with only those raters who correctly rated ≥8 of the 10 videos during the first study, and (3) a public unpaid study (N=115) identical to the first study.

Results: For Study 1, the mean score of the participants who completed all questions was 7.50/10 (SD 1.46). When only analyzing the workers who scored ≥8/10 (n=27/54), there was a weak negative correlation between the time spent rating the videos and the sensitivity (ρ=–0.44, *P*=.02). For Study 2, the mean score of the participants rating new videos was 6.76/10 (SD 0.59). The average deviation between the crowdsourced answers and gold standard ratings provided by two expert clinical research coordinators was 0.56, with an SD of 0.51 (maximum possible SD is 3). All paid crowd workers who scored 8/10 in Study 1 either expressed enjoyment in performing the task in Study 2 or provided no negative comments. For Study 3, the mean score of the participants who completed all questions was 6.67/10 (SD 1.61). There were weak correlations between age and score (*r*=0.22, *P*=.014), age and sensitivity (*r*=–0.19, *P*=.04), number of family members with autism and sensitivity (*r*=–0.195, *P*=.04), and number of family members with autism and precision (*r*=–0.203, *P*=.03). A two-tailed *t* test between the scores of the paid workers in Study 1 and the unpaid workers in Study 3 showed a significant difference (*P*<.001).

Conclusions: Many paid crowd workers on AMT enjoyed answering screening questions from videos, suggesting higher intrinsic motivation to make quality assessments. Paid crowdsourcing provides promising screening assessments of pediatric autism with an average deviation <20% from professional gold standard raters, which is potentially a clinically informative estimate for parents. Parents of children with autism likely overfit their intuition to their own affected child. This work provides preliminary demographic data on raters who may have higher ability to recognize and measure features of autism across its wide range of phenotypic manifestations.

(J Med Internet Res 2019;21(5):e13668) doi:10.2196/13668

KEYWORDS

crowdsourcing; autism; mechanical turk; pediatrics; diagnostics; diagnosis; neuropsychiatric conditions; human-computer interaction; citizen healthcare; biomedical data science; mobile health; digital health

Introduction

Autism spectrum disorder (ASD, or autism) [1] is a complex neurodevelopmental disorder that manifests in children by or before 3 years of age and now impacts 1 in 40 children in the United States [2]. Obtaining a diagnosis of ASD, like that of many other neuropsychiatric conditions, requires long waiting times, often exceeding a year [3,4]. Furthermore, obtaining a professional diagnosis is often prohibitively expensive for much of the global population [5,6]. *Crowdsourcing* provides a scalable alternative to the current diagnostic evaluations for ASD that include extensive and lengthy clinical evaluation by a trained professional and are inaccessible to families in rural areas as well as those with low incomes [7]. Crowdsourcing, broadly defined, is a type of participative online activity in which an entity proposes the voluntary undertaking of a task to a group of individuals [8]. Crowdsourcing can improve the quality and speed of medical research projects by leveraging the crowd for large-scale problem solving, data processing, surveillance/monitoring, and surveying [9]. Thus, crowdsourcing may enable families who are limited by long waiting times, are members of underserved populations in the United States, or are living in countries where accessibility is hampered by available resources to obtain an accurate diagnosis in a timely manner.

Crowdsourcing is increasingly used in health promotion, research, and care [10]. Ben-Sasson et al showed that parents of children suspected to have autism can fill out standardized questionnaires about their own child online to prescreen for autism; a machine learning classifier trained on the questionnaire responses identified 58% to 88% of children at risk for autism [11]. Tariq et al showed that feature tagging by independent nonexpert raters enables rapid machine learning risk prediction for autism by using home videos of <3 minutes in length [12]. However, such approaches have not yet been performed or tested at scale.

Large-scale crowdsourcing can be achieved through online and virtual workforce platforms. Amazon Mechanical Turk (AMT) is a popular crowdsourcing platform that offers a paradigm for engaging a large number of users for short times and low monetary costs [13]. AMT has been successfully used in health care settings. Kuang et al showed that crowd workers on AMT had significantly higher scores when rating pictographs than in-person participants in a hospital [14]. CrowdMed is a software platform that aims to leverage crowdsourcing to help undiagnosed or misdiagnosed patients by allowing them to submit their cases and interact with case solvers to obtain diagnostic outcomes [15]. Such large-scale crowdsourcing systems and paradigms have not yet been tested for their potential in assisting the screening and diagnostic processes for children at risk for developmental delays.

A limiting factor for the crowdsourcing detection of pediatric conditions is the collection of structured data such as video or audio. Recently, an increasing number of mobile health tools are being developed for children with autism [4,16-25], including at least one that has been clinically validated in a randomized controlled trial [26]. These tools not only provide opportunities for better health care but can also be a significant data resource and specifically increase the potential to collect rich, naturalistic behavioral data via structured mobile videos [27,28]. Nazneen et al developed an effective system used in a home setting to capture videos of children with autism, and professional diagnosticians deemed 96% of the collected videos clinically useful for making an autism diagnosis [29]. Voss et al developed a wearable system on Google Glass providing real-time emotion feedback to the child wearer while simultaneously capturing videos through the front-facing camera of the Glass [19,24-26]. Kalantarian et al developed a mobile charades game that promotes facial contact with the parent by the affected child while capturing highly structured videos through the front-facing camera of iOS and Android devices [27,28]. Videos produced by such systems can be fed into a crowdsourcing pipeline for manual labeling of videos that can be analyzed using artificial intelligence and produce rapid screening and diagnosis for children at risk for developmental delay and other conditions marked by behavioral symptoms. Such applications could be particularly valuable because they have the potential to flow more easily into the health care system, given the widespread adoption of mobile devices globally [30,31], enabling easy and free data collection and transfer between families, crowdsourced workforces, and the health care system.

Here, we present a series of three crowdsourced studies which enabled us to (1) test the ability of the crowd to directly identify autism and (2) provide behavioral metrics that could be used for machine learning autism classification based on a short video clip of a child interacting with his/her parent. In Study 1, we evaluated whether a randomly selected set of paid crowd workers could accurately label videos of children interacting with family members as either “Autism” or “Not autism.” In Study 2, we evaluated whether high-scoring crowd workers providing intuitive answers about a disorder would perform well on a different set of videos and be motivated to perform a more thorough task on AMT. In Study 3, we tested how unpaid crowd workers perform when rating videos for diagnostics. We hypothesized that the workers would enjoy the rating task, certain demographics of workers would emerge as high-quality raters, paid and unpaid crowd workers would perform equally well on the same set of videos, high-scoring workers on simple rating tasks would continue to perform well on harder rating tasks, and crowd ratings would approximate a set of “gold standard” ratings from professionals.

Methods

Summary of Studies

We performed three studies that were designed incrementally in response to the results from our prior work, which examined feature tagging by independent nonexpert raters for autism risk prediction using home videos [12]. All three studies were approved by the Stanford Institutional Review Board. Videos for all studies were sourced from publicly available YouTube videos. We searched for videos of children both with and without autism based on the following criteria for the videos: (1) it shows the child’s hands and face; (2) it includes clear opportunities for direct social engagement; and (3) it involves opportunities for the child to use an object such as a utensil, crayon, or toy. We selected a subset of 20 videos for this study balanced by age, gender, and diagnosis. Table 1 provides video demographic statistics of the sets of videos used in all tasks.

Study 1: Paid Crowdsourcing on Mechanical Turk

Overview

In order to evaluate whether a randomly selected set of paid crowd workers could accurately classify videos as either “Autism” or “Not Autism,” we recruited 54 workers on AMT and recorded their demographic traits.

Participants

Workers were paid US $3.50 each to complete the task, which aligns with the California minimum wage payment rate based on our estimate of the time needed to complete each task. To ensure quality, workers were required to have a task approval rate >98% for all requesters’ human intelligence tasks and a total number of approved human intelligence tasks >500. In order to identify any differences in rating ability based on demographic trends, we asked workers for their age and gender, whether the rater is a parent, the number of children the rater knows with autism, the number of family members with autism, the number of affected friends, and whether the rater himself/herself has autism.

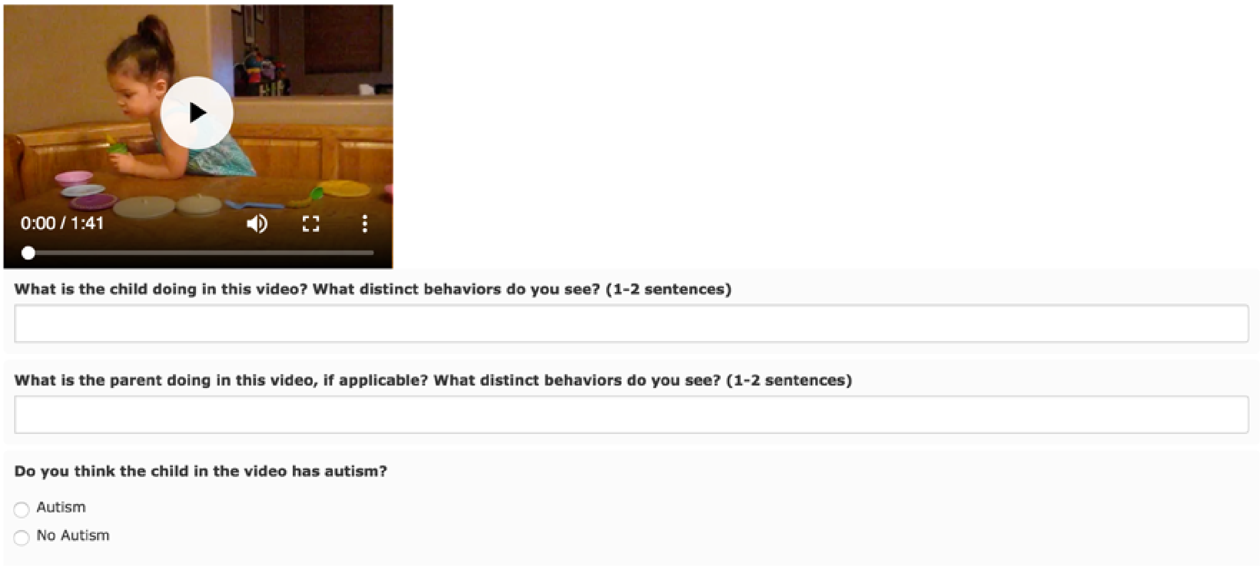
Task

The task consisted of workers viewing and answering questions on 10 videos of a parent interacting with a child. The videos were equally balanced for gender and diagnosis (Table 1). Workers were not required to watch the entire video but were instructed to *“scroll through the video once you get the idea, but watch enough to be able to answer the questions*.” We allowed the workers to skip to other parts of the videos and replay them. As shown in Figure 1, each video was followed by free-response questions asking the rater to briefly describe the activity of the parent and the child in the video. This was used to verify that the workers adequately engaged with the videos before classifying them as “Autism” or “No Autism.” If the answers were logically inconsistent, we discarded the worker’s answers. After answering all questions, we asked the workers a series of demographic questions to help them describe themselves. These questions included age, gender, geographic location, whether the rater is a parent, the number of children the rater knows with autism, the number of family members with autism, the number of affected friends, and whether the rater himself/herself has autism (Multimedia Appendix 1).

Table 1. Summary of the videos used in all three studies.

Studies	Video length, mean (range)	Child age (years), mean (range)	Female, %	Children with autism, %
1, 2, 3	3 minutes 2 seconds (49 s to 6 min 39 s)	3.2 (2-5)	50	50
2	2 minutes 9 seconds (1 min 7 s to 4 min 40 s)	2.9 (2-5)	50	50

Figure 1. An example question set on the paid crowdsourcing Mechanical Turk Study 1 task. Workers answered the same set of questions for 10 separate videos.



What is the child doing in this video? What distinct behaviors do you see? (1-2 sentences)

What is the parent doing in this video, if applicable? What distinct behaviors do you see? (1-2 sentences)

Do you think the child in the video has autism?

○ Autism

○ No Autism

Analysis

We used the Pearson correlation when comparing real numbers to performance metrics and the point biserial correlation when comparing binary variables to performance metrics. In particular, Pearson correlation was used to compare scores, precision, recall, and specificity to time spent, age, number of children known with autism, number of family members with autism, number of friends with autism, number of people known with autism, and number of children known with autism. Point biserial correlation was used to compare scores, precision, recall, specificity, and time spent to whether the rater has autism, whether the rater is a parent, and the gender of the rater. The metrics used were accuracy, precision (true positive/[true positive+false positive]), recall (true positive/[true positive+false negative]), and specificity (true negative/[true negative+false positive]). We analyzed the subset of workers who scored well (≥8/10) in addition to the pool of all workers in order to determine demographic traits specific to high-performing workers.

Study 2: Paid Crowdsourcing With High-Scoring Workers

Overview

In order to evaluate whether high-scoring workers providing intuitive answers about a disorder would perform well on a different set of videos and be motivated to perform a more thorough task on AMT, we conducted a follow-up study with the workers who performed well (scored ≥8/10) in Study 1. The study was divided into two parts: (1) conducting the same task as that in Study 1 but with a different set of 10 videos, and (2) answering a series of 31 multiple-choice questions about specific behaviors of the child for each of the 10 videos from Study 1.

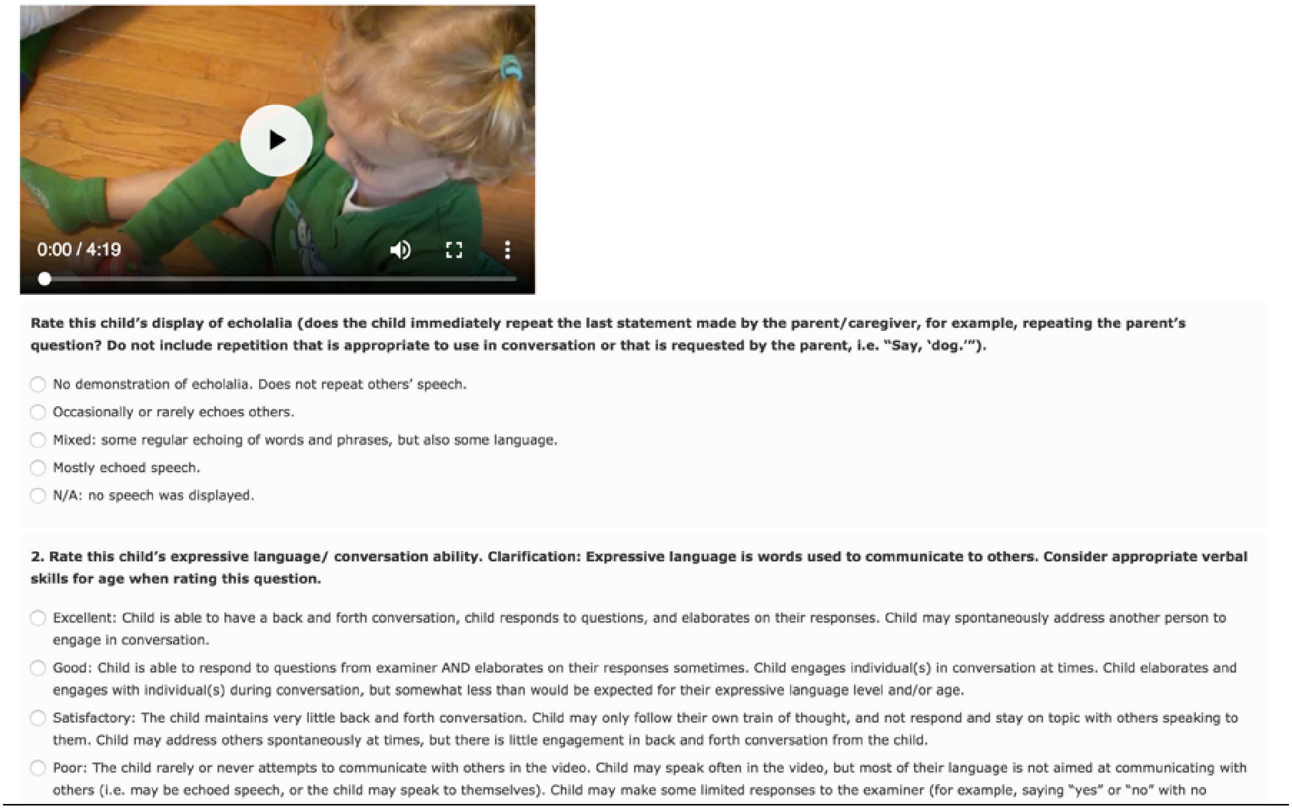
Participants

A total of 27 workers who scored ≥8/10 in Study 1 were successfully recruited to complete an additional set of 11 tasks. We chose to exclude workers who did not perform well in Study 1 because we wanted to filter out workers who did not demonstrate intuitive skill for detecting developmental delays in children. We chose a cutoff of 8/10 because a higher cutoff would not yield a large enough worker pool to recruit from. Workers were recruited by providing a worker bonus of US $0.05, and they were sent a message describing the additional tasks and pay for completion of the tasks.

Tasks

In the first task, the setup was identical to that in Study 1 except that a different set of videos was used (Table 1) . The remaining 10 tasks required workers to answer a series of 31 multiple-choice questions that have been previously shown to have high predictive power for detecting autism through video [12]. The 31 multiple-choice questions target 31 different symptoms of ASD and are written to rate the presence of these symptoms in short videos of children. Each question asks the rater to rate an individual symptom of autism, with the answer choices representing increasing levels of severity; each question therefore serves as a rating scale that allows for quantitative comparisons. The full set of 31 questions can be found in a previous study [12]. To understand the worker satisfaction with the tasks, we asked all workers to provide any free-form comments about the rating experience. Figure 2 shows the user interface for these tasks. The same set of 10 videos used in Study 1 was also used in this series of tasks. For each task, the interface consisted of a single video followed by a series of 31 multiple-choice questions, including the original diagnosis question (“Autism” or “Not autism”) asked in Study 1. Because all the recruited workers had already performed well in Study 1, no verification questions about the video were asked.

Figure 2. Two questions on the paid crowdsourcing Amazon Mechanical Turk Study 2 multiple-choice tasks. Workers were asked to answer 31 multiple-choice questions for a single video per task. There were 10 available identical tasks with different videos.



Rate this child's display of echolalia (does the child immediately repeat the last statement made by the parent/caregiver, for example, repeating the parent's question? Do not include repetition that is appropriate to use in conversation or that is requested by the parent, i.e. "Say, 'dog.'").

○ No demonstration of echolalia. Does not repeat others' speech.
○ Occasionally or rarely echoes others.
○ Mixed: some regular echoing of words and phrases, but also some language.
○ Mostly echoed speech.
○ N/A: no speech was displayed.

2. Rate this child's expressive language/ conversation ability. Clarification: Expressive language is words used to communicate to others. Consider appropriate verbal skills for age when rating this question.

○ Excellent: Child is able to have a back and forth conversation, child responds to questions, and elaborates on their responses. Child may spontaneously address another person to engage in conversation.
○ Good: Child is able to respond to questions from examiner AND elaborates on their responses sometimes. Child engages individual(s) in conversation at times. Child elaborates and engages with individual(s) during conversation, but somewhat less than would be expected for their expressive language level and/or age.
○ Satisfactory: The child maintains very little back and forth conversation. Child may only follow their own train of thought, and not respond and stay on topic with others speaking to them. Child may address others spontaneously at times, but there is little engagement in back and forth conversation from the child.
○ Poor: The child rarely or never attempts to communicate with others in the video. Child may speak often in the video, but most of their language is not aimed at communicating with others (i.e. may be echoed speech, or the child may speak to themselves). Child may make some limited responses to the examiner (for example, saying "yes" or "no" with no

Comparison to Gold Standard Ratings

In order to compare the answers provided by the crowd workers with a "gold standard" rating, we asked two trained clinical research coordinators experienced in working with children with autism and neuropsychiatric disorders to answer all 31 questions for each of the 10 videos that included multiple-choice questions. This rating was used as a baseline to compare the answers from the AMT workers.

Analysis

As in Study 1, Pearson correlation was used to compare scores to time spent, age, number of children known with autism, number of family members with autism, number of friends known with autism, number of people known with autism, and number of children known with autism. Point biserial correlation was used to compare scores to whether the rater has autism, whether the rater is a parent, and the gender of the rater.

Study 3: Public Crowdsourcing Through Citizen Healthcare

Overview

In order to test how unpaid crowd workers perform when rating videos for diagnostics, we developed a public website (videoproject.stanford.edu) for watching the videos and answering questions about the videos. Through pilot testing, we found that unpaid crowd workers are not willing to answer 31 multiple-choice questions for several videos; therefore, we focused on the "Autism or Not" task from Study 1.

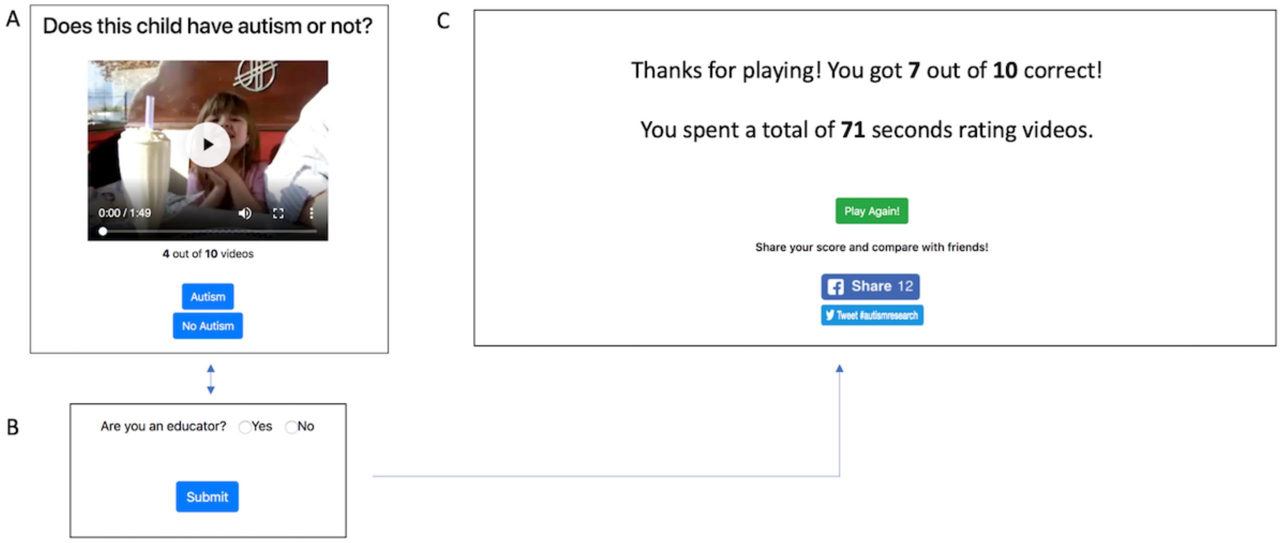
Participants

A total of 115 participants were successfully recruited via our public-facing website (videoproject.stanford.edu) that was distributed via social media shares and online community noticeboards (eg, Nextdoor.com).

Task

When the users navigated to the webpage, they were provided with a video and two buttons allowing them to classify the video as "Autism" or "Not autism" (Figure 3A), as in Study 1. In order to minimize participant dropout due to a long list of demographic questions, we interleaved each question throughout the video-rating process (Figure 3B). After rating all videos and providing all demographic information, the user is directed to a results page where his/her score and total rating time are displayed along with the option to play again and share their results on social media (Figure 3C).

Figure 3. (A) The primary interface for the "citizen healthcare" public crowdsourcing study. Citizen healthcare providers watch a short video and then classify the video as "Autism" or "Not Autism." (B) After rating each video in the "citizen healthcare" public crowdsourcing study, users are asked a single demographic question about themselves. This allows us to collect demographic information without overwhelming the user, which would otherwise lead to lower participant retention rates. (C) At the end of the "citizen healthcare" public crowdsourcing study, users are informed of their score and the time they spent rating. They then have the option to play the game again and share their result on Facebook or Twitter.



Does this child have autism or not?

Are you an educator? ○Yes ○No

Thanks for playing! You got **7** out of **10** correct!

You spent a total of **71** seconds rating videos.

Share your score and compare with friends!

Results

Study 1: Paid Crowdsourcing on Mechanical Turk

The mean score of the participants who completed all questions was 7.5/10 (SD 1.46). Table 2 shows the summary demographics of the crowd workers who completed the task. Most of the demographic trends from the Mechanical Turk cohort were not statistically significant. However, when only analyzing the workers who scored well (scored ≥8/10; N=27/54), there was a weak negative correlation between the time spent rating the videos and the sensitivity (ρ=–0.44, *P*=.02). There were similar almost-significant trends for accuracy (ρ=–0.35, *P*=.07) and precision (ρ=–0.38, *P*=.05). There was no trend for sensitivity or specificity. Table 3 shows the average rating of all video raters.

Table 2. Summary demographics of the crowd workers in Study 1 (N=54).

Demographic	Value
Age, mean (SD)	36.4 (9.0)
With autism, n (%)	3 (5.6)
Is a parent, n (%)	25 (46.3)
Female, n (%)	20 (37.0)
Number of known affected children, mean (SD)	0.7 (0.9)
Number of affected families, mean (SD)	0.4 (0.7)
Number of affected friends, mean (SD)	1.3 (1.2)
Number of total known affected people, mean (SD)	2.3 (3.3)

Table 3. Ratings labeled as “Autism” across all 54 paid crowd workers in Study 1.

Video number	Ratings labeled as “Autism”, %	True rating
1	87	Autism
2	6	Not autism
3	2	Not autism
4	44	Autism
5	81	Autism
6	2	Not autism
7	39	Autism
8	49	Not autism
9	70	Autism
10	2	Not autism

Study 2: Paid Crowdsourcing With High-Scoring Workers

Overview

Table 4 shows the comparison between workers who performed well (≥8/10 videos correctly diagnosed) and poorly (<8/10) in Study 1. There were no statistically significant differences between the two populations except in the mean number of affected children that the worker knew.

Performance on Different Video Sets

The mean score of the crowd workers was 6.76/10 (SD 0.59). Because the study cohort was smaller for Study 2, we did not analyze demographic trends. Instead, we analyzed completion rate, rating trends, and agreement with “gold standard” raters. Table 5 shows the autism classification ratings of all raters in this part of the study. There was significantly more rater agreement than that in Study 1, indicating that crowd workers who perform well on providing diagnoses on one set of videos will also perform well on a different set of videos with similar characteristics.

Worker Satisfaction

None of the workers provided any negative comment about any of the tasks in this study. Several workers had positive comments (Textbox 1).

Worker Motivation

In addition to thanking the researchers for the provided tasks, some workers (4/27) volunteered detailed explanations about the videos and the reasoning behind their ratings. Comments from Video 4 are shown as a representative example in Textbox 2.

Table 4. Comparison of summary demographics of the crowd workers who performed well (≥8/10 videos correctly diagnosed) and poorly (<8/10) in Study 1 (N=27).

Demographic	Performed well (score≥8/10)	Performed poorly (score<8/10)	*P* value
Age, mean (SD)	34.7 (6.5)	38.1 (10.8)	.17
With autism, n (%)	2 (7.4)	1 (3.7)	.56
Is a parent, n (%)	12 (44.4)	13 (48.1)	.79
Female, n (%)	12 (44.4)	8 (29.6)	.27
Number of known affected children, mean (SD)	0.5 (0.7)	1.0 (1.0)	.048
Number of affected families, mean (SD)	0.2 (0.4)	0.5 (0.9)	.09
Number of affected friends, mean (SD)	1.1 (1.3)	1.5 (1.2)	.23
Number of total known affected people, mean (SD)	2.3 (3.9)	2.3 (2.6)	0.97

Table 5. Ratings labeled as “Autism” across all 22 paid crowd workers in the task with a different set of 10 videos.

Video number	Ratings labeled as “Autism”, %	True rating
11	100	Autism
12	0	Not autism
13	43	Autism
14	0	Not autism
15	90	Autism
16	76	Autism
17	90	Autism
18	10	Not autism
19	24	Not autism
20	0	Not autism

Textbox 1. Representative examples of positive comments from crowd workers.

"Well organized and enjoyable.”

"Thank you as always! I appreciate the opportunity and the behavioral learning experience.”

"Fantastic survey. I really hope there are more of these in the future!”

"Thank you so much! So glad to be a part of your studies and hope that this will progress your work in autism.”

"Thank you and look forward to more hits.”

Textbox 2. Explanations for the ratings for Video 4.

"It was a bit hard to hear the other people besides the child, so not sure if they were talking to her when she shouted 'yes.'”

"I want to say this child has Autism, but would like to see some more sensory information before I truly decided. Hence why some questions were N/A.”

"The child seems to be listening to something on earphones. Some part of it seems like she is singing along to it. The limited interaction with parents seems more like distraction by the music than some developmental problem.”

"While I think this girl has some sort of developmental issues, (her playing with the straw, hand motions, fixation on the parents phone at the end), I think it's a stretch to call it autism. I'll do the last one of these later today!”

Comparison with Gold Standard Ratings

The ratings between the two gold standard raters were identical for all videos except for one, where the answers differed by one point for a single question. Across all videos, the average deviation between the average crowdsourced answers and the gold standard ratings was 0.56, with an SD of 0.51. Figure 4 shows the distribution of deviations of all questions for all videos. There were 310 data points (31 questions × 10 workers). We followed up this analysis with a qualitative inspection of all video-question pairs where the average deviation exceeded 1.5.

Analyzing the Underlying Cause of Worker Deviation

A qualitative analysis of the video-question pairs where the average deviation exceeded 1.5 answer choices on the rating scale helped us explore the underlying cause of worker deviation from the gold standard rating. There were 22 such pairs (of a possible 31×10=310 pairs). There were 2 questions, in particular, that had high deviation across multiple videos (Table 6). Questions 13 and 16 were vaguer than the other questions, providing a list of numerous example behaviors within the question. These example behaviors were not exhaustive; therefore, it is possible that some workers only looked for the explicitly listed behaviors without generalizing them. On one video in the dataset, in particular, raters performed poorly on several questions. This video involved a 4-year-old girl with ASD singing songs with her father. It is possible that ASD features are more difficult to distinguish from a video when observing singing rather than natural speech.

Figure 4. A histogram of the AMT worker deviation from the gold standard ratings for all questions and all videos. The maximum possible deviation is 3.0. Most video ratings have a deviation below 1.0, which is an acceptable error. However, several worker responses deviated greatly from the gold standard. AMT: Amazon Mechanical Turk.

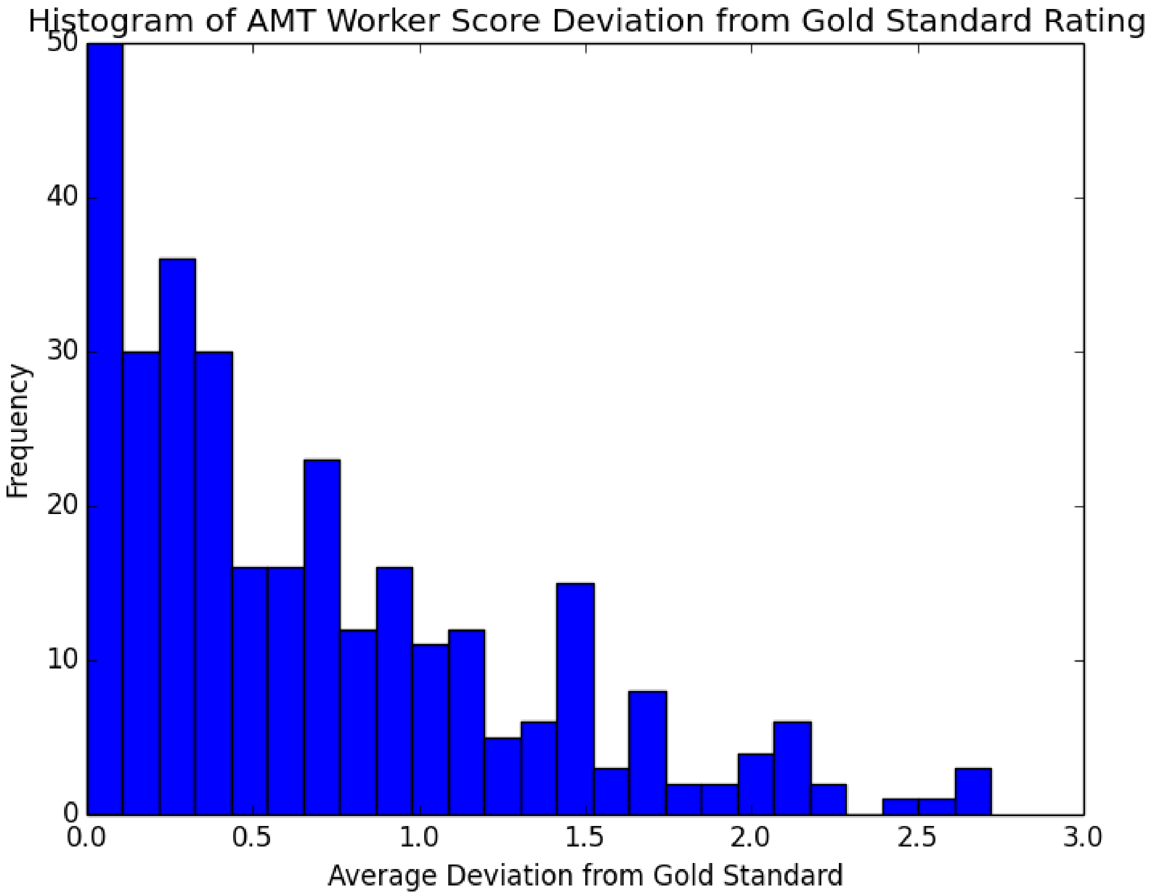


Table 6. Questions where the average worker answer was >1.5/3.0 answer choices away from the gold standard rating for multiple videos.

Question		Number of deviating videos (of 10)
13	Does the child get upset, angry or irritated by particular sounds, tastes, smells, sights or textures?	4
16	Does the child stare at objects for long periods of time or focus on particular sounds, smells or textures, or like to sniff things?	5

Study 3: Public Crowdsourcing Through Citizen Healthcare

There were 145 unique visits to videoproject.stanford.edu. A total of 126 participants provided at least one rating of the series of 10 videos. Of these 126 participants who started the rating process, 115 completed all videos (91.3% retention). The mean score of the participants who completed all questions was 6.67/10 (SD 1.61). The mean score of the paid participants was 7.50 (SD 1.46) for the same set of videos in Study 1 and 6.76 (SD 0.59) in Study 2 for a different set of videos with the high-scoring workers. A two-tailed *t* test between the scores of the paid workers in Study 1 and the unpaid workers in Study 3 showed significant differences (t_{168}=3.37, *P*<.001).

As in Study 1, we analyzed the Pearson correlation when comparing real numbers to scores and the point biserial correlation when comparing binary variables to scores. There were weak correlations between age and score (*r*=0.22, *P*=.02), age and sensitivity (*r*=–0.19, *P*=.04), age and total time spent rating videos (*r*=0.25, *P*=.004), whether the rater is an educator and total time spent rating videos (*r*=–0.185, *P*=.045), number of family members with autism and sensitivity (*r*=–0.195, *P*=.04), and number of family members with autism and precision (*r*=–0.203, *P*=.03).

Discussion

Interpretation of Principal Results

We have demonstrated the feasibility of both paid and volunteer “citizen healthcare” crowd workers to provide pediatric diagnostic information on behavioral disorders based on short video clips. We first ran a study (Study 1) with 54 AMT workers and found that there was a weak negative correlation between the time spent rating the videos and the sensitivity (ρ=–0.44, *P*=.02). We then conducted a follow-up study (Study 2) with the AMT workers who performed well (correctly classified at least 8 of the 10 videos) in Study 1 and received exclusively positive feedback about the tasks; we even received requests for more. This feedback, in conjunction with the high completion rate for the set of tasks, indicates that crowd workers who perform well in simple tasks (eg, from Study 1) are likely to not only participate in but also *enjoy* completing further tasks if they are encouraged about their prior performance and paid sufficiently.

We also found that across all videos, the average deviation between the average crowdsourced answers and the gold standard ratings was 0.56, with an SD of 0.51. Since the scales are from 0 (not severe) to 3 (severe), this deviation indicates that the crowd tends to rate within acceptable error. Most of the deviations fell within 1.0, although there was a nonnegligible number of video questions with a larger SD.

Finally, we ran the procedures from Study 1 on a public website advertised on social media and found weak correlations between certain demographic groups, due, at least in part, to the small sample sizes per category. Larger sample sizes will be required to draw significant conclusions about the inherent accuracy within or across demographic groups. A two-tailed *t* test between the scores of the paid workers in Study 1 and the unpaid workers in Study 3 showed a significant difference (t_{168}=3.37, *P*<.001). This indicates that paid workers will outperform a general unpaid crowd of online citizen workers.

Limitations

A limitation of this work includes the lack of the assessment of this crowdsourced “citizen healthcare model” in a real-world clinical setting. We are working on establishing the infrastructure to test this kind of system prospectively (see Future Work). Our current findings using publicly available YouTube videos and “uploader reported” diagnoses for this initial study lend support to the potential for such future research.

It is unclear whether results from AMT can be generalized to all paid crowdsourcing platforms. It is possible that another paid crowdsourcing platform could yield workers with higher or lower performance than those that chose to participate in our AMT studies. There were 27 well-performing workers who moved on to participate in Study 2, but testing Study 2 procedures with participants who scored <8/10 would provide additional insights into the performance of crowdsourced video raters.

We emphasize that the work performed here is a pilot study for crowdsourcing acquisition of pediatric diagnostic information from an untrained population. In future studies, it will be fruitful to explore a larger diagnostic workforce and replicate the processes described here with independent subsets of the crowd.

In terms of the volunteer-based “citizen healthcare” experiment included in this study, some of the results could have been skewed by our recruiting methodologies. We recruited participants largely via conference presentations and recruitment postings on Nextdoor [32] in the San Francisco Bay Area, California, as well as south Austin, Texas. We were ultimately only able to enroll 71 people to complete the study, but a larger-scale crowdsourced study with broader public recruitment strategies might yield emerging demographic trends that did not arise in this study due to the limited sample size.

Future Work

Future work should examine the potential of crowd workers to provide ratings about other demographic groups such as adults, individuals with other neuropsychiatric disorders, and populations in other geographic regions. Although performing a study on a public cohort of citizen raters scoring 31 multiple questions was not feasible at scale, we believe that future work should explore motivations, through mediums such as gamification, for crowd workers to participate in diagnostic microtasks for free.

Additionally, we hope to assess the feasibility of this pipeline for standard of care practice, where we use the crowd to analyze videos of children referred to developmental specialists by primary care providers. This will not only allow us to better understand the feasibility of using this system in a clinical setting but will also allow us to better assess the validity of the pipeline by utilizing videos of children who receive professional diagnoses. This will permit us to compare diagnostic outcomes from the crowd to those assigned by licensed professionals. In addition, efforts should be made into expanding the source of gold standard ratings to a larger network of expert clinical raters.

Crowdsourcing of rich video data opens the doors to understanding the forms of autism, including the potential contributions from genetics and environment, in part, due to the ability to develop an online community network and a rich digital phenotype for many subjects in a scalable and affordable fashion. Eventually, crowdsourcing could provide scientists with enough data to find the link between genetics and the behaviors present in videos [33]. This work is a step toward this goal.

Using a crowd of raters to answer questions about short structured videos of a child for mobile machine learning-aided detection and diagnosis may help to ameliorate some of the inefficiencies with the current standards of care for autism diagnosis. For families lacking the financial resources to obtain a formal diagnosis, a crowdsourced paradigm like the one tested here could be a viable alternative when provided with a proper system and feature measurement design.

Conclusion

In summary, we have shown that paid crowd workers enjoy answering screening questions from videos, suggesting higher intrinsic motivation for making quality assessments. Paid and vetted crowd workers also showed reasonable accuracy with detection of autism as well as other developmental delays in children between 2 and 5 years of age, with an average deviation <20% from professional gold standard raters, whereas parents of children with autism likely overfit their video assessments to their own affected child. These results show promise for the potential use of virtual workers in developmental screening and provide motivation for future research in paid and unpaid crowdsourcing for the diagnosis of autism and other neuropsychiatric conditions.

Acknowledgments

We thank all the crowd workers and citizen scientists who participated in the studies. These studies were supported by awards to DW by the National Institutes of Health (1R21HD091500-01 and 1R01EB025025-01). Additionally, we acknowledge the support of grants to DW from The Hartwell Foundation, the David and Lucile Packard Foundation Special Projects Grant, Beckman Center for Molecular and Genetic Medicine, Coulter Endowment Translational Research Grant, Berry Fellowship, Spectrum Pilot Program, Stanford's Precision Health and Integrated Diagnostics Center (PHIND), Wu Tsai Neurosciences Institute Neuroscience: Translate Program, and Stanford's Institute of Human Centered Artificial Intelligence as well as philanthropic support from Mr. Peter Sullivan. HK would like to acknowledge support from the Thrasher Research Fund and Stanford NLM Clinical Data Science program (T-15LM007033-35).

Conflicts of Interest

DW is the founder of Cognoa.com. This company is developing digital health solutions for pediatric care. CV, AK, and NH work as part-time consultants to Cognoa.com. All other authors declare no competing interests.

Multimedia Appendix 1

The full list of multiple-choice questions asked on Amazon Mechanical Turk.

[PDF File (Adobe PDF File), 102KB - jmir_v21i5e13668_app1.pdf]

References

1. Goldsmith T, LeBlanc L. Use of technology in interventions for children with autism. Journal of Early and Intensive Behavior Intervention 2004;1(2):166-178. [doi: 10.1037/h0100287]
2. Kogan M, Vladutiu C, Schieve L, Ghandour R, Blumberg S, Zablotsky B, et al. The Prevalence of Parent-Reported Autism Spectrum Disorder Among US Children. Pediatrics 2018 Dec;142(6):e20174161. [doi: 10.1542/peds.2017-4161] [Medline: 30478241]
3. Gordon-Lipkin E, Foster J, Peacock G. Whittling Down the Wait Time: Exploring Models to Minimize the Delay from Initial Concern to Diagnosis and Treatment of Autism Spectrum Disorder. Pediatr Clin North Am 2016 Dec;63(5):851-859 [FREE Full text] [doi: 10.1016/j.pcl.2016.06.007] [Medline: 27565363]
4. Yee H. Mobile technology for children with Autism Spectrum Disorder: Major trends and issues. 2012 Presented at: IEEE Symposium on E-Learning, E-Management and E-Services (IS3e); 2012; Kuala Lumpur, Malaysia p. 1-5.
5. Buescher A, Cidav Z, Knapp M, Mandell D. Costs of autism spectrum disorders in the United Kingdom and the United States. JAMA Pediatr 2014 Aug;168(8):721-728. [doi: 10.1001/jamapediatrics.2014.210] [Medline: 24911948]
6. Leigh P, Du J. Brief Report: Forecasting the Economic Burden of Autism in 2015 and 2025 in the United States. J Autism Dev Disord 2015 Dec;45(12):4135-4139. [doi: 10.1007/s10803-015-2521-7] [Medline: 26183723]
7. Mandell D, Novak M, Zubritsky C. Factors associated with age of diagnosis among children with autism spectrum disorders. Pediatrics 2005 Dec;116(6):1480-1486 [FREE Full text] [doi: 10.1542/peds.2005-0185] [Medline: 16322174]
8. Estellés-Arolas E, González-Ladrón-De-Guevara F. Towards an integrated crowdsourcing definition. Journal of Information Science 2012;38(2):189-200. [doi: 10.1177/0165551512437638]
9. Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, et al. Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review. J Gen Intern Med 2014 Jan;29(1):187-203 [FREE Full text] [doi: 10.1007/s11606-013-2536-8] [Medline: 23843021]
10. Créquit P, Mansouri G, Benchoufi M, Vivot A, Ravaud P. Mapping of Crowdsourcing in Health: Systematic Review. J Med Internet Res 2018 Dec 15;20(5):e187 [FREE Full text] [doi: 10.2196/jmir.9330] [Medline: 29764795]
11. Ben-Sasson A, Robins D, Yom-Tov E. Risk Assessment for Parents Who Suspect Their Child Has Autism Spectrum Disorder: Machine Learning Approach. J Med Internet Res 2018 Dec 24;20(4):e134 [FREE Full text] [doi: 10.2196/jmir.9496] [Medline: 29691210]
12. Tariq Q, Daniels J, Schwartz J, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. PLoS Med 2018 Dec;15(11):e1002705 [FREE Full text] [doi: 10.1371/journal.pmed.1002705] [Medline: 30481180]
13. Kittur A, Chi E, Suh B. Crowdsourcing user studies with Mechanical Turk. 2008 Presented at: SIGCHI conference on human factors in computing systems; 2008; Florence, Italy p. 453-456.
14. Kuang J, Argo L, Stoddard G, Bray B, Zeng-Treitler Q. Assessing Pictograph Recognition: A Comparison of Crowdsourcing and Traditional Survey Approaches. J Med Internet Res 2015 Dec 17;17(12):e281 [FREE Full text] [doi: 10.2196/jmir.4582] [Medline: 26678085]
15. Meyer A, Longhurst C, Singh H. Crowdsourcing Diagnosis for Patients With Undiagnosed Illnesses: An Evaluation of CrowdMed. J Med Internet Res 2016 Jan 14;18(1):e12 [FREE Full text] [doi: 10.2196/jmir.4887] [Medline: 26769236]

16. Ayres K, Mechling L, Sansosti F. The use of mobile technologies to assist with life skills/independence of students with moderate/severe intellectual disability and/or autism spectrum disorders: Considerations for the future of school psychology. Psychol Schs 2013 Jan 24;50(3):259-271. [doi: 10.1002/pits.21673]
17. Bereznak S, Ayres K, Mechling L, Alexander J. Video Self-Prompting and Mobile Technology to Increase Daily Living and Vocational Independence for Students with Autism Spectrum Disorders. J Dev Phys Disabil 2012 Feb 23;24(3):269-285. [doi: 10.1007/s10882-012-9270-8]
18. Daniels J, Haber N, Voss C, Schwartz J, Tamura S, Fazel A, et al. Feasibility Testing of a Wearable Behavioral Aid for Social Learning in Children with Autism. Appl Clin Inform 2018 Dec;9(1):129-140 [FREE Full text] [doi: 10.1055/s-0038-1626727] [Medline: 29466819]
19. Daniels J, Schwartz J, Voss C, Haber N, Fazel A, Kline A, et al. Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism. npj Digital Med 2018 Aug 2;1(1):32. [doi: 10.1038/s41746-018-0035-3]
20. Escobedo L, Nguyen D, Boyd L, Hirano S, Rangel A, Garcia-Rosas D, et al. MOSOCO: a mobile assistive tool to support children with autism practicing social skills in real-life situations. 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2012; Austin, Texas p. 2589-2598.
21. Keshav N, Salisbury J, Vahabzadeh A, Sahin N. Social Communication Coaching Smartglasses: Well Tolerated in a Diverse Sample of Children and Adults With Autism. JMIR Mhealth Uhealth 2017 Sep 21;5(9):e140 [FREE Full text] [doi: 10.2196/mhealth.8534] [Medline: 28935618]
22. Sahin N, Keshav N, Salisbury J, Vahabzadeh A. Second Version of Google Glass as a Wearable Socio-Affective Aid: Positive School Desirability, High Usability, and Theoretical Framework in a Sample of Children with Autism. JMIR Hum Factors 2018 Jan 04;5(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.8785] [Medline: 29301738]
23. Vahabzadeh A, Keshav N, Salisbury J, Sahin N. Improvement of Attention-Deficit/Hyperactivity Disorder Symptoms in School-Aged Children, Adolescents, and Young Adults With Autism via a Digital Smartglasses-Based Socioemotional Coaching Aid: Short-Term, Uncontrolled Pilot Study. JMIR Ment Health 2018 Mar 24;5(2):e25 [FREE Full text] [doi: 10.2196/mental.9631] [Medline: 29610109]
24. Voss C, Washington P, Haber N, Kline A, Daniels J, Fazel A, et al. Superpower glass: delivering unobtrusive real-time social cues in wearable systems. 2016 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; 2016; Heidelberg, Germany p. 1218-1226.
25. Washington P, Wall D, Voss C, Kline A, Haber N, Daniels J, et al. SuperpowerGlass. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-22. [doi: 10.1145/3130977]
26. Voss C, Schwartz J, Daniels J, Kline A, Haber N, Washington P, et al. Effect of Wearable Digital Intervention for Improving Socialization in Children With Autism Spectrum Disorder: A Randomized Clinical Trial. JAMA pediatrics 2019;Online first:online.
27. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, Wall DP. A Gamified Mobile System for Crowdsourcing Video for Autism Research. 2018 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); 2018; New York, NY p. 350-352.
28. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, Wall DP. Guess What? J Healthc Inform Res 2018 Oct 2;3(1):43-66. [doi: 10.1007/s41666-018-0034-9]
29. Nazneen N, Rozga A, Smith C, Oberleitner R, Abowd G, Arriaga R. A novel system for supporting autism diagnosis using home videos: iterative development and evaluation of system design. JMIR mHealth and uHealth 2015;3(2):e68. [Medline: 26085230]
30. Castells M, Fernandez-Ardevol M, Qiu J, Sey A. Mobile Communication and Society: A Global Perspective. Cambridge, Massachusetts: MIT Press; 2009.
31. Mosa A, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
32. Abhyanker R. United States Patent (US 8,863,245). 2014 Oct 14. Nextdoor neighborhood social network method, apparatus, and system URL: https://patentimages.storage.googleapis.com/84/aa/66/de4904c77018ab/US8863245.pdf [accessed 2019-04-20] [WebCite Cache ID 77pRGtvnr]
33. David M, Babineau B, Wall DP. Can we accelerate autism discoveries through crowdsourcing? Research in Autism Spectrum Disorders 2016 Dec;32:80-83. [doi: 10.1016/j.rasd.2016.09.001]

Abbreviations

AMT: Amazon Mechanical Turk
ASD: autism spectrum disorder

Edited by G Eysenbach; submitted 09.02.19; peer-reviewed by E Giannoula; comments to author 03.04.19; revised version received 15.04.19; accepted 16.04.19; published 23.05.19.

Please cite as:
Washington P, Kalantarian H, Tariq Q, Schwartz J, Dunlap K, Chrisman B, Varma M, Ning M, Kline A, Stockham N, Paskov K, Voss C, Haber N, Wall DP
Validity of Online Screening for Autism: Crowdsourcing Study Comparing Paid and Unpaid Diagnostic Tasks
J Med Internet Res 2019;21(5):e13668
URL: http://www.jmir.org/2019/5/e13668/
doi:10.2196/13668
PMID:31124463

Original Paper

Impact of Physician-Patient Communication in Online Health Communities on Patient Compliance: Cross-Sectional Questionnaire Study

Xinyi Lu[1], BEng; Runtong Zhang[1], PhD

[1] School of Economics and Management, Beijing Jiaotong University, Beijing, China

Corresponding Author:
Runtong Zhang, PhD
School of Economics and Management
Beijing Jiaotong University
Shangyuancun No 3, Haidian District
Beijing, 100044
China
Phone: 86 10 51683854
Email: rtzhang@bjtu.edu.cn

Abstract

Background: In China, the utilization of medical resources is tense, and most hospitals are highly congested because of the large population and uneven distribution of medical resources. Online health communities (OHCs) play an important role in alleviating hospital congestions, thereby improving the utilization of medical resources and relieving medical resource shortages. OHCs have positive effects on physician-patient relationships and health outcomes. Moreover, as one of the main ways for patients to seek health-related information in OHCs, physician-patient communication may affect patient compliance in various ways. In consideration of the inevitable development of OHCs, although they have several shortcomings, identifying how physician-patient communication can impact patient compliance is important to improve patients’ health outcomes through OHCs.

Objective: This study aimed to investigate the impact of physician-patient communication on patient compliance in OHCs through the mediation of the perceived quality of internet health information, decision-making preference, and physician-patient concordance, using an empirical study based on the self-determination theory.

Methods: A research model was established, including 1 independent variable (physician-patient communication), 3 mediators (perceived quality of internet health information, decision-making preference, and physician-patient concordance), 1 dependent variable (patient compliance), and 4 control variables (age, gender, living area, and education level). Furthermore, a Web-based survey involving 423 valid responses was conducted in China to collect data, and structural equation modeling and partial least squares were adopted to analyze data and test the hypotheses.

Results: The questionnaire response rate was 79.2% (487/615) and the validity rate was 86.9% (423/487); reliability and validity are acceptable. The communication between physicians and patients in OHCs positively affects patient compliance through the mediation of the perceived quality of internet health information, decision-making preference, and physician-patient concordance. Moreover, physician-patient communication exhibits similar impacts on the perceived quality of internet health information, decision-making preference, and physician-patient concordance. Patients’ decision-making preference shows the weakest impact on patient compliance compared with the other 2 mediators. Ultimately, all 3 mediators play a partially mediating role between physician-patient communication and patient compliance.

Conclusions: We conclude that physician-patient communication in OHCs exhibits a positive impact on patient compliance; thus, patient compliance can be improved by guiding physician-patient communication in OHCs. Furthermore, our findings suggest that physicians can share high-quality health information with patients, discuss benefits, risks, and costs of treatment options with patients, encourage patients to express their attitudes and participate in health-related decision making, and strengthen the emotional connection with patients in OHCs, thereby decreasing patients’ misunderstanding of information and increasing concordance between physicians and patients. OHCs are required to not only strengthen the management of their published health information quality but also understand users’ actual attitudes toward information quality and then try to reduce the gap between the perceived and actual quality of information.

(J Med Internet Res 2019;21(5):e12891) doi:10.2196/12891

KEYWORDS

patient portals; communication; patient compliance; consumer health information; decision making; physician-patient relations; personal autonomy

Introduction

Background

As a type of virtual community, online health communities (OHCs) are developed with the Web 2.0 technology [1,2]. OHCs are platforms for people to communicate with one another regarding health-related topics, thereby promoting the interaction between physicians and patients [3]. OHCs can break through time and space limitations as users can communicate through posts and Web-based messages without meeting each other or chatting in real time [2]. From the perspective of users, OHCs can be divided into 3 categories: (1) for patients to discuss illnesses, share experiences, and exchange information, (2) for physicians to exchange their professional knowledge, and (3) for patients and physicians, specifically, patients can communicate with physicians and seek help from physicians, and physicians can answer patients' questions and publish common health-related knowledge. This paper mainly focuses on the third category and studies the physician-patient communication in OHCs.

In OHCs, patients can conveniently ask for physicians' help anytime and anywhere in 2 main ways: posts and one-to-one communication. Therefore, patients can diagnose some simple symptoms by themselves on the basis of the information obtained from OHCs, and their privacy can be protected as the communication is not face to face [3-5]. In terms of functions, OHCs not only provide health information but also provide social support to patients [6]. For example, some psychological needs, such as self-esteem and self-efficacy, can be satisfied [7]. Patients with chronic diseases, such as diabetes, hypertension, and obesity, especially mental diseases, such as depression and schizophrenia, are more dependent on OHCs because of emotional communication [8-11]. However, OHCs also have several drawbacks. Physicians may be less likely to understand patients' illnesses because of the lack of face-to-face communication. Moreover, the quality of information obtained from OHCs cannot be guaranteed as patients are unable to ensure the identities of physicians, given the zero gatekeeping and zero-cost publishing of the internet [12]. In that case, some patients may be hesitant to seek information or communicate with physicians in OHCs or be hesitant to trust physicians or adopt health-related information provided by physicians in communication. Nevertheless, OHCs are still so popular that an increasing number of patients would like to use OHCs to seek health-related information, to connect with other patients and physicians, and to ask for support [6], because of OHCs' advantages, such as saving queueing time, developing health management, enhancing physician-patient relationships, and improving health service quality. In consideration of the advantages and inevitable development of OHCs, problems of OHCs need to be improved to assist offline treatments and help patients maintain a healthy lifestyle [7]. Communication is an important way for physicians to provide patients services and for patients to seek health information in OHCs [10]; therefore, we intend to further explore physician-patient communication in OHCs for the purpose of improving treatment efficiency.

In China, the utilization of medical resources is tense, and most hospitals are highly congested because of the large population, uneven distribution of medical resources, and low treatment efficiency. Moreover, medical resources cannot meet the daily needs of residents in some regions of China. OHCs can help alleviate hospital congestions, improve the utilization of medical resources, and relieve medical resource shortages to a certain extent [10,13,14]. Specifically, patients are able to diagnose some simple symptoms by themselves and so do not need to go to hospitals frequently and, therefore, their occupancy of medical resources can be decreased. In addition, physicians can publish health-related articles, answer patients' questions, and provide advice in OHCs when they are not busy diagnosing patients. Moreover, patients from some medically underdeveloped areas can also acquire help from physicians through OHCs. Physician-patient communication in OHCs has important effects on physician-patient relationships, patients' satisfaction, and health service accessibility. Atanasova et al [7] summarized that physicians participating in OHCs and communicating with patients can improve patients' satisfaction, enhance patients' confidence in physician-patient relationship, and increase the possibility of using health services. Wu and Lu [15] identified the impact of the service provided by physicians in OHCs on patients' satisfaction and treatment efficiency. Yang et al mainly explored how OHCs can improve treatment efficiency from the perspective of patients' satisfaction. Sarah et al [16] proposed that physicians are required to actively communicate with patients in OHCs to improve patients' satisfaction. Petrič et al [17] found that social process, such as communication in OHCs, can affect the relationship between patients and physicians. In general, physician-patient communication in OHCs can help improve treatment efficiency, as physicians can better serve patients if hospital congestions can be alleviated and medical resources can be redistributed and reused through OHCs. In addition to the physician-patient relationship and patients' satisfaction, patient compliance, which has not been widely studied, is another perspective to improve treatment efficiency through OHCs. Previous studies show that patient compliance can impact the effect of medical regimens and treatments, and high patient compliance is conducive to accelerating patient recovery and increasing treatment efficiency [18-20]. Consequently, the inefficient occupancy of medical resources can be reduced and OHCs can reuse and redistribute these resources. Therefore, this paper intends to discuss how OHCs can influence patient compliance so that we can provide a new way to improve treatment efficiency through OHCs.

OHCs can influence physician-patient relationships [21], and patient compliance is critical in physician-patient relationships and health care. Medical regimens and treatments can be effective if patients can regularly take medicines following prescriptions and keep a healthy lifestyle according to their physicians' recommendations [22]. Moreover, patients with high compliance tend to be relatively healthier than noncompliant ones [23,24]. Effective communication between physicians and patients can considerably help establish physician-patient relationships and promote effective information exchange, which is beneficial for patients obtaining health information, making suitable decisions, and ultimately producing positive results, such as improving compliance [25]. Therefore, the perceived quality of health information may be a mediator between physician-patient communication and patient compliance. Zolnierek and Dimatteo [26] corroborated that physician-patient communication can influence patient compliance through multiple mechanisms. For instance, effective communication between physicians and patients can bring benefits for support, collaborative relationships, patient-centered interviews, and consequently improving patient compliance. Roberts et al [27] concluded that effective physician-patient communication can lead to high patient compliance. Furthermore, Bultman and Svarstad [28] explored the effect of physician-patient communication on patient compliance through the mediation of satisfaction with treatments. Molfenter and Brown [29] found that patients' health beliefs play a vitally mediating role between physician-patient communication and patient compliance. Ultimately, Laugesen et al [30] determined the impact of internet health information on patient compliance and proposed that internet health information can improve the communication between physicians and patients and then improve patient compliance. Therefore, there may be several mediators between physician-patient communication in OHCs and patient compliance. In the context of OHCs, we considered the perceived quality of internet health information as one mediator. The internet provides patients new platforms, such as OHCs, to communicate with physicians and promotes them to participate in making health-related decisions [31]; thus, the decision-making preference may also be a mediator between physician-patient communication in OHCs and patient compliance. In the physician-patient relationship, physician-patient concordance is a critical element. High-quality communication can promote the exchanging and understanding of opinions between physicians and patients; therefore, the opinions of patients and physicians may tend to be identical. Laugesen et al [30] identify the significant impact of physician-patient concordance on patient compliance in the context of the internet. Therefore, we take into account the mediation of physician-patient concordance.

This study aimed to identify the impact of physician-patient communication in OHCs on patient compliance from the perspective of psychology, attempting to guide patient compliance through communication in OHCs. As a complex field, behavioral psychology has received attention from researchers in recent years, especially for its application in the study of health [32]. However, patient compliance, which is a dynamic parameter and may be easy to change because of psychological factors, has not been paid sufficient attention [33,34]. In addition, factors such as basic national conditions and national policies have created a unique medical system in China, which is considerably different from foreign medical systems. Therefore, the theoretical and practical results of OHC services in foreign countries cannot be directly applied to Chinese scenarios. Research on OHCs in China is still in its infancy. In this study, on the basis of the self-determination theory, we intended to explore how physician-patient communication in OHCs affects patient compliance through mediations of the perceived quality of internet health information, decision-making preference, and physician-patient concordance to fill in the gap of research and practice.

Physician-Patient Communication

Collaborative medical interactions exhibit considerable relevance to health care outcomes [35]; thus, physician-patient communication plays an important role in the health care system. Thomas et al [36] claim that physician-patient communication remarkably affects patient response to treatments and, consequently, the quality of diagnoses. Therefore, high-quality physician-patient communication brings benefits to improving health care quality and patients' health outcomes [37,38], and the communication quality has been regarded as one of the critical elements of health literacy. With the development of the internet, a new form of physician-patient communication, internet communication, has emerged. Patients are increasingly willing to communicate with physicians on the Web because of the advantages of physician-patient internet communication [39]. For instance, the internet provides patients a platform to seek health-related information to self-manage and self-monitor their health and diagnose some nonurgent medical problems [40,41]. In that case, patients may not have to wait for a long time for a simple illness, and congestions in hospitals are likely to be alleviated [10,13,14].

OHCs are one of the main channels for patients to communicate with physicians on the Web. On the one hand, patients can engage with physicians without going to hospitals instead of only seeking health-related information from the internet. In addition, patients can also obtain additional information about their physicians if they contact the physicians before going to the hospital, which may reduce the uncertainty and sense of risk [14]. On the other hand, the majority holds the view that the internet helps individuals improve their abilities of communication [42]. Moreover, between patients and physicians, communicating without meeting each other is less likely to cause conflicts, thereby strengthening the physician-patient relationship. Accordingly, patients are likely to be satisfied and trust their physicians [14,43] and may be willing to comply. However, physician-patient communication in OHCs exhibits shortcomings, which may negatively influence patient compliance in turn. For instance, accurately diagnosing patients without face-to-face communication is difficult for physicians, as they cannot observe patients' breath, sound, or facial expressions [40]. In this situation, patients may regard their physicians as unprofessional and then refuse to comply with treatments [44]. Therefore, this paper aimed to explore how to guide physician-patient communication in OHCs to improve patient compliance.

Patient Compliance

Generally, the effect of treatments depends on 2 aspects: (1) whether the treatment proposed by physicians is the correct remedy and (2) whether patients comply with the treatment [30]. Given that we cannot control the degree of physicians' professionalism, we pay increasing attention to patient compliance to improve the curative effects. Patient compliance is used to measure how patients follow medical diagnoses and treatment regimens recommended by their physicians [30], including medicine adherence and maintaining a healthy lifestyle [33]. Moreover, patient compliance plays a vital role in physician-patient relationships and health care systems, especially from the perspective of chronic diseases, whose recovery relies on patients' self-management and self-monitoring [23]. If medical diagnoses and treatment regimens are proper, treatments can be increasingly effective when patients' daily habits are consistent with physicians' advice [22]. Therefore, patients with high compliance are likely to be relatively healthier than those with low compliance [23]. Khera et al [45] and Johal et al [46] validated that a healthy lifestyle is good for preventing cardiovascular diseases. With the increasing number of patients with chronic diseases, patient compliance becomes increasingly important.

High compliance is conducive to patients' health outcomes, whereas low compliance or noncompliance may cause negative consequences related to patients, the economy, and the society. Diseases may be hard to treat if patients refuse to comply with physicians. According to Varleta et al [47], blood pressure may be difficult to control without regularly taking medicines following prescriptions. In that case, medical resources will be wasted, given that they fail to work. Moreover, some types of medicines or therapies may be regarded as invalid, and then their use may be terminated. When this phenomenon occurs in clinical practice, medical productivity may be negatively influenced [48,49]. Furthermore, patient compliance is a dynamic parameter [34]; thus, this study aimed to discuss various methods to improve patient compliance.

Perceived Quality of Internet Health Information

Benefitting from the development of the internet [50], people can gain access to internet-based health-related information that is divided into 2 categories [51-53]: (1) health care information, which is related to diseases, medicines, physicians, hospitals, and therapies and (2) health lifestyle information, which guides not only patients but also all individuals to keep a healthy lifestyle and prevent diseases. Numerous portals have been built by governments, medical institutions, and business corporations to publish various health information on the Web [54], and OHCs are one of the main types of these portals. For specific needs [55], patients enter OHCs to obtain health information in 2 ways: searching for information and communicating with physicians. Physician-patient communication in OHCs is patient-active. If physicians do not respond to patients, it cannot be considered as an entire physician-patient communicating process. Therefore, in this study, patients can be regarded as having obtained health information from physicians as long as they communicate with physicians, regardless of the quality and quantity of information. Compared with the health information from other sources, internet health information can be published on time and be obtained conveniently and quickly because of zero gatekeeping and zero-cost publishing [56]. However, internet health information lacks monitoring, and the quality of internet health information has become a serious problem [57].

Individuals with different health literacy levels exhibit different levels of ability to distinguish the quality of internet health information and perceive different information quality. From the perspective of information users, the perceived quality is different from the actual quality. Sporadically, a person may encounter a piece of high-quality health information related to a specific topic but may consider this information low in quality, in other words, the perceived quality is low. This situation may result from many factors. For instance, the channel that publishes this information may be low in quality and cannot convince users. In addition, the person's cognition on this topic may be wrong; thus, he or she may consider the information to be low in quality.

Decision-Making Preference

The internet provides patients with opportunities to contact physicians and obtain health information. Thus, some patients' health literacy can be improved, and they may be willing to play active roles in health-related decision making [58]. A survey conducted by Jr et al [59] verifies that participating in decision-making activities helps patients achieve improved health outcomes. In addition, Beaver et al [60] proved that patients with psychological diseases are less likely to relapse if they participate in decision making. Therefore, an increasing number of patients are encouraged to be involved in decision making in treatments [59]. However, in the meantime, some patients prefer to rely on their physicians and want physicians to be the main or only decision maker [61]. From the viewpoint of some patients, some physicians may dislike making shared decisions with patients, and therefore, patients cannot participate in decision making; otherwise, physicians may think that their patients are questioning their professions.

Actually, physicians are willing to discuss benefits, risks, and costs of treatment options with patients, encourage patients to positively participate in decision making, and make decisions after considering patients' views [61,62], which is beneficial for the improvement of patient satisfaction, enhancement of physician-patient relationship, and improvement in health care quality [62]. Participating in decision making does not indicate that patients make decisions by themselves regardless of physicians' professional advice. On the contrary, participation requires sufficient communication between physicians and patients [63]. When patients' opinions are contrary to physicians' opinions, physicians need to patiently explain to patients and make efforts to be consistent with patients. Moreover, as the core of patient-centered care [64], shared decision making plays a key role in high-quality physician-patient communication [65]. Even if patients do not intend to make decisions by themselves, they still want to know treatment options from their physicians. Hence, physicians are required to positively understand their patients' preferences of decision making instead of waiting for patients' inquiries.

Patients' preference for participating in making decisions is dynamic [66]. Vogel et al [67] confirmed that lung cancer patients' decision-making preference could be stable in the short term but remarkably change after more than 3 months. A number of researchers carried out studies to identify the factors that influence patients' decision making. Furthermore, Cajita et al [68] explored that some demographic factors (age, gender, education level, and health condition) can affect a patient's decision-making preference. For instance, patients' preference for making health-related decisions will change with age. Moise et al [65] concluded that depressive symptoms have an impact on patients' decision-making preference and that patients with elevated depressive symptoms prefer clinician-directed decision making. In addition, Harvey et al [69] proved that patients' preferences for participating in decision making vary by their circumstances. Therefore, they proposed that physicians can encourage patients to participate actively in making health-related decisions by sharing information and helping process information to reduce conflicts and improve patient compliance. Currently, patients seek health information on the Web and take the internet as a new open channel to communicate with physicians; thus, they are more willing to participate in decision making [31]. Therefore, decision making may be a mediator between physician-patient communication and patient compliance.

Physician-Patient Concordance

Physician-patient concordance indicates that patients and physicians equally discuss treatment options [70] and then reach an agreement in terms of medical diagnoses and treatment regimens [21,71], which is important in physician-patient interactions [72]. According to previous studies, when the opinions of patients and physicians tend to be identical, patients' satisfaction and health care outcomes will be high [72,73]. Patients may be able to manage themselves well and not need further consultation, thereby reducing health care costs [72]. However, a gap about medical knowledge exists between patients and physicians because of the professionalism of medicine; thus, patients' expectations of health outcomes may sometimes be different from those of physicians, and it may be difficult for physicians to propose medical treatments that are consistent with patients' values, goals, and priorities [74]. Accordingly, patients' health outcomes and physician-patient relationships may be negatively affected. Therefore, physician-patient concordance must be improved.

A considerable number of researchers have conducted surveys on physician-patient concordance, comprising its pattern of manifestation, advantages, shortcomings, and influencing factors. Shin et al [73] proved that increased efforts by physicians to understand demands of patients with cancer for the achievement of concordance are beneficial in improving patient compliance. In terms of influencing factors, the impact of patients' sociodemographic characteristics on concordance has been identified, but other factors have not been fully understood. Moreover, Gross et al [72] deduced that physicians maintaining a long-term relationship with patients can help enhance concordance, whereas ignoring patients' accurate demands may not directly affect concordance in the consultation process but may affect follow-up treatments.

Self-Determination Theory

The self-determination theory was proposed by Deci and Ryan [75], and it is the only theory of human motivation that determines autonomy as a human need and has been applied to the fields of health care, education, and organization [76]. According to the self-determination theory [77], when people perceive high autonomy support from social events, their intrinsic motivation of activities will be enhanced [78]. By contrast, perceiving being controlled or forced in social events is more likely to decrease an individual's intrinsic motivation of activities [79]. In the context of health care, Ng et al [76] inferred that patients' health-related behaviors may be positively motivated by their autonomy orientations. Specifically, supporting autonomy for patients can satisfy 3 basic aspects of psychological needs (ie, autonomy, relatedness, and competence) [75,77,80] and then patients will begin their health-related behaviors [76], such as quitting smoking, weight control, and medicine adherence [79]. By contrast, if the satisfaction is low, then self-determined motivation may be reduced, and people may involuntarily behave negatively, such as be noncompliant [80].

Autonomous self-regulation plays a critical role in health care, and Ng et al [76] concluded that the self-determination theory is viable to studying the motivation for health-related behaviors. This paper attempted to explore how physician-patient communication in OHCs can impact patients' motivation and compliance to treatments from the mediating perspective of the perceived quality of internet health information, decision-making preference, and physician-patient concordance by employing the self-determination theory.

Model and Hypotheses

This paper intended to explore the impact of physician-patient communication in OHCs on patient compliance by establishing the research model (see Figure 1), including 1 independent variable (physician-patient communication), 3 mediators (perceived quality of internet health information, decision-making preference, and physician-patient concordance), and 1 dependent variable (patient compliance).

Figure 1. Research model. H1-H6: hypothesis number.

Patient compliance is a dynamic parameter, and noncompliance or low compliance may be involuntary [81]. Therefore, this study focused on patients' autonomy to improve patient compliance. As one of the main contents of physician-patient communication, information delivery is advantageous to obtaining health-related information and knowledge, helping patients understand their health and conditions and improving patient satisfaction with health care services [10]. Lin et al [43] proved that physician-patient communication on the Web, which relies on an internet-based channel, can be more likely to promote patient satisfaction compared with offline communication [39]. In addition, Web-based communication can help improve patients' trust in their physicians and the frequency of using OHCs [14]. High trust and satisfaction with physicians make patients perceive autonomy support from physicians, make patients willing to obtain health information from physicians, and make patients likely to trust this information [55]. In that case, internet health information obtained from physicians in OHCs through communication can be fully processed; thus, patients' psychological demands for the information may be satisfied, and they may perceive high-quality health information. Therefore, we suggested the following hypothesis: (H1) Physician-patient communication has a positive impact on patients' perceived quality of internet health information.

Promoting physician-patient communication means encouraging patients to inform physicians of their health conditions and urging physicians to discuss treatment options with patients instead of making decisions themselves. A patient's preference of participating in making decisions is dynamic [66] and thus may change with the progress of communication. The self-determination theory highlights that people who perceive support of autonomy will likely be motivated to change their behaviors [75,79]. When patients and their physicians positively communicate with each other, patients can feel that physicians provide opportunities for them to participate in decision making. Hence, they will perceive autonomy supported by physicians, and then they may actively participate in making decisions and assisting physicians. By contrast, without autonomy support, patients may be overwhelmed by the feeling of being controlled or coerced, which leads to the weakened intrinsic motivation of activities. Accordingly, we proposed the following hypothesis: (H2) Physician-patient communication has a positive impact on patients' decision-making preference.

Physicians and patients tend to have differences in their cognitions [70], but this divergence may be reduced if they can communicate with each other. Specifically, communication makes physicians understand patients' health conditions, and patients can feel increased autonomy support from their physicians. In that case, the self-determination theory proposes that patients' psychological needs can be satisfied and that they are willing to agree with their physicians. In addition, empathy can be produced between physicians and patients, resulting in high concordance. According to Gross et al [72], patients are more likely to be concordant with their long-term physicians, as the communication may be frequent and positive in this relationship. By contrast, if physicians are unwilling to explain benefits, risks, and costs of treatment options or if patients refuse to describe the details of illness, the quality of communication is bad, which may cause suspicion. Moreover, in such an inefficient treatment environment with low support of autonomy, patients may sometimes question their physicians and be reluctant to cooperate with treatments [75]. Consequently, we suggested the following hypothesis: (H3) Physician-patient communication has a positive impact on physician-patient concordance.

Perceiving internet health information as high in quality may make patients think that their physicians are professional and indeed share health-related information with them, thereby enabling patients to feel being allowed or inspired by physicians to participate in medical decision making, and their perceived autonomy support is increased. Under the guidance of the self-determination theory [75], perceived autonomy support may encourage patients to positively change their medical behaviors, such as complying with treatments. In addition, high-quality information may help patients extend the knowledge and experience related to the importance of self-regulation and self-management [76], which can help improve patient compliance with medical diagnoses and treatment regimens. Laugesen et al [30] identified a weak, indirect effect of internet health information quality on patient compliance through mediations of perceived information asymmetry and physician-patient concordance, whereas Lu et al [33] found an indirect effect of internet health information quality on patient compliance by means of the mediating role of affect-based trust. Therefore, the indirect effect of internet health information on compliance may be associated with the mediators, background, sample, and other possible factors. What we can ensure is that the perceived quality of internet health information indeed has an effect on patient compliance. Therefore, we attempted to examine the direct effect of the perceived quality of internet health information on patient compliance. Hence, we proposed the following hypothesis: (H4) Patients' perceived quality of internet health information has a positive impact on patient compliance.

From the perspective of patients, a high decision-making preference does not mean that they prefer to make decisions by themselves rather than consider physicians' advice. Generally, physicians mainly decide the final medical options, and patients' decision-making preference represents the degrees at which patients participate in making decisions. Highly participating in health-related decision making means patients tend to be more autonomous. According to the self-determination theory, autonomy may encourage patients to behave positively in treatments, such as medicine adherence. Therefore, patients who prefer to participate in decision making are likely to perceive autonomy support from their physicians, which may help enhance their internal sense of health care and develop their ability of self-regulation [76], ultimately improving patient compliance with treatments. Therefore, this situation led us to derive the following hypothesis: (H5) Patients' decision-making preference has a positive impact on patient compliance.

High concordance between physicians and patients is the ideal consequence of treatments. In this situation, patients will highly agree with their physicians; thus, complying with physicians also means following their own choices. Similar to the discussion of the perceived quality of internet health information, patients assume positive attitudes toward their physicians, and they assume a perception of high supportive autonomy that promotes self-regulation of healthy behaviors [76], such as patient compliance, which may be present, according to the self-determination theory. Moreover, Laugesen et al [30] found a direct impact of physician-patient concordance on patient compliance. Hence, we proposed the following hypothesis: (H6) Physician-patient concordance has a positive impact on patient compliance.

Methods

Instrument Development

To guarantee reliability and validity, we adopted the previous scales validated by published works to measure variables in the research model (see Figure 1) with a 5-point Likert-type response format that ranged from *strongly disagree* to *strongly agree*, as shown in Multimedia Appendix 1. In the context of OHCs, this study adapted a 14-item scale from Makoul et al [82] to measure the communication between physicians and patients. The perceived quality of internet health information, which was once examined by Laugesen et al [30], was measured using a 16-item scale. In addition, physician-patient concordance and patient compliance under the background of internet health information were also discussed in the study by Laugesen et al [30], with 2 different 5-item scales. To address the subject of this study, we adapted these 2 scales to measure physician-patient concordance and patient compliance in OHCs. Ultimately, the decision-making preference of patients was measured using a 6-item scale by Aoki et al [83].

The next step involved translating the instrument into Chinese, given that the survey would be conducted in China and our subjects were Chinese individuals who have communicated with physicians in OHCs. Referring to the similar translation process by previous studies [84,85], we first recruited native Chinese speakers with at least a master’s degree and who were good at speaking English and scientific research translation to translate the scales into Chinese, considering the cross-cultural adaptation [86]. Second, we invited individuals with experiences in communicating with physicians in OHCs and who were from different backgrounds of ages, genders, and educational levels to complete the questionnaire, provide recommendations for modifying scales, and then improve comprehensibility, conciseness, appropriateness, and readability. The reverse translation process was the last necessary step, which was important to ensure that our scales were conceptually consistent with the original English version [30,87,88].

Analysis Tool Selection

Structural equation modeling (SEM) is useful in analyzing the causal relationships of research models, including mediators, and accommodating intricate causal networks [30], and it can help incorporate the measurement error and detect effects [89]. This study adopted the partial least squares (PLS)-SEM method to analyze the research model and used SmartPLS version 3.2.8 (SmartPLS GmbH, Bönningstedt, Germany) to analyze the collected data and test hypotheses, drawing lessons from previous studies [30,90].

Data Collection and Respondent Profile

The subjects of this investigation were Chinese individuals who have communicated with physicians in OHCs within the previous month to ensure that they could recall their relevant experiences. With the help of a medical association in China, the formal investigation was conducted in May 2018, and the questionnaires were sent to 615 participants. The participants’ informed consent was secured, and we committed that their privacy would be strictly protected. We used a Web-based platform to create and maintain the questionnaire, and participants also filled this questionnaire through the platform. This platform can help record the completion time of each response; therefore, the response, whose completion time is obviously lower than the average time, is regarded as invalid. In addition, the response that was not completed or that missed at least 1 answer is also invalid. Finally, we received 487 responses, and 423 of them were valid. Therefore, the response and validity rates were 79.2% (487/615) and 86.9% (423/487), respectively. Table 1 shows the demographics of the sample; we found that 59.3% (251/423) participants were aged 20 to 40 years, 53.6% (220/423) of the participants were female, and 51.3% (217/423) of the participants held at least a bachelor’s degree. Thus, more than half of the subjects were young, female, and highly educated, which was consistent with the characteristics of OHCs’ users and met our requirements [2,91,92].

Table 1. Sample demographics (N=423).

Demographic characteristics	n (%)
Age (years)	
<20	19 (4.5)
20-29	127 (30.0)
30-39	124 (29.3)
40-49	97 (22.9)
50-59	49 (11.6)
60 and above	7 (1.7)
Gender	
Male	203 (48.0)
Female	220 (52.0)
Living area	
Urban	240 (56.7)
Rural	183 (43.3)
Education	
Junior middle school	22 (5.2)
High school	60 (14.2)
Junior college	124 (29.3)
Bachelor’s degree	159 (37.6)
Master’s degree	48 (11.3)
Doctorate	10 (2.4)

Results

Data Analysis

To identify the effect of demographic factors on relationships in the research model and to adjust the results, we added age, gender, living area, and education level into the research model as control variables. Although this study used previous validated scales to measure variables, we reevaluated the reliability and validity because of differences in backgrounds and participants. We calculated the Cronbach alpha of each construct using SmartPLS software version 3.2.8, as shown in Table 2, and all values were greater than the cut-off value of .700 [88,93], which indicated a good reliability of scales.

Table 3 provides the composite reliability (CR) and the average variance extracted (AVE) of constructs, and Table 4 shows the correlations between each of the 2 constructs. The convergent validity of scales was acceptable, as all the CR values exceeded the cut-off value of .700 and all the AVE values exceeded the cut-off value of .500. For each construct, the square root of AVE was above each correlation between other construct and itself. Therefore, the discriminant validity was acceptable [94,95].

Table 2. Cronbach alpha of constructs.

Constructs	Cronbach alpha
Physician-patient communication	.909
Perceived quality of internet health information	.919
Decision-making preference	.749
Physician-patient concordance	.750
Patient compliance	.787

Table 3. Composite reliability and average variance extracted.

Construct	Composite reliability	Average variance extracted	Square root of average variance extracted
Physician-patient communication	.940	.527	.726
Perceived quality of internet health information	.947	.528	.727
Decision-making preference	.866	.519	.721
Physician-patient concordance	.869	.570	.755
Patient compliance	.864	.561	.749

Table 4. Correlations between each of the 2 constructs.

Construct	PPCOM[a]	PQIHI[b]	DMP[c]	PPCON[d]	PC[e]
PPCOM	1.000	—[f]	—	—	—
PQIHI	.724	1.000	—	—	—
DMP	.624	.665	1.000	—	—
PPCON	.705	.722	.617	1.000	—
PC	.725	.700	.595	.686	1.000

[a] PPCOM: Physician-patient communication.

[b] PQIHI: Perceived quality of internet health information.

[c] DMP: Decision-making preference.

[d] PPCON: Physician-patient concordance.

[e] PC: Patient compliance.

[f] Not applicable.

Hypothesis Testing

According to the results by SmartPLS 3.2.8, we corroborated that age positively affected patient compliance in OHCs. Specifically, older patients were more willing to comply with medical regimens and treatments recommended by physicians compared with younger patients. In terms of gender, females were more likely to perceive high-quality health information in OHCs and exhibited high compliance with physicians compared with males. Furthermore, the educational level positively affected decision-making preference, indicating that patients who were highly educated would be more likely to participate in health-related decision making. This study used Cohen f^2 [96] to evaluate the effects of control variables, and Table 5 shows the results of the multivariate coefficient of determination (R^2). Ultimately, we contended that control variables had limited or insignificant effects on the research model.

Results of the PLS-SEM can be observed in Figure 2, and Table 6 shows the magnitude and significance of the path coefficients. All 6 hypotheses were supported; specifically, physician-patient communication had positive impacts on perceived quality of internet health information, decision-making preference, and physician-patient concordance. In addition, the perceived quality of internet health information, decision-making preference, and physician-patient concordance all positively affected patient compliance. Table 7 presents the effects of constructs in the research model, which indicates that physician-patient communication had a strong impact, with large effect sizes on the perceived quality of internet health information, decision-making preference, and physician-patient concordance, and the impact of perceived quality of Internet health information, decision-making preference and physician-patient concordance on patient compliance were all weak with small effect sizes.

Table 5. Multivariate coefficient of determination (R^2) results.

Variables	R^2		Control variable effects		
	With control variables	Without control variables	ΔR^{2a}	f^{2b}	Effects
Perceived quality of internet health information	0.532	0.524	0.008	0.017	Insignificant
Decision-making preference	0.394	0.389	0.005	0.008	Insignificant
Physician-patient concordance	0.501	0.496	0.005	0.010	Insignificant
Patient compliance	0.588	0.570	0.018	0.044	Small

[a] ΔR^2: $R^2_{\text{with control variables}} - R^2_{\text{without control variables}}$.

[b] f^2: Cohen f^2.

Figure 2. Research model with path coefficients. H1-H6: hypothesis number.

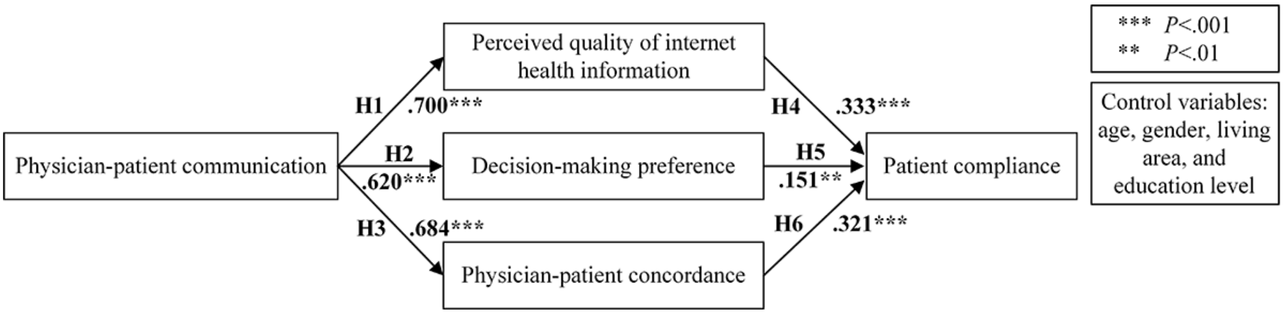


Table 6. Hypothesis testing.

Hypothesis	Path coefficient	t test	P value
Physician-patient communication has a positive impact on patients' perceived quality of internet health information	.700	18.693	<.001
Physician-patient communication has a positive impact on patients' decision-making preference	.620	16.629	<.001
Physician-patient communication has a positive impact on physician-patient concordance	.684	19.677	<.001
Patients' perceived quality of internet health information has a positive impact on patient compliance	.333	4.569	<.001
Patients' decision-making preference has a positive impact on patient compliance	.151	3.002	.003
Physician-patient concordance has a positive impact on patient compliance	.321	3.951	<.001

Table 7. Partial least squares effect size analysis.

Constructs	R^{2a}		ΔR^{2b}	f^{2c}	Effect size
	In	Out			
Patient compliance					
Perceived quality of internet health information	0.588	0.545	0.043	0.104	Small
Decision-making preference	0.588	0.577	0.011	0.027	Small
Physician-patient concordance	0.588	0.543	0.045	0.109	Small
Perceived quality of internet health information					
Physician-patient communication	0.532	0.094	0.438	0.936	Large
Decision-making preference					
Physician-patient communication	0.394	0.050	0.344	0.568	Large
Physician-patient concordance					
Physician-patient communication	0.501	0.083	0.418	0.838	Large

[a] R^2: Multivariate coefficient of determination.

[b] ΔR^2: $R^2_{\text{with control variables}} - R^2_{\text{without control variables}}$.

[c] f^2: Cohen f^2.

Table 8. Path coefficients by bootstrapping.

Effect	Path coefficients (SD)	*P* value
Direct effects		
PPCOM[a]→PQIHI[b]	0.703 (0.038)	.000
PPCOM→DMP[c]	0.623 (0.038)	.000
PPCOM→PPCON[d]	0.687 (0.035)	.000
PQIHI → PC[e]	0.215 (0.066)	.001
DMP→PC	0.094 (0.045)	.04
PPCON→PC	0.209 (0.082)	.010
PPCOM→PC	0.339 (0.067)	.000
Total effects		
PPCOM→PC	0.693 (0.035)	.000

[a] PPCOM: Physician-patient communication.

[b] PQIHI: Perceived quality of internet health information.

[c] DMP: Decision-making preference.

[d] PPCON: Physician-patient concordance.

[e] PC: Patient compliance.

To further evaluate the mediating effects in the research model, we conducted an additional analysis using the bootstrapping method (n=5000, 95% CI). As shown in Table 8, the total effect of physician-patient communication on patient compliance was significant, and the direct effects of physician-patient communication on the 3 mediators and of the 3 mediators on patient compliance were all significant. Therefore, we used the Sobel test to assess the mediating role played by the 3 mediators between physician-patient communication and patient compliance. For the perceived quality of internet health information, the value of $Z_{\text{perceived quality of internet health information}}$ was 3.028, which was significantly greater than 1.960, indicating that the mediation of perceived quality of internet health information was significant. In addition, for decision-making preference, the value of $Z_{\text{decision-making preference}}$ was 2.072, which was significantly greater than 1.960, indicating that the mediation of decision-making preference was significant. Furthermore, for physician-patient concordance, the value of $Z_{\text{physician-patient concordance}}$ was 2.528, which was significantly greater than 1.960, indicating that the mediation of physician-patient concordance was significant. We can thus conclude that the perceived quality of internet health information, decision-making preference, and physician-patient concordance all played a partially mediating role between physician-patient communication and patient compliance, given that the direct effect of physician-patient communication on patient compliance was significant.

Discussion

Principal Findings

This study is the first that explores the impact of physician-patient communication in OHCs on patient compliance, and it makes theoretical contributions and practical implications for future studies on physician-patient communication and for guiding patient compliance through OHCs from the perspective of psychology. First, we constructed a research model to clarify the mechanisms through which physician-patient communication in OHCs impacts patient compliance by employing the self-determination theory. Previous studies have mainly focused on the relationship between offline physician-patient communication and physician-patient relationship and health outcomes, whereas the impact of communication in OHCs on patient compliance remains to be more focused, as OHCs are still in the stage of development, especially in China. Therefore, this study enriches theoretical researches on OHCs and patient compliance and improves the deficiencies of studies on strengthening patient compliance through communication in OHCs in China. In addition, this study used the self-determination theory to promote hypotheses and identify the motivation of patient compliance from the perspective of OHCs, and it enriches the application of the self-determination theory in the field of health-related behavior. The communication between physicians and patients in OHCs indirectly and positively affects patient compliance through the mediations of the perceived quality of internet health information, decision-making preference, and physician-patient concordance. Therefore, physician-patient communication in OHCs is beneficial for improving patient compliance.

Second, path coefficients from physician-patient communication to the 3 mediators are similar, and physician-patient communication just has a slightly stronger impact on the perceived quality of internet health information compared with decision-making preference and physician-patient concordance. Laugesen et al [30] validated that internet health information quality exerted a weak impact on physician-patient concordance. We speculated that physician-patient communication may have an indirect impact on physician-patient concordance through the mediation of perceived quality of internet health information; thus, the direct impact of physician-patient communication on physician-patient concordance was slightly weaker than that of physician-patient communication on the perceived quality of internet health information. The main purpose of patients using OHCs is to seek health-related information, including physicians, therapies, medicine, and other medical knowledge. As an important form of using OHCs, the communication of patients with physicians significantly affects patients' perceived quality of internet health information. Moreover, we identified the significant impact of the perceived quality of internet health information on patient compliance, which was supported by Laugesen et al [30], who also claimed that high-quality internet health information can help enhance patient compliance. Furthermore, physicians are required to share high-quality information with their patients and guarantee patients' perceived quality of information through communication in OHCs. OHCs must focus on the quality of their published health information. On the one hand, OHCs should strengthen the management of internet health information quality, which involves not only information itself but also users who publish and share information with other users. On the other hand, OHCs can conduct investigations to obtain feedback from their users so that they can understand the gap between the actual and perceived quality of information and then make efforts to improve the perceived information quality.

Third, the path coefficient from a patient's decision-making preference to patient compliance is the smallest among relationships from the 3 mediators to patient compliance, and the path coefficients from the other 2 mediators to patient compliance are similar, implying that the impact of patient's decision-making preference on patient compliance is weaker than the impact of the perceived quality of internet health information and physician-patient concordance on patient compliance. This finding is similar to the results of Laugesen et al [30], who confirmed that physician-patient concordance strongly, positively, and directly impacts patient compliance. Moreover, they did not determine the strong impact of internet health information quality on patient compliance, which is different from our finding. We speculated that this is because we discussed the direct impact of the perceived quality of internet health information, whereas Laugesen et al [30] considered the indirect impact. In addition, we narrowed the research background into OHCs. Findings suggest that physician-patient concordance is also a significant perspective, which promotes patient compliance by improving physician-patient communication in addition to patients' perceived quality of internet health information. Physicians can seek concordance with patients in treatments to improve patient compliance. For instance, physicians can explain benefits, risks, and costs of treatment options [68] to patients, as a gap of medical knowledge may exist between physicians and patients; it may be useful to reduce the difference in cognition and increase concordance. Encouraging patients to tell their health conditions and express their attitudes toward treatments can help physicians learn additional information about patients and propose suitable recommendations, which may achieve an increased sense of identity from patients. In addition, physicians are encouraged to strengthen the emotional connection with patients to provide emotional autonomy support, which can satisfy patients' psychological needs and improve patient compliance [75].

Finally, compared with the perceived quality of internet health information and physician-patient concordance, decision-making preference shows the weakest effect on the relationship between physician-patient communication and patient compliance, but it can also be a perspective to improve patient compliance. To maintain high compliance, physicians can encourage their patients to participate in decision making in OHCs. Although patients may be unable to assist in making any decision, physicians are required to inform patients of medical options and understand patients' real ideas related to decision making.

In that case, patients can perceive the support of autonomy in treatments and think that the decision that they follow is partly made by themselves and then be likely to comply with the treatment.

Limitations

Several limitations and prospects in this study must be considered. First, this study used the perceived quality of internet health information, decision-making preference, and physician-patient concordance as mediators, and other variables can be discussed in future studies. Second, the development of health care is special in China because of its large population and uneven distribution of medical resources. The effect of OHCs on health care in China may be different from that in other countries. Therefore, the similarities and differences between China and other countries can be explored in further studies. Third, this study only collected data through a cross-sectional survey once; hence, we were unable to dynamically capture the changes of participants' attitudes toward all variables. Finally, we only matched the sample with the characteristics of typical OHCs' users but did not consider the Chinese population. We originally intended to consider Chinese census data but found it difficult because of China's large population. Ultimately, we believe that future studies may be able to address this issue.

Conclusions

This study indicates that physician-patient communication in OHCs positively impacts patient compliance through mediations of the perceived quality of internet health information, decision-making preference, and physician-patient concordance. In our research model, physician-patient communication shows similar effects on the perceived quality of internet health information, decision-making preference, and physician-patient concordance, and patient's decision making has the weakest impact on patient compliance compared with the other 2 mediators. In terms of the mediation, all 3 mediators play a partially mediating impact on the relationship between physician-patient communication and patient compliance. In addition, these findings suggest the following: (1) physicians can share high-quality health information with patients, ask patients' real opinions about information, and make efforts to decrease patients' misunderstanding of information; (2) physicians can discuss benefits, risks, and costs of treatment options with patients, encourage patients to express their attitudes and participate in health-related decision making, and strengthen the emotional connection with patients to provide emotional autonomy support in OHCs; and (3) OHCs can not only strengthen the management of their published health information quality but also understand users' actual attitudes toward information quality and then try to reduce the gap between the perceived and actual quality of information.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (key program; grant number 71532002), National Social Science Foundation of China (major project; grant number 18ZDA086).

Authors' Contributions

All 2 authors contributed to this study. XL and RZ conceived and designed the study, developed the research model, designed the questionnaire, conducted data collection and analysis, and drafted as well as modified the manuscript. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Measurement instruments.

[DOC File, 116KB - jmir_v21i5e12891_app1.doc]

References

1. Luo P, Chen K, Wu C, Li Y. Exploring the social influence of multichannel access in an online health community. J Assoc Inf Sci Technol 2018 Jan;69(1):98-109. [doi: 10.1002/asi.23928]
2. Yan Z, Wang T, Chen Y, Zhang H. Knowledge sharing in online health communities: a social exchange theory perspective. Inform Manage 2016 Jul;53(5):643-653. [doi: 10.1016/j.im.2016.02.001]
3. Lamberg L. Online empathy for mood disorders: patients turn to internet support groups. J Am Med Assoc 2003 Jun 18;289(23):3073-3077. [doi: 10.1001/jama.289.23.3073] [Medline: 12813097]
4. Jiang S, Street RL. Pathway linking Internet health information seeking to better health: a moderated mediation study. Health Commun 2017 Dec;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
5. Wu B. Patient continued use of online health care communities: web mining of patient-doctor communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]

6. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
7. Atanasova S, Kamin T, Petrič G. Exploring the benefits and challenges of health professionals' participation in online health communities: emergence of (dis)empowerment processes and outcomes. Int J Med Inform 2017 Feb;98:13-21. [doi: 10.1016/j.ijmedinf.2016.11.005] [Medline: 28034408]
8. Lu Y, Wu Y, Liu J, Li J, Zhang P. Understanding health care social media use from different stakeholder perspectives: a content analysis of an online health community. J Med Internet Res 2017 Apr 7;19(4):e109 [FREE Full text] [doi: 10.2196/jmir.7087] [Medline: 28389418]
9. Yan L, Tan Y. Feeling blue? Go online: an empirical study of social support among patients. Inform Syst Res 2014 Dec;25(4):690-709. [doi: 10.1287/isre.2014.0538]
10. Yang H, Guo X, Wu T. Exploring the influence of the online physician service delivery process on patient satisfaction. Dec Support Syst 2015 Oct;78:113-121. [doi: 10.1016/j.dss.2015.05.006]
11. Leung L. Loneliness, social support, and preference for online social interaction: the mediating effects of identity experimentation online among children and adolescents. Chinese J Comm 2011 Dec;4(4):381-399. [doi: 10.1080/17544750.2011.616285]
12. Cacciatore MA, Anderson AA, Choi D, Brossard D, Scheufele DA, Liang X, et al. Coverage of emerging technologies: a comparison between print and online media. New Media Soc 2012 Mar 21;14(6):1039-1059. [doi: 10.1177/1461444812439061]
13. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
14. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Dec;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
15. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
16. Van Oerle S, Lievens A, Mahr D. Value co-creation in online healthcare communities: the impact of patients’ reference frames on cure and care. Psychol Mark 2018 May 23;35(9):629-639. [doi: 10.1002/mar.21111]
17. Petrič G, Atanasova S, Kamin T. Impact of social processes in online health communities on patient empowerment in relationship with the physician: emergence of functional and dysfunctional empowerment. J Med Internet Res 2017 Dec 13;19(3):e74. [doi: 10.2196/jmir.7002] [Medline: 28288953]
18. Tsimmerman Ya S, Vologzhanina LG. [Compliance of patients with prescribed medication as an important factor of treatment efficiency]. Klin Med (Mosk) 2015 Jul;93(3):5-13. [Medline: 26168597]
19. Bichara L, Aragón ML, Brandão GA, Normando D. Factors influencing orthodontic treatment time for non-surgical Class III malocclusion. J Appl Oral Sci 2016 Sep;24(5):431-436 [FREE Full text] [doi: 10.1590/1678-775720150353] [Medline: 27812612]
20. Remor E. [Assessment of adherence to antiretroviral therapy in HIV patients]. Psicothema 2002 May;14(2):262-267 [FREE Full text]
21. Erdem SA, Harrison-Walker LJ. The role of the Internet in physician–patient relationships: the issue of trust. Bus Horiz 2006 Sep;49(5):387-393. [doi: 10.1016/j.bushor.2006.01.003]
22. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]
23. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. Arch Intern Med 1993 Aug 23;153(16):1863-1868. [Medline: 8250647]
24. Marck CH, De Livera AM, Brown CR, Neate SL, Taylor KL, Weiland TJ, et al. Health outcomes and adherence to a healthy lifestyle after a multimodal intervention in people with multiple sclerosis: three year follow-up. PLoS One 2018;13(5):e0197759 [FREE Full text] [doi: 10.1371/journal.pone.0197759] [Medline: 29791509]
25. Matusitz J, Spear J. Effective doctor-patient communication: an updated examination. Soc Work Public Health 2014;29(3):252-266. [doi: 10.1080/19371918.2013.776416] [Medline: 24802220]
26. Zolnierek KB, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834 [FREE Full text] [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]
27. Roberts L, Cornell C, Bostrom M, Goldsmith S, Ologhobo T, Roberts T, et al. Communication skills training for surgical residents: learning to relate to the needs of older adults. J Surg Educ 2018;75(5):1180-1187. [doi: 10.1016/j.jsurg.2018.02.005] [Medline: 29609892]
28. Bultman DC, Svarstad BL. Effects of physician communication style on client medication beliefs and adherence with antidepressant treatment. Patient Educ Couns 2000 May;40(2):173-185. [Medline: 10771371]

29. Molfenter TD, Brown RL. Effects of physician communication and family hardiness on patient medication regimen beliefs and adherence. Gen Med (Los Angel) 2014 Jan;2:136 [FREE Full text] [doi: 10.4172/2327-5146.1000136] [Medline: 25593977]
30. Laugesen J, Hassanein K, Yuan Y. The impact of Internet health information on patient compliance: a research model and an empirical study. J Med Internet Res 2015;17(6):e143 [FREE Full text] [doi: 10.2196/jmir.4333] [Medline: 26068214]
31. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
32. Swindell JS, McGuire AL, Halpern SD. Beneficent persuasion: techniques and ethical guidelines to improve patients' decisions. Ann Fam Med 2010;8(3):260-264 [FREE Full text] [doi: 10.1370/afm.1118] [Medline: 20458111]
33. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship between internet health information and patient compliance based on trust: empirical study. J Med Internet Res 2018 Aug 17;20(8):e253 [FREE Full text] [doi: 10.2196/jmir.9364] [Medline: 30120087]
34. Umaki TM, Umaki MR, Cobb CM. The psychology of patient compliance: a focused review of the literature. J Periodontol 2012 Apr;83(4):395-400. [doi: 10.1902/jop.2011.110344] [Medline: 21819248]
35. Umar N, Litaker D, Schaarschmidt M, Peitsch WK, Schmieder A, Terris DD. Outcomes associated with matching patients' treatment preferences to physicians' recommendations: study methodology. BMC Health Serv Res 2012 Jan 3;12:1 [FREE Full text] [doi: 10.1186/1472-6963-12-1] [Medline: 22214259]
36. Campbell TA, Auerbach SM, Kiesler DJ. Relationship of interpersonal behaviors and health-related control appraisals to patient satisfaction and compliance in a university health center. J Am Coll Health 2007;55(6):333-340. [doi: 10.3200/JACH.55.6.333-340] [Medline: 17517545]
37. Kreps G, O'Hair D. Communication and Health Outcomes. Cresskill, NJ: Hampton Pres; 1995.
38. Lipscomb J, Snyder CF. The outcomes of cancer outcomes research: focusing on the National Cancer Institute's quality-of-care initiative. Med Care 2002 Jun;40(6 Suppl):III3-II10. [Medline: 12064755]
39. Adler KG. Web portals in primary care: an evaluation of patient readiness and willingness to pay for online services. J Med Internet Res 2006;8(4):e26 [FREE Full text] [doi: 10.2196/jmir.8.4.e26] [Medline: 17213045]
40. Huang L, Hsu C, Ko H, Kuo F. Influential factors on physicians' behaviour towards the medical web messaging board: a qualitative exploration. Int J Electron Healthc 2007;3(2):220-231. [doi: 10.1504/IJEH.2007.013102] [Medline: 18048271]
41. Car J, Sheikh A. Email consultations in health care: 2--acceptability and safe application. Br Med J 2004 Aug 21;329(7463):439-442 [FREE Full text] [doi: 10.1136/bmj.329.7463.439] [Medline: 15321903]
42. Harris I. The Harris Poll. 2000. Harris Interactive/ARiA Marketing Healthcare Satisfaction Study URL: http://www.harrisinteractive.com/news/downloads/HarrisAriaHCSatRpt.PDF [accessed 2006-08-25] [WebCite Cache ID 5IOYLVTUb]
43. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
44. Breen R, Thornhill JT. Noncompliance with medication for psychiatric disorders. CNS Drugs 1998;9(6):457-471. [doi: 10.2165/00023210-199809060-00004]
45. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N Engl J Med 2016 Dec 15;375(24):2349-2358 [FREE Full text] [doi: 10.1056/NEJMoa1605086] [Medline: 27959714]
46. Johal S, Jamsen KM, Bell JS, Mc Namara KP, Magliano DJ, Liew D, et al. Do statin users adhere to a healthy diet and lifestyle? The Australian diabetes, obesity and lifestyle study. Eur J Prev Cardiol 2017 Dec;24(6):621-627. [doi: 10.1177/2047487316684054] [Medline: 28326830]
47. Varleta P, Acevedo M, Akel C, Salinas C, Navarrete C, García A, et al. Mobile phone text messaging improves antihypertensive drug adherence in the community. J Clin Hypertens (Greenwich) 2017 Dec;19(12):1276-1284 [FREE Full text] [doi: 10.1111/jch.13098] [Medline: 28941056]
48. Hensher DA, Rose JM, Greene WH. Applied Choice Analysis, Second Edition. Cambridge: Cambridge University Press; 2015.
49. Brown MT, Bussell J, Dutta S, Davis K, Strong S, Mathew S. Medication adherence: truth and consequences. Am J Med Sci 2016 Apr;351(4):387-399. [doi: 10.1016/j.amjms.2016.01.010] [Medline: 27079345]
50. Pang PC, 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 6;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
51. Dobransky K, Hargittai E. Inquiring minds acquiring wellness: uses of online and offline sources for health information. Health Commun 2012;27(4):331-343. [doi: 10.1080/10410236.2011.585451] [Medline: 21932982]
52. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
53. Suri VR, Majid S, Chang Y, Foo S. Assessing the influence of health literacy on health information behaviors: a multi-domain skills-based approach. Patient Educ Couns 2016 Dec;99(6):1038-1045. [doi: 10.1016/j.pec.2015.12.017] [Medline: 26794667]

54. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
55. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: an empirical analysis of a national cancer-related survey. Dec Support Syst 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
56. Li X, Zhang G. Perceived credibility of Chinese social media: toward an integrated approach. Int J Public Opin Res 2017 Feb 6;30(1):79-101. [doi: 10.1093/ijpor/edw035]
57. Maloney EK, D'Agostino TA, Heerdt A, Dickler M, Li Y, Ostroff JS, et al. Sources and types of online information that breast cancer patients read and discuss with their doctors. Palliat Support Care 2015 Apr;13(2):107-114 [FREE Full text] [doi: 10.1017/S1478951513000862] [Medline: 24182945]
58. Zhang Y, Su H, Shang L, Li D, Wang R, Zhang R, et al. Preferences and perceived involvement in treatment decision making among Chinese patients with chronic hepatitis. Med Decis Making 2011;31(2):245-253. [doi: 10.1177/0272989X10375990] [Medline: 20709961]
59. Hart A, Smith WR, Tademy RH, McClish DK, McCreary M. Health decision-making preferences among African American men recruited from urban barbershops. J Natl Med Assoc 2009 Jul;101(7):684-689. [Medline: 19634589]
60. Beaver K, Bogg J, Luker KA. Decision-making role preferences and information needs: a comparison of colorectal and breast cancer. Health Expect 1999 Dec;2(4):266-276 [FREE Full text] [Medline: 11281903]
61. Cox K, Britten N, Hooper R, White P. Patients' involvement in decisions about medicines: GPs' perceptions of their preferences. Br J Gen Pract 2007 Oct;57(543):777-784 [FREE Full text] [Medline: 17925134]
62. Wei S, Chen F, Chen H, Guo Y, Hui D, Yennurajalingam S, et al. Patients' and family members' decision-making and information disclosure preferences in a single-center survey in China: a pilot study. Am J Hosp Palliat Care 2016 Sep;33(8):733-741. [doi: 10.1177/1049909115588302] [Medline: 26019263]
63. Bruera E, Willey JS, Palmer JL, Rosales M. Treatment decisions for breast carcinoma: patient preferences and physician perceptions. Cancer 2002 Apr 1;94(7):2076-2080 [FREE Full text] [Medline: 11932912]
64. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ 1995 May 1;152(9):1423-1433 [FREE Full text] [Medline: 7728691]
65. Moise N, Ye S, Alcántara C, Davidson KW, Kronish I. Depressive symptoms and decision-making preferences in patients with comorbid illnesses. J Psychosom Res 2017 Dec;92:63-66 [FREE Full text] [doi: 10.1016/j.jpsychores.2015.12.001] [Medline: 26682488]
66. Jabbour J, Dhillon HM, Shepherd HL, Sundaresan P, Milross C, Clark JR. The relationship between role preferences in decision-making and level of psychological distress in patients with head and neck cancer. Patient Educ Couns 2018 Oct;101(10):1736-1740. [doi: 10.1016/j.pec.2018.05.023] [Medline: 29866431]
67. Vogel B, Bengel J, Helmes A. Information and decision making: patients' needs and experiences in the course of breast cancer treatment. Patient Educ Couns 2008 Apr;71(1):79-85. [doi: 10.1016/j.pec.2007.11.023] [Medline: 18191933]
68. Cajita MI, Whitehouse E, Budhathoki C, Hodgson N. Association between internet use and decision-making preference in older adults. Gerontechnology 2016;14(2):97-104 [FREE Full text] [doi: 10.4017/gt.2016.14.2.008.00] [Medline: 28190988]
69. Harvey R, Kazis L, Lee A. Decision-making preference and opportunity in VA ambulatory care patients: association with patient satisfaction. Res Nurs Health 1999 Feb;22(1):39-48. [Medline: 9928962]
70. Ikeda Y, Sasaki T, Kuwahata S, Imamura M, Tanoue K, Komaki S, et al. Questionnaire survey from the viewpoint of concordance in patient and physician satisfaction concerning hypertensive treatment in elderly patients - patients voice study. Circ J 2018 Mar 23;82(4):1051-1061 [FREE Full text] [doi: 10.1253/circj.CJ-17-1015] [Medline: 29445065]
71. Kerse N, Buetow S, Mainous AG, Young G, Coster G, Arroll B. Physician-patient relationship and medication compliance: a primary care investigation. Ann Fam Med 2004;2(5):455-461 [FREE Full text] [Medline: 15506581]
72. Gross K, Schindler C, Grize L, Späth A, Schwind B, Zemp E. Patient-physician concordance and discordance in gynecology: do physicians identify patients' reasons for visit and do patients understand physicians' actions? Patient Educ Couns 2013 Jul;92(1):45-52. [doi: 10.1016/j.pec.2013.02.003] [Medline: 23481216]
73. Shin DW, Kim SY, Cho J, Sanson-Fisher RW, Guallar E, Chai GY, et al. Discordance in perceived needs between patients and physicians in oncology practice: a nationwide survey in Korea. J Clin Oncol 2011 Nov 20;29(33):4424-4429. [doi: 10.1200/JCO.2011.35.9281] [Medline: 22010016]
74. Schubart JR, Toran L, Whitehead M, Levi BH, Green MJ. Informed decision making in advance care planning: concordance of patient self-reported diagnosis with physician diagnosis. Support Care Cancer 2013 Feb;21(2):637-641. [doi: 10.1007/s00520-012-1654-x] [Medline: 23179490]
75. Deci E, Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior. New York: Plenum Press; 1985.
76. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
77. Deci E, Ryan R. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
78. Deci E, Ryan R. The Handbook of Self-Determination Research. Rochester: University of Rochester Press; 2002:141-159.

79. Wild TC, Cunningham JA, Ryan RM. Social pressure, coercion, and client engagement at treatment entry: a self-determination theory perspective. Addict Behav 2006 Oct;31(10):1858-1872. [doi: 10.1016/j.addbeh.2006.01.002] [Medline: 16480834]
80. Edmunds J, Ntoumanis N, Duda JL. Adherence and well-being in overweight and obese patients referred to an exercise on prescription scheme: a self-determination theory perspective. Psychol Sport Exer 2007 Sep;8(5):722-740. [doi: 10.1016/j.psychsport.2006.07.006]
81. Barbouni A, Nalmpanti M, Gennimata D, Theodoridis D, Merakou K. Beliefs and practices of Greek doctors in relation to patients' adherence to antihypertensive medication. J Hum Hypertens 2017 May;31(5):341-346. [doi: 10.1038/jhh.2016.84] [Medline: 28032629]
82. Makoul G, Krupat E, Chang C. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. Patient Educ Couns 2007 Aug;67(3):333-342. [doi: 10.1016/j.pec.2007.05.005] [Medline: 17574367]
83. Aoki A, Suda A, Nagaoka S, Takeno M, Ishigatsubo Y, Ashizawa T, et al. Preferences of Japanese rheumatoid arthritis patients in treatment decision-making. Mod Rheumatol 2013 Sep;23(5):891-896. [doi: 10.1007/s10165-012-0761-3] [Medline: 23053717]
84. Bu X, You L, Li Y, Liu K, Zheng J, Yan T, et al. Psychometric properties of the Kessler 10 Scale in Chinese parents of children with cancer. Cancer Nurs 2017;40(4):297-304. [doi: 10.1097/NCC.0000000000000392] [Medline: 27257800]
85. Xiao Y, Li T, Xiao L, Wang S, Wang S, Wang H, et al. The Chinese version of instrument of professional attitude for student nurses (IPASN): assessment of reliability and validity. Nurse Educ Today 2017 Feb;49:79-83. [doi: 10.1016/j.nedt.2016.11.013] [Medline: 27889582]
86. Wong WS, Chen PP, Chow YF, Wong S, Fielding R. A study of the reliability and concurrent validity of the Chinese version of the pain medication attitude questionnaire (ChPMAQ) in a sample of Chinese patients with chronic pain. Pain Med 2016 Jun;17(6):1137-1144. [doi: 10.1093/pm/pnv086] [Medline: 26843593]
87. Gefen D, Straub D, Boudreau M. Structural equation modeling and regression: guidelines for research practice. Commun Assoc Inform Syst 2000 Dec;4(7):1-77 [FREE Full text]
88. Buysse HE, Coorevits P, Van Maele G, Hutse A, Kaufman J, Ruige J, et al. Introducing telemonitoring for diabetic patients: development of a telemonitoring 'Health Effect and Readiness' Questionnaire. Int J Med Inform 2010 Aug;79(8):576-584. [doi: 10.1016/j.ijmedinf.2010.05.005] [Medline: 20599161]
89. Sardeshmukh SR, Vandenberg RJ. Integrating moderation and mediation: a structural equation modeling approach. Organ Res Methods 2016 Jan 6;20(4):721-745. [doi: 10.1177/1094428115621609]
90. Wiertz C, de Ruyter K. Beyond the call of duty: why customers contribute to firm-hosted commercial online communities. Organ Stud 2016 Jun 30;28(3):347-376. [doi: 10.1177/0170840607076003]
91. Audrain-Pontevia A, Menvielle L. Do online health communities enhance patient-physician relationship? An assessment of the impact of social support and patient empowerment. Health Serv Manage Res 2018 Aug;31(3):154-162. [doi: 10.1177/0951484817748462] [Medline: 29280679]
92. Huh J, Kwon BC, Kim S, Lee S, Choo J, Kim J, et al. Personas in online health communities. J Biomed Inform 2016 Oct;63:212-225. [doi: 10.1016/j.jbi.2016.08.019] [Medline: 27568913]
93. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro C, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. Sleep Med Rev 2016 Feb;25:52-73. [doi: 10.1016/j.smrv.2015.01.009] [Medline: 26163057]
94. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Market Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
95. Sobral M, Costa M, Schmidt L, Martins M. Corrigendum: COMPI fertility problem stress scales is a brief, valid and reliable tool for assessing stress in patients seeking treatment. Hum Reprod 2017 Feb;32(2):375-382. [doi: 10.1093/humrep/dew315] [Medline: 27979919]
96. Cohen J. Statistical power analysis for the behavioral sciences. J Am Stat Assoc 1988;2(4):19-74. [doi: 10.1016/B978-0-12-179060-8.50012-8]

Abbreviations

AVE: average variance extracted
CR: composite reliability
OHC: online health community
PLS: partial least squares
SEM: structural equation modeling

Edited by G Eysenbach; submitted 22.11.18; peer-reviewed by J Laugesen, H Wu; comments to author 17.01.19; revised version received 14.03.19; accepted 06.04.19; published 13.05.19.

Please cite as:
Lu X, Zhang R
Impact of Physician-Patient Communication in Online Health Communities on Patient Compliance: Cross-Sectional Questionnaire Study
J Med Internet Res 2019;21(5):e12891
URL: http://www.jmir.org/2019/5/e12891/
doi:10.2196/12891
PMID:31094342

Original Paper

A Virtual Agent to Support Individuals Living With Physical and Mental Comorbidities: Co-Design and Acceptability Testing

Katherine Easton[1,2], BSc, MRES, PhD; Stephen Potter[1,2], BSc, MSc, PhD; Remi Bec[3,4], MA, PhD; Matthew Bennion[2,5,6], BEng, MSc, PhD; Heidi Christensen[2,5], MSc, PhD; Cheryl Grindell[3], BSc, MSc; Bahman Mirheidari[5], BSc, MSc, PhD; Scott Weich[1], MSc, MD; Luc de Witte[1,2], MD, PhD; Daniel Wolstenholme[3], BSc, MMSc; Mark S Hawley[1,2], BSc, PhD

[1]School of Health and Related Research, The University of Sheffield, Sheffield, United Kingdom

[2]Centre for Assistive Technology and Connected Healthcare, The University of Sheffield, Sheffield, United Kingdom

[3]Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber, Royal Hallamshire Hospital, National Institute of Health, Sheffield, United Kingdom

[4]Lab4Living, Art and Design Research Centre, Sheffield Hallam University, Sheffield, United Kingdom

[5]Department of Computer Science, The University of Sheffield, Sheffield, United Kingdom

[6]Department of Psychology, The University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Katherine Easton, BSc, MRES, PhD
School of Health and Related Research
The University of Sheffield
217 Portobello
Sheffield, S1 4DA
United Kingdom
Phone: 44 1142220690
Email: k.a.easton@sheffield.ac.uk

Abstract

Background: Individuals living with long-term physical health conditions frequently experience co-occurring mental health problems. This comorbidity has a significant impact on an individual's levels of emotional distress, health outcomes, and associated health care utilization. As health care services struggle to meet demand and care increasingly moves to the community, digital tools are being promoted to support patients to self-manage their health. One such technology is the autonomous virtual agent (chatbot, conversational agent), which uses artificial intelligence (AI) to process the user's written or spoken natural language and then to select or construct the corresponding appropriate responses.

Objective: This study aimed to co-design the content, functionality, and interface modalities of an autonomous virtual agent to support self-management for patients with an exemplar long-term condition (LTC; chronic pulmonary obstructive disease [COPD]) and then to assess the acceptability and system content.

Methods: We conducted 2 co-design workshops and a proof-of-concept implementation of an autonomous virtual agent with natural language processing capabilities. This implementation formed the basis for video-based scenario testing of acceptability with adults with a diagnosis of COPD and health professionals involved in their care.

Results: Adults (n=6) with a diagnosis of COPD and health professionals (n=5) specified 4 priority self-management scenarios for which they would like to receive support: at the time of diagnosis (*information provision*), during acute exacerbations (*crisis support*), during periods of low mood (*emotional support*), and for general self-management (*motivation*). From the scenario testing, 12 additional adults with COPD felt the system to be both acceptable and engaging, particularly with regard to internet-of-things capabilities. They felt the system would be particularly useful for individuals living alone.

Conclusions: Patients did not explicitly separate mental and physical health needs, although the content they developed for the virtual agent had a clear psychological approach. Supported self-management delivered via an autonomous virtual agent was acceptable to the participants. A co-design process has allowed the research team to identify key design principles, content, and functionality to underpin an autonomous agent for delivering self-management support to older adults living with COPD and potentially other LTCs.

(J Med Internet Res 2019;21(5):e12996) doi:10.2196/12996

KEYWORDS

COPD; chronic obstructive pulmonary disease; mental health; comorbidity; chronic illness; self-management; artificial intelligence; virtual systems; computer-assisted therapy; chatbot; conversational agent

Introduction

Comorbid Mental Health and Physical Long-Term Conditions

In the past decade, it has been acknowledged in research [1], government policy [2] and clinical practice [3] that physical conditions, such as respiratory and cardiovascular disease, and mental health conditions, including anxiety and depression, frequently coexist. It is estimated that 30% of those with 1 or more physical long-term conditions (LTCs) in England will have a comorbid mental health condition [3,4]. This is particularly the case for patients with a diagnosis of chronic obstructive pulmonary disease (COPD), a patient population that has a higher prevalence of anxiety and depression [5] and which is up to 10 times more likely to experience panic attacks than the general population [6]. These mental health conditions are distressing, have a clinically significant impact on health-related quality of life, and can lead to further health complications and greater consumption of costly health care [1,3,7]. For those living with COPD, management of their condition means maintenance of mental health as much as of physical health.

Unfortunately, the presence of mental health conditions in individuals with a physical LTC can go undetected and untreated. The addition of a physical disease process can compound mental health problems [8-10], and many individuals with LTCs do not have access to mental health care once clinically identified. Despite mental health problems accounting for 23% of the disease burden in the United Kingdom, only 11% of the UK National Health Service (NHS) budget is allocated to mental health services [11]. The high demand for services and a lack of trained health professionals able to offer quality care mean waiting times can be long [12,13]. Accordingly, the issues of management of both physical and mental health for patients with comorbidities is an imperative for research, policy, and clinical practice.

Self-Management

Patient health care is increasingly moving into the community, with greater emphasis placed on supporting patients to *take control* of their own health (patient activation [14,15]). Self-management of health includes maintaining positive emotional well-being, eating a healthy diet, exercising, optimizing medication, monitoring symptoms, coping with changes in symptoms, and knowing when to seek help [15]. To do this, patients require techniques to draw upon and support to develop the self-confidence and skills required. A review of 228 systematic reviews of self-management interventions indicates that there is little consistent evidence for the appropriate duration, mode of delivery, or specific components of such interventions [16]. Psychological approaches that include techniques such as relaxation training, mindfulness, motivational interviewing, and cognitive restructuring appear to be of benefit in helping patients improve their self-management skills [16]. In addition, health education provision, biofeedback techniques, and physical activity have been found to improve emotional well-being for patients with a range of LTCs [15,16]. Interactive Web-based self-management programs and telehealth initiatives could be particularly useful elements of self-management support for patients, irrespective of the diagnosis [15]. Regardless of the actual techniques and tools used, it is clear that any truly comprehensive and successful self-management intervention for LTCs must address mental health as well as physical health and should see the care of the 2 as forming a holistic *care package* for the patient [1].

Technology and Health Care

Commonplace digital technology consumer products such as mobile phones, tablets, laptops, and wearable devices have the potential to support and expand the delivery of health services in a more accessible and efficient manner than traditional face-to-face, 9 am to 5 pm, weekday delivery of care. Patients can use these products to access, for example, Web-based health information, self-help materials, electronic therapies, blended care, peer support, and one-to-one consultations [17]. For some patients, the greater convenience, accessibility, and availability of digital services outweigh the advantages of face-to-face contact with health professionals [18]. However, for current digital health technology to be evaluated as engaging, as acceptable by patients, and as clinically effective as traditional face-to-face delivery of care, a blended or guided approach, which includes aspects of face-to-face or telephone support, is recommended rather than the use of digital tools in isolation from human contact [19-21]. It is possible that more sophisticated digital health interventions may address these shortcomings if they can replicate key elements of person-to-person interaction.

Artificial Intelligence and the Provision of Therapy

A promising approach to the development of more sophisticated digital health services lies in the use of artificial intelligence (AI) to build computer systems that are able to deliver therapeutic services with a certain degree of autonomy [22]. In such systems, the role of AI is to process the user's input data (usually in the form of written or spoken natural language utterances) and then to select or construct the corresponding appropriate responses.

In some respects, the use of AI-based systems in this way represents a natural development: as conventional psychotherapy typically relies on a series of language-based interactions with a therapist, any system that can input and output the symbols that constitute that language—in other words, that has the potential to communicate with the patient—represents, in theory at least, a platform for the delivery of that therapy. Digital computers (and networks of digital computers, such as the Web) constitute one such platform, and as researchers in AI have devoted much effort over the years to emulating human natural language processing on digital computers, the use of AI to deliver therapy readily suggests itself. (Indeed, one of the earliest and most famous AI systems, ELIZA, dating from the 1960s, emulated, albeit in a somewhat tongue-in-cheek fashion, a *talking cure* therapist.) Moreover, computer programs and services promise to overcome some of the physical limitations of a human therapist: they can be duplicated rapidly and almost without limit, and they are not fixed to a physical location and never tire of operation, being available (almost) wherever and whenever they are needed.

In recent years, several such *autonomous virtual agents* have been developed to support the delivery of psychological therapies, incorporating automated responses to text-based chat [23,24]. However, being text based, these therapies lack the naturalness and immediacy of spoken language and can exclude those not skilled in typing, and it can be difficult to convey the subtleties of emotion or attitude. It seems likely that more inclusive and sophisticated therapeutic agents will have to incorporate spoken language interfaces.

Avachat: A Virtual Agent for Long-Term Condition Self-Management

The aim of the work presented here was to explore the acceptability of an autonomous agent for supporting people with comorbid physical LTCs and mental health problems. The exemplar LTC chosen in this instance was COPD because of the high reported co-occurrence of common mental health conditions. We hypothesized a virtual agent system, called *Avachat* (a portmanteau of *avatar* and *chat*, but also a serendipitous near-homophone of *have a chat*), that would offer users acceptable support and guidance based on self-management principles [15]. This system was to be structured around a persona or character, christened *Ava*, which would act as a focus for the user's interactions with the system and with which the user would—perhaps—form something akin to a therapeutic relationship. We imagined that Ava would come to personify the support mechanisms, with a visible (onscreen) and audible presence, and that its users would interact with it through natural language, but beyond this, we made few assumptions as to its nature, intending that this should emerge during a co-design process.

Methods

Design

Working with people with lived experience of COPD and health care professionals and using existing paper-based self-management materials as a starting point, we had the following objectives:

1. To co-design the *core functionality* of the system by identifying common problems faced when (self-)managing COPD, along with the solutions that were thought most appropriate or useful. These problems were then prioritized according to their impact on everyday life.
2. To elicit prospective users' *operational requirements* for the system, including interface features and visual appearance and delivery modalities, and, based on these, the co-design as far as possible of the corresponding features of the system.
3. To develop a prototype implementation that would meet the core functionality and operational requirements as comprehensively as possible given the constraints on resources (time, personnel, and available technology). The prototype was to be developed rapidly to allow its use in a subsequent iteration of the co-design process.
4. To test the *acceptability* of the design concepts with a wider group of participants.

These methods were adapted from the work of the UK National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber [25] *Translating Knowledge into Action* theme. They encourage a creative co-design approach to enable meaningful participation by all key stakeholders: an approach based on the principles of service design [26] and the UK Design Council's *double diamond* model of the design process [27]. The approach involves divergent and subsequent convergent thinking phases to, in the first instance, open up the space of challenges before honing the problem definition and then to explore potential ideas before focusing on a workable solution to the problem in hand. If all these processes are done with the involvement of all key stakeholders, the approach becomes coproductive by definition [26]. The explicit use of creative methods in coproduction addresses many of the challenges of coproduction, namely imbalances or deficits of power, language, trust, and time. It fosters an inclusive nonhierarchical environment [28] that allows all participants to be recognized as experts and produces results in the form of visible and tangible outputs owned by those who created them [29].

Participants

We aimed to recruit up to 10 individuals in total to 2 co-design workshops, drawn from the sets of patients with lived experience of COPD and of health professionals working with this patient group. In addition, we aimed to recruit a further 10 patients to take part in subsequent video-based scenario testing. In more detail, eligible participants were to be either of the following:

- An individual with a self-reported diagnosis of COPD or chronic lung disease, medically stable, and with a self-reported experience of emotional symptoms in relation to their LTC. A diagnosed mental health condition was not an inclusion criterion for taking part; however, as noted previously, mental health conditions in patients with physical comorbidities are common and often undiagnosed.
- Eligible health professionals were those working with this patient group in some capacity such as a general practitioner (GP; the doctor who will be responsible for treating all common medical conditions and for referring patients for urgent or specialist treatment), mental health professional, specialist nurse, occupational therapist, physiotherapist, or pulmonary rehabilitation specialist.

Individuals with lived experience were identified through the local British Lung Foundation *Breathe Easy* support group. A researcher (KE) attended the support group sessions and presented the research to members, along with information sheets. Interested individuals emailed the study lead (KE) to sign up for the workshops. Our professional contacts from NHS Trusts in and around the Yorkshire region of the United Kingdom were used to obtain the health professional sample: a snowball sampling technique was adopted, with senior managers forwarding an invitation email to relevant staff.

Data Collection

Workshop 1

The first workshop was run to consider initial user requirements for the prototype system. Activities were conducted in small groups with time for feedback and reflection after each activity. This workshop focused on gaining a shared understanding of the lived experience of COPD by:

- Mapping out *a day in the life* journey together with a participating health professional. Sticker icons were used as prompts to help tell each person’s story and *missing sentence* prompt cards (“I wish I had known that…”; “It would have been good to have…”; “If only…”; and “It helped to have…”) were used to help participants reflect on the positive and negative impacts of their physical and mental health on their day. This process addressed Objective (1), namely the elicitation and prioritization of problems in need of solutions.
- Showing participants examples of existing virtual artificial agents to critique. Mood boards were populated to define the ideal set of features for Ava and Avachat. The look, sound, format, and potential features of Ava and Avachat more generally were explored to address Objective (2), eliciting the operational requirements of the system.

Workshop 2

The aim of this workshop was to design and develop how Avachat, in the guise of Ava, would deliver the core functionality identified in the first workshop. This would address Objective (2) eliciting operational requirements. Activities including the following:

- Adopt and elaborate one of a number of vignettes (Figure 1)—each corresponding to an archetypal person living with COPD—using techniques such as role play to allow participants to step out of their particular contexts and move toward a shared, empathic understanding of different potential users’ needs, experiences, behaviors, and goals.
- The vignettes then informed the development of *scripts* or sample interactions with Ava in a particular scenario, each related to one of the core functionalities of the system. Each script constituted an essentially linear dialogue, with some limited branching to explore particularly problematic or interesting alternatives. The scripts helped refine the nature of the support that would be expected of the system as well as define aspects such as safeguards (in what circumstances it was advisable and acceptable to deviate from the script to alert a GP or caregiver, for instance), the style of language and speech that should be adopted by Ava.
- Feedback on the vignettes (Figure 2) was used to gather further information for the Ava prototype development.

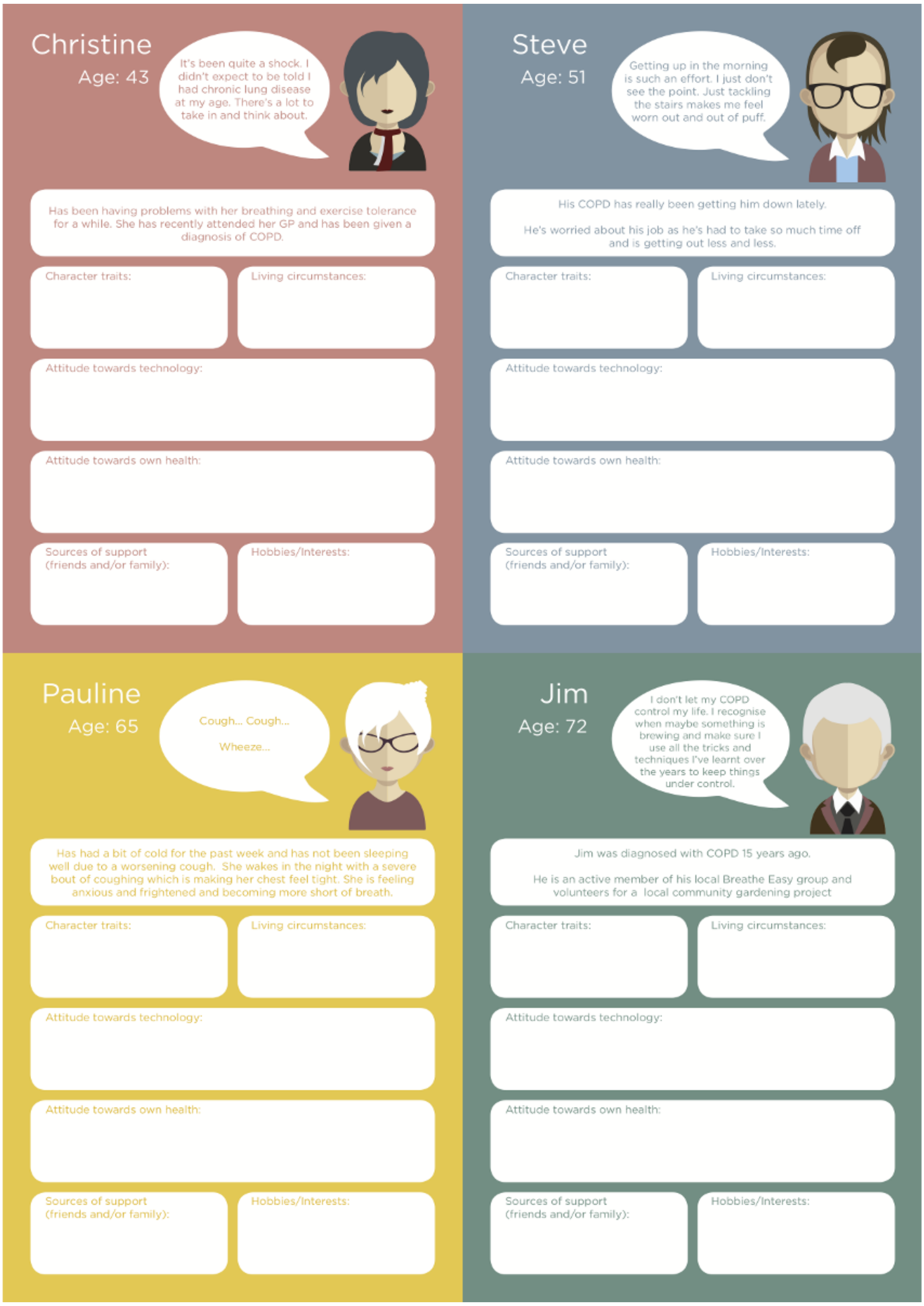
Video-Based Scenario Testing of Acceptability

Following the workshops, we incorporated some of the more achievable user requirements and content into a semiautonomous prototype Avachat system. On the basis of the scripts developed during Workshop 2, we also developed a short screenplay that introduced Ava and the concept of Avachat and depicted how it might provide support to its user in typical domestic situations. This screenplay was filmed in a living laboratory facility, with a volunteer person with COPD taking the role of the user and simulated interactions with Ava (with an offscreen human operator acting as a *Wizard of Oz* to provide appropriate responses). This material was edited into a short video (under 8 min), which was then used as the basis for scenario testing with a sample of patients to determine acceptability of the application and content. We explored participants’ views and opinions and asked them to self-complete the System Usability Scale (SUS) questionnaires [30]. The measure has 10 items on usability, complexity, need for support or expert knowledge, integrity, and consistency. Each question is rated on a Likert scale from 1 to 5 (strongly disagree to strongly agree).

Data Analysis

Workshop activities were conducted on paper and were additionally captured in photograph and video format. Data from Workshop 1 on the challenges of daily living were analyzed using content analysis techniques [31] to determine the scenarios around which to develop the core functionality of Avachat. In addition to providing further operational requirements, outputs from Workshop 2 were synthesized using thematic analysis [32] to develop a set of 4 vignettes (case scenarios) that would serve as exemplars for interactions with Ava. We used 2 of these in the screenplay that was filmed and then during the video-based scenario testing. The discussions from the scenario testing were transcribed verbatim and presented as collections of conversation around a priori themes of design, functionality, and content. The completed SUS questionnaires were scored as stipulated by the developers of the instrument. Higher scores are indicative of a more user-friendly system. Although SUS was originally developed to compare alternative systems and their interfaces for performing the same task, its developers argue that, as some sort of absolute rating, a score of 68 represents an average score across all systems.

Figure 1. Persona worksheets. COPD: chronic obstructive pulmonary disease; GP: general practitioner.



Christine
Age: 43

It's been quite a shock. I didn't expect to be told I had chronic lung disease at my age. There's a lot to take in and think about.

Has been having problems with her breathing and exercise tolerance for a while. She has recently attended her GP and has been given a diagnosis of COPD.

Character traits:
Living circumstances:
Attitude towards technology:
Attitude towards own health:
Sources of support (friends and/or family):
Hobbies/Interests:

Steve
Age: 51

Getting up in the morning is such an effort. I just don't see the point. Just tackling the stairs makes me feel worn out and out of puff.

His COPD has really been getting him down lately.

He's worried about his job as he's had to take so much time off and is getting out less and less.

Character traits:
Living circumstances:
Attitude towards technology:
Attitude towards own health:
Sources of support (friends and/or family):
Hobbies/Interests:

Pauline
Age: 65

Cough... Cough...

Wheeze...

Has had a bit of cold for the past week and has not been sleeping well due to a worsening cough. She wakes in the night with a severe bout of coughing which is making her chest feel tight. She is feeling anxious and frightened and becoming more short of breath.

Character traits:
Living circumstances:
Attitude towards technology:
Attitude towards own health:
Sources of support (friends and/or family):
Hobbies/Interests:

Jim
Age: 72

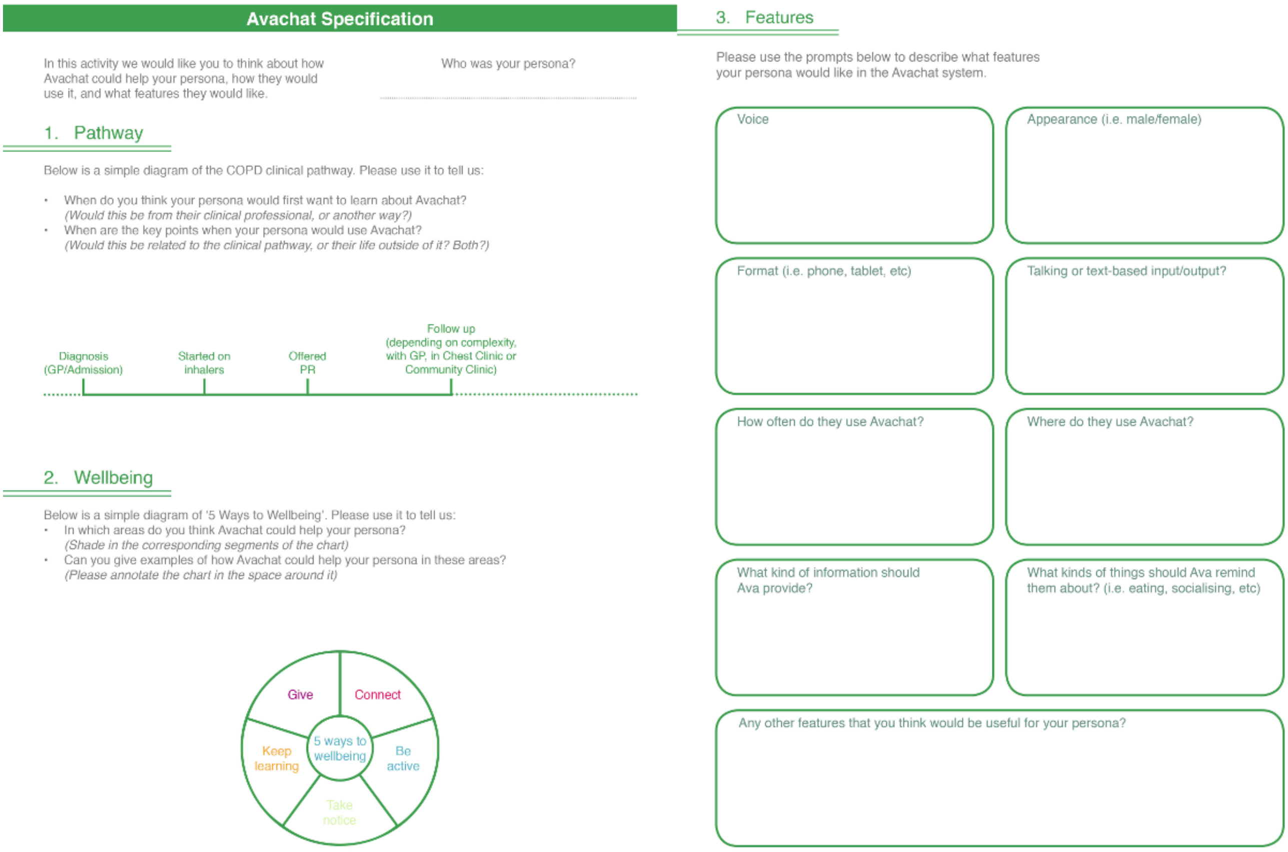
I don't let my COPD control my life. I recognise when maybe something is brewing and make sure I use all the tricks and techniques I've learnt over the years to keep things under control.

Jim was diagnosed with COPD 15 years ago.

He is an active member of his local Breathe Easy group and volunteers for a local community gardening project

Character traits:
Living circumstances:
Attitude towards technology:
Attitude towards own health:
Sources of support (friends and/or family):
Hobbies/Interests:

Figure 2. Vignette feedback sheets. COPD: chronic obstructive pulmonary disease; GP: general practitioner; PR: pulmonary rehabilitation.



Avachat Specification

In this activity we would like you to think about how Avachat could help your persona, how they would use it, and what features they would like.

Who was your persona?

1. Pathway

Below is a simple diagram of the COPD clinical pathway. Please use it to tell us:

- When do you think your persona would first want to learn about Avachat? *(Would this be from their clinical professional, or another way?)*
- When are the key points when your persona would use Avachat? *(Would this be related to the clinical pathway, or their life outside of it? Both?)*

2. Wellbeing

Below is a simple diagram of '5 Ways to Wellbeing'. Please use it to tell us:

- In which areas do you think Avachat could help your persona? *(Shade in the corresponding segments of the chart)*
- Can you give examples of how Avachat could help your persona in these areas? *(Please annotate the chart in the space around it)*

3. Features

Please use the prompts below to describe what features your persona would like in the Avachat system.

Voice

Appearance (i.e. male/female)

Format (i.e. phone, tablet, etc)

Talking or text-based input/output?

How often do they use Avachat?

Where do they use Avachat?

What kind of information should Ava provide?

What kinds of things should Ava remind them about? (i.e. eating, socialising, etc)

Any other features that you think would be useful for your persona?

Results

Workshop 1

Characteristics of the Sample

The workshop was run in July 2017 and lasted 5 hours. Overall, 5 individuals with a diagnosis of COPD and a carer of an individual with COPD agreed to attend (n=6), of whom 5 attended (male n=2 and female n=3), with 1 dropping out because of ill health (male). Participants had a median age of 82.5 years (range 69 to 86 years) and had been diagnosed with COPD for an average of 10.5 years. No medical data on the stage of COPD were available. One participant was currently using oxygen daily. All participants had previously attended pulmonary rehabilitation courses. A total of 6 health professionals attended, all female, working across general practice, secondary care respiratory services, and specialist mental health care.

Activities Data

A Day in the Life

All participants were medically stable. None of the participants disclosed COPD exacerbations or acute mental health problems. Conversation focused on ways in which they prevented low mood and social isolation. From the *a day in the life* exercise, the group identified positive and negative aspects of their lives. Figure 3 shows the range of activities described by participants. Household chores and social interactions occupied the majority of midday activities for participants. Evenings were quieter, with the group stating that they rarely went out of the home in the evening, often feeling very tired as the day progressed. However, the group reported that at times they lacked *motivation* (intrinsic or extrinsic) to follow through on planned household chores or social activities. Sleep was often interrupted at times because of their medical condition or associated medication (frequent need to urinate in the night, leg cramps, and breathing difficulties). Episodes of low mood were reported, particularly low mood on waking.

Missing Sentences Task

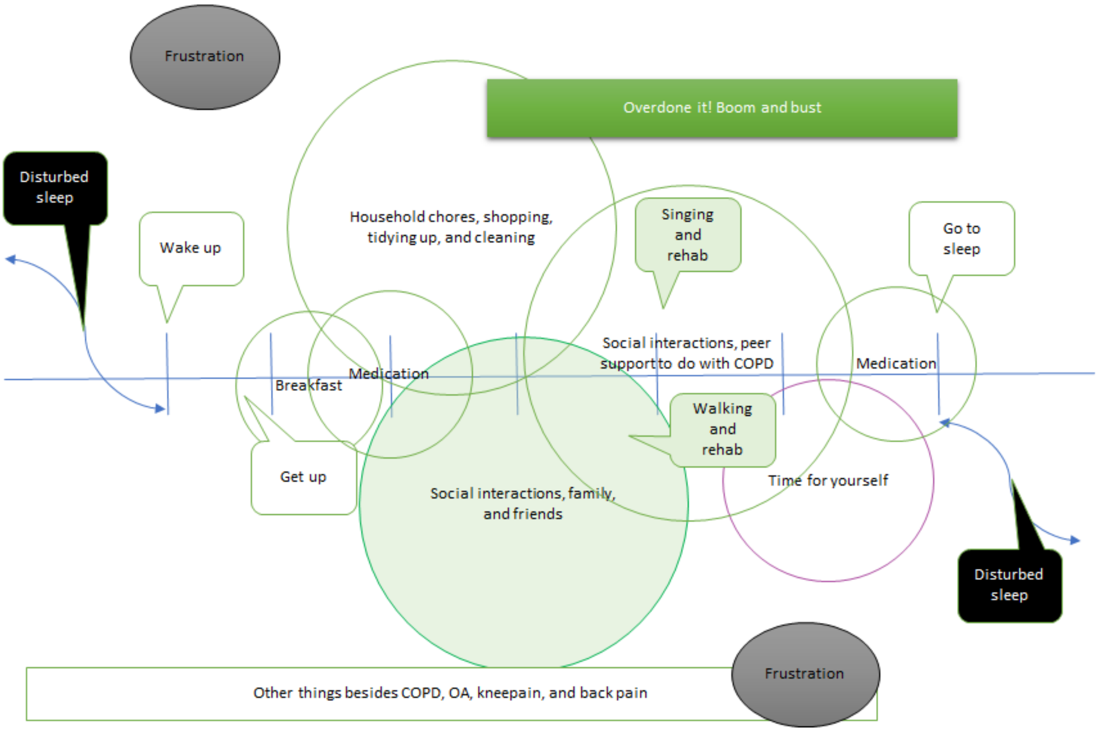
The missing sentences task highlighted that additional support and *information provision* in the early stages of the disease would have been beneficial:

> *It would have been useful to have been assigned to the respiratory nurse earlier on in my treatment, [to provide] general advice on medication/holidays/ support.*
>
> *I think there should be more info for newly diagnosed patients of what services are available.*

Members of the group felt that advice and support should be *informal* and that, ideally, they would have liked one-to-one support:

> *I think that information given by professionals should be less professional [in style] and more humanised.*
>
> *If only there were individual staff for each patient to be around whenever needed.*

Figure 3. A day in the life of a chronic obstructive pulmonary disease patient. COPD: chronic obstructive pulmonary disease; OA: osteoarthritis.



In addition to this, the need for support in *crisis* situations or during acute exacerbations was highlighted:

> *I think people should be able to get more instant support, especially when feeling stressed.*
>
> *I needed more support when my husband had a severe exacerbation/ panic attack. I called 999 [the UK emergency telephone number]. I need some strategies.* [Carer]
>
> *I wish I had someone around to help me when it [breathlessness] hits you.*

Additional data from the missing sentence activity in Workshop 1 were related to the techniques that participants used to maintain or increase their *emotional well-being*. These resembled the New Economics Foundation's *Five ways to wellbeing* [33] and were included in the Avachat scripts (Table 1).

Table 1. The *Five ways to wellbeing* and the data to support its application in Avachat. The terms in italics are the 5 recommended actions to foster one's own well-being.

Components of well-being	Evidenced from the missing sentence activity
Give....; Your time, words, presence	"I have always been busy, driving the bus for the local care centre, being in the buddy system to support people....sitting in the clinics to encourage people to sign up for pulmonary rehab"; "....choir, breathe easy treasurer, meetings etc. But I am glad I do it. The benefits are wonderful"
Be active....; Do what you can, enjoy what you do, move your body	"I enjoyed the feel-good factor after exercising"; "I think GPs should introduce exercise as a treatment option earlier"; "I feel much more positive, healthy and confident since starting pulmonary rehab and breathe easy"
Keep learning....; Embrace new experiences, take opportunities	"There is a video now to help people with COPD understand more about the service—this needs sharing"; "I think there should be more info for newly diagnosed patients of what services are available"; "It was good to join the breathe easy choir and enjoy all the songs we sing from the shows etc. Good for the lungs and great for the spirits"
Connect....; Talk, listen, share, feel connected	"I enjoy hearing the positivity of the people with COPD—and how well they cope with what can be a difficult condition" (health professional); "It's good to get perspective of other people's experiences"; "I try and see someone every day, at the shops, playing bridge, at my friend's for dinner...."
Take notice....; Notice small things	"I think people should carry on doing things. I like listening to the radio and watching the garden, the plants, the grass....I work hard to keep it nice. I'm in no rush"

Outputs From Workshop 1

Care Pathway Scenarios

The data generated from the *a day in the life* task and the missing sentences suggest 4 scenarios in which Avachat could support patients. These are at the time of diagnosis (*information provision*), during acute exacerbations (*crisis support*), during periods of low mood (*emotional support*), and for general self-management (*motivation*). Ideas for other features of the system included reminders for medication and appointments, and providing access to clinical results.

Avatar User—Specifications

Figure 4 illustrates the data generated from the critiques of existing artificial agents and mood board activities that participants undertook in Workshop 1. It shows the look, sound, format, and potential features of Ava that were considered important to the participants. Participants stated that they would like Ava to be multimodal and accessed via connected consumer devices such as their computer, tablet, mobile phone, or television. They wanted it to be recognizably human in appearance, to be able to personalize Ava to some degree (eg, to alter its *gender* and accent) and to have options for voice- or text-only communication. Participants spoke about the desire for a conversational discourse but also education provision (the predominantly one-way relay of information).

Script Content for Avachat

Draft scripts were developed for an introduction to Ava and the 4 scenarios. The content for the scripts was derived from clinical information (British Lung Foundation patient literature), behavior change theory (motivational interviewing), emotional well-being advice (New Economics Foundation’s *Five ways to wellbeing*), and peer-driven content (participants). This content was used as a basis for further elaboration of the scenarios by the participants in Workshop 2.

Workshop 2

Characteristics of the Sample

The workshop was held in October 2017, and 5 patients who attended the first workshop also signed up for this second workshop. Of these, 4 attended (male n=2 and female n=2), with 1 drop out because of ill health (male). The median age of the participants was 73 years (range 66 to 80 years), and the average duration of diagnosis was 11.2 years. No data on comorbidities or stage of COPD were collected. Overall, 4 health professionals attended, all of whom had also attended Workshop 1, and all were female, working across secondary care respiratory services and specialist mental health care.

Activities Data

Scenarios

Scenarios for exploration using vignettes included at time of diagnosis (*information provision*), during acute exacerbations (*crisis support*), during periods of low mood (*emotional support*), and for general self-management (*motivation*). Figure 5 illustrates the content of a persona focusing on the crisis support scenario.

Figure 4. User specifications for Avachat (avatar graphic shown courtesy of Bot Libre). COPD: chronic obstructive pulmonary disease; GP: general practitioner.

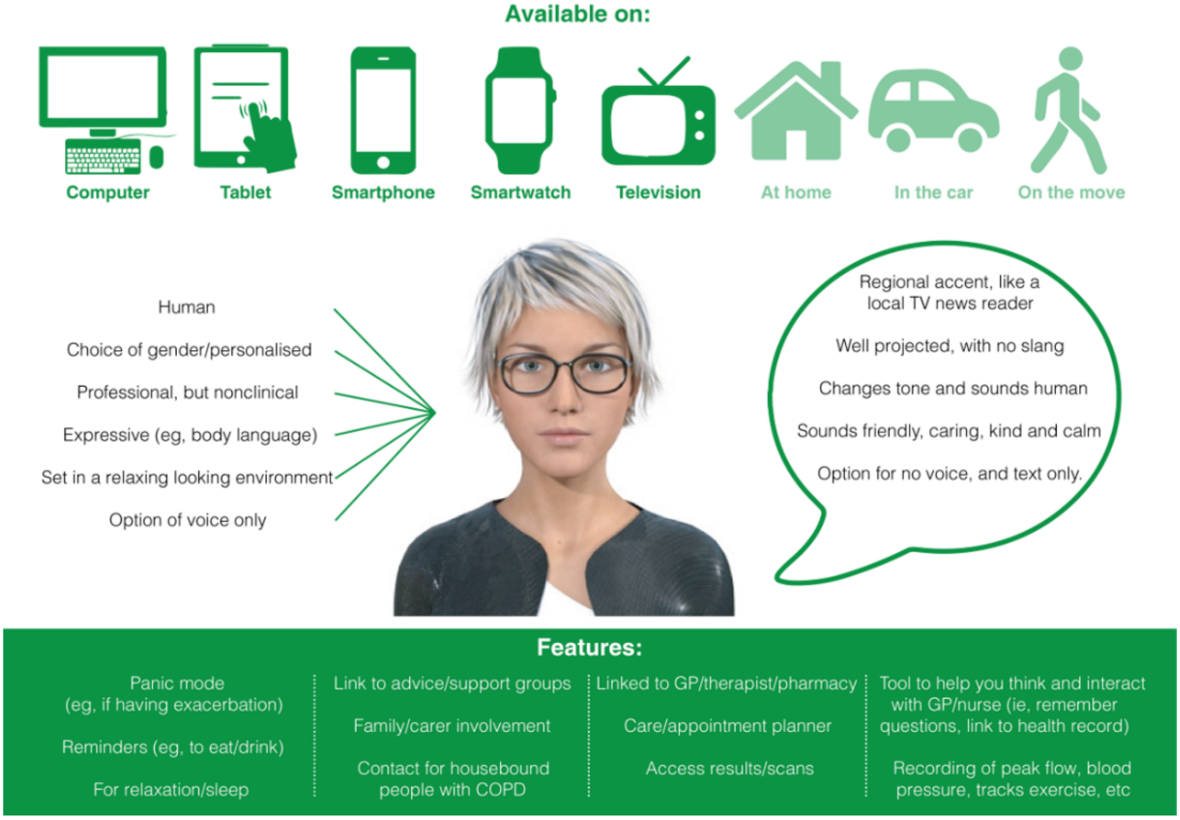
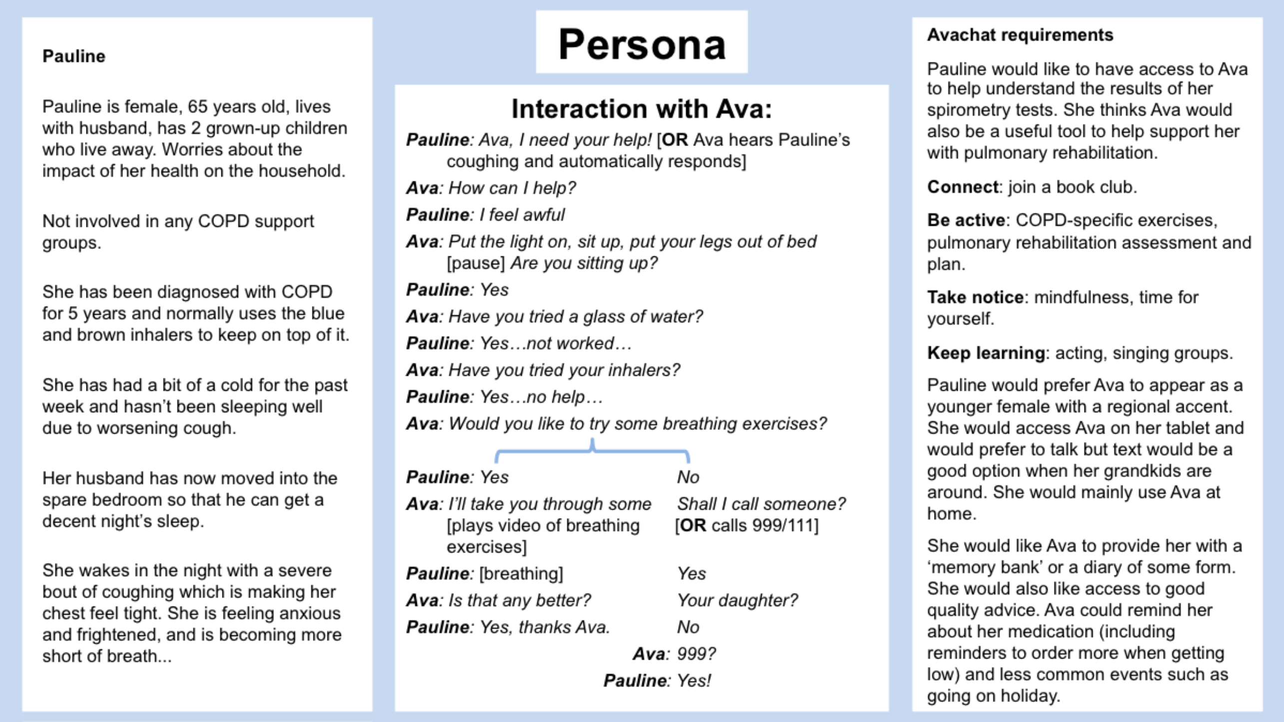


Figure 5. Persona crisis support. COPD: chronic obstructive pulmonary disease; 999 and 111 are the UK telephone numbers for life-threatening and non-life-threatening emergencies respectively.



Pauline

Pauline is female, 65 years old, lives with husband, has 2 grown-up children who live away. Worries about the impact of her health on the household.

Not involved in any COPD support groups.

She has been diagnosed with COPD for 5 years and normally uses the blue and brown inhalers to keep on top of it.

She has had a bit of a cold for the past week and hasn't been sleeping well due to worsening cough.

Her husband has now moved into the spare bedroom so that he can get a decent night's sleep.

She wakes in the night with a severe bout of coughing which is making her chest feel tight. She is feeling anxious and frightened, and is becoming more short of breath...

Persona

Interaction with Ava:

Pauline: *Ava, I need your help!* [**OR** Ava hears Pauline's coughing and automatically responds]

Ava: *How can I help?*

Pauline: *I feel awful*

Ava: *Put the light on, sit up, put your legs out of bed* [pause] *Are you sitting up?*

Pauline: *Yes*

Ava: *Have you tried a glass of water?*

Pauline: *Yes…not worked…*

Ava: *Have you tried your inhalers?*

Pauline: *Yes…no help…*

Ava: *Would you like to try some breathing exercises?*

Pauline: *Yes*	*No*
Ava: *I'll take you through some* [plays video of breathing exercises]	*Shall I call someone?* [**OR** calls 999/111]
Pauline: [breathing]	*Yes*
Ava: *Is that any better?*	*Your daughter?*
Pauline: *Yes, thanks Ava.*	*No*

Ava: *999?*

Pauline: *Yes!*

Avachat requirements

Pauline would like to have access to Ava to help understand the results of her spirometry tests. She thinks Ava would also be a useful tool to help support her with pulmonary rehabilitation.

Connect: join a book club.

Be active: COPD-specific exercises, pulmonary rehabilitation assessment and plan.

Take notice: mindfulness, time for yourself.

Keep learning: acting, singing groups.

Pauline would prefer Ava to appear as a younger female with a regional accent. She would access Ava on her tablet and would prefer to talk but text would be a good option when her grandkids are around. She would mainly use Ava at home.

She would like Ava to provide her with a 'memory bank' or a diary of some form. She would also like access to good quality advice. Ava could remind her about her medication (including reminders to order more when getting low) and less common events such as going on holiday.

Output from Workshop 2

Technical Development

We developed a responsive interface that presented Ava to the user via a variety of devices and allowed spoken or written interactions with the system. In addition, there was a second interface intended for a remote human *wizard*, hidden from the user: this interface would allow an interaction episode to be initiated and subsequently would display the user's utterances and allow the wizard to choose or type an appropriate response and then transmit it to the user interface, where it would be both spoken aloud and displayed onscreen. (It was not considered necessary at this stage to handle more than 1 simultaneous user: we assumed there would be 1 user and 1 wizard per Avachat system.)

The scripts developed during the workshops were encoded internally as simple state-transition models that were used to present preformulated response alternatives as the user and Avachat proceeded through the script. If the user deviated from the main narrative of the script in a way that had not been envisaged, the wizard could fall back on a default *safeguarding* script, involving an emergency call to a relative or for an ambulance, but the wizard was otherwise left to his or her own devices. In this manner, the encoded scripts represented much of the Avachat system's *knowledge* about the support that should be offered to users and, more implicitly, about appropriate ways of offering this support and interacting in an appropriate manner with the users.

Both user and wizard interfaces were implemented for Web browsers (using JavaScript). In addition to providing portability (the interfaces run on any computer with a modern browser installed), this meant that the underlying data communications between the 2 could be facilitated with the use of the internet and conventional HTTP protocols. Avatar graphics and animations for a range of different characters were licensed from a commercial company (Bot Libre) [34]. An initial implementation in Java provided a bespoke Web server for connecting the interfaces as well as the underlying script-processing mechanisms for the wizard. This effectively allows the wizard system and Web server to be run from a single laptop for maximum physical portability, allowing the system to be demonstrated in a variety of contexts. However, after the first workshop, and for greater convenience, the system was reimplemented in Python to run on a conventional, fixed-internet protocol (IP) Web server hosted at the University of Sheffield, which ensured that the system was always available and required a minimum of technical knowledge and time to run. In addition, in the course of reimplementation, and again for convenience, the speech recognition facilities of the system were changed from being based on the open-source Kaldi toolkit [35] to using those built into Google's freely available Chrome browser.

Although this development process resulted in a system that was functional, in practice, unless the user conformed quite closely to one of the developed scripts, a significant amount of input was required of the wizard. Moreover, as it was to be expected that this input would often be health related, unless the wizard had experience of the LTC (as either a person living with the condition or as a health care professional), there was every likelihood that incorrect information or unhelpful (or even unsafe) advice might be shared. As a consequence, rather than attempting *live* interactive sessions with users of the system to explore acceptability, it was instead decided to use a video showing a carefully plotted *safe* interaction.

Wizard of Oz Video-Based Scenario Testing

As described above, we filmed 2 of the elaborated scenarios with a Wizard of Oz implementation of Avachat: during acute exacerbations (*crisis support*) and for general self-management (*motivation*). The film can be viewed in Multimedia Appendix 1. We then screened this video with 8 patients who had a diagnosis of COPD or were a carer of a spouse with COPD: 5 males and 3 females. Overall, 3 of these participants had been part of the co-design process, 5 were novel and new to the concept of Avachat. The median age of the group was 71 years (range 56 to 86 years). The stage of COPD was not reported; however, 1 participant was prescribed oxygen daily. No other comorbidities were reported. All but 1 of the participants used both a smartphone and tablet computer/laptop in their homes. After viewing the film, participants discussed the acceptability of Avachat with regard to design, functionality, and content.

Design

All members of the group apart from 1 reported generally positive feedback on the look and feel (and sound) of the Ava character:

> *I can't believe it really. Impressive. (I can't believe)...that she's not real.* [Participant 1]
>
> *Nice looking girl isn't she.* [Participant 2]
>
> *She says she's here for you—like she will help, doesn't talk down to you.* [Participant 3]
>
> *I think she's lovely. I think she would say the right things and be helpful, just based on the way she looks.* [Participant 1]

The inferred informal interaction with Ava, as demonstrated in the film, was mentioned, as was her role as a companion as opposed to health professional:

> *I like that she says she's not a health professional, more of a companion or assistant but she has information to help.* [Participant 3]

There was some clarity needed on whether Ava was a real doctor or an advisor:

> *She's not medically qualified is she—she just gives advice?* [Participant 4]
>
> *But she says that at the start doesn't she, that she isn't a human or doctor, that she is giving medical advice but not like a real doctor.* [Participant 1]

One participant reported that the technology was not sophisticated enough to appear *real* to them; however, another member of the group felt this could happen with more traditional technologies:

> *I find it off-putting that her mouth doesn't match her words.* [Participant 5]
>
> *Yes but you can get that on the TV even.* [Participant 2]

Functionality

Participants' responses supported the workshop findings that end users would like Ava to be multimodal:

> *I like her on a tablet.* [Participant 3]
>
> *What if you haven't got one? How about a TV?—a lot of people have a TV, and in the bedroom.* [Participant 4]
>
> *I like the laptop because someone else might be using the TV.* [Participant 1]
>
> *Can't you use any? Any at any time—phone, TV, tablet, laptop?* [Participant 2]
>
> *If you were housebound and not up to date with technology then TV would be the most useful and easiest to use and in the bedroom.* [Participant 5]
>
> *Out and about, it's good to have somethings that's mobile.* [Participant 2]
>
> *It could be something that you wear—like on your wrist like a Fitbit—instant access. Then she is there whenever you need her.* [Participant 5]
>
> *Some aren't interested even bit one bit in technology. It would have to be TV for them* [Participant 8]
>
> *Choice is good—individual.* [Participant 6]

In addition, 3 female participants discussed being about to talk to Ava and use both speech and text to communicate:

> *How about like the Amazon Alexa—bit like a radio: talk and it talks back. They can do things for you—scary.* [Participant 1]
>
> *Once you've seen her on an image it would be okay to just have the voice—I would like both.* [Participant 8]
>
> *If you could take her places—like to the doctors' she could record what is said and help you and then you could talk to ava about what was discussed. We've seen her talking but you text her—that would suit me more than anything.* [Participant 3]
>
> *I can't text so we need options.* [Participant 7]

Internet-of-things applications were touched upon by 3 participants:

> *Oh look, she said the air quality in the room is poor and humid, that's great to know.* [Participant 5]
>
> *We could have all sorts attached, oximeter, to check oxygen levels.* [Participant 1]
>
> *If Ava could be wired into someone's home to check you are okay, to listen to the coughing...could you have a pop up on your phone to ask if you are okay "do you want to speak to me?"* [Participant 5]
>
> *I like that she can see the temperature, humidity in the room etc...It helps you figure out why you might be coughing.* [Participant 6]

Moreover, 2 members of the group liked the idea that Ava could encourage you to call emergency services when appropriate:

> *People are scared of ringing the emergency services.* [Participant 5]
>
> *Ava could say you need to ring emergency services but sometimes people just need to be told to do that.* [Participant 1]

Participants were asked if they had any concerns relating to data, with respect to sharing data or where that data may go to once Avachat had received it. No members of the group voiced concerns in this area and 3 participants felt that sharing of data was commonplace in today's society:

> *We're not bothered, they [the government] know more about us than we do.* [Participant 2]
>
> *We are thinking it's all safe.* [Participant 6]
>
> *It's all on a database.* [Participant 3]

Content

When discussing the scenario of motivational support, the group thought this would be particularly useful for those living alone:

> *Brilliant that, a good idea.* [Participant 4]
>
> *Ava encouraged him to go out. If there is no one there to push you, you feel ill and your mood drops you don't want to know. It can happen quite quickly.* [Participant 5]
>
> *As you get older you don't have the social life. You don't go out in the evenings even. You don't go to the pub as you can't drive and drink.* [Participant 2]
>
> *You always had someone around you in the past and small children can lift your spirits, but if you don't have that...mine [children] are miles away. If you need someone quickly...the nearest one is 3 hours away, but if you needed them, and they can't always drop everything.* [Participant 3]
>
> *If you have a partner, they can encourage you.* [Participant 6]
>
> *It's a lonely life if you don't have anyone. That can be very depressing. It's different if you have a partner, you can talk to them and they can talk to you and help in many ways.* [Participant 2]

One member also felt Ava could be useful for people who do not want to burden their partners or who do not feel comfortable doing so:

> *Your partner doesn't always understand and you don't want to burden them with things.* [Participant 8]

All participants reflected on the experience of acute COPD exacerbation and were particularly positive about this aspect of support from Avachat:

> *I were in on my own once, and I had this bad panic do, lucky my sister came, but I couldn't get to the phone. I was right as rain at first. I had a coughing do and I must have panicked. It calms me down....the thought I could have someone to talk to, that's good.* [Participant 2]
>
> *It certainly helps to know that someone is there.* [Participant 8]

After the screening, each participant was asked to imagine using the Avachat system in their own contexts, and, based on this, to complete a SUS questionnaire. Overall, 62% (5/8) of the SUS scores were above the score of 68 (with scores of 72.5, 75, 75, 80, and 85), with the other 38% (3/8) participants rating the system as follows: 52.2, 62.5, and 65. This gives the system an overall median rating of 73.75. 50% (4/8) of the participants strongly agreed with the statement "I think that I would like to use this system frequently," and the other 50% (4/8) were unsure. Moreover, 88% (7/8) of the participants strongly agreed that the system was easy to use, possibly as the video shows a seamless use of the agent in practice. All but 1 of the participants strongly agreed that they would feel confident using the system. Views on whether they would need the support of a technical person to be able to use the system were split: half of the participants strongly disagreed that they would and half strongly agreed that they would. Again, with the statement "I needed to learn a lot of things before I could get going with this system," 38% (3/8) participants strongly disagreed, 38% (3/8) strongly agreed, and 25% (2/8) were neutral.

Discussion

Principal Findings

Our aim was to explore the acceptability of both the form and the content of a computerized virtual agent with natural language capabilities intended to support self-management for people with comorbid LTCs and mental health problems. Through co-design with people living with COPD, we identified and prioritized 4 scenarios in which the system may be applied: at the time of diagnosis (*information provision*), during acute exacerbations (*crisis support*), during periods of low mood (*emotional support*), and for general self-management (*motivation*). The participants did not separate mental and physical health needs during the co-design process, although the generated content utilized psychological approaches to supporting patients to self-manage. Triangulation of clinically accurate information, behavior change techniques, emotional well-being advice, and peer-driven support appears to be the preferred content by patients. The results suggest that, for this small sample of older adults with comorbidities, it would be acceptable to receive both self-management support and support for acute exacerbations from an AI-based virtual agent.

The user requirements for the system were identified and many have been incorporated into a prototype system called Avachat. Specifically, we determined that the virtual agent must have the ability to understand and respond with both text and speech and must be accessible through a variety of domestic computing devices. The ability to interact with other digital services and domestic devices such as calendars, smart home technology, and medical peripherals is desirable. The system should be personalizable and its personification (Ava) should be emotionally expressive, and the system would be enriched by the ability to detect emotion (distress, fatigue, and irritation) in speech if it is to engage with users in a natural manner.

A blended approach to care has been shown to be most effective and acceptable for patients when managing both physical and emotional health [19]; however, as demand for support outstrips supply, it is crucial to explore what elements of person-to-person care could be provided using digital and AI-based systems. Indeed, a James Lind Alliance Priority Setting Partnership on digital technology and mental health, tasked with identifying the top 10 research questions posed by patients and clinicians, asks "Can common elements of therapy that come from person-to-person interactions be maintained with digital technology?" (priority 8) [36]. Our work speaks to this research question, in providing some evidence that a virtual agent could replicate certain aspects of traditional peer support incorporating behavior change techniques. With this patient group, such support is generally used to help patients better self-manage their conditions and decrease adverse health outcomes such as exacerbations, unnecessary emergency department attendance, and hospital admissions. Of course, patient outcomes are outside the scope of this work (and the capabilities of the current technology fall short of being able to provide such support in reality), so validation of the benefits for patients is work still to be done.

This empathetic ability of an automated system to detect and react appropriately to the nonverbal cues present in a person's spoken utterances is likely to be crucial for the therapeutic relationship between agent and patient. Although not convenient for all people and in all situations, spoken language is the most immediate communication medium we possess and provides the basis for many conventional psychological or *talking* therapies. *Closed-loop* language interfaces with computers, which are able to both process the user's speech and generate an appropriate spoken response, are becoming more sophisticated and are already found in popular consumer digital devices such as Apple's Siri and Amazon's Alexa (and, hence, are more and more a feature that users expect, at least in the English-speaking world; at the time of writing, the provision for other languages depends both on the number of speakers it has and the size of the consumer market they represent). Embodied virtual agents based on AI technology are already used in health-related software tools and apps, including a handful of mental health-focused products [37-42]. These have rarely been tested with samples of older adults (where most comorbidities are evident) [37,38]. A recent scoping review of conversational agents in clinical psychology [38] indicates that the ability of unconstrained natural language input capabilities for health-related purposes is an emerging field of research, although a computer agent that is able to maintain its side of a rich, open-ended dialogue necessary for the delivery of most psychotherapies is still some way off [43]. Even if not yet a practical proposition, we assume that users will soon expect the primary interaction modality of any autonomous therapeutic agent to be spoken natural language.

Participants in this study readily accepted Ava as a *human*, referring to it as *she*; they expressed a liking for her in terms of appearance, while the way in which she interacted with them suggests she *understands* the situation. Taken together, this suggests that the potential exists for the development of a relationship of sorts between human and autonomous agents, a suggestion supported by a recent study investigating the use of text-based AI to deliver psychological therapy [42].

Limitations

Although based on input from those with lived experience of COPD and self-management, our findings were limited by the small sample of white, British, medically stable, regional participants recruited. None of the participants reported a current mental health problem but all referred to experiencing episodes of low mood or worry in the past. Furthermore, their existing engagement with the support groups through which they were recruited and their ability to attend the workshops suggest that they are representative of the type of patients who already self-manage their health. Although this meant that they could help us to identify the *best practice* with regard to self-management, it does raise doubts about whether the principles identified hold equally for those who cope less well or who are less experienced in self-management. In the next phase of our research, we will use more in-depth methods to engage hard-to-reach individuals, including those who are medically unstable, house bound, or from black and ethnic minority groups, those experiencing mental health conditions, or those for whom English is not their first language.

There was relatively little input from the health practitioners compared with patients during the workshops. This allowed the patient's voice to dominate; however, it is not clear whether the health professionals agreed with or supported the final scenarios as a clinical priority or the content as best practice. Additional research is required to explore this stakeholder perspective and to place the system appropriately in the clinical pathway/patient journey or redesign that pathway, as appropriate.

This was not a software development project, although the research team did include software engineers, and as a result the amount of effort that could be assigned to development was limited. Our focus was the rapid development of an Avachat system that incorporated the user requirements as they emerged from the workshops; this enabled us to present the system to the participants to further refine their requirements and help elicit their wider opinions about the role of such technologies in their lives. This approach is promoted in the literature [38]. The system did not need to be fully functional (and, in any case, given the current level of technology available to us this would have been impossible); however, we wanted to produce a system that was as functional as possible and, hence, as accurate a reflection as possible of the short-to-medium-term potential of the technology. A purely imaginary requirements-gathering exercise for such a system would be difficult to constrain and would be likely to lead to a wholly impractical and unachievable specification. These considerations influenced many of the design and implementation choices for the system.

A point worthy of reflection is that rarely were the possible negative consequences of the technology discussed during the co-design and evaluation process. We did not raise these issues, preferring to allow the discussions to be participant-led and to specifically raise issues of duty of care, for example. Equally, consideration was given neither to the possibility of incorrect or inappropriate advice being given by the system nor to the wider ethical implications for the nature and standard of care that users would receive. It seems as though it was assumed that the system would be operate *perfectly*, without error or ambiguity and always in the best interests of the patient, and with no associated reduction in existing care. It is possible that, as consumers, we simply cannot fully understand the complex and dispersed data architectures and the complex distribution of roles, responsibilities, rights, and obligations that underlie newest digital technologies.

Conclusions

Our research focused on co-designing the content and application of a conversational virtual agent to support self-management of physical and mental health comorbidities for patients with a diagnosis of COPD. We identified the underlying design concepts that would make the system useful and acceptable to the target population. Technological developments such as the automated delivery of therapy and health care support pose significant and difficult questions for our societies. Every new application of technology is accompanied by risks, and in this case, those risks touch upon the things we prize most dearly: health, social interaction, nature of work, and fair and appropriate distribution of limited resources. Although these developments can sometimes seem inevitable, the form that they take is not predetermined. Early stage co-design efforts such as those reported here represent a mechanism by which patient groups and their caregivers can have a voice in the process, and at a stage when not all the significant decisions have already been made.

Future research is planned to further develop and evaluate the system for a range of patient samples and care settings, in particular, crisis support. Clinical applications are broad. The benefits of a straightforward, low-cost system could include increased self-efficacy and self-confidence, reduced social isolation, and improved illness knowledge. Clinical gains could include fewer clinical exacerbations and a decrease in conventional health care service utilization, with an attendant reduction in costs.

Acknowledgments

This research was funded and supported by the NIHR Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber. The views and opinions expressed are those of the authors and are not necessarily those of the NHS, NIHR, or Department of Health, United Kingdom. The authors would like to thank the Sheffield Breathe Easy group, the patients, and the clinicians involved for taking part and sharing their expertise. The authors are grateful for the use of the Home Lab facilities of the Centre for Assistive Technology and Connected Healthcare at the University of Sheffield.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Avachat film used for video-based scenario testing.

[MOV File, 159MB - jmir_v21i5e12996_app1.mov]

References

1. Naylor C, Parsonage M, McDaid D, Knapp M, Fossey M, Galea A. Long-Term Conditions and Mental Health: The Cost of Co-Morbidities. London: The King's Fund and Centre for Mental Health; 2012.
2. Department of Health. No Health Without Mental Health: A Cross-Government Mental Health Outcomes Strategy for People of All Ages. London: TSO; 2011.
3. National Institute for Health and Care Excellence. 2009. Depression in adults with a chronic physical health problem: recognition and management URL: https://www.nice.org.uk/guidance/cg91 [accessed 2019-05-01] [WebCite Cache ID 781XPd7cr]
4. Cimpean D, Drake R. Treating co-morbid chronic medical conditions and anxiety/depression. Epidemiol Psychiatr Sci 2011 Jun;20(2):141-150. [Medline: 21714361]
5. Hynninen KM, Breitve MH, Wiborg AB, Pallesen S, Nordhus IH. Psychological characteristics of patients with chronic obstructive pulmonary disease: a review. J Psychosom Res 2005 Dec;59(6):429-443. [doi: 10.1016/j.jpsychores.2005.04.007] [Medline: 16310027]
6. Livermore N, Sharpe L, McKenzie D. Panic attacks and panic disorder in chronic obstructive pulmonary disease: a cognitive behavioral perspective. Respir Med 2010 Sep;104(9):1246-1253 [FREE Full text] [doi: 10.1016/j.rmed.2010.04.011] [Medline: 20457513]
7. 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 8;370(9590):851-858. [doi: 10.1016/S0140-6736(07)61415-9] [Medline: 17826170]
8. Easton K, Coventry P, Lovell K, Carter LA, Deaton C. Prevalence and measurement of anxiety in samples of patients with heart failure: meta-analysis. J Cardiovasc Nurs 2016;31(4):367-379 [FREE Full text] [doi: 10.1097/JCN.0000000000000265] [Medline: 25930162]
9. Cepoiu M, McCusker J, Cole MG, Sewitch M, Belzile E, Ciampi A. Recognition of depression by non-psychiatric physicians--a systematic literature review and meta-analysis. J Gen Intern Med 2008 Jan;23(1):25-36 [FREE Full text] [doi: 10.1007/s11606-007-0428-5] [Medline: 17968628]

10. Katon W. Clinical and health services relationships between major depression, depressive symptoms, and general medical illness. Biol Psychiatry 2003 Aug 1;54(3):216-226. [Medline: 12893098]
11. The King's Fund. Has the government put mental health on an equal footing with physical health? URL: https://www.kingsfund.org.uk/projects/verdict/has-government-put-mental-health-equal-footing-physical-health [accessed 2019-05-01] [WebCite Cache ID 74I4Fw8cp]
12. Mental Health Taskforce to the NHS in England. NHS England. 2016. The five year forward view for mental health URL: https://www.england.nhs.uk/publication/the-five-year-forward-view-for-mental-health/ [accessed 2019-05-01] [WebCite Cache ID 782ZZ8KQo]
13. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
14. Hibbard J, Gilburt H. The King's Fund. 2014. Supporting people to manage their health: an introduction to patient activation URL: https://www.kingsfund.org.uk/publications/supporting-people-manage-their-health [accessed 2019-05-01] [WebCite Cache ID 74I74v4YK]
15. de Silva D. Evidence: Helping People Help Themselves. London: The Health Foundation; 2011.
16. The Evidence Centre. National Voices.: National Voices; 2014. Supporting self-management: a summary of the evidence URL: https://www.nationalvoices.org.uk/publications/our-publications/supporting-self-management [accessed 2019-05-01] [WebCite Cache ID 74I6iuVaS]
17. Cotton R, Irwin J, Wilkins A, Young C. The NHS Confederation.: The NHS Confederation; 2014. The future's digital: Mental health and technology URL: https://www.nhsconfed.org/resources/2014/09/the-future-s-digital-mental-health-and-technology [accessed 2019-05-01] [WebCite Cache ID 74I6pnoGD]
18. Bee P, Lovell K, Lidbetter N, Easton K, Gask L. You can't get anything perfect:. Soc Sci Med 2010 Oct;71(7):1308-1315. [doi: 10.1016/j.socscimed.2010.06.031] [Medline: 20675026]
19. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
20. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
21. Knowles S, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. BMJ Open 2015 Nov 30;5(11):e008581 [FREE Full text] [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]
22. Luxton DD, editor. Artificial Intelligence in Behavioural and Mental Healthcare. London: Academic Press; 2015.
23. Fitzpatrick K, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
24. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. Behav Cogn Psychother 2014 Nov;42(6):731-746. [doi: 10.1017/S135246581300060X] [Medline: 23899405]
25. Collaboration for Leadership in Applied Health Research and Care. National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber URL: http://clahrc-yh.nihr.ac.uk/home [accessed 2019-05-01] [WebCite Cache ID 782dZVFXO]
26. Stickdorn M, Schneider J, editors. This Is Service Design Thinking. Amsterdam: BIS; 2011.
27. McNabola A, Moseley J, Reed B, Bisgaard T, Dorthe Josiassen A, Melander C, et al. Design Council.: Design Council; 2013. Design for public good URL: https://www.designcouncil.org.uk/resources/report/design-public-good [accessed 2019-05-01] [WebCite Cache ID 782eVPAbX]
28. Holmes BJ, Best A, Davies H, Hunter D, Kelly MP, Marshall M, et al. Mobilising knowledge in complex health systems: a call to action. Evid Policy 2017 Aug 14;13(3):539-560. [doi: 10.1332/174426416X14712553750311]
29. Langley J, Wolstenholme D, Cooke J. 'Collective making' as knowledge mobilisation: the contribution of participatory design in the co-creation of knowledge in healthcare. BMC Health Serv Res 2018 Jul 25;18(1):585 [FREE Full text] [doi: 10.1186/s12913-018-3397-y] [Medline: 30045726]
30. Brooke J. SUS: A 'quick and dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation In Industry. London: Taylor & Francis Ltd; 1996.
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Weber R. Basic Content Analysis. Newbury Park, CA: Sage Publications; 1990.
33. New Economics Foundation. Five ways to emotional wellbeing: The evidence URL: https://neweconomics.org/2008/10/five-ways-to-wellbeing-the-evidence [accessed 2019-05-01] [WebCite Cache ID 74I7a0kQM]
34. Bot Libre: Free open artificial intelligence for everyone. URL: https://www.botlibre.com/ [accessed 2019-05-01] [WebCite Cache ID 782gYuMTW]

35. Kaldi. URL: http://kaldi-asr.org/ [accessed 2019-05-01] [WebCite Cache ID 782go1GV7]
36. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
37. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
38. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 Dec 9;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
39. Rus-Calafell M, Jamieson-Craig T, Garety PA, Ward T, McCrone P, Emsley R, et al. Reducing the frequency and severity of voices: the avatar clinical trial. Eur Psychiatr 2014;29:1. [doi: 10.1016/S0924-9338(14)77707-7]
40. Ortiz A, del Puy Carretero M, Oyarzun D, Yanguas JJ, Buiza C, Gonzalez MF, et al. Elderly users in ambient intelligence: does an avatar improve the interaction? In: Stephanidis C, Pieper M, editors. Universal Access In Ambient Intelligence Environments: 9th Ercim Workshop On User Interfaces For All, Königswinter, Germany, September 27-28, 2006, Revised Papers (Lecture Notes In Computer Science). Berlin: Springer; 2007.
41. Carrasco E, Epelde G, Moreno A, Ortiz A, Garcia I, Buiza C, et al. Natural interaction between avatars and persons with Alzheimer's disease. In: Miesenberger K, Klaus J, Zagler W, Karshmer A, editors. Computers Helping People With Special Needs. ICCHP 2008. Lecture Notes in Computer Science, vol 5105. Berlin: Springer; 2008.
42. Ly K, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
43. Fonseca A. National Elf Service: The Mental Elf. 2017 Dec 7. The chatbot therapist: can we mimic therapy support with a conversational agent? URL: https://www.nationalelfservice.net/treatment/digital-health/the-chatbot-therapist-can-we-mimic-therapy-support-with-a-conversational-agent-digimhweek/ [accessed 2019-05-01]

Abbreviations

AI: artificial intelligence
COPD: chronic obstructive pulmonary disease
GP: general practitioner
LTC: long-term condition
NHS: National Health Service
NIHR: National Institute for Health Research
SUS: system usability scale

Edited by G Eysenbach; submitted 29.11.18; peer-reviewed by L Garvin, K Blondon; comments to author 27.02.19; revised version received 13.03.19; accepted 24.03.19; published 30.05.19.

Please cite as:
Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, Mirheidari B, Weich S, de Witte L, Wolstenholme D, Hawley MS
A Virtual Agent to Support Individuals Living With Physical and Mental Comorbidities: Co-Design and Acceptability Testing
J Med Internet Res 2019;21(5):e12996
URL: http://www.jmir.org/2019/5/e12996/
doi:10.2196/12996
PMID:31148545

Review

Psychosocial Health Interventions by Social Robots: Systematic Review of Randomized Controlled Trials

Nicole Lee Robinson[1,2], PhD; Timothy Vaughan Cottier[2], BBehavSc(Hons); David John Kavanagh[2], PhD

[1]Australian Centre for Robotic Vision, Brisbane, Australia
[2]Queensland University of Technology, Brisbane, Australia

Corresponding Author:
Nicole Lee Robinson, PhD
Australian Centre for Robotic Vision
2 George Street
Brisbane, 4000
Australia
Phone: 61 7 3138 0607
Email: n7.robinson@qut.edu.au

Abstract

Background: Social robots that can communicate and interact with people offer exciting opportunities for improved health care access and outcomes. However, evidence from randomized controlled trials (RCTs) on health or well-being outcomes has not yet been clearly synthesized across all health domains where social robots have been tested.

Objective: This study aimed to undertake a systematic review examining current evidence from RCTs on the effects of psychosocial interventions by social robots on health or well-being.

Methods: Medline, PsycInfo, ScienceDirect, Scopus, and Engineering Village searches across all years in the English language were conducted and supplemented by forward and backward searches. The included papers reported RCTs that assessed changes in health or well-being from interactions with a social robot across at least 2 measurement occasions.

Results: Out of 408 extracted records, 27 trials met the inclusion criteria: 6 in child health or well-being, 9 in children with autism spectrum disorder, and 12 with older adults. No trials on adolescents, young adults, or other problem areas were identified, and no studies had interventions where robots spontaneously modified verbal responses based on speech by participants. Most trials were small (total N=5 to 415; median=34), only 6 (22%) reported any follow-up outcomes (2 to 12 weeks; median=3.5) and a single-blind assessment was reported in 8 (31%). More recent trials tended to have greater methodological quality. All papers reported some positive outcomes from robotic interventions, although most trials had some measures that showed no difference or favored alternate treatments.

Conclusions: Controlled research on social robots is at an early stage, as is the current range of their applications to health care. Research on social robot interventions in clinical and health settings needs to transition from exploratory investigations to include large-scale controlled trials with sophisticated methodology, to increase confidence in their efficacy.

(J Med Internet Res 2019;21(5):e13203) doi:10.2196/13203

KEYWORDS

social robot; healthcare; treatment; therapy; autism spectrum disorder; dementia

Introduction

Background

In recent years, we have seen exciting developments in the application of robotics to medical treatments. Medical robot–assisted surgery in operating theaters enhances patient outcomes of surgical procedures in orthopedics, radiosurgery, and neurology [1-6]. Exoskeleton devices work to enhance strength or improve movement for patients suffering from traumatic brain and spinal cord injury, disability such as stroke and multiple sclerosis, and rehabilitation treatment [7-9]. Surgical and rehabilitative robotics offer distinct advances in their exceptional ability to augment treatment practices to enhance patient outcomes but are restricted to a highly specific field of medical assistance. This leaves other health care services untouched by the potential benefits that robotics may offer for health care professionals and their patients, including psychosocial interventions for health or well-being.

At first thought, robotics for psychosocial interventions may seem counterintuitive. If, as has been argued, the therapeutic relationship is key to positive treatment outcomes [10], how could a robot perform such a task? However, there are precedents for such a role. Strong positive effects have been obtained from digital mental health programs for anxiety and depression (eg, g=0.80 and ds=0.49-1.14) [11,12] and small-to-moderate effects for alcohol use [13,14]. Although having a therapist or coach to guide the use of these interventions assists in maintaining engagement and appears to give somewhat better outcomes [15], significant effects can also be obtained by self-guided programs [11,12,16], and their low unit cost means that self-guided programs are easier to take to scale and have superior cost-effectiveness as numbers increase [17]. In some direct comparisons, coached and self-guided programs have even been able to achieve similar treatment outcomes [18]. Although self-guided programs achieve these effects without a therapeutic relationship, we argue that they potentially satisfy other elements of a therapeutic alliance such as perceived safety and consistency with personal goals and avoid many negative effects of face-to-face therapy, such as perceived judgment or stigma. Interestingly, the scores on therapeutic alliance measures from users of self-guided programs can be quite high [19]. It is plausible that a social robot (a robot that can communicate and interact with people) could offer education, model some skills, and deliver a fixed intervention program. As we discuss later in the paper, even more sophisticated therapy may be offered in the future, with emerging developments in robotic technology.

Existing Reviews

Identified systematic reviews are for social robot interventions in highly specialized areas of elderly care [20,21] and autism spectrum disorder (ASD) [22,23]. They contain a mixture of experimental methodologies such as single subject, quasi-experimental, cross-sectional without control, and free interaction. Mixed-trial designs have generally been considered acceptable when evaluating the initial prospect of a novel intervention [24,25]. However, robotic interventions do require critical evaluation using a series of high-quality trial designs to demonstrate sufficient evidence to achieve effective health outcomes. Current reviews that contain mixed-experiment designs present a limitation around the conclusive nature of identified experimental studies, especially when appraising the use of robotic interventions in routine clinical practice. A high-impact method of clinical trial design involves a randomized controlled trial (RCT), which reduces the influence of bias and confounds on trial outcomes by scientifically rigorous methods of intervention testing to assess treatment benefits [26]. Several systematic reviews using RCTs in surgical robotic interventions have been published [27,28], but evidence has yet to be synthesized for RCTs for social robots to deliver psychological interventions for health or well-being across all age groups.

Aim

The aim of this systematic review was to undertake a comprehensive examination of existing RCTs on the use of social robots to deliver psychosocial interventions for health or well-being. The review is timely, given the fast pace of developments in robotics, the rapid uptake of social robots that is likely to occur, and the wide-ranging nature of their potential applications to improved health care and self-management support.

Methods

Literature Review and Selection of Trials

This systematic review protocol followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for its search, screening, and evaluation processes [29]. Database searches were conducted in Medline, PsycInfo, ScienceDirect, Scopus, and Engineering Village in November 2018. Health, psychology, engineering, and computer science databases were chosen to maximize the chance of identifying published trials that fulfilled the search criteria. Each search used (Title: Robot*) AND (Abstract: Health* OR Anxi* OR Depress* OR Distress* OR Disorder* OR Autis* OR Dement*) AND (Abstract: Therap* OR Behav* OR Treat* OR Intervention* OR Counsel* OR Psychosocial OR Psychotherap*) AND (Title:Abstract:Key: Random*). Medline and PsycInfo searches used Boolean and phrase search modes and included all the results for source types and years. ScienceDirect and Scopus searches were refined to include all years with no document type exclusions in the result searches. Engineering Village contained the GeoRef, Inspec, and Compendex databases, including all years and no document type exclusions. The identified papers from the databases were supplemented by backward and forward searches (ie, checking titles in reference lists and citations of identified papers for any additional studies).

Inclusion and Exclusion

The eligible trials for the review (1) used a social robot to deliver a psychosocial intervention for health or well-being (ie, one that used verbal communication or other social interaction) and (2) examined the effects of at least 2 conditions in an RCT over at least 2 measurement occasions. A social robot was defined as a humanoid or nonhumanoid robot that could communicate or interact with people using verbal or nonverbal communication or both. These robots could vary from ones with rudimentary abilities (eg, minor motor movements and no communicative speech) to ones with advanced communicative abilities designed to present the illusion of social intelligence. The included trials could use robots operated by staff using Wizard of Oz controls, given that the outcomes of robot-delivered interventions were of greater interest than the ability of robots to deliver interventions autonomously. However, trials using technological agents without embodiment (eg, chatbots or avatars) were excluded, as were ones using robotic devices without communicative abilities, such as prosthetic devices and teleoperated, surgical, and exoskeletal robots.

The papers could have been published in any year. In recognition of the acceptability of conference proceedings as publication outlets for engineering and computer science, the papers could be in a peer-reviewed journal or conference proceedings. Multiple papers on different aspects of a single trial were all used to provide information, but if multiple papers presented the same material, the most complete and current report was selected for evaluation and review.

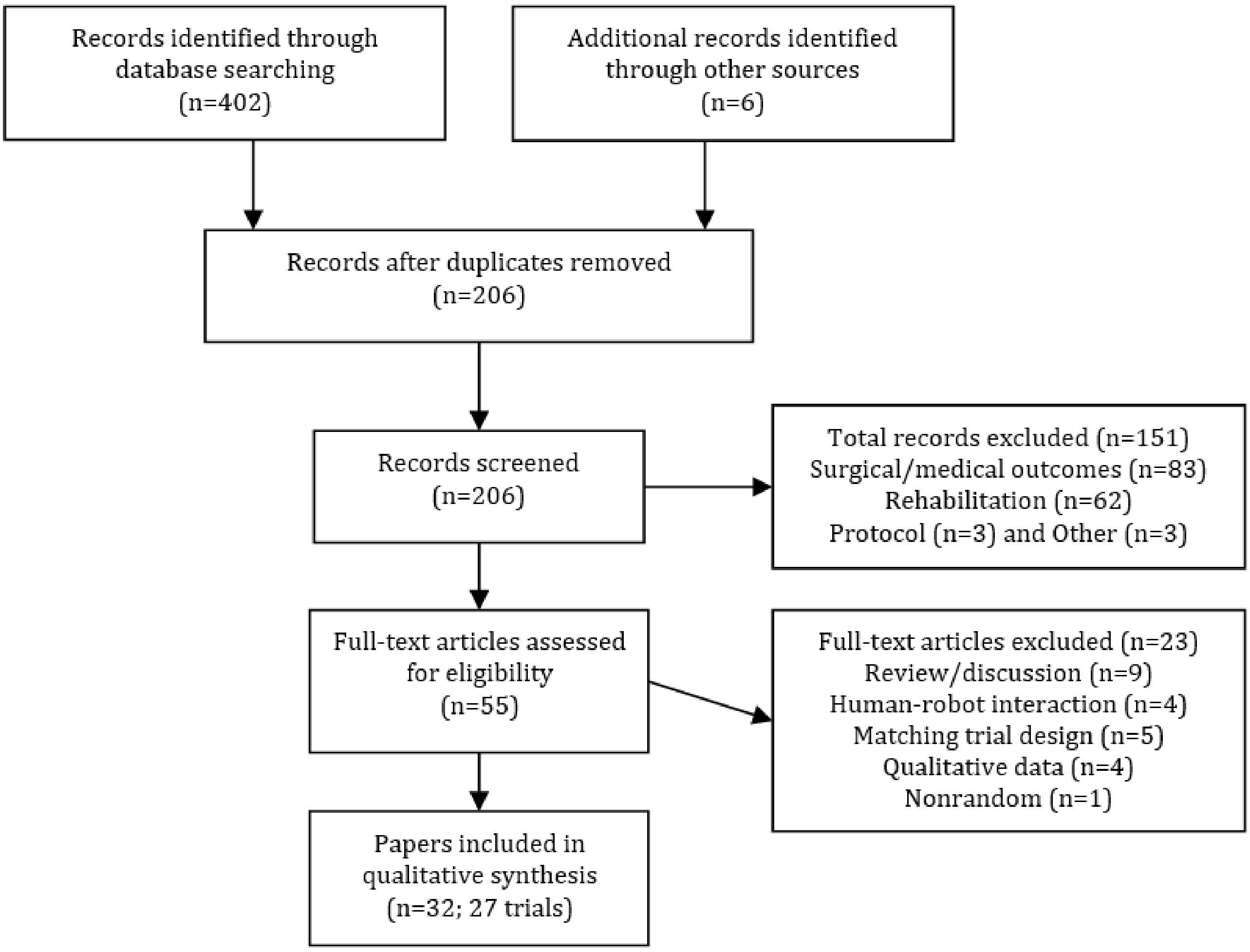
Data Extraction and Analysis

Data extraction was conducted in November 2018 by NLR and TVC and reviewed for consistency and accuracy by all 3 authors. All eligible papers were extracted directly from academic databases. The authors were not contacted to provide additional data or unpublished results. Trial extraction involved an initial screen, assessing titles for relevance. Selected papers were further appraised using abstracts, and papers that appeared to meet the criteria were independently reviewed for eligibility and coding in the presented tables by NLR and DJK. In cases of any disagreement in inclusion or coding, the point was checked in the paper and consensus was reached on the final decision. Human-robot interaction factors such as acceptability, likability, and trust of the robot were not reviewed in detail as that lay outside this review. The presented results in the tables include statistically significant ones only: other listed measures were not used as outcomes or did not give significant results.

Results

The initial search identified 402 records from 5 databases, plus 6 that were identified through forward and backward searches. Identification of duplicates using title, year, and authors resulted in 202 records being removed, leaving 206 for screening. A total of 151 were excluded based on irrelevant titles (eg, surgical and medical trials, exoskeletons, protocols, and economic analyses) and 23 were excluded based on a detailed examination of the abstract and full text, leaving a total of 27 trials for full evaluation. The details of reasons for exclusion are in Figure 1.

Figure 1. Systematic review flow chart.



Included Trials

Overall, 6 of the 27 included trials (22%) addressed child health or well-being, 9 (33%) were on children with ASD, and 12 (45%) were on older adults and focused on cognitive or psychological functioning. The most commonly used robots were the NAO humanoid from Softbank Robotics [30], and the PARO harp seal companion robot from PARO Robots [31] (7 trials), although 13 other robots were each used in at least 1 trial. Owing to the wide range of measures and the limited consistency in the presentation of results in different trials, it was not possible to conduct a meta-analysis or to report effect sizes in a standard manner. Accordingly, the systematic review is descriptive.

Child Interventions for Health or Well-Being

The 6 trials on children's health or well-being are summarized in Table 1. All had the individual child as the unit of randomization, but only Beran et al [32] and Jibb et al [33] reported computerized or Web-based randomization. Sample sizes ranged from 5 to 57 (median=34). Participants were aged from 4 to 14 years (median reported average=9.9 years), and samples were drawn from Canada, the Netherlands, and Iran. Durations of studies ranged from 1 to 18 weeks (median=4.5 weeks), and no trials had a follow-up assessment. Only Jibb et al [33] reported blind observational coding. All used NAO robots, with preprogramming [32,33] or Wizard-of-Oz individualization [34,35]. The number of treatment sessions ranged from 1 to 10 (median=3).

One study [35] demonstrated a significantly greater rise in diabetes knowledge when a social robot administered a diabetes quiz to children with type 1 diabetes compared with a usual care control. A more personal robot elicited greater pleasure and feelings of self-determination from the participants during the final session, but there were no differences between the robot types on diabetic knowledge. In addition, 3 studies obtained reductions in negative emotions when a social robot was used to assist needle insertion or to address emotional responses in oncology patients or children learning a foreign language [32,36,37]. Less pain was reported about needle insertion in the study by Beran et al [32] and less avoidance to the needle insertion in both the studies by Beran et al [32] and Jibb et al [33]. None of the studies assessed sustained changes in distress, quality of life, health-related behavior, or health outcomes. Although this set of studies provided some evidence in favor of robot use in children’s well-being, conclusions were limited by infrequent blind assessment and a lack of follow-up data or information on behavioral or functional impacts. Research in this area is at a very early stage.

Children Interventions for Autism Spectrum Disorder

Overall, 9 trials of robot interventions for children aged 4 to 12 years with ASD or pervasive developmental disorder were identified (Table 2). In addition, 6 trials randomized individuals to conditions, 2 randomized to condition order, and 1 had cluster randomization. None of the trials reported independent randomization, and only 2 trials [38-41] reported using computer-generated randomization. Moreover, 3 studies [40,42,43] had a single-blind assessment, and all but 1 study [44] reported that the reliability of observations was confirmed against another rater (in that study, a single rater was used, but reliability was established in training).

Table 1. Child interventions delivered via a robot for health or well-being.

Author	Sample	Design, conditions (n)[a]	Duration	Measures	Outcomes[b]
Beran et al [32]	57 Canadian vaccination patients (30 male, 53%), aged 4-9 years (mean 6.9, SD 1.3)	Robot CBT (28): Distraction before, during, and after injection; control (29): standard nurse administration	1 session[c] at vaccination	Faces Pain Scale-Revised (FPS-R); Behavioral Approach—Avoidance Distress Scale (BAADS)	Robot Cognitive-Behavior Therapy versus control during session: <pain (FPS-R) from parent, child*, nurse*, and researcher and* <BAADS Distress**, Avoidance***
Blanson Henkemans et al [34]	5 Dutch type 1 diabetes patients (3 male, 60%) aged 9-12 years (mean 10.2, SD 1.3)	Game-like quizzes (10 out of 20 of the questions on diabetes each session); personal robot (3): eyes in favorite color, used child's name, mentioned child's favorite activity, asked opinion of game, if wanted to keep playing, etc; and neutral robot (2): no personalization	3 sessions (45, 45, and 30 min) at 2-3 week intervals	Type 1 diabetes knowledge; health-related quality of life; and Mind Youth Questionnaire (MY-Q)	Across conditions, Pre to Session 3[d]: > correct diabetes questions*
Alemi and Meghdari [36]	46 Iranian female students aged 12-13 years, with beginners' level English	Individual randomization to classes; robot-assisted language learning (RALL; 30 in 2 groups); and control: teacher only (1 group of 16)	10 × 1-hour sessions over 5 weeks	Foreign Language Classroom Anxiety Scale (FLCAS); attitude questionnaire	RALL versus control at 5 weeks: >FLCAS (less anxiety)*
Alemi et al [37]	11 Iranian oncology patients (1 male, 9%) aged 7-12 years (mean 9.5, SD 1.6)	Social robot-assisted therapy (SRAT, 6): robot took roles of doctor, chemo-hero, nurse, cook, ill kid; shared hopes and dreams, said goodbye and control: psychologist only (same content; 5)	8 sessions[c] over 1 month	Multidimensional Anxiety Children Scale (MASC); Children's Depression Inventory (CDI); and Children's Inventory of Anger (CIA)	SRAT versus control, pre versus 1 month: >falls in anxiety (MASC)**, depression (CDI)*, and anger (CIA)*
Blanson Henkemans et al [35]	28 Dutch type 1 diabetes patients (13 male, 46%) aged 7-14 years (mean 11.0, SD=1.7)	Diabetes education quizzes; Personal robot (9): as per Blanson Henkemans et al [34]; neutral robot (8): no personalization; and control (11): no robot or quiz	Robot groups: 3 sessions (50, 40, and 40 min), 6 weeks apart	Diabetes knowledge; quiz rounds decided to play, desire to play in a fourth session, rated pleasure; behavior during interaction; and Basic Need Satisfaction in Relationships Scale	Combined robot groups versus control[e, f]: >correct diabetes knowledge questions after Session 3;*** personal versus neutral robot: >quiz rounds in Session 3; > number electing to play a fourth session; > on some positive behaviors during some sessions (eg, smiling at the robot in all sessions); > perceived self-determination on BSNR* during Session 3
Jibb et al [33]	40 Canadian cancer patients (24 male, 60%) aged 4-9 years (mean 6.2, SD 1.5)	Cognitive-behavioral robot ("MEDiPORT", 19): supportive statements, deep breathing exercises; active distraction Robot (21): Introduction statement, dancing moves while singing	1 session at subcutaneous needle insertion appointment	BAADS; Face Pain—Revised (FPS-R); Children's Fear Scale; and Acceptability questionnaire (Likert and free text)	Active distraction robot: < avoidance during nurse movement toward child**, at needle insertion;** and < parent-rated acceptability of time to conduct needle insertion*

[a] All studies used the NAO robot, and all were randomized controlled trials with individual randomization. Numbers are at allocation.

[b] Effects on measures not reported under results were not statistically significant.

[c] Duration was not reported.

[d] Differences between effects of the 2 conditions were only reported descriptively.

[e] Results reported on 27 patients (1 neutral robot participant dropped out before session 1).

[f] Personal versus neutral robot effect for knowledge not reported.

*P<.05.

**P<.01.

*** P≤.001.

Table 2. Child interventions delivered via robot for autism spectrum disorder.

Author	Sample	Design, conditions (n)[a]	Duration	Measures	Outcomes[b]
Kim et al [45]	24 US children with autism-spectrum disorder (ASD); 21 male, 88%) aged 4-12 years (mean 9.4, SD 2.4); Autism Diagnostic Observation Schedule (ADOS): 20 met criteria for autism, and 4 for autism spectrum disorder	Random order within subject: Pleo robot interaction, adult interaction and computer game	1 session: 3 × 6-min interactions, each separated by 6 min of interview and play	Verbalization (number of utterances produced)	Robot segment: > total speech versus adult*, Computer game***; >speech to confederate versus Adult*, computer game***; > speech to Pleo than computer game***, Pleo versus Adult not significant (ns)
Huskens et al [46][c]	6 Dutch males with ASD aged 8-12 years (mean 10.50, SD 1.37); all had Social Communication Questionnaire (SCQ) >15 (range 18-32)	Random order within subject: robot: NAO; human trainer; robot and human made statements inviting a question and performed requested actions (eg, dance)	Introduction to robot (2 sessions); baseline—4 robot, 4 human 10-mi training sessions; and follow-up 2 weeks after last training	Question-asking (number of self-initiated questions) in 3-5 × 10-min sessions with human assessor at baseline and follow-up	Both conditions, baseline session 1 versus intervention and follow-up[d]: >correct questions during training, maintained at follow-up
Pop et al [47]	20 Romanian children (sex not stated) with ASD aged 4-9 years; no significant between-group differences on Children's Autism Rating Scale (CARS)	Randomization in clusters of 3: story telling; Probo robot-assisted therapy social stories (SS-RAT, 7); computer-presented social stories (SS-PC, 6); and control (7)	SS-RAT and SS-PC: 6 sessions[e]; control: 4 × 10-min observations on different days	Social expression (degree of prompt required for social response)	SS-RAT versus control at post[e]: > social expression*; (SS-PC versus control ns)
Peca, Simut [48]	27 Romanian children[f] (22 male, 82%), 18 with ASD, 9 with pervasive developmental disorder (PDD), aged 4.5-8 years (mean 6.2, SD 1.0). No significant between-group difference in mean ADOS (Robonova: 15.00; adult: 15.09)	Contingent (imitating child) and non-contingent play, with: Robonova robot (12) and adult (9)	1 session: 2 × 80-second segments[g] separated by a 5-min pause	Social intention (eye gaze, positive affect, initiations, intention testing, tests per initiation frequency); contingent (mirrored behavior)	Robot versus adult: > eye gaze*** (contingency ns)
Srinivasan et al [38,39,41,49]	36 US children with ASD (32 male, 89%) aged 5-12 years (mean 7.6, SD 2.2) ADOS-2 range 6-10 (means—Robio: 8.5, rhythm: 7.9, and control: 8.4)	Robot (12): NAO and Rovio, whole-body imitation and interpersonal synchrony games; rhythm (12): human, singing and whole-body imitation games; and control (SC, 12): tabletop activities (academic, communication, and fine motor)	32 sessions over 8 weeks (post at 10 weeks)	Joint Attention Test (JTAT); social verbalization; imitation, praxis, interpersonal synchrony; Bruinicks-Oseretsky Test of Motor Proficiency (BOT); Repetitive and maladaptive behaviors; and Affective states	Robot versus control[h]: >attention to human partner, elsewhere***[i]; <attention to objects***; > spontaneous human attention***; > self-directed vocalization**; < _human social vocalization_ ***; < _spontaneous human social vocalization_*; < sensory behaviors in late session**; > _negative affect_*; < _interested affect_*; _and_ < _fine motor control at Post_*. Robot versus Rhythm, during session[h]: <_attention to human partners and elsewhere_ ***[i]; < _spontaneous human attention_ ***; > _self-directed vocalization_**; < _spontaneous human social vocalization_*; _and_ < _positive affect in mid & late sessions_**. Group x Early, Mid, Late Session: Words in response to questions*** (only Rhythm rose). Robot, pre and post: > body co-ordination* and > imitation**. Robot, early versus late session: <positive affect* and > time in-synchrony*

Author	Sample	Design, conditions (n)[a]	Duration	Measures	Outcomes[b]
Costescu et al [44]	27 Romanian children with ASD (20 male, 74%) aged 6-12 years (mean 8.7, SD 1.8); ADOS-Generic (mean 10.32)	Robot-enhanced therapy (14, RET)[j]: My Keepon, distinguishing emotions from 15 social situations; discussion: cognitions, emotions and behavior connections; adaptive strategies for anger, self-control and control (n=15, standard care [SC])	RET: 6 × 2-hour weekly group sessions	Frequencies of correct strategies in a social situation; rational or irrational beliefs; adaptive behaviors; and emotional intensity	RET versus control, post (controlling pre): >rational beliefs** and <(negative) emotion intensity***
Yun et al [40]	15 Korean males with ASD aged 4-7 years (mean 5.8, SD 0.9). No significant between-group differences on ADOS subscales or current SCQ (lifetime SCQ higher** and IQ lower* in robot group)	Social skills training (eye contact and reading emotions). Robot (8): iRobiQ (4 weeks), CARO (4 weeks) and human trainer (7)	8 × weekly 30-40 min sessions (post at week 9)	Autism Diagnostic Observation Schedule (ADOS, by blind rater); Vineland Adaptive Behavior Scale (Korean version); Social Communication Questionnaire; Social Responsiveness Scale; and Child Behavior Checklist (Korean version, CBCL)	No differences robot, human; both (versus pre): < (better) ADOS Play*; <CBCL Internalizing at post* (Depression and Anxiety*, Withdrawal* subscales); >frequency of eye contact, Session 8*; >recognition accuracy of most difficult facial emotions by Session 4*
So et al [42]	13 Hong Kong children (10 males, 77%) with ASD aged 6-12 (mean 9.0, SD 2.4) ADOS scores not reported (nr)	NAO Robot (7); control (6): educational videos; for both, phase I: Recognize 8 gestures; phase 2: Produce 8 gestures	In each 6-week phase: 4 × 30-min sessions over 2 weeks; tests pre, post, and 2 -week follow-up	Phase 1: Recognize gestures; phase 2: Produce gestures; tested on 2 trained gestures, 2 untrained; 20% of ratings rescored by a blind rater	Phase 1 recognition, robot versus control: pre and post: >scores on trained***, generalized***, human-to-human*** gestures; post follow-up: ns; phase 1 production, robot versus control: pre and post: >scores on trained**, generalized[k], human-to-human ns; post follow-up: ns

Author	Sample	Design, conditions (n)[a]	Duration	Measures	Outcomes[b]
So et al [43]	45 Hong Kong (Cantonese-speaking) children (36 males, 80%), aged 4-6 years; 30 with ASD (3 female): intervention (mean 5.8, SD 0.8) waitlist (mean 5.7, SD 0.4); 15 age-matched controls (6 female) (mean 5.3, SD 0.7); and ASD severity nr	NAO robot demonstrates and elicits gestures while narrating stories; intervention (15); waitlist (15); age-matched, no ASD control (15)	Over 9 weeks: 4 × 30-min training sessions for 14 gestures (2 sessions per week); tests at pre, post, 2-week follow-up (2 test sessions each)	Gestural production in training, novel stories (10 seconds to respond, prompt and another 10 seconds if no response); gestural recognition; psychoeducational—third edition; Bruininks-Oseretsky Test of Motor Proficiency 2nd Edition (BOT); and Attention Network Task (ANT)	Gestural production (pre, post, follow-up), controlling for language and developmental age, BOT, ANT, gestural recognition[k]: Group × Time***, Group × Training and Novel ***, Group × Time ×Training and Novel*** control > Intervention*, Wait List* at Pretrained: intervention > waitlist (post***, follow-up***); > control (Post***, Follow-up*); -Novel: Intervention > Wait List (Post***, Follow-up**); = Age-matched controls; Intervention versus Wait List (Pre to Follow-up) with covariates as above: Verbal imitation: Group × Time* (only Intervention group increasing*)

[a] Randomized controlled trial with the individual participant as the unit of randomisation unless labeled otherwise.

[b] Effects on measures not reported under results were not statistically significant. Some results that did not involve the robot condition are omitted. Results where the robot did significantly worse than the comparison condition are italicized.

[c] Differences between effects of the 2 conditions were only reported descriptively.

[d] Analyses of changes within conditions are reported separately, as are effects for each individual.

[e] Total period of training and timing of post not reported.

[f] An additional 6 children were excluded because they refused to undertake the tasks.

[g] The paper refers to the session segments as sessions.

[h] Results from these studies were incompletely reported, and some reporting is ambiguous. Effects are across sessions unless otherwise stated.

[i] The attention target analysis appears inappropriate (only the robot group could have attention to the robot, affecting analysis of condition effects). “Elsewhere” is attention other than to the human partner, robot, or objects.

[j] Analyses were on 12 RET (2 withdrew); 15 control participants.

[k] Recoding for gestural appropriateness rather than strict accuracy was interpreted as supporting these results, but only gave Group effects (using pre and follow-up only).

*$P<.05$.

**$P<.01$.

***$P\leq.001$.

Sample sizes ranged from 6 to 45 (median=24). However, only 3 recent papers reported some blind coding of observations [40,42,43]. The studies had 1 to 32 sessions (median=4), and study durations ranged from 0 to 14 weeks (median=9 weeks, 2 studies were unclear). In addition, 4 studies [38,41-43,46] reported a follow-up, all were of only 2 weeks. All but the small study by Huskens et al [46] presented the results against a comparison condition: they reported the results within each condition and within each participant.

Participants were aged from 4 to 12 years (median reported mean age=8.7 years). Samples were drawn from the United States (2 studies), Romania (3 studies), Hong Kong (2 studies), and the Netherlands and Korea (1 study each). The nature and roles of the robot were also diverse. Some studies used the robot as an assistive tool to therapist interventions [47], whereas in others, it was the primary method of therapy delivery [40]. All but 3 studies (78%) had a researcher who was operating the robot using Wizard-of-Oz control. The remaining studies used a set program where the researcher could pause the program if needed [42,43] or where limited branching was produced by eye contact or by the researcher pressing a button to record the child’s response [40]. The number of treatment sessions ranged from 1 to 32 (median=4) over a period of 1 to 8 weeks (median=2 weeks, 2 unknown). Despite the small sample sizes in these trials, positive effects were found on several measures; although inspection of Table 2 shows that differential results on many measures were not statistically significant. In relation to changing beliefs, a robot to deliver therapy increased the presence of more rational beliefs [44]. For improving emotional affect, the identified trials resulted in decreased negative emotion intensity [44] and lower scores on depression, anxiety, and withdrawal subscales after treatment [40]. Social behavior improvements were present with increased eye contact [40], gaze frequency in the direction of the interaction partner [48], increased levels of social expression [47], higher total number of produced verbal utterances [45], recognition accuracy of facial emotions [40], and number of correct questions [46], as well as improved gestural recognition [42] and production [43]. Robots could achieve greater effects than a standard care control [44,47], educational videos [42], and a computer game [45].

Effects of the robot versus a human trainer were typically the same [40] or superior on at least 1 measure [48]. The 1 exception was a study by Srinavasan et al [38,39], where average effects of the robot were never superior to the human, and the human condition did better on several social indices. However, that study confounded the actor (human versus robot) and intervention content (eg, the human condition used singing and the robot one did not). Furthermore, calculation of the focus of attention was affected by the fact that a focus on the robot was not counted as attention to the social partner, and the superior result on fine motor skills in the control group may be ascribed to a difference in fine versus gross motor tasks in the 2 conditions rather than to use a robot per se. The observation of less interested affect and more negative affect in the robot condition than in controls deserves further attention, although it appears inconsistent with positive effects on negative moods that were seen in the studies by Costescu et al [44] and Yun et al [40]. Overall, the results by Srinavasan et al appear at odds with those from other trials and are subject to methodological limitations.

In summary, the strengths of these trials included their substantiation of interobserver reliability and the fact that a third had some blind assessment, and 4 trials had a follow-up assessment (albeit only 2 weeks later). Social robots for young people with ASD appear to have positive outcomes, although studies with larger samples and longer follow-ups are needed to build confidence in the strength and sustained maintenance of these effects.

Interventions for Older Adults

Overall, 12 trials of robot interventions for older adults were identified (Table 3). Most aimed to improve cognitive and or psychological functioning or neural integrity, although 1 focused on self-management of chronic obstructive pulmonary disease (COPD) [50]. Where mean ages of participants were reported, they ranged from 67.4 to 85.3 years (median=84 years). In addition, 4 studies were from New Zealand, 2 from Australia, 2 from the United States, and 1 study from Korea, Norway, Spain, and Japan each. Moreover, 8 trials were conducted in residential facilities, but 4 used an intervention in the home or day care center. Furthermore, 8 trials randomized individuals (2 of these to a random order of conditions); 4 had cluster randomization (1 to a random order).

Table 3. Adult interventions delivered via robot.

Author	Sample	Design, Conditions (n)[a]	Duration	Measures	Outcomes[b]
Banks et al [51]	40 US residents of long-term care facilities scoring ≥24 on the Mini-Mental State Examination (MMSE) and ≥30 on University of California, Los Angeles (UCLA) Loneliness Scale (age and gender nr)	Interactions with robot (15)[c]-AIBO robot dog or living dog (15); control (13)—no intervention	Robot, dog: 8 weekly 30-min sessions	UCLA Loneliness scale; Lexington Attachment to Pets Scale	Robot or dog versus control, pre and post: >fall in loneliness* (robot=dog)
Tanaka et al [52][d]	34 female Japan residents aged 66-84 years	Kabochan Nodding Communication ROBOT (18): Communicate by talking and nodding and control (16): same robot but no talking or nodding	Robot at home for 8 weeks	MMSE; Cognistat test; Blood and saliva samples; Accelerated plethysmography; Questionnaire: Appetite (visual analogue scale) sleep; depressive symptoms (Geriatric Depression Scale [GDS-15]); Activities of daily living—Tokyo Metropolitan Institute of Gerontology Index of Competence	Communication Robot: > MMSE score after 8 weeks**; > Verbal memory after 8 weeks*; > Everyday/concrete judgements after 8 weeks*; > Attenuation of fatigue compared with control*; > Enhancement of motivation compared with control**; and >healing compared with control*
Robinson et al [53]	40 New Zealand retirement home residents (13 men, 33%) aged 55-100 years	Robot (20)[e]: PARO—interactions with robot and control (20): alternate activities)	Robot: 2 group sessions per week for 3 months	UCLA Loneliness scale; GDS-15; and Quality of Life for Alzheimer's Disease (QoL-AD)	Robot versus control, pre and post: >fall in loneliness*
Moyle et al [54]	18 Australian residential aged care residents (sex not stated) aged ≥ 65 (mean 85.3, SD 8.4)	Within-participant crossover design (random order); Robot first (1 group of 9): PARO -discovery, emotional response, discussion about PARO, touching PARO and control first (9): Being read to, looking at pictures, discussion of readings	Each phase: 3 × 45-min 9-member sessions per week over 5 weeks; 3-week washout between phases	Quality of Life for Alzheimer's Disease Scale; Rating Anxiety in Dementia Scale (RAID—self-reported and Proxy); Apathy Evaluation Scale; Geriatric Depression Scale; Revised Algase Wandering Scale Nursing Home version; Observed Emotion Rating Scale (OERS); (Assessors independent—unclear if blind)	Robot versus control after intervention (reporting range of Cohen *d*)[f]: >Quality of Life (0.6 to 1.3); < anxiety on RAID Proxy version (−0.4 to −0.3) but greater on RAID (0.4 to 0.4), OERS (0.5 to 0.7)[f]; > OERS sadness (0.4 to 0.6), pleasure (0.7 to 0.7)[f]
Broadbent et al [55]	29 New Zealand retirement village residents (14 male, 48%) aged 72-94 years (mean 85.2, SD 5.1)	Within-participant crossover design (random order)[g]; iRobiQ or Cafero robot at home versus control—measured blood pressure and pulse oximetry, had music and quotes; iRobiQ: also medication reminders, alert to nurse if not taken or said unwell; and Cafero: cognitive exercises, village map, and calendar reminder	2 × 6-week periods with 18-day washout	Geriatric Depression Scale; Health-related Quality of Life; and Medication Adherence Report Scale (Single-blind assessment)	iRobiQ or Cafero versus control pre and post: not significant (ns)

Author	Sample	Design, Conditions (n)[a]	Duration	Measures	Outcomes[b]
Kim et al [56]	85 Korean community residents (25 male, 29%) aged > 60 (mean 67.4) with MMSE Korean version > 26 (mean 29)	All: 10 hour dementia prevention education on before baseline; cognitive training: robot (24)[h]: Silbot and Mero -17 training programs with individual rewards immediately after smart pad answers; winner of day, month; traditional cognitive training (24)[h]: question and answer display; nonrandom control (37), no training)	Education: 2 hours per day over a week; cognitive training: 60 × 90-min 8 - member sessions over 12 weeks	MRI cortical thickness, intracerebral volume, structural connectivity; Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog); Cambridge Neuropsychological Test Automated Battery; Delayed Matching to Sample; Pattern Recognition Memory (PRM); Paired Associates Learning; Spatial Working Memory; Stockings of Cambridge (SOC); Reaction Time; Rapid Visual Information Processing (Blind scoring of all assessments)	Cognitive training versus control, pre and post: <reduction in cortical thickness*, nodal strength*, global efficiency*, clustering coefficient* > executive function (SOC)*** (robot=traditional); robot versus traditional: <cortical thinning in right and left anterior cingulate, areas of right inferior temporal cortices***; > nodal strength, left rectus gyrus***; and < improved on ADAS-Cog* and PRM*
Valenti Soler et al [57]	Spanish nursing home patients with dementia; phase 1: 101 adults (12 male, 11.8%) aged 58-100 (mean 84.7); phase 2: 110 adults (11 male, 10.0%) aged 59-101 (mean 84.7)	Cluster randomization by living unit; all: training—for example, identifying numbers, words, colors; use of everyday objects; sensory stimulation; phase 1: assisted by PARO (33), NAO (30), and control (38); 9-month washout; phase 2[i]: Assisted by PARO (42), dog (36), control (32); Day Care (Nonrandom); phase 1: assisted by NAO (20); 9-month washout; and phase 2[i]: assisted by PARO (17)	30-40-min group or individual sessions × 2 days per week × 3 months	Global Deterioration Scale; Severe Mini Mental State Examination; Mini Mental State Examination (MMSE); Neuropsychiatric Inventory (NPI); Quality of Life in Late-stage Dementia (QUALID); Apathy Scale for Institutionalized Patients with Dementia Nursing (Home version; APADEM); Apathy Inventory (single-blind assessments)	Nursing home phase 1; NAO versus control, pre and post: > reduction in APADEM total*, Cognitive inertia subscale*: > reduction in NPI apathy/indifference*; _worse delusions*; < (worse) mental state (MMSE)_*; PARO versus control, pre-post: > reduction in APADEM total*_; > NPI Irritability/lability_* phase 2 PARO versus control: > quality of life (QUALID)*_; > NPI hallucinations*, irritability/lability_** Day Care phase 1 (NAO): < NPI total**, Irritability/lability* phase 2 (PARO): ns
Jøranson et al [58,59]	60 Norwegian nursing home patients (10 male, 33%), aged 62-95 (mean 84) with dementia or MMSE < 25	Cluster randomization of 10 living units; robot (30)[j]: PARO; control (30)[j]: SC	Robot: 2 × 30-min sessions (≤6 members) per week for 12 weeks; tested at pre, post, 3-month follow-up	Brief Agitation Rating Scale (BARS, interrater reliability reported); Cornell Scale for Symptoms of Depression in Dementia (Norwegian, CSDD); Medication; QUALID	Robot versus control, Pre and Follow-up[j]: < agitation (BARS)*; < depression (CSDD)*; > quality of life (QUALID, severe dementia patients only)*; robot versus control, Pre and Post: <medication, severe dementia patients only*
Liang et al [60]	30 New Zealand dyads: patients with dementia (11 male, 36%), aged 67-98; caregivers (4 male, 13%), aged 30-86)	Robot (15)[k]: PARO, at day care and at home and control (15)[k]: SC	Robot, over 6 weeks: 2-3 × 30 - min sessions per week (day care, groups of 3 -6) and ad lib at home; tested at pre, 6, 12 weeks	Behavioral, affective, and social responses during sessions; Blood pressure; salivary cortisol; Addenbrookes Cognitive Examination (NZ version); CSDD; Neuropsychiatric Inventory Brief Questionnaire Form; Cohen-Mansfield Agitation Inventory (Short Form); and Hair cortisol	Robot versus control, Pre, 6 and 12 weeks: > drop in depressive symptoms (CSDD) but increase between 6 to 12 weeks (interaction effect*). Robot versus control, during sessions: >Happy, smiling facial expressions (Agitation, social interactions ns)

Author	Sample	Design, Conditions (n)[a]	Duration	Measures	Outcomes[b]
Petersen et al [61]	61 US patients in assisted living memory care units with mild-moderate dementia (14 male, 23%) aged ≥ 60 (mean 83.4)	Cluster randomization by coin toss: Robot (35): PARO and control (26): SC activities	Both: 3 × 20-min sessions per week (6 members) for 12 weeks	Global Deterioration Scale (interrater reliability reported) RAID; CSDD; Galvanic skin response (GSR); pulse rate; pulse oximetry; and medication doses	Robot versus control, pre and post: > rise?[l] in anxiety (RAID)**, depression (CSDD)***; > rise?[l] in GSR***, pulse oximetry***; > fall in pulse rate***; and >fall in doses of pain medication*** and behavior medication***
Broadbent et al [50]	60 New Zealand patients (aged between 16-90 (mean 69.8, 62% female) with chronic obstructive pulmonary disease (COPD), recruited at inpatient discharge	Robot (30)[m], iRobi at home: weekly clinical assessments; reminders to take medication, inhalers, do rehab exercises; education in videos, pop-up messages; "I am feeling unwell" button (initiating clinical assessment, message to staff); display trends in status, adherence. Linked to SmartInhaler alert to staff if missed medications, exercise 3 times. Phone calls to follow-up alerts, remind to use robot and control (30)[m], SC	4-month robot use	Quality of life—Clinical COPD Questionnaire; medication adherence—Medication Adherence Report Scale—and Frequency of rehabilitation exercise	Robot versus control, pre and post (controlling for comorbidities, past hospitalizations): hospitalizations (primary outcome) ns; > self-reported medication adherence* (electronic inhaler only before covariates); > self-reported rehab exercises***; robot versus control: <direct cost (saving NZ$1152; *d*=.27), total hospitalization cost (saving NZ$1579; *d*=.27)

Author	Sample	Design, Conditions (n)[a]	Duration	Measures	Outcomes[b]
Moyle et al [62] and Jones et al [63]	415[n] Australian residential patients with dementia (100 male, 24.1%) aged >60 years (mean 84-86 in each condition)	Cluster randomization (N facilities, participants); PARO (9, 138)[n]; Plush toy—PARO with robot features disabled (10, 140)[n]; SC (9, 137)[n]	PARO and Plush Toy: 3 × 15-min individual, non-facilitated sessions for 10 weeks (ie, 30 total) and assessed at pre and weeks 1, 10, 15 (post)	Positive behavioral engagement, mood states and agitation (video observation); Cohen-Mansfield Agitation Inventory-Short Form; Rowland Universal Dementia Assessment Scale; Multicultural Cognitive Assessment Scale; Using SenseWear Professional 8.0 activity armband: Day and nighttime motor activity (steps, hours of physical activity) and hours lying down, asleep, and awake	PARO versus Plush Toy, pre and post[o]: > verbal* (.011), visual engagement***; < steps in day*, nightime*[p]; < hours physical activity*[p]. PARO versus SC, pre and post[o]: > neutral* and pleasure** affect; < agitation**; < steps in day*[p]. PARO and Plush Toy versus SC, pre and post[o]: > neutral affect**

[a] Randomized controlled trial with the individual participant as the unit of randomization unless labeled otherwise. Numbers are at allocation.

[b] Effects on measures not reported under results were not statistically significant. Some results that did not involve the robot condition are omitted. Results where the robot did significantly worse than the comparison condition are italicized.

[c] Analyzed 13 robot, 13 dog participants.

[d] Random assignment matched for age and MMSE score.

[e] Analyzed 17 robot (3 died), 17 control participants (2 died, 1 moved away).

[f] Text says the amount of missing data was large, and no substitution for missing data was made. However, tables give an n of 18. Analyzed by standardized mean difference between scores after each intervention. Results with Cohen *d* ≥0.3 are displayed (range in brackets).

[g] Number in each order not reported. Individual randomization, but mentions 2 participants who were married and living together.

[h] Excluded 2 robot, 1 traditional participant from MRI analyses (similarity index <0.5).

[i] Some overlap of phase 2 participants from phase 1. Loss to analyses: nursing home phase 2 dog (1); day care phase 1 (2), phase 2 (2).

[j] Lost 2 robot, 4 control participants who died; 1 robot participant withdrew. However, analyses used intention to treat (by imputation, mixed models).

[k] Analyses on 13 PARO, 11 SC participants.

[l] All of the results are described in the text as greater improvements in the robot condition, but mean changes presented in Table 2 on the RAID, CSDD, and GSR show larger positive post minus pre changes in the robot condition. That would indicate greater deterioration. A question mark is used to highlight the issue.

[m] Hospitalizations were reported as intention to treat (omitting 1 who died) and per protocol. Most other results (referred to as *intention to treat*): were on 25 robot participants (3 withdrew, 2 died), 26 controls (1 withdrew, 1 did not complete follow-up assessments, 2 died). Electronic inhaler results were on 18 robot, 25 control participants.

[n] All allocated participants were in analyses. Losses to assessment postallocation included PARO: 7 deceased, 1 relocated; Plush toy: 14 deceased, 1 palliative care; SC: 5 deceased, 1 palliative care, and 1 relocated.

[o] Secondary analyses examined effects at weeks 1 and 5.

[p] Interpreted as a positive outcome because of association of physical activity with agitation.

**P*<.05.

***P*<.01.

****P*≤.001.

The use of a robot had an impact on emotions, such as achieving an increase in neutral and pleasure affect [62,64], and happy, smiling facial expressions [60], with a decrease in depressive symptoms [58,59] and loneliness scores [51,53]. Reported increases were found in quality-of-life measures [54,57], but for another sample, this was only present for severe dementia patients [58,59]. Cognitive functioning could also be improved in areas such as reducing cortical thickness and improving executive functioning [56]. The use of a robot helped to reduce agitation [58,62,64] and increase other behaviors, such as verbal memory [52]. Robots also assisted with a fall in pulse rate, doses of pain medication and behavior medication [61], increase in self-reported medication adherence, and rehabilitation exercises, with substantial cost saving [50].

In addition, 6 studies reported using independent or blind randomization [50,54,56,58,60,62,63]—in the case of Broadbent et al [50], controlling for ethnicity and gender. Robinson et al [53] reported using a random list generator but did not state other details. In some other trials, the randomization was basic (eg, Petersen et al [61] used a coin toss and Valenti Soler et al [57], a die) or potentially problematic (eg, Broadbent et al [55] randomly allocated at least 1 couple in a study with participant randomization). Where block randomization was used, only Jøranson et al [58] appeared to have analyzed for cluster effects. Kim et al [56] had an additional nonrandom control group, and Valenti Soler et al [57] also reported on a nonrandom study.

The trial durations ranged from 8 to 24 weeks (median=13 weeks). A total of 3 trials had a follow-up assessment: 2 for 5 to 6 weeks [60,62,63] and 1 for 3 months [58]. In addition, 4 trials had a blind assessment of observations [55-57,62,63], and interrater reliability on at least 1 key measure was reported in 3 trials [59,61-63]. Sample sizes ranged from 18 to 415 (median=50). Moreover, 8 trials reported intention-to-treat results on at least some measures data, and retention was generally high. Some data appeared to conflict with the described results in the study by Petersen et al [61].

Identified trials often involved the PARO robot, which resembles a baby harp seal, and is designed to mimic animal behaviors, but avoids attendant risks of injury or infection. PARO is furry; responds to touch, sound, light, posture, and temperature; and has a diurnal rhythm and some interaction capability [31,65]. As a result, it can be reliably used for responsive interactions without Wizard-of-Oz control. However, similar to animals, it cannot verbally communicate, thus limiting the range of social interactions it can undertake. These trials aimed to elicit behavioral, affective, and social responses to improve mood and cognitive functioning.

Table 3 demonstrates that the robotic interventions typically resulted in better cognitive or neural functioning, reduced distress, or better quality of life, although (as may be expected in dementia) positive cognitive or neural outcomes sometimes involved less decline rather than greater improvement [56]. Overall, 2 sets of results were particularly notable: a lesser reduction in cortical thickness and in global efficiency in the trial by Kim et al [56] and (despite the lack of significant differences in hospitalizations) reduced direct costs from treatment of COPD in the study by Broadbent et al [50]. However, consistent with the trials on ASD, many measures did not show differential changes from the robot intervention.

Even more importantly, as shown in the italicized results in Table 3, some trials found that the robot condition had inferior results on some measures. Examples include inconsistent results on emotions in the study by Moyle et al [54] and on cognitive tests in the study by Kim et al [56], and some negative effects on symptoms including irritability/lability in the study by Valenti Soler et al [57]. Importantly, a set of predictive analyses that was undertaken by Jones et al [63] within the PARO condition of the recent study by Moyle et al [62] and Jones et al [63] showed that more positive and visual engagement with PARO was seen in participants with low levels of agitation at baseline. Mild cognitive impairment predicted greater visual engagement with the robot and more pleasure at week 10. These results suggest that positive effects from PARO may predominantly occur in less severely affected participants. A cost-effectiveness study on the recent study by Mervin et al [66] found that the PARO robot gave a slightly lower incremental cost-effectiveness over usual care in reducing agitation than did the Plush Toy (the PARO with disabled robotic features). However, neither cost was substantial.

Ineligible Trials

Inspection of excluded trials also offers some important insights into the state of current research on robotics in health care. Medical trials retain dominance with 145 excluded papers (83 papers on surgical, rehabilitative, and exoskeletal applications and 62 on rehabilitation and gait training) and that research shows greater maturity than in other health care domains. Other social robot studies placed greater focus on the acceptability of social robots or evaluation of different robot characteristics, demonstrating that potential applications and intervention elements are still being identified [67]. Excluded papers also included trial protocols [68,69], qualitative studies [70], and reviews [71].

Discussion

Principal Findings

In comparison with trials on surgical or other medical applications of robots, trials on improving health, well-being, or psychological interventions using social robots are very few; are limited to the contexts of child health, ASD, and older adults; and, as a group, are relatively unsophisticated. Conversely, the contexts where these robots have been trialed present significant challenges for both treatment and research.

In light of those challenges, it is encouraging that all of the reviewed papers provided some evidence of positive effects from an intervention using a social robot, even though several found that some measures showed no differential effects or favored alternate treatments. Furthermore, some null results may be regarded as a positive finding. For example, the fact that Huskens et al [46] and Yun et al [40] found no difference between results from a human and a robot trainer may presage greater cost-effectiveness, if a human operator was not required in similar future applications. Some effects that appeared negative could be attributed to methodological issues with the trial, such as the confounding of content and delivery in the study by Srinivasan et al [39]. However, some other results were more disquieting, such as increased negative affect and less interested affect than controls in the study by Srinivasan et al [39], increases in sadness and (on 1 measure) anxiety in the study by Moyle et al [54], and some worsening of symptoms in the study by Valenti Soler et al [57]. Whether social robots sometimes trigger distress or other negative emotions in dementia needs further examination, as do the ways any negative effects may be avoided or reduced and whether interventions should primarily target people with less agitation or cognitive impairment. Although some of these results were inconsistent with those from other studies and others may have been because of uncontrolled factors, they require further attention to see if they are replicated and identify factors that are triggering them. For example, some participants with ASD or dementia may find specific social robots or robotic behaviors anxiety provoking or greater effort may be needed to acclimatize them to the novelty of interacting with a robot.

Generalizability of benefits from the robotic intervention outside the context-specific tasks was not tested in most trials. For example, child social skills’ training was not explored beyond the context of the session to investigate its translatable impact on other interactions, such as with adults outside of the research team or other children. Only 6 trials (22%) had any follow-up at all, and among these 6, none had a follow-up assessment beyond 12 weeks (range=2-12 weeks, median=5). Most trials were small (median N=34), with only 2 trials having more than 100 participants. A single-blind assessment of observations was reported in just 8 trials (30%). Overall, more recent trials tended to have superior methodological quality, especially in the older adults’ group of studies. However, all 3 areas had significant thematic limitations.

Limitations

The limitations of the review include its inability to assess average effect sizes because of the wide range of measures and reporting methods in the identified papers. The restriction of trials to the ones reported in English may have missed some trials, and some other databases or search terms may have identified further trials, although the use of forward and backward searches should have reduced the risk of missed trials. Given a publication bias toward significant results, there may be other trials with less positive outcomes than the ones reported. The review is also limited by its focus on published papers and contacting authors may have clarified some methodological features that were unreported or ambiguous. Large-scale RCTs are often subject to extensive testing and recruitment timeframes, and therefore, eligible trials may still be in operation at the time of publication. In addition, trials are restricted by the limited number of social robots available to them, constraining researchers to shape their interventions to fit onto the current capabilities of the robot. This could have severely limited the prospect of large sample sizes because of the low numbers of social robots available. Moreover, researchers may not have been able to create and deploy sophisticated health, well-being, or psychological interventions to the same standard as other digital programs onto the robot because of the current software or hardware constraints.

Opportunities for Further Applications

There were good reasons why the initial studies on social robots have focused on the above participant groups. The NAO robot that has been used in child health is a similar size to a young child and has movements and features that children have generally regarded as acceptable [72], in addition to receiving high acceptability ratings for its application to treatments [73]. Children with ASD tend to have difficulties with human interaction and appropriate expression of emotion, and the robots that have been trialled in that context offer simplified versions of both social interactions and emotional expression that make them particularly suitable for this group [74,75]. The PARO robots used in older age groups are similar in appearance, texture, and behavior to a small, friendly animal and imitate animal-assisted therapy, which is already an established approach for dementia [76], without presenting physical health risks. Thus, social robots in these contexts build on well-established theory and research applications or provide a digital spin to current treatment practice. However, it remains surprising that health-related interventions by social robots do not appear to have been trialled in other problem domains or in adolescents or young adults. Such trials would be timely and important.

Advances in technology equip robots with the capability to conduct health care–related tasks in fields beyond the scope of children, ASD, and elderly care applications. For example, robot capability to respond to information and deliver a structured conversational exchange related to a health care service. These capabilities could translate into health tasks to support patients during their visit, such as verbal discussion of an appropriate homework task, provision of health education during a consultation time, or disclosure of sensitive medical information. Traditional therapeutic elements can also be performed by a robot using these techniques, such as positive verbal affirmations, and providing customized coaching based on interpreting and responding to multiple signal inputs, such automated physiological recordings, self-reported data, or verbal reports on the participant’s progress [77]. Other applications may include encouragement and coaching support for Web-based or digital interventions to increase adherence and impact, as a partial or full substitute for human coaching, which has increased adherence and produced stronger treatment effects in some trials [15]. Robots have already demonstrated an ability to incorporate physical monitoring and service alerts [50], which could in the future reduce response times and hospitalization rates and improve outcomes. Integration of health service robots with personal robots and other digital devices could enhance the transfer of within-session gains to the natural environment, through additional verbal and enactive rehearsal and cueing of behaviors at times in which they are most needed.

Further technological advancements could also progress social robot capabilities beyond limited and constrained tasks. For instance, the use of natural language processing can capture and interpret key elements of human speech [78], and computer vision can recognize faces and activities and detect changes in physiological arousal and emotions [79,80]. Advances in robotic technology are also likely to encompass greater ease of use by nontechnical experts and improved reliability, robustness, and autonomy. Ultimately, these improvements will obviate the need for teleoperation or monitoring by a health professional or trained staff member, which currently occurs in Wizard-of-Oz–style studies, where the role of the robot currently exceeds its technical ability for autonomous operation. Enhancements will also improve the robots’ capacity to deal with uncertain or unpredictable factors in treatment sessions, such as modifying the content of responses and the focus of the session, based on new information from the participant or making a timely response to verbal or nonverbal cues that might indicate disinterest, uneasiness, or annoyance. Such advancements would more accurately emulate the nature of evidence-based face-to-face treatments.

Advances of this kind may be seen by some therapists as a threat. However, we argue that they are more likely to provide exciting opportunities to supplement and augment the impact of standard treatment, reducing time on routine and low-intensity tasks, including information provision, standardized assessment, and treatments that are effective when applied through a highly specified protocol, and increasing the retention and generalization of insights or skills that have been gained within sessions. Practitioners may then focus on more personally satisfying and challenging work, including their relationship with the client; enhancing and maintaining motivation; collaborative goal setting and planning; and addressing severe, complex, or co-occurring problems. However, this proposed outcome remains dependent on continued efforts to develop and test digital health interventions that can be reliably and safely delivered by social robots as their hardware and software progressively increases in sophistication. Although these developments will raise ethical and practical issues and will require careful monitoring of any negative outcomes or potential risks, their benefits to clients and to the cost-effectiveness of health services may be substantial.

Conclusions

The evidence for health, well-being, and psychosocial interventions that are delivered by social robots remains at an early stage, with few trials being identified. The methodological quality in many trials was reduced by their small sample sizes, an absence of independent randomization, blind assessment or follow-up, and their somewhat rudimentary statistical analyses. However, the higher quality of some recent trials gives cause for optimism that some current and emerging trials will meet more rigorous methodological standards and will steadily move from a focus on efficacy to examining effectiveness in routine care. Progressive reductions in the cost of social robots and improvements in their accessibility for purchase will also make it easier to conduct trials, as will the further development of standard program routines that enable their intuitive and flexible use.

Overall, the initial evidence from clinical trials is promising but is not universally positive. Importantly, some studies show some increases in negative affect or psychiatric symptoms, suggesting that their use with some patient groups may not be indicated or may need more preparation before an intervention is attempted. As yet, there is no evidence that treatment gains from these interventions can be sustained over a follow-up period of more than 3 months or (apart from some trials in aged care facilities) that these interventions can be taken to scale. Nor, as yet, is there evidence on the effects of psychosocial interventions by robots on health and well-being outside the 3 contexts reviewed in this paper.

Currently, trials have used robots with limited capabilities (eg, the PARO fur seal), programmed a robot such as NAO with limited or no branching, or have used Wizard-of-Oz control. None of these approaches fully capitalizes on the potential that robots may have for improving the cost-effectiveness and reach of clinical services, once further development of their hardware and software provides more advanced and reliable social capability.

Each of these limitations is likely to be addressed over the coming years. The true potential for social robots to improve the impact, reach, and cost-effectiveness of health care will then be much clearer than at present.

Acknowledgments

NLR was supported by the Australian Postgraduate Award on behalf of the Department of Education and Training and by a Research Grant from the State of Queensland acting through the Department of Science, Information Technology and Innovation. The funding bodies had no role in the study design, extraction, analysis or interpretation of the review, writing of the manuscript, or the decision to submit the paper for publication.

Authors' Contributions

NLR created the systematic review protocol; conducted the extraction, preparation, and analysis of the trials; and wrote the initial manuscript. TVC assisted with the systematic review extraction, preparation, and analysis and reviewed the manuscript drafts. DJK advised on the protocol and trial extraction, conducted an independent analysis of identified trials, and contributed to manuscript revision. All authors contributed to the draft and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Kasina H, Raju Bahubalendruni MV, Botcha R. Robots in medicine: past, present and future. IJMMME 2017;7(4):44-64. [doi: 10.4018/IJMMME.2017100104]
2. Lanfranco AR, Castellanos AE, Desai JP, Meyers WC. Robotic surgery: a current perspective. Ann Surg 2004 Jan;239(1):14-21. [doi: 10.1097/01.sla.0000103020.19595.7d] [Medline: 14685095]
3. Hwang YB, Cheung YJ, Lee MK, Kim HK, Cho HH, Kim JH, et al. Comparing with open surgery, robot-assisted laparoscopic adenomyomectomy is a feasible option of uterus-sparing surgery. J Minim Invasive Gynecol 2016 Dec;23(7):S246-S247. [doi: 10.1016/j.jmig.2016.08.798]
4. O'Malley DM, Smith B, Fowler JM. The role of robotic surgery in endometrial cancer. J Surg Oncol 2015 Dec;112(7):761-768. [doi: 10.1002/jso.23988] [Medline: 26385835]
5. Yamaguchi T, Kinugasa Y, Shiomi A, Tomioka H, Kagawa H, Yamakawa Y. Robotic-assisted vs conventional laparoscopic surgery for rectal cancer: short-term outcomes at a single center. Surg Today 2016 Aug;46(8):957-962. [doi: 10.1007/s00595-015-1266-4] [Medline: 26482845]
6. Tam MS, Kaoutzanis C, Mullard AJ, Regenbogen SE, Franz MG, Hendren S, et al. A population-based study comparing laparoscopic and robotic outcomes in colorectal surgery. Surg Endosc 2016 Feb;30(2):455-463. [doi: 10.1007/s00464-015-4218-6] [Medline: 25894448]
7. Bogue R. Robotic exoskeletons: a review of recent progress. Ind Rob 2015 Jan 19;42(1):5-10. [doi: 10.1108/IR-08-2014-0379]

8. Guo Z, Yu H, Yin YH. Developing a mobile lower limb robotic exoskeleton for gait rehabilitation. J Med Devices 2014 Aug 19;8(4):1-6. [doi: 10.1115/1.4026900]
9. Schwartz I, Meiner Z. Robotic-assisted gait training in neurological patients: who may benefit? Ann Biomed Eng 2015 May;43(5):1260-1269. [doi: 10.1007/s10439-015-1283-x] [Medline: 25724733]
10. Lambert MJ, Barley DE. Research summary on the therapeutic relationship and psychotherapy outcome. Psychotherapy (Chic) 2001;38(4):357-361. [doi: 10.1037/0033-3204.38.4.357]
11. 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]
12. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
13. White A, Kavanagh D, Stallman H, Klein B, Kay-Lambkin F, Proudfoot J, et al. Online alcohol interventions: a systematic review. J Med Internet Res 2010;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]
14. Sundström C, Blankers M, Khadjesari Z. Computer-based interventions for problematic alcohol use: a review of systematic reviews. Int J Behav Med 2017 Oct;24(5):646-658 [FREE Full text] [doi: 10.1007/s12529-016-9601-8] [Medline: 27757844]
15. Sundström C, Gajecki M, Johansson M, Blankers M, Sinadinovic K, Stenlund-Gens E, et al. Guided and unguided internet-based treatment for problematic alcohol use-a randomized controlled pilot trial. PLoS One 2016 Jul;11(7):e0157817 [FREE Full text] [doi: 10.1371/journal.pone.0157817] [Medline: 27383389]
16. 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]
17. Axelsson E, Andersson E, Ljótsson B, Hedman-Lagerlöf E. Cost-effectiveness and long-term follow-up of three forms of minimal-contact cognitive behaviour therapy for severe health anxiety: Results from a randomised controlled trial. Behav Res Ther 2018 Dec;107:95-105. [doi: 10.1016/j.brat.2018.06.002] [Medline: 29936239]
18. Freyer-Adam J, Baumann S, Haberecht K, Tobschall S, Bischof G, John U, et al. In-person alcohol counseling versus computer-generated feedback: results from a randomized controlled trial. Health Psychol 2018 Dec;37(1):70-80. [doi: 10.1037/hea0000556] [Medline: 28967769]
19. Clarke J, Proudfoot J, Whitton A, Birch M, Boyd M, Parker G, et al. Therapeutic alliance with a fully automated mobile phone and web-based intervention: secondary analysis of a randomized controlled trial. JMIR Ment Health 2016 Feb 25;3(1):e10 [FREE Full text] [doi: 10.2196/mental.4656] [Medline: 26917096]
20. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. BMJ Open 2018 Dec 12;8(2):e018815 [FREE Full text] [doi: 10.1136/bmjopen-2017-018815] [Medline: 29440212]
21. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. J Am Med Dir Assoc 2012 Feb;13(2):114-120.e1. [doi: 10.1016/j.jamda.2010.10.002] [Medline: 21450215]
22. Begum M, Serna RW, Yanco HA. Are robots ready to deliver autism interventions? A comprehensive review. Int J of Soc Robotics 2016 Mar 16;8(2):157-181. [doi: 10.1007/s12369-016-0346-y]
23. Pennisi P, Tonacci A, Tartarisco G, Billeci L, Ruta L, Gangemi S, et al. Autism and social robotics: a systematic review. Autism Res 2016 Feb;9(2):165-183. [doi: 10.1002/aur.1527] [Medline: 26483270]
24. Stables RH. Observational research in the evidence based environment: eclipsed by the randomised controlled trial? Heart 2002 Feb;87(2):101-102 [FREE Full text] [Medline: 11796537]
25. Black N. Why we need observational studies to evaluate the effectiveness of health care. Br Med J 1996 May 11;312(7040):1215-1218 [FREE Full text] [Medline: 8634569]
26. Akobeng AK. Understanding randomised controlled trials. Arch Dis Child 2005 Aug;90(8):840-844 [FREE Full text] [doi: 10.1136/adc.2004.058222] [Medline: 16040885]
27. Steffens D, Thanigasalam R, Leslie S, Maneck B, Young JM, Solomon M. Robotic surgery in uro-oncology: a systematic review and meta-analysis of randomized controlled trials. Urology 2017 Aug;106:9-17. [doi: 10.1016/j.urology.2017.03.015] [Medline: 28336286]
28. Shen Z, Sun Z. Systematic review and meta-analysis of randomised trials of perioperative outcomes comparing robot-assisted versus open radical cystectomy. BMC Urol 2016 Sep 23;16(1):59 [FREE Full text] [doi: 10.1186/s12894-016-0177-z] [Medline: 27664079]
29. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
30. SoftBank Robotics. 2018 Jan 04. NAO6 URL: https://www.softbankrobotics.com/emea/en/nao [accessed 2018-12-18] [WebCite Cache ID 74kL0tFld]
31. PARO Robots. 2014. PARO therapeutic robot URL: http://www.parorobots.com/ [accessed 2018-12-17] [WebCite Cache ID 74jeUipLk]

32. Beran TN, Ramirez-Serrano A, Vanderkooi OG, Kuhn S. Reducing children's pain and distress towards flu vaccinations: a novel and effective application of humanoid robotics. Vaccine 2013 Jun 07;31(25):2772-2777. [doi: 10.1016/j.vaccine.2013.03.056] [Medline: 23623861]
33. Jibb LA, Birnie KA, Nathan PC, Beran TN, Hum V, Victor JC, et al. Using the MEDiPORT humanoid robot to reduce procedural pain and distress in children with cancer: a pilot randomized controlled trial. Pediatr Blood Cancer 2018 Sep;65(9):e27242. [doi: 10.1002/pbc.27242] [Medline: 29893482]
34. Blanson Henkemans OA, Bierman BP, Janssen J, Neerincx MA, Looije R, van der Bosch H, et al. Using a robot to personalise health education for children with diabetes type 1: a pilot study. Patient Educ Couns 2013 Aug;92(2):174-181. [doi: 10.1016/j.pec.2013.04.012] [Medline: 23684366]
35. Henkemans OA, Bierman BP, Janssen J, Looije R, Neerincx MA, van Dooren MM, et al. Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. Int J Hum Comput Stud 2017 Oct;106:63-76. [doi: 10.1016/j.ijhcs.2017.06.001]
36. Alemi M, Meghdari A, Ghazisaedy M. The impact of social robotics on L2 learners’ anxiety and attitude in English vocabulary acquisition. Int J Soc Robot 2015 Feb 4;7(4):523-535. [doi: 10.1007/s12369-015-0286-y]
37. Alemi M, Ghanbarzadeh A, Meghdari A, Moghadam LJ. Clinical application of a humanoid robot in pediatric cancer interventions. Int J of Soc Robotics 2015 Mar 18;8(5):743-759. [doi: 10.1007/s12369-015-0294-y]
38. Srinivasan SM, Eigsti I, Neelly L, Bhat AN. The effects of embodied rhythm and robotic interventions on the spontaneous and responsive social attention patterns of children with Autism Spectrum Disorder (ASD): A pilot randomized controlled trial. Res Autism Spectr Disord 2016 Jul;27:54-72 [FREE Full text] [doi: 10.1016/j.rasd.2016.01.004] [Medline: 27453721]
39. Srinivasan SM, Park IK, Neelly LB, Bhat AN. A comparison of the effects of rhythm and robotic interventions on repetitive behaviors and affective states of children with Autism Spectrum Disorder (ASD). Res Autism Spectr Disord 2015 Oct 01;18:51-63 [FREE Full text] [doi: 10.1016/j.rasd.2015.07.004] [Medline: 26251668]
40. Yun S, Choi J, Park S, Bong G, Yoo H. Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system. Autism Res 2017 Jul;10(7):1306-1323. [doi: 10.1002/aur.1778] [Medline: 28464438]
41. Srinivasan SM, Kaur M, Park IK, Gifford TD, Marsh KL, Bhat AN. The effects of rhythm and robotic interventions on the imitation/praxis, interpersonal synchrony, and motor performance of children with Autism Spectrum Disorder (ASD): a pilot randomized controlled trial. Autism Res Treat 2015;2015:736516 [FREE Full text] [doi: 10.1155/2015/736516] [Medline: 26793394]
42. So W, Wong MK, Lam CK, Lam W, Chui AT, Lee T, et al. Using a social robot to teach gestural recognition and production in children with autism spectrum disorders. Disabil Rehabil Assist Technol 2018 Dec;13(6):527-539. [doi: 10.1080/17483107.2017.1344886] [Medline: 28673117]
43. So W, Wong MK, Lam W, Cheng C, Yang J, Huang Y, et al. Robot-based intervention may reduce delay in the production of intransitive gestures in Chinese-speaking preschoolers with autism spectrum disorder. Mol Autism 2018;9:34 [FREE Full text] [doi: 10.1186/s13229-018-0217-5] [Medline: 29796238]
44. Costescu CA, Vanderborght B, David DO. Robot-Enhanced CBT for dysfunctional emotions in social situations for children with ASD. J Evid Based Psychot 2017 Sep 01;17(2):119-132. [doi: 10.24193/jebp.2017.2.7]
45. Kim ES, Berkovits LD, Bernier EP, Leyzberg D, Shic F, Paul R, et al. Social robots as embedded reinforcers of social behavior in children with autism. J Autism Dev Disord 2013 May;43(5):1038-1049. [doi: 10.1007/s10803-012-1645-2] [Medline: 23111617]
46. Huskens B, Verschuur R, Gillesen J, Didden R, Barakova E. Promoting question-asking in school-aged children with autism spectrum disorders: effectiveness of a robot intervention compared to a human-trainer intervention. Dev Neurorehabil 2013 Oct;16(5):345-356. [doi: 10.3109/17518423.2012.739212] [Medline: 23586852]
47. Pop CA, Simut RE, Pintea S, Saldien J, Rusu AS, Vanderfaeillie J, et al. Social robots vs computer display: does the way social stories are delivered make a difference for their effectiveness on ASD children? J Educ Comput Res 2014 Apr 04;49(3):381-401. [doi: 10.2190/EC.49.3.f]
48. Peca A, Simut R, Pintea S, Vanderborght B. Are children with ASD more prone to test the intentions of the robonova robot compared to a human? Int J of Soc Robotics 2015 Aug 23;7(5):629-639. [doi: 10.1007/s12369-015-0317-8]
49. Srinivasan SM, Eigsti IM, Gifford T, Bhat AN. The effects of embodied rhythm and robotic interventions on the spontaneous and responsive verbal communication skills of children with Autism Spectrum Disorder (ASD): A further outcome of a pilot randomized controlled trial. Res Autism Spectr Disord 2016 Jul;27:73-87 [FREE Full text] [doi: 10.1016/j.rasd.2016.04.001] [Medline: 27668011]
50. Broadbent E, Garrett J, Jepsen N, Li OV, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
51. Banks MR, Willoughby LM, Banks WA. Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. J Am Med Dir Assoc 2008 Mar;9(3):173-177. [doi: 10.1016/j.jamda.2007.11.007] [Medline: 18294600]
52. Tanaka M, Ishii A, Yamano E, Ogikubo H, Okazaki M, Kamimura K, et al. Effect of a human-type communication robot on cognitive function in elderly women living alone. Med Sci Monit 2012 Sep;18(9):CR550-CR557 [FREE Full text] [Medline: 22936190]

53. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
54. Moyle W, Cooke M, Beattie E, Jones C, Klein B, Cook G, et al. Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial. J Gerontol Nurs 2013 May;39(5):46-53. [doi: 10.3928/00989134-20130313-03] [Medline: 23506125]
55. Broadbent E, Peri K, Kerse N, Jayawardena C, Kuo I, Datta C. Robots in older people's homes to improve medication adherence and quality of life: A randomised cross-over trial. In: Social Robotics.: International Conference on Social Robotics; 2014 Presented at: Social Robotics 6th International Conference; October 27-29, 2014; Sydney, Australia p. 64-73. [doi: 10.1007/978-3-319-11973-1_7]
56. Kim GH, Jeon S, Im K, Kwon H, Lee BH, Kim GY, et al. Structural brain changes after traditional and robot-assisted multi-domain cognitive training in community-dwelling healthy elderly. PLoS One 2015;10(4):e0123251 [FREE Full text] [doi: 10.1371/journal.pone.0123251] [Medline: 25898367]
57. Valentí Soler M, Agüera-Ortiz L, Olazarán Rodríguez J, Mendoza Rebolledo C, Pérez Muñoz A, Rodríguez Pérez I, et al. Social robots in advanced dementia. Front Aging Neurosci 2015;7:133 [FREE Full text] [doi: 10.3389/fnagi.2015.00133] [Medline: 26388764]
58. Jøranson N, Pedersen I, Rokstad AM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. J Am Med Dir Assoc 2015 Oct 01;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]
59. Jøranson N, Pedersen I, Rokstad AM, Ihlebaek C. Change in quality of life in older people with dementia participating in Paro-activity: a cluster-randomized controlled trial. J Adv Nurs 2016 Dec;72(12):3020-3033. [doi: 10.1111/jan.13076] [Medline: 27434512]
60. Liang A, Piroth I, Robinson H, MacDonald B, Fisher M, Nater UM, et al. A pilot randomized trial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc 2017 Oct 01;18(10):871-878. [doi: 10.1016/j.jamda.2017.05.019] [Medline: 28668664]
61. Petersen S, Houston S, Qin H, Tague C, Studley J. The utilization of robotic pets in dementia care. J Alzheimers Dis 2017;55(2):569-574 [FREE Full text] [doi: 10.3233/JAD-160703] [Medline: 27716673]
62. Moyle W, Jones CJ, Murfield JE, Thalib L, Beattie ER, Shum DK, et al. Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. J Am Med Dir Assoc 2017 Sep 01;18(9):766-773 [FREE Full text] [doi: 10.1016/j.jamda.2017.03.018] [Medline: 28780395]
63. Jones C, Moyle W, Murfield J, Draper B, Shum D, Beattie E, et al. does cognitive impairment and agitation in dementia influence intervention effectiveness? Findings from a cluster-randomized-controlled trial with the therapeutic robot, PARO. J Am Med Dir Assoc 2018 Jul;19(7):623-626. [doi: 10.1016/j.jamda.2018.02.014] [Medline: 29656838]
64. Moyle W, Jones C, Murfield J, Thalib L, Beattie E, Shum D, et al. Effect of a robotic seal on the motor activity and sleep patterns of older people with dementia, as measured by wearable technology: a cluster-randomised controlled trial. Maturitas 2018 Apr;110:10-17. [doi: 10.1016/j.maturitas.2018.01.007] [Medline: 29563027]
65. Pfadenhauer M, Dukat C. Robot caregiver or robot-supported caregiving? The performative deployment of the social robot PARO in dementia care. Int J Soc Robot 2015 Jan 30;7(3):393-406. [doi: 10.1007/s12369-015-0284-0]
66. Mervin MC, Moyle W, Jones C, Murfield J, Draper B, Beattie E, et al. The cost-effectiveness of using PARO, a therapeutic robotic seal, to reduce agitation and medication use in dementia: findings from a cluster-randomized controlled trial. J Am Med Dir Assoc 2018 Jul;19(7):619-622.e1. [doi: 10.1016/j.jamda.2017.10.008] [Medline: 29325922]
67. Thodberg K, Sørensen LU, Videbech PB, Poulsen PH, Houbak B, Damgaard V, et al. behavioral responses of nursing home residents to visits from a person with a dog,a robot seal or a toy cat. Anthrozoös 2016 Mar 08;29(1):107-121. [doi: 10.1080/08927936.2015.1089011]
68. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: study protocol for a randomized controlled trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
69. Mengoni SE, Irvine K, Thakur D, Barton G, Dautenhahn K, Guldberg K, et al. Feasibility study of a randomised controlled trial to investigate the effectiveness of using a humanoid robot to improve the social skills of children with autism spectrum disorder (Kaspar RCT): a study protocol. BMJ Open 2017 Dec 22;7(6):e017376 [FREE Full text] [doi: 10.1136/bmjopen-2017-017376] [Medline: 28645986]
70. Moyle W, Bramble M, Jones CJ, Murfield JE. "She Had a Smile on Her Face as Wide as the Great Australian Bite": a Qualitative Examination of Family Perceptions of a Therapeutic Robot and a Plush Toy. Gerontologist 2019 Dec 09;59(1):177-185. [doi: 10.1093/geront/gnx180] [Medline: 29165558]
71. Broekens J, Heerink M, Rosendal H. Assistive social robots in elderly care: a review. Gerontechnology 2009;8(2):94-103. [doi: 10.4017/gt.2009.08.02.002.00]
72. Belpaeme T, Baxter P, Greeff J, Kennedy J, Read R, Looije R. Child-robot interaction: perspectives and challenges. In: Social Robotics. 2013 Presented at: Social Robotics 5th International Conference; October 27-29, 2013; Bristol, UK p. 452-459. [doi: 10.1007/978-3-319-02675-6_45]

73. Al-Taee MA, Kapoor R, Garrett C, Choudhary P. Acceptability of robot assistant in management of type 1 diabetes in children. Diabetes Technol Ther 2016 Dec;18(9):551-554. [doi: 10.1089/dia.2015.0428] [Medline: 27479198]
74. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng 2012;14:275-294. [doi: 10.1146/annurev-bioeng-071811-150036] [Medline: 22577778]
75. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord 2012 Dec;6(1):249-262 [FREE Full text] [doi: 10.1016/j.rasd.2011.05.006] [Medline: 22125579]
76. Perkins J, Bartlett H, Travers C, Rand J. Dog-assisted therapy for older people with dementia: a review. Australas J Ageing 2008 Dec;27(4):177-182. [doi: 10.1111/j.1741-6612.2008.00317.x] [Medline: 19032618]
77. Rabbitt SM, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded use. Clin Psychol Rev 2015 Feb;35:35-46. [doi: 10.1016/j.cpr.2014.07.001] [Medline: 25462112]
78. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing [Review Article]. IEEE Comput Intell Mag 2018 Aug;13(3):55-75. [doi: 10.1109/MCI.2018.2840738]
79. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018;2018:7068349 [FREE Full text] [doi: 10.1155/2018/7068349] [Medline: 29487619]
80. Tarnowski P, Kolodziej M, Majkowski A, Rak R. Emotion recognition using facial expressions. Procedia Comput Sci 2017;108:1175-1184. [doi: 10.1016/j.procs.2017.05.025]

Abbreviations

ASD: autism spectrum disorder
COPD: chronic obstructive pulmonary disease
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.12.18; peer-reviewed by E Broadbent, K Vänni, O Korn, I de la Torre, L Desideri; comments to author 08.02.19; revised version received 22.03.19; accepted 09.04.19; published 10.05.19.

Please cite as:
Robinson NL, Cottier TV, Kavanagh DJ
Psychosocial Health Interventions by Social Robots: Systematic Review of Randomized Controlled Trials
J Med Internet Res 2019;21(5):e13203
URL: http://www.jmir.org/2019/5/e13203/
doi:10.2196/13203
PMID:31094357

Original Paper

The Role of Social Interactions in Motor Performance: Feasibility Study Toward Enhanced Motivation in Telerehabilitation

Roni Barak Ventura[1], MS; Shinnosuke Nakayama[1], PhD; Preeti Raghavan[2], MD; Oded Nov[3], PhD; Maurizio Porfiri[1,4], PhD

[1]Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn, NY, United States
[2]Department of Rehabilitation Medicine, New York University School of Medicine, New York, NY, United States
[3]Department of Technology Management and Innovation, New York University Tandon School of Engineering, Brooklyn, NY, United States
[4]Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY, United States

Corresponding Author:
Maurizio Porfiri, PhD
Department of Mechanical and Aerospace Engineering
New York University Tandon School of Engineering
6 MetroTech Center
Brooklyn, NY, 11201
United States
Phone: 1 646 997 3681
Fax: 1 646 997 3532
Email: mporfiri@nyu.edu

Abstract

Background: Robot-mediated telerehabilitation has the potential to provide patient-tailored cost-effective rehabilitation. However, compliance with therapy can be a problem that undermines the prospective advantages of telerehabilitation technologies. Lack of motivation has been identified as a major factor that hampers compliance. Exploring various motivational interventions, the integration of citizen science activities in robotics-based rehabilitation has been shown to increase patients' motivation to engage in otherwise tedious exercises by tapping into a vast array of intrinsic motivational drivers. Patient engagement can be further enhanced by the incorporation of social interactions.

Objective: Herein, we explored the possibility of bolstering engagement in physical therapy by leveraging cooperation among users in an environmental citizen science project. Specifically, we studied how the integration of cooperation into citizen science influences user engagement, enjoyment, and motor performance. Furthermore, we investigated how the degree of interdependence among users, such that is imposed through independent or joint termination (JT), affects participation in citizen science-based telerehabilitation.

Methods: We developed a Web-based citizen science platform in which users work in pairs to classify images collected by an aquatic robot in a polluted water canal. The classification was carried out by labeling objects that appear in the images and trashing irrelevant labels. The system was interfaced by a haptic device for fine motor rehabilitation. We recruited 120 healthy volunteers to operate the platform. Of these volunteers, 98 were cooperating in pairs, with 1 user tagging images and the other trashing labels. The other 22 volunteers performed both tasks alone. To vary the degree of interdependence within cooperation, we implemented independent and JTs.

Results: We found that users' engagement and motor performance are modulated by their assigned task and the degree of interdependence. Motor performance increased when users were subjected to independent termination (P=.02), yet enjoyment decreased when users were subjected to JT (P=.005). A significant interaction between the type of termination and the task was found to influence productivity (P<.001) as well as mean speed, peak speed, and path length of the controller (P=.01, P=.006, and P<.001, respectively).

Conclusions: Depending on the type of termination, cooperation was not always positively associated with engagement, enjoyment, and motor performance. Therefore, enhancing user engagement, satisfaction, and motor performance through cooperative citizen science tasks relies on both the degree of interdependence among users and the perceived nature of the task. Cooperative citizen science may enhance motivation in robotics-based telerehabilitation, if designed attentively.

(J Med Internet Res 2019;21(5):e12708) doi:10.2196/12708

KEYWORDS

citizen science; social interactions; telerehabilitation; physical therapy

Introduction

Background

Debilitating neurological diseases such as stroke require intensive, repetitive, and high-frequency physical therapy for maximum recovery of motor function and self-reliance [1-3]. However, costly resources and limited rehabilitation personnel make rehabilitation unavailable to the majority of patients. Furthermore, disability often encumbers mobility, preventing patients from leaving their homes and frequenting the therapists' office [4]. Therefore, reaching their full recovery potential is greatly contingent upon performing self-directed physical therapy with limited professional feedback.

Robot-Mediated Telerehabilitation

Several rehabilitation robots have been developed for delivery of exercise programs for the upper limb, including the MIT-Manus (Massachusetts Institute of Technology) [5,6], Gentle/S (University of Reading) [7], ArmIn (ETH Zurich) [8], and Mirror Image Movement Enabler (Stanford University) [9]. Ultimately, robotic rehabilitation devices aim to administer and monitor exercise for the arm with reproducible high-intensity and high-dosage sensorimotor therapy while collecting pertinent data for assessment by a medical professional [10,11].

The major hurdles in the widespread adoption of these robotics-based technologies are costs and user-friendliness, whereby these devices often have prohibitive costs to the general public and require some form of technological proficiency that may be beyond the typical background of patients or even therapists [12,13]. To fill these gaps, several studies have explored the feasibility of delivering rehabilitation treatments using low-cost, off-the-shelf gaming systems such as the Microsoft Kinect and PlayStation EyeToy [14-17]. Gaming controllers are intuitive to users, easy to repurpose, and more affordable, thereby offering a promising means for accessible home-based telerehabilitation. Gaming controllers can also measure motor performance objectively, toward remote assessment of patient status and progress by physicians [18].

As an example, the Novint Falcon can detect subtle differences in the kinematics of healthy and affected individuals [19] and evaluate patients' motor learning as their rehabilitation progresses [20], through measurements of mean speed, peak speed, and path length traversed. These metrics have been previously used and upheld in robotic telerehabilitation of the upper limb to rapidly assess physical effort and movement accuracy and smoothness [21,22]. Other metrics can be used to assess patients' motor performance, including range of motion [8,23], coordination [24,25], and amount of force exerted [14,26].

Adherence to Rehabilitation Regimen

Although the mechanical framework of telerehabilitation has been successfully implemented in homes, patients often fail to comply with their home-based physical therapy, primarily because of the lack of motivation [27,28]. Acknowledging that sustained engagement is a prerequisite for successful outcomes, a large body of research has studied motivational interventions through game designs toward overcoming noncompliance in telerehabilitation [29-33]. For example, in an experimental study, Colombo et al [31] simulated a video game experience by displaying performance scores to improve motivation and adherence to the physical regimen. Similarly, in the study by Nijenhuis [29], a motivational messaging system was introduced to encourage future engagement in training sessions. Other studies have considered the use of serious games that do not aim primarily at entertainment for enhancing the physical rehabilitation experience [34-37]. For instance, Jonsdottir et al [38] have demonstrated the feasibility and efficacy of Rehab@Home, a therapeutic framework that simulates daily life activities in a virtual environment using Kinect. This gaming system was shown to increase gross motor function and improve patients' experience and perception of health in patients with multiple sclerosis.

Overall, these studies have demonstrated that gamification increases engagement in rehabilitation exercises [37,39]. Yet, the full capacity of supplemental motivational interventions remains largely untapped as designers rarely emphasize the users' intellect and interest to maintain prolonged engagement. Particularly in the context of rehabilitation, the age group of the majority of patients may not be conducive to the use of typical computer games that target young gamers [40,41]. Aiming to address the differential motivations of the elderly, Flores et al [27] pinpointed the gaming design criteria catering to the needs of both young and elderly users, which include (1) consideration of decreased sensorimotor abilities, (2) cognitively challenging elements, and (3) some degree of socialization.

Citizen Science and Telerehabilitation

Following this line of work, we have previously demonstrated the potential utility of citizen science in increasing engagement and enjoyment in rehabilitation through the systematic interaction of environmental citizen science and robotics-based low-cost telerehabilitation technologies [19,42,43]. Citizen science projects address a wide range of scientific fields of inquiry. For example, on Stardust@Home, users can review images of an aerogel that was sent to outer space and flag traces of interstellar dust trapped in it [44]. In a different project, Foldit, volunteers fold virtual proteins and produce novel models of protein structures [45]. Citizen science could be a passive undertaking whereby citizen scientists lend computational power while their computers are idle [46,47]. The activities are not restricted to desktop computers and may also take place outdoors, where volunteers report of animal sightings or record air and water quality using their mobile phones [46,47].

In general, in citizen science projects, members of the public execute scientific tasks in authentic research projects, led by professional scientists and otherwise [46,47]. The contribution of volunteers typically involves data collection or data analysis and does not require specific expertise [47,48]. The motivation ascribed to citizen science projects is majorly intrinsic as participation is intellectually stimulating and promotes learning [48,49]. Moreover, the completion of individual tasks requires a small time commitment, allowing users to contribute at their own pace. Therefore, citizen science inherently satisfies the criteria identified by Flores et al [27], with the exception of social interaction, which must be separately addressed through new design interventions.

Social Interactions as a Motivational Driver

Personal and social support, whether provided by practitioners, family, or friends, has been found to increase patients' motivation toward performing exercise at home and to improve their mental well-being [50,51]. In the context of telerehabilitation, socially assistive robots, affording social interaction with patients while relaying treatments, were created [52]. The mere interaction with inanimate socially assistive robots was demonstrated to increase patients' engagement in therapy and alleviate their feelings of stress and depression [52-54]. Building on this evidence, interhuman social interactions were introduced between the patient and their practitioner [52,55,56] and subsequently extended to include interactions with relatives and friends, and even strangers [30,52,57-59]. In all cases, patients expressed a strong preference to perform exercise with another person rather than alone and with a human partner rather than a virtual one [30,57,60]. Moreover, social interactions were demonstrated to improve motor performance [61]. Yet, the context in which social interactions were studied is largely limited to games.

Whether social interactions could benefit or hamper the success of citizen science–based rehabilitation treatments remains elusive. It is known that social presence alone should enhance user engagement and prolong participation in Web-based platforms through social comparison [62-65]. However, working in a team may also lead to the opposite outcome, whereby users could reduce their participation in an activity because of diffusion of responsibility, a sociopsychological phenomenon observed when an individual is less likely to assume responsibility of action in the presence of other individuals [66]. Diffusion of responsibility is moderated by several factors, including anonymity [67], group size [68], and division of labor [69], which can all be found in citizen science [70]. As a result, it is difficult to predict whether including social elements in citizen science–based rehabilitation could produce the sought motivational factor advocated in the study by Flores et al [27] or, instead, produce an adverse social phenomenon through diffusion of responsibility.

Objectives

In this study, we sought to fill this gap in knowledge by examining the influence of computer-mediated cooperation on the engagement and motor performance of users involved in a rehabilitation exercise that integrated environmental citizen science and robotics-based technologies. We hypothesized that introducing cooperative tasks into citizen science would motivate users to extend their contribution by increasing the amount of scientific data they collect or analyze (productivity) and the time they spend performing the scientific task (persistence). This hypothesis rests on previous evidence that both of these measures are positively associated with motivation in goal-related activities [71]. In addition to increasing engagement, we expected that the integration of cooperation would improve users' motor performance, reflected by their exertion of higher levels of physical effort. Finally, we hypothesized that the level of improvement in engagement and motor performance would be modulated by varying the degree of independence between paired users, whereby strengthening the interdependence between them would mitigate diffusion of responsibility.

To test our hypotheses, we created a novel, dedicated interface for Brooklyn Atlantis—a local citizen science project for environmental monitoring of the highly polluted Gowanus canal, located in Brooklyn, New York [72]. Our system enabled users to analyze pictures of the canal taken by an aquatic robot using a low-cost haptic controller, whose potential use in rehabilitation treatments on patients has been previously demonstrated by our group and other researchers [19]. Using the system, a pair of volunteers was presented with a list of descriptive keywords that may describe the objects in images. The volunteers sorted a list of labels together where one user allocates labels that describe objects in the image while the other discards irrelevant labels. Here, we report results for the effect of using the platform and the collaborative procedure on healthy people.

Methods

Hardware and Software

All activities were performed using the Novint Falcon game controller, a low-cost haptic controller capable for use in 3 dimensions (Figure 1). The Novint Falcon offers translational hand movement with 3 degrees of freedom: left-right (x-axis), up-down (y-axis), and push-pull (z-axis). This device was demonstrated to provide effective fine-motor hand rehabilitation [73]. The system, developed using Unity 3D (Unity Technologies), displayed a 360° image of the Gowanus Canal on a computer screen (Figure 2). To explore the image, users pressed the middle button on the controller continuously and moved the controller in the general direction they wanted to rotate the view. A reproduction of the Novint Falcon interface was continuously shown on the screen, as a reference for the function of each button (Figure 2).

Movement was implemented in spherical coordinates, whereby motion of the controller along the x-axis (Figure 1) translated into azimuthal rotation (turning right or left in Figure 2) and motion of the controller along the y-axis (Figure 1) translated into elevational rotation (turning up or down in Figure 2). As the motion along the push-pull axis did not convey a meaningful function (zoom was not offered), a highly resistive force was applied in this direction to prevent motion. In addition, visual feedback was added to the system, such that the image of the Gowanus Canal would fade in response to motion along the z-axis, either pushing or pulling. The deviation from the z-axis was further conveyed through a black circle and radiating cone, portraying where the user is facing and how far off the axis they are located (Figure 2).

As part of the citizen science image classification project, 2 tasks were implemented in the system. The first task consisted of tagging objects observed in the images using labels from a list, located on the right of the 360° image (green panel in Figure 2). The second task entailed eliminating labels from the list that were not in the image by allocating them to the trash bin, located on the right of the list of labels (yellow panel in Figure 2). When 2 users performed the tasks together, each user had independent control over exploration of the 360° image. Social cues were conveyed through the system as cooperating users could view their partner's actions in real time. That is, the user assigned with the tagging task could see a list of the eliminated labels forming below the trash bin (Figure 2). Similarly, the user assigned with the trashing task could see the labels assigned to the image (Figure 2). The presence of a peer was further made evident by highlighting a label in red, indicating that it was selected by the peer. An illustration of the setup is depicted in Figure 3.

Users were able to select labels by pressing the right or left button on the controller, depending on their dexterity. Once a button was pressed, the label was tethered to the cursor and effectively dragged by it. To deselect a label, or to release it at a desired location, users pressed the controller button again. Once assigned, labels were replaced by others from a predetermined sequence of 49 labels, all of which were previously contributed by citizen scientists in Brooklyn Atlantis in our previous research [19,74-77]. To maintain fluidity of the image classification process and avoid oversaturation of images with tags, an image was replaced by another image after 5 tags were assigned. A tag counter was displayed on the screen to inform the user of the number of tags assigned to the current image (Figure 2).

Users were able to terminate the activity by pressing the red *Quit* button on the screen (Figure 2). To simulate the different levels of interdependence in the cooperation between individuals, 2 types of termination were considered in the experiments: independent termination (IT), whereby users could continue contributing to the project even after their peer had quit, and joint termination (JT) whereby termination by 1 user ceased the session for the other user too.

Figure 1. The Novint Falcon with the designated axes of motion.

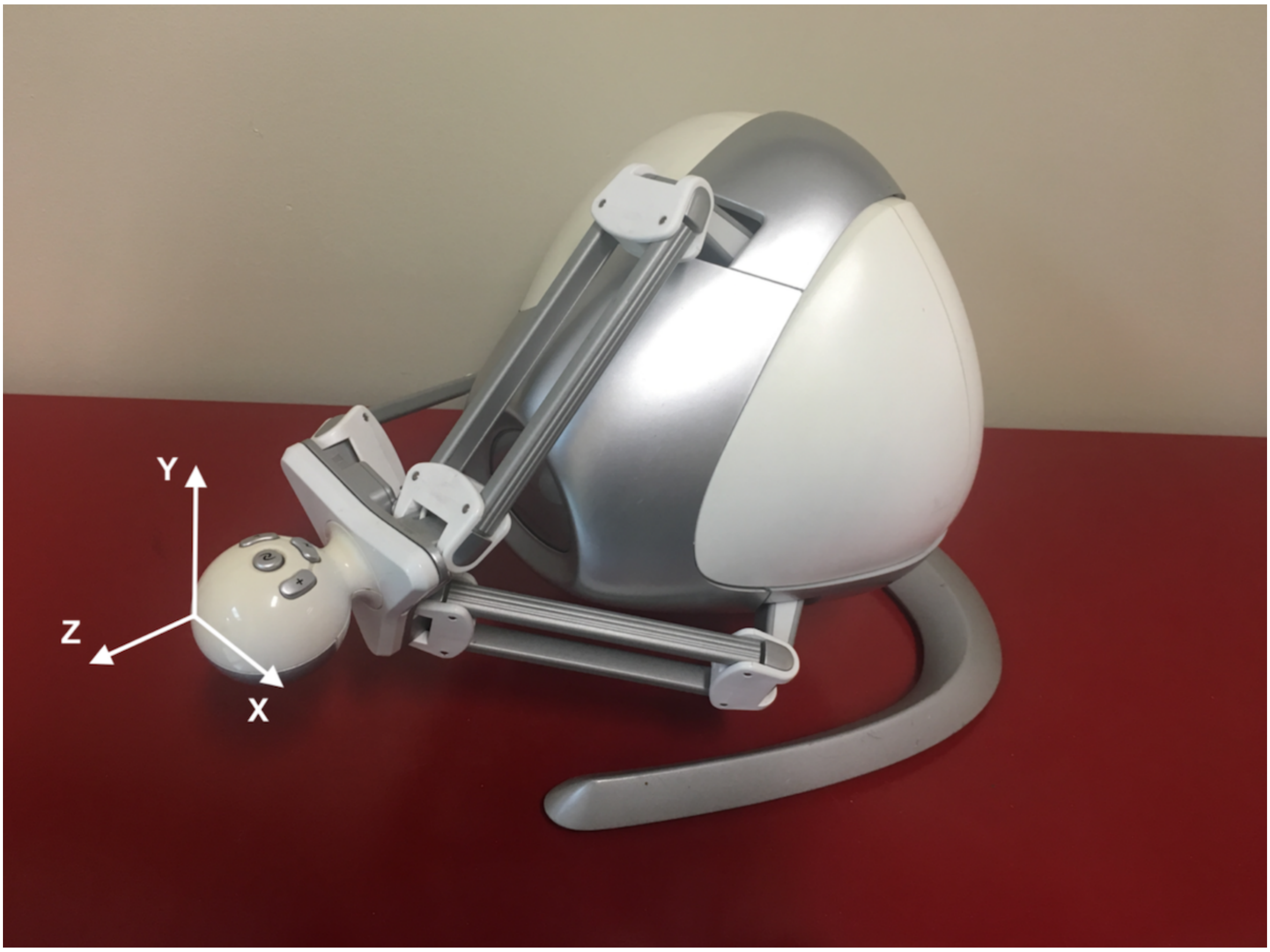
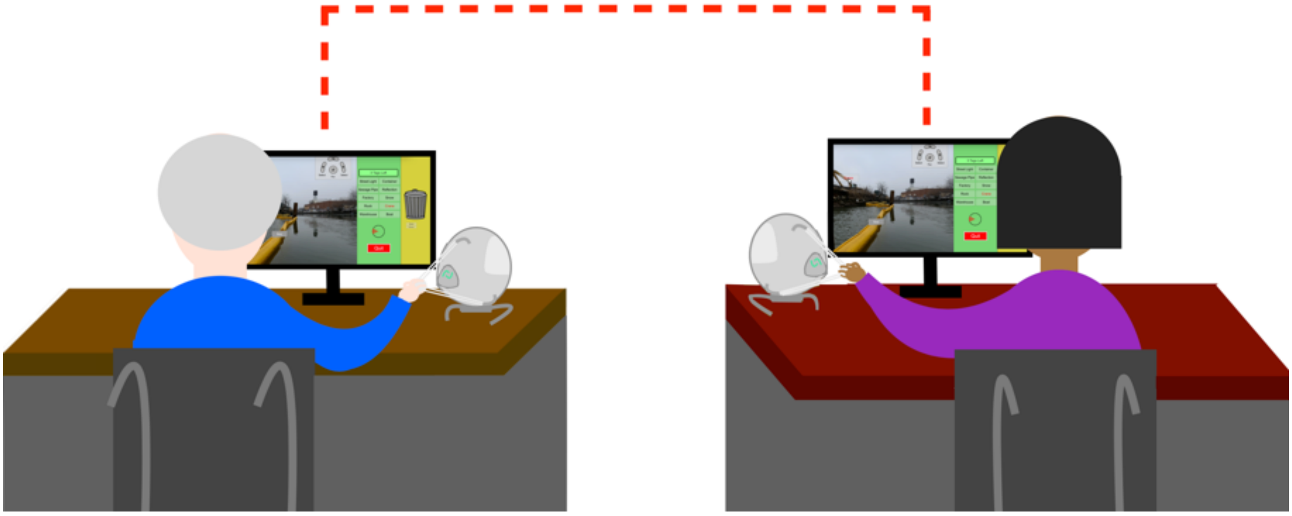


Figure 2. A screenshot of the user interface. On the left of the screenshot is a 360° image of the Gowanus canal. The user's cursor is placing the label "Crane" onto the image, while a tag containing the word "Buoy" has already been placed. A reproduction of the Novint Falcon controller with a description of the function of each button is located on the upper right corner of the image. In the green panel, a counter of the number of labels that are yet to be assigned to the current image is displayed at the top. Below the counter, there is a list of 10 labels. The label "Crane" is highlighted in red as it is currently selected by the user. Below the list of labels is a visual feedback that represents deviation from the z-axis. A Quit button is situated at the bottom of the green panel. In the yellow panel, there is a garbage bin for eliminating labels that do not describe objects in the current image. The labels below it, "Robot" and "Person", have been eliminated by the user.

Figure 3. Schematic of two cooperating users classifying images remotely from two different computers in separate rooms.



Experimental Procedure

This study was carried out in compliance with the institutional review board (IRB) at New York University (IRB FY2016-184). Overall, 120 members of the university community, with a mean age of 27.36 (SD 8.28) years, were recruited and subjected to one of 3 conditions (Table 1): IT (50 volunteers), JT (48 volunteers), and control (22 volunteers). Although control subjects performed both tagging and trashing, volunteers in IT and JT cooperatively carried out the activity such that half of the volunteers (25 and 24 volunteers in IT and JT, respectively) performed only tagging and the other half (25 and 24 volunteers in IT and JT, respectively) performed only trashing. In the control condition, the volunteer was able to withdraw from the activity at any time.

Table 1. A summary of the experimental conditions tested.

Condition and task assignment	Cooperation	Number of volunteers
Control		
Tagging and trashing	Absent	22
Independent termination		
Tagging	Present	25
Trashing	Present	25
Joint termination		
Tagging	Present	24
Trashing	Present	24

Recruitment and experimental procedures were standardized through scripts and a PowerPoint presentation. We recruited volunteers in public spaces on campus. During recruitment, we verbally introduced potential participants on campus to the notion of citizen science following a script. Once recruited, paired volunteers were brought into 2 separate private rooms to simulate Web-based cooperation as envisioned in future application within robotics-based telerehabilitation. They did not know who their peer was.

All participants were subjected to the same experimental protocol. Before beginning the experiment, participants were given an overview of the Gowanus Canal and Brooklyn Atlantis using a PowerPoint presentation. Through the presentation, within cooperative conditions, IT and JT, participants were notified that they will be working together with a peer and were instructed to complete their assigned task only. They were informed that they may withdraw at any point they would like and whether their withdrawal will terminate their peer's participation (JT) or not (IT). Upon signing a consent form, the participants underwent a tutorial teaching them how to use the Novint Falcon and the system. After the tutorial, they were connected with their peer and began carrying out their tasks. Users who were subjected to the control condition carried out both tasks, tagging and trashing. Users who were subjected to cooperative conditions were randomly assigned to one of the 2 tasks. They performed the exercise until they pressed the quit button. After quitting, the participants rated their experience on a 7-point Likert scale in response to the statements "I enjoyed this activity" and "This activity was fun." Once the volunteers submitted their answers, the experiment was concluded.

Data Collection and Analysis

Data Acquisition

For each user, 3 datasets were created. The first dataset documented information on tag allocation, including tag content, time of allocation, and allocating user identity number. The second dataset recorded users' scores of enjoyment. The third dataset consisted of the Novint Falcon controller position in 3D space, recorded at a sampling rate of 60 positions per second. The collected data were used to quantify user engagement, enjoyment, and motor performance.

Data Processing

User engagement was evaluated through their productivity and persistence [71]. Productivity was measured as the number of labels processed by the user. Persistence was measured as the time spent performing the activity [71]. Users' enjoyment was evaluated from surveys. Interrater reliability was validated using the Cronbach alpha [78]. Enjoyment was scored by averaging the ratings on the multiple questions for each user, linearly scaling between 0 (Likert scale 1) and 1 (Likert scale 7), and normalizing using an arcsine transformation by considering the proportional nature of the variable [79].

The trajectory of the controller was examined from the recording of consecutive points in space over time. A total of 3 motion metrics were evaluated from the trajectory, namely, the controller's mean speed, peak speed, and path length. For each trial, the instantaneous speed was estimated using a backward Euler scheme on the sampled positions from the haptic device. The mean speed was computed by averaging instantaneous values over the whole trajectory and the peak speed as the maximum value among the 90th percentile from the trajectory [80]. The path length was measured as the sum of distances between pairs of consecutive data points.

Statistical Analysis

The influence of cooperation on engagement, enjoyment, and motor performance was investigated by fitting each variable into a generalized linear-mixed effects model [81], specifying condition (3 levels: control, IT, and JT) as an independent variable and both pair identity and task assignment (tagging and trashing) as random effects (R *lme4* package version 1.1-15 [82]). To improve the normality of the model residual, we specified a Gaussian family with a log link for persistence and enjoyment, a Poisson family with a log link for productivity, and a gamma family with a log link for motor performance. The significance of the influence of conditions was tested using a likelihood ratio test, comparing the model against a null model in the absence of the condition as the independent variable. When significant effect was found, post hoc analysis was performed using the Dunnett test (R *multcomp* package version 1.4-8 [83]).

Next, we evaluated the influence of the modality through which cooperation was implemented on engagement, enjoyment, and motor performance. Specifically, we fitted each variable into a generalized linear mixed-effects model, specifying condition

(2 levels: IT and JT), task assignment (2 levels: tagging and trashing), and the interaction between them as independent variables, and pair identity as a random effect. The same error family as the previous model was used for the corresponding variable. To test the significance of the interaction term, the full model was tested against a null model without the interaction using a likelihood ratio test. In case a significant interaction was found, the difference between tasks was further examined within each condition, by specifying task as an independent variable and pair identity as a random effect. In case the interaction was not significant, we removed the interaction from the full model, and the effects of condition and task were tested using a likelihood ratio test, individually, comparing against a null model.

Although not part of our original hypotheses, we also tested the influence of social presence on individual speed performance. In each pair of cooperating peers in IT, the trajectory of the more persistence was partitioned into 2 parts, before and after their peer had quit. Users' mean and peak speeds were fitted into separate generalized linear mixed-effects models, specifying the time partition (2 levels: before and after) as an independent variable and user identity as a random effect. A gamma family with log link was specified to normalize the model residual. Users' speeds before peer withdrawal were compared with their speeds following peer withdrawal using a likelihood ratio test, comparing the model against a null model in the absence of the time partition as the independent variable.

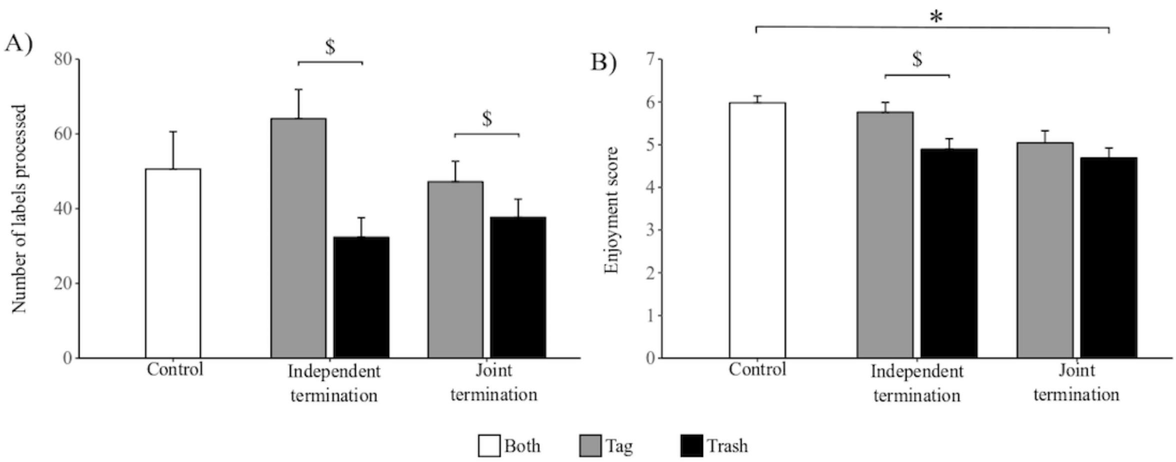
For all statistical tests, we set the level of significance at alpha=.05.

Results

Influence of Cooperation on Engagement

On average, users processed (tagged or trashed) a mean of 46.35 (SD 3.16) labels, spending 16.37 (SD 0.63) min. Neither productivity nor persistence were found to differ among conditions (χ^2_2=0.1 *P*=.92 and χ^2_2<0.1; *P*=.79, respectively; Figure 4). However, the level of enjoyment was found to vary among conditions (χ^2_2=10.5; *P*=.005; Figure 4), with JT users rating the activity significantly lower than control users (z=3.25; *P*=.002). By contrast, IT users did not rate the activity significantly different from control users (z=1.94; *P*=.08).

Figure 4. Engagement of users in the activity. A) number of labels processed by participants in each condition. B) rate of enjoyment for each condition. The vertical lines represent standard errors. *: statistically different means among conditions. $: statistically different means among tasks.



Influence of Cooperation on Motor Performance

With regard to motor performance, the mean speed did not differ among conditions (χ^2_2=2.6; *P*=.26). Contrarily, we determined a significant variation of peak speed among conditions (χ^2_2=7.7; *P*=.02; Figure 5). Although post hoc comparisons failed to identify a significant difference between JT and control conditions (z=0.11; *P*=.99), we registered a significant difference between IT users and control users (z=2.44; *P*=.02). The path length was not significantly different among conditions (χ^2_2=4.1; *P*=.12).

Influence of Modality on Engagement

Testing for the influence of cooperation modality on productivity, we found a significant interaction between condition and task (χ^2_2=43.1; *P*<.001). When investigating the effect of task assignment in each condition, we found a significant difference in productivity between the tasks in both conditions, IT (χ^2_1=265.7; *P*<.001) and JT (χ^2_1=25.5; *P*<.001). We failed to identify a significant interaction between condition and task with regard to persistence (χ^2_1=3.3; *P*=.06). An interaction between condition and task was not found in enjoyment as well (χ^2_1=1.1; *P*=.29). Enjoyment was significantly different between the tasks (χ^2_1=7.7; *P*=.005), whereas condition failed to reach significance (χ^2_1=2.8; *P*=.09).

Influence of Modality on Motor Performance

With regard to the performance metrics, a significant interaction between condition and task was found to influence path length (χ^2_1=6.3; *P*=.01), mean speed (χ^2_1=7.4; *P*=.006), and peak speed (χ^2_1=25.8; *P*<.001; Figure 5). For path length, we found a significant difference between task assignment in IT condition

(χ^2_1=11.3; *P*<.001), whereas we did not find one in JT condition (χ^2_1<0.1; *P*=.90). For mean speed, a significant difference between tasks was found in IT condition (χ^2_1=6.2; *P*=.01) but not in JT condition (χ^2_1=2.4; *P*=.12). Finally, for peak speed, a significant difference was found between tasks in both IT and JT conditions (χ^2_1=16.1; *P*<.001 and χ^2_1=11.0; *P*<.001, respectively).

In IT, volunteers significantly reduced their mean and peak speeds following their peer's quitting (z=2.97; *P*=.002 and z=3.30; *P*<.001, respectively; Figure 6).

Figure 5. Motor metrics. A) mean speed in each condition, B) peak speed for each condition, C) path length traversed by the controller in each conditions. The vertical lines represent standard errors. * represents statistically different means among conditions. $ represents statistically different means among tasks.

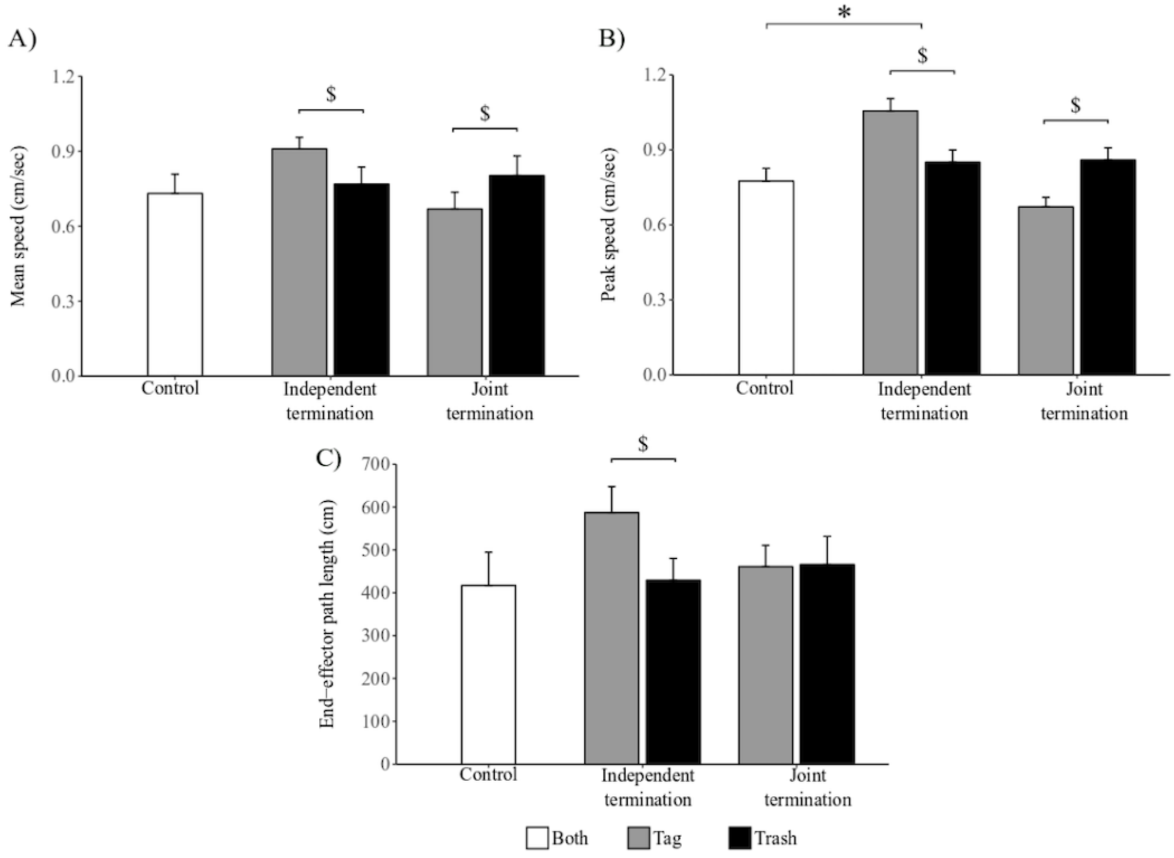
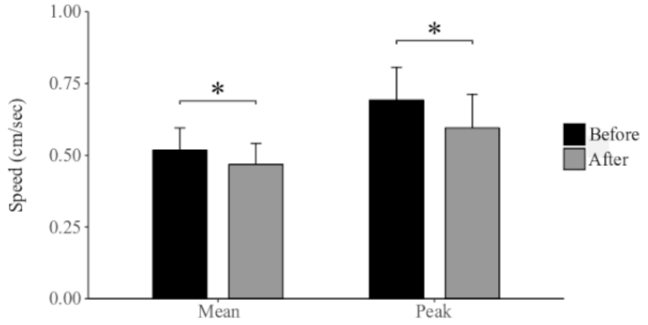


Figure 6. Differences in mean and peak speeds of the more persistent users in condition IT, before and after their peer has withdrawn. The vertical lines represent standard errors. * represents statistically different means among conditions.



Discussion

Principal Findings

Citizen science is an effective means for improving rehabilitation treatments. Patients undergoing physical rehabilitation have shown a strong preference toward exercise embedded with citizen science and were more likely to repeat it at the cost of their time commitment [19]. Although social interactions hold potential to further increase patients' engagement in rehabilitation [30,57], the modality in which they are framed could widely shape the outcomes of the treatment [84]. In this study, we attempted to elucidate the influence of computer-mediated cooperation on motor performance during a citizen science activity, mediated by a low-cost haptic device.

We designed a series of experiments simulating an authentic telerehabilitation setting, where participants remotely cooperated in the analysis of environmental images using low-cost haptic devices. From survey instruments and direct measurements of motor activities, we sought to quantify the potential effect of cooperation in telerehabilitation. Our results indicate that 2 elements, interdependence and task assignment, can influence the effects of cooperation on engagement and motor performance, thereby offering a potential means for improving rehabilitation treatments. However, the way in which these 2 variables interact to influence the response of the subjects may challenge one's intuition.

In partial contrast with our hypotheses, we did not find that cooperative division of labor, such that each user is assigned to a different task toward a shared goal [85], is always conducive to higher engagement. Although we did not register a difference in the level of engagement between control users and users who cooperated via IT, we found an expected reduction in enjoyment for users who cooperated via JT. In our experimental design, JT was implemented to promote interdependence between users, which we had initially identified as a key factor to mitigate diffusion of responsibility and coerce users to persist for a longer period of time [86,87]. However, it is likely that JT was accompanied by other confounding factors, contributing to a reduction in enjoyment.

It is tenable that we inadvertently introduced random termination in the trials where users faced uncertainty with regard to the timing of termination, as the other player could terminate the task at any time. This uncertainty about the horizon of the relationship with the other player was shown in previous research to impact the degree of cooperativeness of players negatively [88,89], where the more termination becomes likely, the less cooperation has been observed [90]. In game theory, random termination has been shown to discount the payoff of players' actions [88,89] such that players would try to avoid losses and become less cooperative. Such weakening of cooperation also results in lower levels of enjoyment and satisfaction [91]. Similar dynamics likely emerged in the proposed citizen science–based telerehabilitation activity, thereby calling for future research to explore alternative strategies that could promote interdependence between users. For instance, we could attempt at a priori identifying a predetermined length for the trials, by matching patients undergoing a similar rehabilitation therapy.

Our findings also provided insight into the role of social interactions on motor performance. Measuring relevant kinematic variables is central to the notion of telerehabilitation, whereby supplying care providers with clinical information will enable them to track patients' status and adjust their rehabilitation program remotely and efficiently [18]. Ideally, care providers could also infer abnormal, compensatory movement from the data and instruct patients to correct it [18]. Mean speed, peak speed, and path length traversed [19,92] have been used in robotic telerehabilitation of the upper limb as indicators of motion quality [14,21,22]. In addition to the evaluation of motor performance, the physical effort exerted by an individual can also be linked to their motivation to perform the exercise task. Investing greater effort to complete a challenging task often leads to self-determined behavior, resulting in a sense of competence [93] and an increase in intrinsic motivation [94].

Although we expected that users would improve their motor performance because of cooperation, we found a modest reduction in motor performance similar to the discussed reduction in enjoyment. More specifically, motor performance of users cooperating via JT was similar to control with regard for all the selected metrics, whereas IT resulted in higher values of the peak speed relative to control. It is possible that random termination could explain the observed difference, where a user would not invest the same effort when faced with the potential that his/her work could be vanished because of exogenous termination of the activity by the peer. Future research should seek to explore alternative modalities to favor cooperation, without challenging enjoyment and effort that are key to the success of rehabilitation. The notion of setting intervals for the exercises could be a viable approach, whereby it could mitigate the harmful effects of random termination, while leveraging the beneficial role of cooperation. In fact, analyzing time variations of motor performance of users who cooperated via IT, we discovered that the speed of the more persistent users in IT significantly reduced following their peer's withdrawal. This confirms the intuition that social interaction should prompt individuals to exert more effort in their task, thereby calling for future studies to engineer social interactions toward improving recovery and patients' self-perception of physical capacity [95,96].

In addition to the type of termination, we found that task assignment can modulate user engagement. We found that cooperating users in IT condition assigned with tagging were more engaged than their peers who were assigned with trashing. The difference in engagement may be attributed to the perceived nature of the task a user has been assigned to. In fact, engagement is positively associated with the identifiability of the share an individual contributes to the groupwork [87,97], that is, individuals whose contribution is more valued and recognized by group peers are more likely to be motivated to perform their task. Conversely, individuals whose contribution is less important and recognized by others in the group are less motivated to perform their task. Therefore, the dissimilarity between tagging and trashing tasks could lead to unequal levels of engagement, with tagging users being more engaged than trashing users. This observation calls for further studies in which targeted design interventions will be explored to investigate differences between cooperation and collaboration, where individuals are assigned to the same task. It is tenable that cooperating individuals should perform better when assigned with a common group task rather than interdependent tasks [84]. Working in a collaborative setting where they fulfill identical functions jointly in support of a shared goal [85] could bolster team cohesion and lead to even higher performance and satisfaction among users [98].

Evidence shows that group cohesiveness can be improved with increasingly overt sharing of information, leading to greater engagement in group tasks [99]. Recently, we showed that the mere presentation of *social foot prints*, digital cues that suggest the presence of other Web users can be used to increase the amount and duration of physical exercise during citizen science activities [64]. In a different study, using a virtual peer operating in open- and closed-loop paradigms, we demonstrated that bidirectional flow of social information can substantially increase user contribution to a citizen science project [75]. Seemingly, as more social presence is conveyed between Web users, an increasingly trusting climate is created, conducive to cooperation [100]. In future studies, one could explore how sharing of personal information such as age, location, and interests can impact trust and engagement in Web-based citizen science telerehabilitation [101].

Citizen science contributes to a sense of community in cooperative telerehabilitation through the introduction of scientific content. Unlike gaming-based motivational interventions, this study capitalizes on human intellect as an intrinsic motivator. Previously, we had shown that users prefer to perform an exercise associated with scientific content [19,92]. Beyond the intellectual stimulus, citizen science adds virtue and a sense of contribution to the activity, which is not found in the majority of serious games. Web-based social platforms where individuals share personal values often create communities with which contributors can identify [48]. In telerehabilitation in particular, patients can benefit from such an environment that could alleviate the isolation many of them experience [102,103].

Limitations

Although our work brings forward evidence in favor of the use cooperative citizen science in rehabilitation, it comes with a number of limitations. First, the difference in engagement that we observed among conditions was moderate. It is possible that because citizen science is inherently engaging [92], additional motivational interventions such as social interactions offer only a weak enhancement to engagement, limited by a ceiling effect. To surpass such a ceiling effect, it would be beneficial to explore the role of cooperation in citizen science–based rehabilitation in a longitudinal study, where persistence is measured by the frequency patients choose to engage in exercise and productivity is measured as aggregated contribution.

Second, we studied motor performance using the Novint Falcon, a haptic device which is no longer being produced. However, the fine-motor tasks imparted by the Novint Falcon can be achieved using other haptic devices. For instance, surgical delta robots such as the Force Dimension and Phantom offer movement with 6 degrees of freedom and can apply a comparable amount of force feedback [104-106].

Third, in this study, we recruited healthy subjects from New York University campuses. Our findings may be narrowly generalizable as the sample consists of healthy individuals from the Brooklyn area with access to high education. Although we drew our sample from different programs of the university, the volunteers may have distinctly different interests and motivations from the typical patient undergoing rehabilitation. For example, the participants in this study may have greater interest in science or in restoration of the Gowanus Canal than the average person. Offering a wider range of citizen science projects to choose from based on personal interests might further improve enjoyment, engagement, and motor performance. Future research with patients from diverse backgrounds in a clinical setting will elucidate effects of this work on clinical outcomes.

Conclusions

We offer evidence for the utility of cooperation in improving engagement in citizen science–based telerehabilitation. Citizen science can offer intellectual stimulus and a community for patients to engage with and relate to. It attends the needs of more patients, including those who are less interested in traditional gaming [27,37], thereby extending the benefits of adherence to home-based physical therapy to a larger population. The value of this study can be expanded to other domains that rely on user participation and engagement, including Web-based consumer platforms [107], social networks [108], crowdsourcing efforts [109], and general game design [91]. Ultimately, we anticipate our approach will be translated into low-cost technology for telerehabilitation and help patients reach their full potential recovery.

Acknowledgments

This study was supported by the National Science Foundation under award number CBET-1604355. RBV would like to acknowledge the Mitsui USA Foundation for their generous support. All the authors would like to thank M Nadini for helping conduct experiments and S Richmond and TJ Tolbert for software development.

Authors' Contributions

RBV, ON, and MP conceived and designed the experiments. RBV performed the experiments. RBV, SN ON, and MP analyzed the data. RBV, SN, ON, and MP wrote the manuscript. RBV, SN, PR, ON, and MP reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Bütefisch C, Hummelsheim H, Denzler P, Mauritz KH. Repetitive training of isolated movements improves the outcome of motor rehabilitation of the centrally paretic hand. J Neurol Sci 1995 May;130(1):59-68. [doi: 10.1016/0022-510X(95)00003-K] [Medline: 7650532]

2. Harris JE, Eng JJ, Miller WC, Dawson AS. A self-administered Graded Repetitive Arm Supplementary Program (GRASP) improves arm function during inpatient stroke rehabilitation: a multi-site randomized controlled trial. Stroke 2009 Jun;40(6):2123-2128. [doi: 10.1161/STROKEAHA.108.544585] [Medline: 19359633]
3. Volpe BT, Lynch D, Rykman-Berland A, Ferraro M, Galgano M, Hogan N, et al. Intensive sensorimotor arm training mediated by therapist or robot improves hemiparesis in patients with chronic stroke. Neurorehabil Neural Repair 2008;22(3):305-310 [FREE Full text] [doi: 10.1177/1545968307311102] [Medline: 18184932]
4. Kroll T, Jones GC, Kehn M, Neri MT. Barriers and strategies affecting the utilisation of primary preventive services for people with physical disabilities: a qualitative inquiry. Health Soc Care Community 2006 Jul;14(4):284-293. [doi: 10.1111/j.1365-2524.2006.00613.x] [Medline: 16787479]
5. Krebs HI, Ferraro M, Buerger SP, Newbery MJ, Makiyama A, Sandmann M, et al. Rehabilitation robotics: pilot trial of a spatial extension for MIT-Manus. J Neuroeng Rehabil 2004 Oct 26;1(1):5 [FREE Full text] [doi: 10.1186/1743-0003-1-5] [Medline: 15679916]
6. Fasoli SE, Krebs HI, Stein J, Frontera WR, Hogan N. Effects of robotic therapy on motor impairment and recovery in chronic stroke. Arch Phys Med Rehabil 2003 Apr;84(4):477-482. [doi: 10.1053/apmr.2003.50110] [Medline: 12690583]
7. Coote S, Murphy B, Harwin W, Stokes E. The effect of the GENTLE/s robot-mediated therapy system on arm function after stroke. Clin Rehabil 2008 May;22(5):395-405. [doi: 10.1177/0269215507085060] [Medline: 18441036]
8. Nef T, Mihelj M, Colombo G, Riener R. ARMin - Robot for rehabilitation of the upper extremities. In: Proceedings of the 2006 IEEE International Conference on Robotics and Automation. 2006 Presented at: IEEE International Conference on Robotics and Automation; May 15-19, 2006; Orlando, Florida, USA. [doi: 10.1109/ROBOT.2006.1642181]
9. Lum PS, Burgar CG, Loos MV, Shor PC, Majmundar M, Yap R. MIME robotic device for upper-limb neurorehabilitation in subacute stroke subjects: a follow-up study. J Rehabil Res Dev 2006;43(5):631-642. [doi: 10.1682/JRRD.2005.02.0044] [Medline: 17123204]
10. Laut J, Porfiri M, Raghavan P. The present and future of robotic technology in rehabilitation. Curr Phys Med Rehabil Rep 2016 Dec;4(4):312-319 [FREE Full text] [doi: 10.1007/s40141-016-0139-0] [Medline: 28603663]
11. Krebs H, Dipietro L, Levy-Tzedek S, Fasoli S, Rykman-Berland A, Zipse J, et al. A paradigm shift for rehabilitation robotics. IEEE Eng Med Biol Mag 2008 Jul;27(4):61-70 [FREE Full text] [doi: 10.1109/MEMB.2008.919498] [Medline: 19004697]
12. Romer GR, Stuyt HJ, Peters A. Cost-savings and economic benefits due to the Assistive Robotic Manipulator (ARM). 2005 Presented at: International Conference on Rehabilitation Robotics; 28 June-1 July, 2005; Chicago, Illinois, USA. [doi: 10.1109/ICORR.2005.1501084]
13. Wootton R, Hebert MA. What constitutes success in telehealth? J Telemed Telecare 2001;7(Suppl 2):3-7. [doi: 10.1258/1357633011937245] [Medline: 11747643]
14. Reinkensmeyer DJ, Pang CT, Nessler JA, Painter CC. Web-based telerehabilitation for the upper extremity after stroke. IEEE Trans Neural Syst Rehabil Eng 2002 Jun;10(2):102-108. [doi: 10.1109/TNSRE.2002.1031978] [Medline: 12236447]
15. Lange B, Chang C, Suma E, Newman B, Rizzo AS, Bolas M. Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. Conf Proc IEEE Eng Med Biol Soc 2011;2011:1831-1834. [doi: 10.1109/IEMBS.2011.6090521] [Medline: 22254685]
16. Weiss PL, Rand D, Katz N, Kizony R. Video capture virtual reality as a flexible and effective rehabilitation tool. J Neuroeng Rehabil 2004 Dec 20;1(1):12 [FREE Full text] [doi: 10.1186/1743-0003-1-12] [Medline: 15679949]
17. Rand D, Kizony R, Weiss PTL. The Sony PlayStation II EyeToy: low-cost virtual reality for use in rehabilitation. J Neurol Phys Ther 2008 Dec;32(4):155-163. [doi: 10.1097/NPT.0b013e31818ee779] [Medline: 19265756]
18. Perry JC, Ruiz-Ruano JA, Keller T. Telerehabilitation: toward a cost-efficient platform for post-stroke neurorehabilitation. IEEE Int Conf Rehabil Robot 2011;2011:5975413. [doi: 10.1109/ICORR.2011.5975413] [Medline: 22275616]
19. Laut J, Cappa F, Nov O, Porfiri M. Increasing patient engagement in rehabilitation exercises using computer-based citizen science. PLoS One 2015;10(3):e0117013 [FREE Full text] [doi: 10.1371/journal.pone.0117013] [Medline: 25793867]
20. Scalona E, Martelli F, Del PZ, Palermo E, Rossi S. A novel protocol for the evaluation of motor learning in 3D reching tasks using Novint Falcon. 2018 Presented at: IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics; August 27-29, 2018; Enschede, Netherlands. [doi: 10.1109/BIOROB.2018.8487735]
21. Bosecker C, Dipietro L, Volpe B, Krebs HI. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. Neurorehabil Neural Repair 2010 Jan;24(1):62-69 [FREE Full text] [doi: 10.1177/1545968309343214] [Medline: 19684304]
22. Zollo L, Rossini L, Bravi M, Magrone G, Sterzi S, Guglielmelli E. Quantitative evaluation of upper-limb motor control in robot-aided rehabilitation. Med Biol Eng Comput 2011 Oct;49(10):1131-1144. [doi: 10.1007/s11517-011-0808-1] [Medline: 21792622]
23. Colombo R, Pisano F, Micera S, Mazzone A, Delconte C, Carrozza MC, et al. Robotic techniques for upper limb evaluation and rehabilitation of stroke patients. IEEE Trans Neural Syst Rehabil Eng 2005 Sep;13(3):311-324. [doi: 10.1109/TNSRE.2005.848352] [Medline: 16200755]
24. Raghavan P, Santello M, Gordon AM, Krakauer JW. Compensatory motor control after stroke: an alternative joint strategy for object-dependent shaping of hand posture. J Neurophysiol 2010 Jun;103(6):3034-3043 [FREE Full text] [doi: 10.1152/jn.00936.2009] [Medline: 20457866]

25. Squeri V, Zenzeri J, Morasso P, Basteris A. Integrating proprioceptive assessment with proprioceptive training of stroke patients. IEEE Int Conf Rehabil Robot 2011;2011:5975500. [doi: 10.1109/ICORR.2011.5975500] [Medline: 22275696]
26. Boissy P, Bourbonnais D, Carlotti MM, Gravel D, Arsenault BA. Maximal grip force in chronic stroke subjects and its relationship to global upper extremity function. Clin Rehabil 1999 Aug;13(4):354-362. [doi: 10.1191/026921599676433080] [Medline: 10460123]
27. Flores E, Tobon G, Cavallaro E, Cavallaro F, Perry J, Keller T. Improving patient motivation in game development for motor deficit rehabilitation. 2008 Presented at: International Conference on Advances in Computer Entertainment Technology; December 3-5, 2008; Yokohama, Japan p. 381-384. [doi: 10.1145/1501750.1501839]
28. Conraads VM, Deaton C, Piotrowicz E, Santaularia N, Tierney S, Piepoli MF, et al. Adherence of heart failure patients to exercise: barriers and possible solutions: a position statement of the Study Group on Exercise Training in Heart Failure of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2012 May;14(5):451-458 [FREE Full text] [doi: 10.1093/eurjhf/hfs048] [Medline: 22499542]
29. Nijenhuis SM, Prange GB, Amirabdollahian F, Sale P, Infarinato F, Nasr N, et al. Feasibility study into self-administered training at home using an arm and hand device with motivational gaming environment in chronic stroke. J Neuroeng Rehabil 2015 Oct 09;12:89 [FREE Full text] [doi: 10.1186/s12984-015-0080-y] [Medline: 26452749]
30. Novak D, Nagle A, Keller U, Riener R. Increasing motivation in robot-aided arm rehabilitation with competitive and cooperative gameplay. J Neuroeng Rehabil 2014 Apr 16;11:64 [FREE Full text] [doi: 10.1186/1743-0003-11-64] [Medline: 24739255]
31. Colombo R, Pisano F, Mazzone A, Delconte C, Micera S, Carrozza MC, et al. Design strategies to improve patient motivation during robot-aided rehabilitation. J Neuroeng Rehabil 2007 Feb 19;4:3 [FREE Full text] [doi: 10.1186/1743-0003-4-3] [Medline: 17309790]
32. Zimmerli L, Jacky M, Lünenburger L, Riener R, Bolliger M. Increasing patient engagement during virtual reality-based motor rehabilitation. Arch Phys Med Rehabil 2013 Sep;94(9):1737-1746. [doi: 10.1016/j.apmr.2013.01.029] [Medline: 23500181]
33. Vong SK, Cheing GL, Chan F, So EM, Chan CC. Motivational enhancement therapy in addition to physical therapy improves motivational factors and treatment outcomes in people with low back pain: a randomized controlled trial. Arch Phys Med Rehabil 2011 Feb;92(2):176-183. [doi: 10.1016/j.apmr.2010.10.016] [Medline: 21272712]
34. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Dec 20;11:CD008349. [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
35. Taylor MJ, Griffin M. The use of gaming technology for rehabilitation in people with multiple sclerosis. Mult Scler 2015 Apr;21(4):355-371. [doi: 10.1177/1352458514563593] [Medline: 25533296]
36. Burke JW, McNeill M, Charles D, Morrow P, Crosbie J, McDonough S. Serious games for upper limb rehabilitation following stroke. : IEEE; 2009 Presented at: Conference in Games and Virtual Worlds for Serious Applications; March 23-24, 2009; Coventry, United Kingdom p. 103-110. [doi: 10.1109/VS-GAMES.2009.17]
37. Rego P, Moreira P, Reis L. Serious games for rehabilitation: a survey and a classification towards a taxonomy. : IEEE; 2010 Presented at: Iberian Conference on Information Systems and Technologies; June 16-19, 2010; Santiago de Compostela, Spain URL: https://ieeexplore.ieee.org/document/5556674
38. Jonsdottir J, Bertoni R, Lawo M, Montesano A, Bowman T, Gabrielli S. Serious games for arm rehabilitation of persons with multiple sclerosis. A randomized controlled pilot study. Mult Scler Relat Disord 2018 Jan;19:25-29. [doi: 10.1016/j.msard.2017.10.010] [Medline: 29112939]
39. Burke JW, McNeill MD, Charles DK, Morrow PJ, Crosbie JH, McDonough SM. Optimising engagement for stroke rehabilitation using serious games. Vis Comput 2009 Aug 27;25(12):1085-1099. [doi: 10.1007/s00371-009-0387-4]
40. Awad M, Ferguson S, Craig C. Designing games for older adults: an affordance based approach. : IEEE; 2014 Presented at: IEEE International Conference on Serious Games and Applications for Health (SeGAH); May 14-16, 2014; Rio de Janeiro, Brazil. [doi: 10.1109/SeGAH.2014.7067103]
41. Garcia MJ, Felix NK, Lawrence E. Serious games to improve the physical health of the elderly: a categorization scheme. CENTRIC 2011; 2011 Presented at: The International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services (CENTRIC); October 23-28, 2011; Barcelona, Spain p. 64-71 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.9722&rep=rep1&type=pdf
42. Palermo E, Laut J, Nov O, Cappa P, Porfiri M. A natural user interface to integrate citizen science and physical exercise. PLoS One 2017;12(2):e0172587 [FREE Full text] [doi: 10.1371/journal.pone.0172587] [Medline: 28231261]
43. Laut J, Henry E, Nov O, Porfiri M. Development of a mechatronics-based citizen science platform for aquatic environmental monitoring. IEEE/ASME Trans. Mechatron 2014 Oct;19(5):1541-1551. [doi: 10.1109/TMECH.2013.2287705]
44. Westphal A, Butterworth A, Snead C, Craig N, Anderson D, Jones S, et al. National Aeronautics and Space Administration. 2005. Stardust@ home: a massively distributed public search for interstellar dust in the stardust interstellar dust collector URL: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050180792.pdf [accessed 2018-11-06] [WebCite Cache ID 73jPyJPWP]
45. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. Nature 2010 Aug 05;466(7307):756-760 [FREE Full text] [doi: 10.1038/nature09304] [Medline: 20686574]

46. Silvertown J. A new dawn for citizen science. Trends Ecol Evol 2009 Sep;24(9):467-471. [doi: 10.1016/j.tree.2009.03.017] [Medline: 19586682]
47. Haklay M. Citizen Science and Policy: A European Perspective. Washington, DC: The Woodrow Wilson Center, Commons Lab; 2015.
48. Nov O, Arazy O, Anderson D. Dusting for science: motivation and participation of digital citizen science volunteers. 2011 Presented at: iConference 2011; February 08-11, 2011; Seattle, Washington, USA p. 68-74. [doi: 10.1145/1940761.1940771]
49. Nov O, Arazy O, Anderson D. Crowdsourcing for science: understanding and enhancing SciSourcing contribution. 2010 Presented at: 2010 Workshop on the Changing Dynamics of Scientific Collaborations; February 7, 2010; Savannah, Georgia, USA URL: http://faculty.poly.edu/~onov/Nov%20Arazy%20Anderson%20CSCW%202010%20workshop.pdf
50. Lewinter M, Mikkelsen S. Patients' experience of rehabilitation after stroke. Disabil Rehabil 2009 Jul 28;17(1):3-9. [doi: 10.3109/09638289509166621]
51. Friedland J, McColl M. Social support intervention after stroke: results of a randomized trial. Arch Phys Med Rehabil 1992 Jun;73(6):573-581. [Medline: 1622308]
52. Tapus A, Ţăpuş C, Matarić MJ. User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. Intel Serv Robotics 2008 Feb 13;1(2):169-183. [doi: 10.1007/s11370-008-0017-4]
53. Tapus A, Mataric M, Scassellati B. Socially assistive robotics [Grand Challenges of Robotics]. IEEE Robotics and Automation 2007 Mar;14(1):35-42. [doi: 10.1109/MRA.2007.339605]
54. Matarić MJ, Eriksson J, Feil-Seifer DJ, Winstein CJ. Socially assistive robotics for post-stroke rehabilitation. J Neuroeng Rehabil 2007 Feb 19;4:5 [FREE Full text] [doi: 10.1186/1743-0003-4-5] [Medline: 17309795]
55. Johnson MJ, Feng X, Johnson LM, Ramachandran B, Winters JM, Kosasih JB. Robotic systems that rehabilitate as well as motivate: three strategies for motivating impaired arm use. : IEEE; 2006 Presented at: IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics; February 20-22, 2006; Pisa, Italy. [doi: 10.1109/BIOROB.2006.1639095]
56. Carignan CR, Krebs HI. Telerehabilitation robotics: bright lights, big future? J Rehabil Res Dev 2006;43(5):695-710. [doi: 10.1682/JRRD.2005.05.0085]
57. Gorsic M, Novak D. Design and pilot evaluation of competitive and cooperative exercise games for arm rehabilitation at home. : IEEE; 2016 Dec Presented at: IEEE Engineering in Medicine and Biology Society (EMBC); August 16-20, 2016; Orlando, Florida, USA p. 4690-4694. [doi: 10.1109/EMBC.2016.7591774]
58. Baur K, Schättin A, de Bruin ED, Riener R, Duarte JE, Wolf P. Trends in robot-assisted and virtual reality-assisted neuromuscular therapy: a systematic review of health-related multiplayer games. J Neuroeng Rehabil 2018 Nov 19;15(1):107 [FREE Full text] [doi: 10.1186/s12984-018-0449-9] [Medline: 30454009]
59. Dithmer M, Rasmussen JO, Grönvall E, Spindler H, Hansen J, Nielsen G, et al. "The Heart Game": using gamification as part of a telerehabilitation program for heart patients. Games Health J 2016 Feb;5(1):27-33 [FREE Full text] [doi: 10.1089/g4h.2015.0001] [Medline: 26579590]
60. Weibel D, Wissmath B, Habegger S, Steiner Y, Groner R. Playing online games against computer- vs human-controlled opponents: effects on presence, flow, and enjoyment. Comput Human Behav 2008 Sep;24(5):2274-2291. [doi: 10.1016/j.chb.2007.11.002]
61. Johnson MJ, Loureiro RC, Harwin WS. Collaborative tele-rehabilitation and robot-mediated therapy for stroke rehabilitation at home or clinic. Intel Serv Robotics 2008 Jan 23;1(2):109-121. [doi: 10.1007/s11370-007-0010-3]
62. Shen J. Social comparison, social presence, and enjoyment in the acceptance of social shopping websites. Journal Electron Commer Re 2012;13(3):198-212 [FREE Full text]
63. Festinger L. A theory of social comparison processes. Hum Relat 2016 Apr 22;7(2):117-140. [doi: 10.1177/001872675400700202]
64. Wu J, Wang S, Tsai H. Falling in love with online games: the uses and gratifications perspective. Comput Human Behav 2010;26(6):1862-1871. [doi: 10.1016/j.chb.2010.07.033]
65. Nakayama S, Tolbert T, Nov O, Porfiri M. Social information as a means to enhance engagement in citizen science ‐ based telerehabilitation. J Assoc Inf Sci Technol 2018 Dec 23;70(6):587-595. [doi: 10.1002/asi.24147]
66. Darley JM, Latané B. Bystander intervention in emergencies: diffusion of responsibility. J Pers Soc Psychol 1968 Apr;8(4):377-383. [Medline: 5645600]
67. Schwartz SH, Gottlieb A. Bystander anonymity and reactions to emergencies. J Pers Soc Psychol 1980 Sep;39(3):418-430. [Medline: 7431204]
68. Freeman S, Walker MR, Borden R, Latané B. Diffusion of responsibility and restaurant tipping: cheaper by the bunch. Pers Soc Psychol Bull 1975;1(4):584-587. [doi: 10.1177/014616727500100407]
69. Kelman HG. Violence without moral restraint: reflections on the dehumanization of victims and victimizers. J Soc Issues 1973;29(4):25-61. [doi: 10.1111/j.1540-4560.1973.tb00102.x]
70. Nov O, Arazy O, Anderson D. Technology-mediated citizen science participation: a motivational model. 2011 Presented at: International Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Catalonia, Spain URL: http://faculty.poly.edu/~onov/Nov_Arazy_Anderson_Citizen_Science_ICWSM_2011.pdf

71. Touré-Tillery M, Fishbach A. How to measure motivation: a guide for the experimental social psychologist. Soc Personal Psychol Compass 2014 Jul 02;8(7):328-341. [doi: 10.1111/spc3.12110]
72. Laut J, Porfiri M, Nov O. Brooklyn Atlantis: a robotic platform for environmental monitoring with community participation. 2014 Presented at: International Conference of Control, Dynamic Systems, and Robotics; May 15-16, 2014; Ottawa, Ontario, Canada p. 961-968 URL: https://avestia.com/CDSR2014_Proceedings/papers/96.pdf
73. Lange B, Flynn S, Rizzo A. Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation. Phys Ther Rev 2013 Jul 19;14(5):355-363. [doi: 10.1179/108331909X12488667117258]
74. Laut J, Cappa F, Nov O, Porfiri M. Increasing citizen science contribution using a virtual peer. J Assoc Inf Sci Technol 2016 Aug 16;68(3):583-593. [doi: 10.1002/asi.23685]
75. Nov O, Laut J, Porfiri M. Using targeted design interventions to encourage extra-role crowdsourcing behavior. J Assn Inf Sci Tec 2015 May 05;67(2):483-489. [doi: 10.1002/asi.23507]
76. Laut J, Cappa F, Nov O, Porfiri M. A model for citizen scientist contribution in an image tagging task. 2016 Presented at: Dynamic Systems and Control Conference; October 12-14, 2016; Minneapolis, Minnesota, USA. [doi: 10.1115/DSCC2016-9686]
77. Nakayama S, Diner D, Holland JG, Bloch G, Porfiri M, Nov O. The influence of social information and self-expertise on emergent task allocation in virtual groups. Front Ecol Evol 2018 Feb 21;6:-. [doi: 10.3389/fevo.2018.00016]
78. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
79. Warton DI, Hui FK. The arcsine is asinine: the analysis of proportions in ecology. Ecology 2011 Jan;92(1):3-10. [Medline: 21560670]
80. Kaushik K, Sharifi E, Young S. Computing performance measures with national performance management research data set. Transport Res Rec 2015 Nov;2529:10-26. [doi: 10.3141/2529-02]
81. Zuur A, Ieno E, Walker N, Saveliev A, Smith G. Mixed Effects Models and Extensions in Ecology With R. New York, New York, USA: Springer; 2019.
82. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):-. [doi: 10.18637/jss.v067.i01]
83. Bretz F, Hothorn T, Westfall P. Multiple Comparisons Using R. Boca Raton, Florida, USA: Chapman And Hall/CRC; 2019.
84. Jones SC, Vroom VH. Division of labor and performance under cooperative and competitive conditions. J Abnorm Soc Psychol 1964;68(3):313-320. [doi: 10.1037/h0042378]
85. Roschelle J, Teasley S. The construction of shared knowledge in collaborative problem solving. Comput Support Collab Learn 1995:69-97 [FREE Full text] [doi: 10.1007/978-3-642-85098-1_5]
86. Johnson D, Johnson R, Holubec E. The New Circles of Learning: Cooperation in the Classroom and School. Minnesota, USA: Association for Supervision & Curriculum Development; 1994.
87. Wagner JA. Studies of individualism-collectivism: effects on cooperation in groups. Acad Manag J 1995 Feb 01;38(1):152-173. [doi: 10.2307/256731]
88. Pevnitskaya S, Ryvkin D. Environmental context and termination uncertainty in games with a dynamic public bad. Envir Dev Econ 2012 Nov 27;18(01):27-49. [doi: 10.1017/S1355770X12000423]
89. Camerer C. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton, New Jersey, USA: Princeton University Press; 2019.
90. Bó PD. Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. Am Econ Rev 2005 Nov;95(5):1591-1604. [doi: 10.1257/000282805775014434]
91. Sweetser P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. Comput Entertain 2005 Jul 01;3(3):1-24. [doi: 10.1145/1077246.1077253]
92. Cappa P, Clerico A, Nov O, Porfiri M. Can force feedback and science learning enhance the effectiveness of neuro-rehabilitation? An experimental study on using a low-cost 3D joystick and a virtual visit to a zoo. PLoS One 2013;8(12):e83945 [FREE Full text] [doi: 10.1371/journal.pone.0083945] [Medline: 24349562]
93. Liukkonen J, Barkoukis V, Watt A, Jaakkola T. Motivational climate and students' emotional experiences and effort in physical education. J Educ Res 2010 Jun 16;103(5):295-308. [doi: 10.1080/00220670903383044]
94. Deci EL, Ryan RM. The "What" and "Why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
95. Ouellette MM, LeBrasseur NK, Bean JF, Phillips E, Stein J, Frontera WR, et al. High-intensity resistance training improves muscle strength, self-reported function, and disability in long-term stroke survivors. Stroke 2004 Jun;35(6):1404-1409. [doi: 10.1161/01.STR.0000127785.73065.34] [Medline: 15105515]
96. Weiss A, Suzuki T, Bean J, Fielding RA. High intensity strength training improves strength and functional performance after stroke. Am J Phys Med Rehabil 2000;79(4):369-376. [Medline: 10892623]
97. Shaw DM. Size of share in task and motivation in work groups. Sociometry 1960 Jun;23(2):203. [doi: 10.2307/2785683]
98. Van den Bossche P, Gijselaers W, Segers M, Kirschner P. Social and cognitive factors driving teamwork in collaborative learning environments. Small Group Res 2006;37(5):490-521. [doi: 10.1177/1046496406292938]

99. Mesmer-Magnus JR, Dechurch LA. Information sharing and team performance: a meta-analysis. J Appl Psychol 2009 Mar;94(2):535-546. [doi: 10.1037/a0013773] [Medline: 19271807]
100. Friedman B, Khan PH, Howe DC. Trust online. Commun ACM 2000;43(12):34-40. [doi: 10.1145/355112.355120]
101. Robert LP, You S. Are you satisfied yet? Shared leadership, individual trust, autonomy, and satisfaction in virtual teams. J Assoc Inf Sci Technol 2017 Nov 23;69(4):503-513. [doi: 10.1002/asi.23983]
102. Ijsselsteijn W, Nap HH, de Kort Y, Poels K. Digital game design for elderly users. 2007 Presented at: FuturePlay; November 14-17, 2007; Toronto, Canada p. 17-22. [doi: 10.1145/1328202.1328206]
103. Eggermont S, Vandebosch H, Steyaert S. Towards the desired future of the elderly and ICT: policy recommendations based on a dialogue with senior citizens. Poiesis Prax 2006 May 25;4(3):199-217. [doi: 10.1007/s10202-005-0017-9]
104. Brokaw EB, Lum PS, Cooper RA, Brewer BR. Using the Kinect to limit abnormal kinematics and compensation strategies during therapy with end effector robots. 2013 Jun Presented at: IEEE International Conference on Rehabilitation Robotics (ICORR); June 24-26, 2013; Seattle, Washington, USA p. 6650384. [doi: 10.1109/ICORR.2013.6650384]
105. Massie TH, Salisbury JK. The PHANTOM haptic interface: a device for probing virtual objects. 1994 Presented at: Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems; November, 1994; Chicago, Illinois, USA URL: https://alliance.seas.upenn.edu/~medesign/wiki/uploads/Courses/Massie94-DSC-Phantom.pdf
106. Helmer P, Conti F, Grange S, Rouller P. Google Patents. 2012. Haptic device gravity compensation URL: https://patents.google.com/patent/US20100019890 [accessed 2019-04-24] [WebCite Cache ID 77s9rMRBH]
107. Füller J, Bartl M, Ernst H, Mühlbacher H. Community based innovation: how to integrate members of virtual communities into new product development. Electron Commer Res 2006 Jan;6(1):57-73. [doi: 10.1007/s10660-006-5988-7]
108. Sparrowe RT, Liden RC, Wayne SJ, Kraimer ML. Social networks and the performance of individuals and groups. Acad Manag J 2001 Apr 01;44(2):316-325. [doi: 10.2307/3069458]
109. Kittur A, Nickerson JV, Bernstein M, Gerber E, Shaw A, Zimmerman J, et al. The future of crowd work. In: Proceedings of the 2013 conference on Computer supported cooperative work. 2016 Dec Presented at: Conference on Computer supported cooperative work; February 23-27, 2013; San Antonio, Texas, USA p. 4690-4694 URL: http://europepmc.org/abstract/MED/28269319 [doi: 10.1145/2441776.2441923]

Abbreviations

IRB: institutional review board
IT: independent termination
JT: joint termination

Edited by G Eysenbach; submitted 06.11.18; peer-reviewed by S Rossi, C Nebeker, L Hassan; comments to author 30.01.19; revised version received 12.02.19; accepted 17.02.19; published 15.05.19.

Please cite as:
Barak Ventura R, Nakayama S, Raghavan P, Nov O, Porfiri M
The Role of Social Interactions in Motor Performance: Feasibility Study Toward Enhanced Motivation in Telerehabilitation
J Med Internet Res 2019;21(5):e12708
URL: http://www.jmir.org/2019/5/e12708/
doi:10.2196/12708
PMID:31094338

Original Paper

The Deep Learning–Based Recommender System “Pubmender” for Choosing a Biomedical Publication Venue: Development and Validation Study

Xiaoyue Feng[1,2], PhD; Hao Zhang[1], BSc; Yijie Ren[2], BSc; Penghui Shang[2], BSc; Yi Zhu[2], BSc; Yanchun Liang[1,3], PhD; Renchu Guan[1,3*], PhD; Dong Xu[1,2,4*], PhD

[1]Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun, China

[2]Department of Electric Engineering and Computer Science, University of Missouri, Columbia, MO, United States

[3]Zhuhai Sub Laboratory, Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education, Zhuhai College, Jilin University, Zhuhai, China

[4]Christopher S Bond Life Sciences Center, University of Missouri, Columbia, MO, United States

[*]these authors contributed equally

Corresponding Author:
Dong Xu, PhD
Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education
College of Computer Science and Technology
Jilin University
2699 Qinjin Street
Changchun, 130012
China
Phone: 86 13944088266
Email: xudong@missouri.edu

Abstract

Background: It is of great importance for researchers to publish research results in high-quality journals. However, it is often challenging to choose the most suitable publication venue, given the exponential growth of journals and conferences. Although recommender systems have achieved success in promoting movies, music, and products, very few studies have explored recommendation of publication venues, especially for biomedical research. No recommender system exists that can specifically recommend journals in PubMed, the largest collection of biomedical literature.

Objective: We aimed to propose a publication recommender system, named Pubmender, to suggest suitable PubMed journals based on a paper’s abstract.

Methods: In Pubmender, pretrained word2vec was first used to construct the start-up feature space. Subsequently, a deep convolutional neural network was constructed to achieve a high-level representation of abstracts, and a fully connected softmax model was adopted to recommend the best journals.

Results: We collected 880,165 papers from 1130 journals in PubMed Central and extracted abstracts from these papers as an empirical dataset. We compared different recommendation models such as Cavnar-Trenkle on the Microsoft Academic Search (MAS) engine, a collaborative filtering–based recommender system for the digital library of the Association for Computing Machinery (ACM) and CiteSeer. We found the accuracy of our system for the top 10 recommendations to be 87.0%, 22.9%, and 196.0% higher than that of MAS, ACM, and CiteSeer, respectively. In addition, we compared our system with Journal Finder and Journal Suggester, which are tools of Elsevier and Springer, respectively, that help authors find suitable journals in their series. The results revealed that the accuracy of our system was 329% higher than that of Journal Finder and 406% higher than that of Journal Suggester for the top 10 recommendations. Our web service is freely available at https://www.keaml.cn:8081/.

Conclusions: Our deep learning–based recommender system can suggest an appropriate journal list to help biomedical scientists and clinicians choose suitable venues for their papers.

(J Med Internet Res 2019;21(5):e12957) doi:10.2196/12957

KEYWORDS

recommender system; deep learning; convolutional neural network; biomedical literature; PubMed

Introduction

Background

With the fast-growing research activities, more biomedical papers are being published in thousands of journals worldwide. For example, PubMed Central (PMC) has 5.2 million papers and 7409 journals covering biomedical and life sciences [1]. Although these publications play a major role in disseminating research outcome, the growth of journal publications imposes a challenge for selections of appropriate publication venues. It is vital that authors submit to the right journal that meets the journal scope and provides sound reviews. It is equally important that they reach their intended audience and obtain a large number of citations [2]. However, researchers are unfamiliar with all the journals related to their work for choosing the most suitable one for submitting a paper. Moreover, different publication scopes of journals and research interests of reviewers and editors may affect the decision of a submitted manuscript. If the submitted paper cannot meet the interests of a publication venue and its editors and reviewers, it may lead to rejection, delay, or less readership. An appropriate recommender system can help solve this problem.

Recommender systems have been proven to serve as an effective method for decision making in many areas such as music, movies, and information media choices [3-6]. The well-known techniques of recommender systems are content-based recommendation [7,8], collaborative filtering recommendation [4,9], and hybrid recommendation [6,10]. Content-based recommender systems recommend an item to a user based on a description of the item. Collaborative filtering methods and hybrid methods may outperform the content-based recommendations by applying user data, if available. However, after the user privacy issue of Facebook in 2018 and the introduction of European Union General Data Protection Regulation, user data are no longer easy to obtain. Moreover, in many domains, especially in material recommendation, there are no user data available for collaborative filtering methods at the beginning [11], which is regarded as a cold-start problem. Content-based recommendations do not need any user information and are more suitable for solving these problems [12].

Based on the content-based recommendation strategy, several attempts have been made to create recommender systems for medical applications and scientific literature. Using geotagged mobile search logs, Agarwal et al [13] adopted a Random Forest model to predict medical visits. Using topic, writing style, author information, citation information, abstract, and title as information items, latent Dirichlet allocation [14] and k-nearest neighbor [15] were used to classify the scientific literature for recommendation [2,12,16,17]. Luong et al [18] used the coauthors' network as advanced information to recommend a publication venue. Beel et al [19] conducted a literature survey on recommender systems, exploring their methods, evaluation measurements, and datasets. For most of these recommender systems, the high-dimensional and sparse matrix computation is a critical problem [20].

Because of the mismatches caused by ambiguity in text comparisons, the content-based recommendation approach may cause a high error rate [21]. Recently, due to the ability of discovering intricate structures and deep semantics in high-dimensional data, deep learning methods have succeeded in many areas and recently been proposed to build recommender systems for both collaborative filtering and content-based approaches. Hinton et al [22] proposed restricted Boltzmann machines for modeling tabular or count data as a collaborative filtering model on the Netflix data set. McAuley et al [23] proposed an image-based recommendation, which adopted a deep learning model to extract image features. Van den Oord et al [24] applied a deep convolutional neural network (CNN) to predict latent factors from music audios for music recommendation. Wang et al [11] proposed a collaborative deep learning model to jointly perform deep representation learning for the content information and collaborative filtering of a rating matrix. However, to the best of our knowledge, these deep learning techniques have not been used in any biomedical literature recommender system.

Most current venue recommendation studies focus on computer science and technology, but not on the biomedical field. Biomedical sciences are highly interdisciplinary and often link to engineering, medicine, biology, physics, psychology, etc, thereby serving more journals and more diverse topics than any other field. Hence, the development of a recommender system is more essential and challenging for the biomedical sciences than any other discipline. Furthermore, previous recommender systems were based on shallow machine learning methods and social networks. They were generally keyword-based methods and did not take semantics into account. In addition, the few existing systems only focus on journals under a certain organization, such as Elsevier, IEEE, and Springer, instead of PubMed.

Aim

In contrast to our previous study on computer science publication recommendations using conventional machine learning approaches [12], we proposed a deep learning–based recommender system for biomedical publication venues, named Pubmender. Due to the copious vocabulary of biomedical literature, the traditional vector space model can lead to high-dimensional and sparse problems. To address this issue, dimensionality reduction methods are needed before learning the pattern. Moreover, initializing text matrix by pretrained word embedding is more beneficial for training neural networks than random initialized embedding [25]. Accordingly, we applied a word2vec model for our study instead of using the conventional vector space model employed in our previous publication venue recommender system. In addition, deep learning models are able to learn multiple-level abstract representations of data with syntactic and semantic information, since more abstract concepts can be constructed with multiple processing layers [26]. We applied the deep learning approach to provide recommendations of journals for biomedical researchers. Unlike shallow learning, the state-of-the-art embedding method and deep CNN in Pubmender were trained from 837,882 papers in 1130 biomedical journals. This method can help researchers find a variety of choices, without being limited to their own knowledge of journals.

Methods

Pubmender System

Figure 1 shows the architecture and workflow of our Pubmender system. It consists of user interface, data preprocessing, abstract representation, classification, and ranking phase.

The user interface obtains the input data (an abstract submitted by a user) and presents the recommendation results to the user. The data acquirement is followed by data preprocessing and information extraction. At the start of our deep learning model, the abstract representation phase converts an abstract to a vector. The original abstract vector is a concatenation of pretrained word vectors. Subsequently, deep CNN is applied to train the model to achieve high-level abstract representation. A three-layer fully connected network with a softmax operation is applied to classify papers based on the obtained abstract vectors. The recommendation list of the top N journals obtained from the ranking phase is presented to the user.

Data Preprocessing Methods

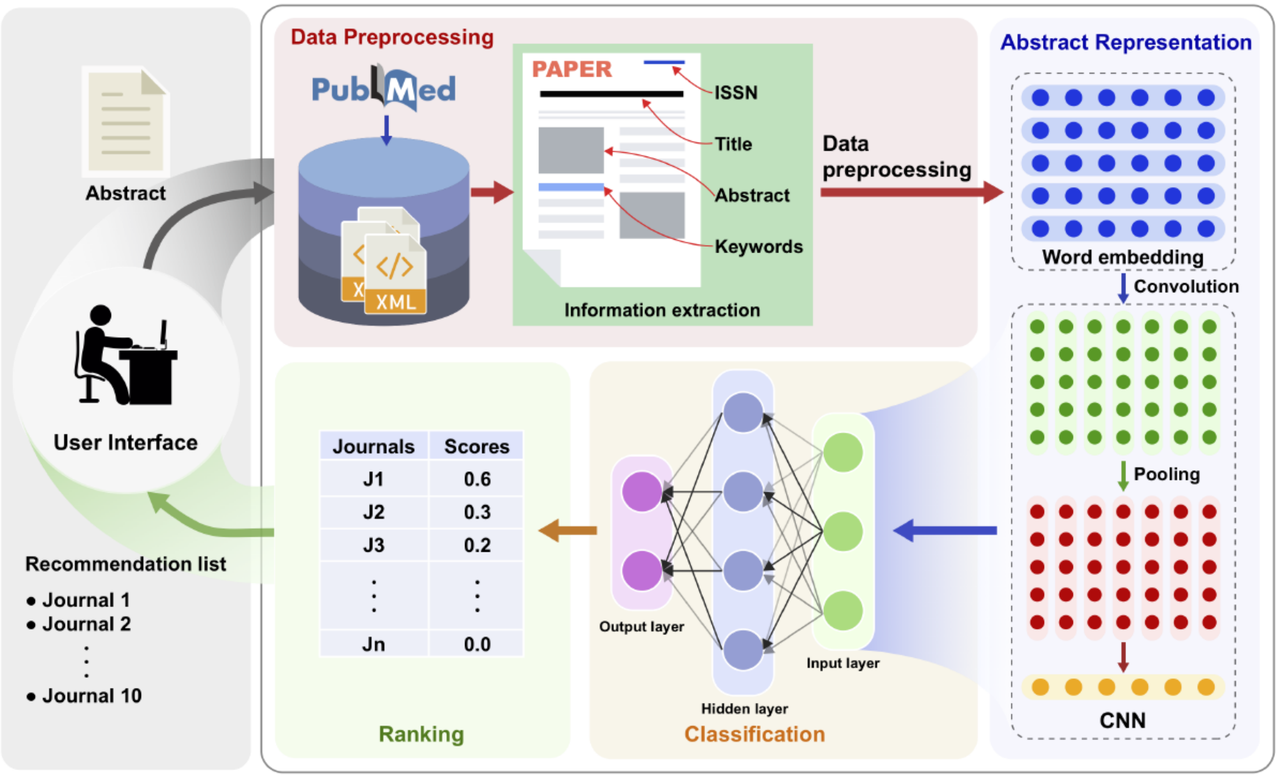
The data were downloaded from the File Transfer Protocol service of PubMed Central (PMC) [27], containing 1,534,649 papers. Based on the journal list of PMC, we selected normal journals deposited under full participation or the US National Institutes of Health portfolio mode, excluding records labeled “Predecessor,” “No New Content,” and “Now Select.” Papers from Jan 2007 to Apr 2017 were selected. Papers with no abstracts or with fewer than 200 characters in abstracts were deleted. Journals containing fewer than 100 papers were also removed. Finally, 880,165 papers in the XML format from 1130 journals were used in our study.

Each PMC file is a semistructured XML document and contains various tags, such as <title>, <abstract>, and <issn>. We extracted the content in <abstract>, <ISSN>, and <pub-date> fields from the raw XML files. Then, pissn and eissn in the ISSN field were replaced by “LocatorPlus ID,” which is the unique identification for a journal in the US National Library of Medicine catalog. After extraction, each abstract was stored in a corresponding file. Natural Language Toolkit was adopted to operate word segmentation [28].

Abstract Representation

In Pubmender, the recommendation task is formulated into a multilabel classification problem, where the text representation and classification methods are critical. For abstracts, we originally embedded abstracts with pretrained word vectors. Thereafter, the original embeddings were fed into CNN to achieve more abstract representation as explained below.

Figure 1. Architecture of our Pubmender system. CNN: convolutional neural network; ISSN: International Standard Serial Number.



Journals	Scores
J1	0.6
J2	0.3
J3	0.2
⋮	⋮
Jn	0.0

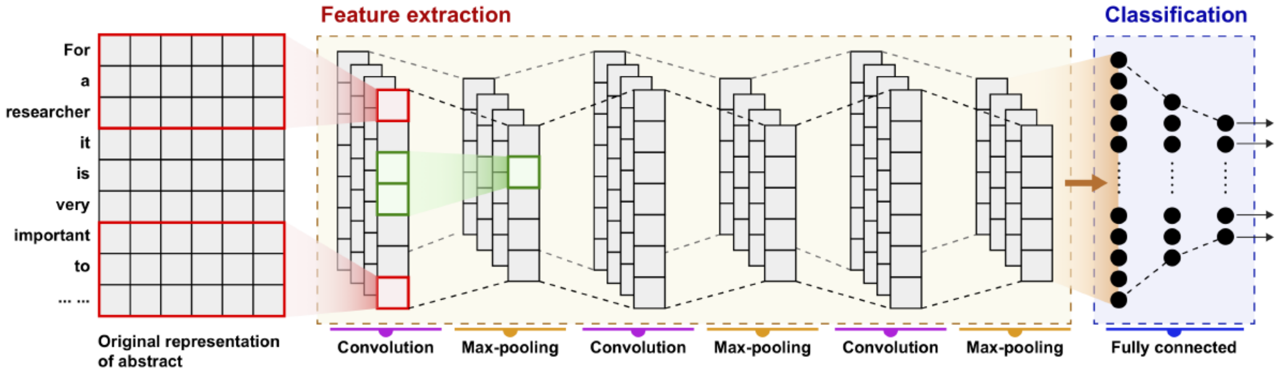
Let $\mathbf{v}_i \in \mathrm{R}^k$ be the k-dimensional word vector corresponding to the i-th word in the abstract A. An original representation of A is represented as a matrix $\mathbf{V}=\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}^T$, which is the concatenation of the words' vectors. Due to the different sizes of abstracts, we set m as the maximum count of words in an abstract. A padding operation with zeros was adopted for input with fewer than m words in an abstract and a tail truncation operation for more than m words. The vectors of words adopt pretrained vectors using word embedding and are induced from the PubMed abstracts and PubMed Central full text. The word2vec tool [29] was adopted for word embedding using the skip-gram model with a window size of h, hierarchical softmax training, and a frequent word subsampling threshold to create k-dimensional vectors. Word vectors are initialized by zeros if they are not in the pretrained vocabulary. Finally, the representation of an abstract is matrix $\mathbf{V}$ with a dimensionality of $m*k$. It was used as the input to feed to the next step. To achieve more abstract and semantic features, we adopted CNN to extract semantic information.

Figure 2 shows the structure of our deep CNN model. There are three convolutional and max-pooling layers in CNN, one fully connected layer, one hidden layer, and one softmax layer for classification. For an abstract, $A(w_1,w_2,\ldots,w_n)$ with w_i represents the i-th word and $\mathbf{v}_i \in \mathrm{R}^k$ is the k-dimensional word vector corresponding to word w_i. The abstract is represented as $\mathbf{v}_{1:m}=\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_m$ (1), where is the concatenation operator, m is the maximum length of abstracts (a scalar), and $\mathbf{v}_{i:i+j}$ refers to the vector of concatenation of the words $w_i,w_{i+1},\ldots,w_{i+j}$. The first convolutional layer performs as a one-dimensional convolution operation on sliding windows of h_1 words to produce a phrase feature. For example, a feature c_{ji} is generated from a window of words $\mathbf{v}_{i:i+h_1-1}$ by $c_{ji}=g(f_j \blacksquare \mathbf{v}_{i:i+h_1-1}+\mathbf{b}_1)$ (2). Here, $\mathbf{b}_1 \in \mathrm{R}$ is a bias term and g is a nonlinear function such as rectified linear unit (ReLu). $f_j \in \mathrm{R}^{k \times h_1}$ is the j-th convolutional kernel, whose shape is $k \times h_1$, where k is the dimension of word vectors and h_1 is the window size. This kernel is applied to each possible window of words in the abstract $\{\mathbf{v}_{1:h_1},\mathbf{v}_{2:h_1+1},\ldots,\mathbf{v}_{m-h_1+1:m}\}$ to produce a feature map $\mathbf{C}_j=[c_{j1}, c_{j2}\ldots,c_{j,m-h_1+1}]$ (3) with $\mathbf{C}_j \in \mathrm{R}^{m-h_1+1}$.

If there are r_1 convolutional kernels, then is the result of the first convolution operation on $\mathbf{V}$. The pooling operation is then carried out on $\mathbf{C}^{(1)}$. Its function is to progressively reduce the spatial size of the representation to extract the key features and reduce the number of dimensions in the network. The pooling layer operates independently on every depth slice of the input and resizes it spatially, using the max-pooling operation [30] in every two-unit window for each $\mathbf{C}_{\mathbf{j}}^{(1)}$. $\mathbf{P}^{(1)}$, described below, is the result of the max-pooling operation: (4), where j is the j-th filter of the convolutional operation (5).

The second and third convolutional and pooling layers work the same as Equations (2) and (5). Following the three convolutional and pooling operations is the fully connected layer. Here, the input is represented with a more abstract feature , where r_3 is the number of third-layer convolutional filters. The three convolutional and pooling operations indicate a phrase-level feature, a sentence-level feature, and an abstract-level feature. The algorithm of abstract embedding is listed in Textbox 1.

Figure 2. The structure of our deep convolutional neural network model.



Textbox 1. Abstract embedding algorithm.

Input:

- Embedding the abstract A to matrix $\mathbf{V}=\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}^T$
- r_t is the number of convolutional filters of layer t, where t=1, 2, 3
- h_t is the convolutional window size of layer t, where t=1, 2, 3

Output:

Procedure:

```
1.
2.  for $t$=1, 2, 3
3.      for $j$=1, 2,…$r_t$
4.          for $i$=1, 2,…$m$–$h_t$+1
5.
6.          End for
7.          for $i$=1, 2,… ($m$–$h_t$+1)/2
8.
9.          End for
10.
11.     End for
12.
13. End for
```

Softmax Classification

A fully connected softmax layer is the last layer of Pubmender. Given the training sample, A, where T is the number of possible labels, z is the class score for the sample, and the estimated probabilities $S_j\in[0,1)$ for each label $j\in\{1,2\ldots T\}$ the softmax formula is:

We trained the entire model by minimizing the cross-entropy error defined as (7), where Y is the true classification output. This is a one-hot encoding of size T, where all elements except one are 0, and one element is 1. This element marks the correct class for the data classified. We employed the optimizer Adam to learn the model parameters, which is a variant of stochastic gradient descent [31].

Results

Datasets

After data preprocessing, there were 880,165 preprocessed papers from PMC in 1130 open-access journals from Jan 2007 to Apr 2017. The "LocatorPlus ID" assigned to each journal by PMC is regarded as the classification label of a paper. We generated four data sets based on these papers. The first data set included all papers from 2007 to 2016, which was used to choose the feature representation method and train the prediction models. Papers in 2017 formed the second data set, which was used as the test set to verify Pubmender's performance. The last two datasets chose papers from publications in Elsevier and Springer from 2017, which were used to compare our Pubmender with Journal Finder and Suggester. The statistics of the first dataset are described in Table 1.

One of the biggest challenges of these datasets is that the data distribution is highly imbalanced. In the first dataset, 60 journals published more than 2000 papers, while 740 journals published fewer than 400 papers. The number of papers in "PLOS One" was 153,608, which is larger than the number in other journals, based on its extensive and comprehensive scope. "Scientific Reports" ranked second, with 37,864 papers published, while "Horticulture Research" only published 100 papers. The average paper count was 741, and 934 journals had fewer than that number.

Table 1. Details of the first dataset (Jan 2007 to Dec 2016).

Statistic	Number of journals[a]	Number of papers[b]
Size		
100≤x[c]≤400	740	157,038
400<x≤2000	330	259,676
2000<x≤10,000	55	195,426
>10,000	5	225,742
Total	1130	837,882
Maximum class size	1	153,608
Minimum class size	4	100
Average class size	N/A[d]	741

[a] This represents the total number of journals in this range.

[b] This represents the total number of papers published in all journals in this range.

[c] x represents the number of papers published in one journal.

[d] N/A: not applicable.

Parameters and Measurements

For CNN, three convolutional and three pooling operations were adopted. The pretrained word vectors generated by word2vec, available from Evex [29], were used. The window size h was 5 and threshold was 0.001. The dimension of a pretrained vector was 200, that is, k=200. The length of abstract had a fixed size m, which is the maximum number of the words that most abstracts contain. Table 2 shows the word statistic details of the papers. With the statistics, only 43,328 of 837,882 papers (5%) contained abstracts with more than 350 words. Therefore, we chose m=350. A zero-padding operation was applied for abstracts with fewer than 350 words, together with a tail-truncation operation for abstracts containing more than 350 words.

The convolutional operation parameters are listed in Table 3. The activation function adopted the ReLU function. In the pooling layer, the size of max-pooling filters was two, applied with a stride of two down samples. The parameters of the following layers (pooling layers) had the same parameter settings. Normalization and dropout strategies were used in the fully connected layer. Rate of dropout was 0.2 and L_2 normalization was adopted.

Evaluation of Recommendation Results

Toy Experiment

We designed a toy experiment to validate the deep learning method. In the first dataset, 421,168 papers were chosen from 60 journals with more than 2000 papers. The training set and test set contained 37,951 (90%) and 4,217 (10%) papers, respectively. We selected bi-directional long short-term memory (Bi-LSTM) and fastText [32] as comparison models for Pubmender. Bi-LSTM represents the recurrent neural network model with the max-pooling operation from a previous study [33]. Pretrained word vectors, generated by word2vec from a previous study [29], were used as the input original word vectors for fastText and Pubmender.

To evaluate the performance of our system, top-N accuracy was adopted as a measurement, which is defined as the probability that the expected label is in the top N predicted classes. For top-N, if the journal containing the abstract is among the top N ranked journals, the classification is correct. The symbol acc@N represents the accuracy of top-N, N=1, N=3, and N=5. The comparison of accuracy is listed in Table 4, which shows that both deep learning approaches outperformed fastText in all the three measurements. The accuracy of Bi-LSTM is nearly the same as that of Pubmender. However, the running time of Pubmender was 2660 abstracts per second, which is 78% faster than Bi-LSTM (1495 abstracts/second), and Bi-LSTM needs more memory.

Table 2. Word statistics of abstracts.

Size	Number of abstracts
20≤x[a]<50	25,499
50≤x<100	76,614
100≤x<150	139,420
150≤x<200	227,993
200≤x<250	191,156
250≤x<300	87,597
300≤x<350	46,275
x>350	43,328

[a] x denotes the number of words in the abstract.

Table 3. Hyperparameters of convolutional operation.

Convolutional layer	Convolution kernel count	Window size
First	256	3
Second	128	4
Third	96	5

Table 4. Accuracy of Bi-LSTM, fastText, and Pubmender. Italicized values indicate the best results. acc@N represents the accuracy for top-N selection.

Methods	acc@1	acc@3	acc@5
fastText	0.66	0.86	0.92
Bi-LSTM[a] (max-pooling)	0.71	0.90	0.95
Pubmender	*0.72*	*0.92*	*0.96*

[a] **Bi-LSTM:** bi-directional long short-term memory.

First Dataset Result

For the first dataset, there were 837,882 papers in 1130 journals from 2007 to 2016. The training, validation, and test sets contained 670,306 (80%), 83,788 (10%), and 83,788 (10%) randomly selected papers, respectively. The results of and comparisons with previous work are provided in Table 5.

The other two systems were from three widely used digital libraries: Association for Computing Machinery (ACM) [16], CiteSeer [16], and Microsoft Academic Search (MAS) [2]. The results from Table 5 show that Pubmender achieved the best performance. The proposed system can achieve 0.50 on acc@1 and 0.86 on acc@10. Our system improved performance by 225% over CiteSeer and MAS in terms of acc@5, and by 87% over MAS and 196% over CiteSeer in terms of acc@10. The system described by Yang and Davidson [16] used topic and writing-style information, and the system described by Medvet et al [2] used the abstracts and titles. However, our Pubmender obtained the best accuracy by using abstracts only.

To present the ability of handling imbalanced data of Pubmender, we divided the test set into four classes (tiny, small, medium, and large) according to the paper counts of different journals. From Table 6, for the tiny set, Pubmender achieved 0.27 accuracy on the acc@1 and 0.54 on acc@5, which are greater than the accuracy on acc@5 and acc@10 (in Table 5) from MAS and CiteSeer, respectively. The accuracy of the top-10 (acc@10) of a large set reached 0.98. In the paper by Medvet et al [2], 58,466 papers were partitioned almost uniformly into 300 conferences from the MAS. In CiteSeer [16], 35,020 selected papers were published across 739 venues, each of which had at least 20 papers. The average number of papers for each venue was 47. Therefore, the CiteSeer dataset is almost balanced. In contrast, the imbalance of our data, as shown in Table 1, was very critical. The sizes of classes in our dataset ranged from 100 to 153,608 papers; for example, the number of papers in “PLOS One” was 153,608, which is 270 times the average number of papers in all journals. Compared with balanced data, the classification of critically imbalanced data was a complex problem to tackle. For this problem, our model achieved satisfactory results.

Table 5. Accuracy of the classification by Pubmender and other systems. Italicized values indicate the best results. acc@N represents the accuracy for top-N selection.

Methods	Paper count	Journal count	acc@1	acc@3	acc@5	acc@10
Pubmender	837,882	1130	*0.50*	*0.71*	*0.78*	*0.86*
MAS[a] [2]	58,466	300	—[b]	—	0.24	0.46
ACM[c] [16]	172,890	2197	—	—	0.56	0.70
CiteSeer [16]	35,020	739	—	—	0.24	0.29

[a] MAS: Microsoft Academic Search.

[b] Experimental evaluation is not available.

[c] ACM: Association for Computing Machinery.

Table 6. Pubmender accuracy at top N(@N) of imbalance class data. acc@N represents the accuracy for top-N selection.

Paper count range	acc@1	acc@3	acc@5	acc@10	Paper count
Tiny	0.27	0.44	0.54	0.66	16,337
Small	0.43	0.63	0.72	0.82	26,259
medium	0.62	0.81	0.88	0.94	19,588
Large	0.66	0.91	0.96	0.98	22,579
All	0.50	0.71	0.78	0.86	84,763

Moreover, excluding accuracy, we choose precision, recall, and the F1-score as measurements. For an individual class $\mathbf{C}_i$, the assessment is defined by tp_i (true positives), fp_i (false positives), fn_i (false negatives), and tn_i (true negatives). Accuracy, precision, and recall are calculated from the counts for $\mathbf{C}_i$. Quality of the overall classification is evaluated in two ways: macro-averaging and micro-averaging. The macro-average is the average of the same measures calculated for all classes. With the sum of counts to obtain cumulative tp, fp, tn, and fn, micro-average metrics are calculated [34]. The following equations show how the desired results are individually achieved:

We listed macro-average and micro-average metrics in Table 7. Macro-averaging treats all classes equally, while micro-averaging favors bigger classes. From Table 7, it can be seen that when the number of recommended journals increases, the probability of capturing the real journal also increases. Therefore, the recall is increased step by step from top-1 to top-10. With the growth in the number of recommended journals, the number of falsely selected journals is also growing, which results in a decrease in precision. The F1-score favors a balanced view.

New Data Verification

To show the performance on new data, 42,283 papers from January 2017 to April 2017 were extracted to make further predictions. This comprised 1321 journals, some of which did not appear in the first dataset. These unseen journals increased the difficulty of prediction. The accuracy of Pubmender on the top-1, 3, 5, and 10 was 0.39, 0.61, 0.68, and 0.76, respectively. The accuracies on acc@5 and acc@10 were 183% and 162% higher than those of CiteSeer, respectively. From these results, we conclude that our proposed recommender system achieves a satisfactory result, even for new data that may not belong to the same data distribution.

Comparison with Journal Finder and Journal Suggester

Journal Finder is provided by Elsevier for recommending Elsevier journals [35]. There are 45 Elsevier journals in our dataset. Five of them with more papers were selected. The paper counts were 582 for *Medicine*, 193 for *Data in Brief*, 124 for *NeuroImage: Clinical*, 117 for *Redox Biology*, and 87 for *Preventive Medicine Reports*.

Elsevier's Journal Finder requires input of the title and abstract of submitted paper, and fields of research. The titles and abstracts were extracted from XML files and then fed into Journal Finder. We chose fields of research in "Engineering," "GeoSciences," "Life and Health Science," and "Chemistry." The results are listed in Table 8 and show that our system is much better than Journal Finder.

Table 7. Macro-average and Micro-average metrics for recommendation results.

Metrics	Macro-average			Micro-average		
	Precision	Recall	F1	Precision	Recall	F1
Top-1	0.38	0.32	0.33	0.50	0.50	0.50
Top-3	0.37	0.50	0.41	0.45	0.71	0.55
Top-5	0.35	0.59	0.42	0.42	0.78	0.55
Top-10	0.32	0.70	0.42	0.38	0.86	0.53

Table 8. Comparison between Pubmender and Journal Finder. Italicized values indicate the best results. acc@N represents the accuracy for top-N selection.

Systems	acc@1	acc@3	acc@5	acc@10
Pubmender	*0.62*	*0.75*	*0.84*	*0.90*
Journal Finder	0.05	0.12	0.13	0.21
Improvement (%)	1140	525	546	329

Table 9. Comparison between Pubmender and Journal Suggester. Italicized values indicate the best results. acc@N represents the accuracy for top-N selection.

Systems	acc@1	acc@3	acc@5	acc@10
Pubmender	*0.57*	*0.81*	*0.87*	*0.91*
Journal Suggester	0.11	0.15	0.17	0.18
Improvement (%)	418	440	412	406

Journal Suggester [36], recommends journals published by Springer. Journal Suggester also requires input of the title and abstract, and field of research. We chose “Biomedicine” as the field of research. There are 14 journals from Springer in our dataset, and seven of them were chosen for comparison based on a significant number of papers in each journal — *Cell Death & Disease*, *Malaria Journal*, *Nanoscale Research Letters*, *Nature Communications*, *Parasites & Vectors*, *Scientific Reports*, and *Trials*. Each journal chose the top 100 papers according to the size of the XML files. The results are listed in Table 9. Again, our system was much better than Journal Suggester.

Discussion

In this study, Pubmender was proposed to recommend a biomedical publishing venue to user. CNN was used to obtain the abstract representation. Our results show the performance of the system.

Principal Results

For biomedical publications, our Pubmender system is the first recommender system with word embedding and deep learning models. It achieves 87.0%, 22.9%, and 196.0% higher accuracy than recommender systems on MAS, ACM, and CiteSeer, respectively. In addition, the experiment results also revealed that the accuracy of our system was superior to that of Journal Finder and Journal Suggester. Our web service is freely available online [37].

Comparison with Prior Work

Because no paper has been published about biomedical venue recommendations, we cannot perform any exact comparison with previous work. However, some publishers provide tools to help authors choose suitable journals. We chose two tools provided by Elsevier and Springer for comparison. The first one is Journal Finder provided by Elsevier for recommending journals of Elsevier. Pubmender achieved a much higher accuracy than Journal Finder on four metrics. For example, on acc@1, the accuracy of our system reached 0.62 and Journal Finder was given an accuracy rating of 0.05, with 1140% improvement; on acc@10, the accuracy of our system reached 0.84, which is 546% higher than that of Journal Finder. Pubmender also significantly outperformed another tool, Journal Suggester.

Conclusions

In this study, we proposed a biomedical publishing venue recommender system—Pubmender. In this system, an abstract is first represented by a vector using the composition of pretrained word vectors. Subsequently, a deep CNN architecture is designed to represent and classify the submitted abstract. The original vectors are converted into more abstract feature vectors containing semantic information using deep CNNs, which overcome the sparse high-dimensional problem.

The experimental results showed that our proposed system achieves more successful performance than that of MAS, ACM, CiteSeer, Journal Finder, and Journal Suggester. Even for journals containing a small number of abstracts, the performance of Pubmender was satisfactory, because Pubmender’s high-level representation method catches more semantic and structural information from the abstract.

Acknowledgments

This work was partially supported by the US National Institutes of Health (Grant R35-GM126985), the National Natural Science Foundation of China (Grants 61602207 and 61572228), Zhuhai Premier-Discipline Enhancement Scheme (Grant 2015YXXK02), Key Technological Research Projects in Jilin Province (Grant 20190302107GX), and Guangdong Premier Key-Discipline Enhancement Scheme (Grant 2016GDYSZDXK036).

Authors' Contributions

RG, DX, and XF envisioned the study. HZ and PS performed the modeling experiments. YR performed the data preprocessing. YZ and BH built the website. XF drafted the paper; YL, RG, and DX made critical revisions. RG and DX jointly supervised this project as co-corresponding authors. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. PMC: US National Library of Medicine and National Institutes of Health. URL: https://www.ncbi.nlm.nih.gov/pmc/ [accessed 2018-11-26] [WebCite Cache ID 74E8cB679]
2. Medvet E, Bartoli A, Piccinin G. Publication Venue Recommendation Based on Paper Abstract. In: Proceedings of the IEEE 26th International Conference on Tools with Artificial Intelligence. 2014 Nov Presented at: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence; November 10-12, 2014; Limassol, Cyprus p. 1004-1010. [doi: 10.1109/ICTAI.2014.152]
3. Bell R, Koren Y. Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights. In: Proceedings of the Seventh IEEE International Conference on Data Mining 2007. 2007 Presented at: Seventh IEEE International Conference on Data Mining 2007; Oct 28-31, 2007; Omaha, NE p. 43-52. [doi: 10.1109/ICDM.2007.90]
4. Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 2003 Jan;7(1):76-80. [doi: 10.1109/MIC.2003.1167344]
5. Mobasher B, Dai H, Luo T, Nakagawa M. Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. Data Min Knowl Discov 2002 Jan;6(1):61-82. [doi: 10.1023/A:1013232803866]
6. Francesco R, Rokach L, Shapira B. Introduction to Recommender Systems Handbook. In: Recommender Systems Handbook. Boston, MA: Springer; 2019.
7. Brusilovsky P, Kobsa A, Nejdl W. Content-Based Recommendation Systems. In: The Adaptive Web: Methods And Strategies Of Web Personalization (lecture Notes In Computer Science). Berlin, Heidelberg, Germany: Springer; 2019.
8. Balabanović M, Shoham Y. Fab: content-based, collaborative recommendation. Commun ACM 1997 Mar;40(3):66-72. [doi: 10.1145/245108.245124]
9. Sarwar B, Karypis G, Konstan J, Riedl J. Item-based Collaborative Filtering Recommendation Algorithms. In: Proceedings of the 10th international conference on World Wide Web. 2001 Presented at: The 10th international conference on World Wide Web; May 01-05, 2001; Hong Kong p. 285-295. [doi: 10.1145/371920.372071]
10. Salter J, Antonopoulos N. CinemaScreen Recommender Agent: Combining Collaborative and Content-Based Filtering. IEEE Intell. Syst 2006 Jan;21(1):35-41. [doi: 10.1109/MIS.2006.4]
11. Wang H, Wang N, Yeung D. Collaborative Deep Learning for Recommender Systems. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2015 Aug Presented at: The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015 Aug 10-13; Sydney, NSW, Australia p. 1235-1244. [doi: 10.1145/2783258.2783273]
12. Wang DH, Liang YC, Xu D, Feng XY, Guan RC. A content-based recommender system for computer science publications. Knowledge-Based Systems 2018 Oct;157:1-9. [doi: 10.1016/j.knosys.2018.05.001]
13. Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis. J Med Internet Res 2016 Sep 21;18(9):e251 [FREE Full text] [doi: 10.2196/jmir.6240] [Medline: 27655225]
14. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J Mach Learn Res 2003 Jan;3:993-1022 [FREE Full text]
15. Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. The American Statistician 1992 Aug;46(3):175-185. [doi: 10.1080/00031305.1992.10475879]
16. Yang ZH, Davison BD. Venue Recommendation: Submitting Your Paper with Style. In: Proceedings of the 11th International Conference on Machine Learning and Applications. 2012 Presented at: The 11th International Conference on Machine Learning and Applications; 2012 Dec 12-15; Boca Raton, FL p. 681-686. [doi: 10.1109/ICMLA.2012.127]

17. Küçüktunç O, Saule E, Kaya K, Çatalyürek Ü. Arxiv.org. 2012 May 05. Recommendation on Academic Networks using Direction Aware Citation Analysis URL: https://arxiv.org/pdf/1205.1143.pdf [accessed 2019-05-02] [WebCite Cache ID 784Anxdv1]
18. Luong H, Huynh T, Gauch S, Do L, Hoang K. Publication Venue Recommendation Using Author Network's Publication History. In: Pan JS, Chen SM, Nguyen NT, editors. Intelligent Information and Database Systems. ACIIDS 2012.Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2012:426-435.
19. Beel J, Gipp B, Langer S, Breitinger C. Research-paper recommender systems: a literature survey. Int J Digit Libr 2016 Nov;17(4):305-338. [doi: 10.1007/s00799-015-0156-0]
20. Zhao Z, Huang L, Wang C, Lai J, Yu P. Low-Rank and Sparse Matrix Completion for Recommendation. In: Proceedings of International Conference on Neural Information Processing. 2017 Presented at: International Conference on Neural Information Processing; Nov 14-18, 2017; Guangzhou, China p. 3-13.
21. Luong H, Huynh T, Gauch S, Hoang K. Exploiting social networks for publication venue recommendations. In: Proceedings of the International Conference on Knowledge Discovery and Information Retrieval. 2012 Presented at: 4th International Conference on Knowledge Discovery and Information Retrieval; Oct 4-7, 2012; Barcelona, Spain p. 239-245.
22. Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann Machines for Collaborative Filtering. In: Proceedings of the 24th International Conference on Machine Learning. 2007 Presented at: The 24th International Conference on Machine Learning; Jun 20-24, 2007; Corvalis, Oregon p. 791-798. [doi: 10.1145/1273496.1273596]
23. McAuley J, Targett C, Shi Q, van den Hengel A. Image-Based Recommendations on StylesSubstitutes. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2015 Presented at: The 38th International ACM SIGIR Conference on Research and Development in Information Retrieval; Aug 09-13, 2015; Santiago, Chile p. 43-52. [doi: 10.1145/2766462.2767755]
24. van den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation. In: Proceedings of the 27th Conference on Neural Information Processing Systems. 2013 Presented at: the 27th Conference on Neural Information Processing Systems; 2013 Dec 05-10; Lake Tahoe, CA, USA p. 2643-2651 URL: http://papers.nips.cc/paper/5004-deep-content-based-music-recommendation.pdf
25. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research 2011 Aug;12:2493-2537 [FREE Full text]
26. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013 Aug;35(8):1798-1828. [doi: 10.1109/TPAMI.2013.50] [Medline: 23787338]
27. NCBI. FTP Service URL: https://www.ncbi.nlm.nih.gov/pmc/tools/ftp/ [accessed 2019-05-02] [WebCite Cache ID 784GfhZXX]
28. Natural Language Toolkit. URL: https://www.nltk.org/ [accessed 2018-11-26] [WebCite Cache ID 74E8QTItd]
29. EVEX. URL: http://evexdb.org/pmresources/vec-space-models/ [accessed 2018-11-26] [WebCite Cache ID 74E8pH7A4]
30. Hu B, Lu Z, Li H, Chen Q. Convolutional Neural Network Architectures for Matching Natural Language Sentences. In: Proceeding of 27th Conference on Neural Information Processing Systems. 2014 Presented at: 27th Conference on Neural Information Processing Systems; Dec 8-13, 2014; Montreal, Canada p. 2042-2050 URL: http://papers.nips.cc/paper/5550-convolutional-neural-network-architectures-for-matching-natural-language-sentences.pdf
31. Kingma D, Ba J. ArXiv.org. 2014 Dec 22. Adam: A Method for Stochastic Optimization URL: https://arxiv.org/abs/1412.6980 [accessed 2019-05-02] [WebCite Cache ID 7846ZaGfU]
32. GitHub. 2018. FastText URL: https://github.com/facebookresearch/fastText [accessed 2018-11-27] [WebCite Cache ID 74E8vJ40r]
33. Conneau A, Kiela D, Schwenk H, Barrault L, Bordes A. Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. In: Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: EMNLP 2017: Conference on Empirical Methods in Natural Language Processing; Sep 7-11, 2017; Copenhagen, Denmark p. 670-680 URL: https://www.aclweb.org/anthology/D17-1070
34. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Information Processing & Management 2009 Jul;45(4):427-437. [doi: 10.1016/j.ipm.2009.03.002]
35. Elsevier. Journal Finder URL: https://journalfinder.elsevier.com/ [accessed 2018-11-26] [WebCite Cache ID 74E93AryP]
36. Springer Nature. Journal Suggester URL: https://journalsuggester.springer.com/ [accessed 2018-11-26] [WebCite Cache ID 74E9DA7tZ]
37. Pubmender: Deep Learning Based Recommender System for Biomedical Publication Venue. URL: https://www.keaml.cn:8081/ [accessed 2019-05-07] [WebCite Cache ID 78CMVBjkL]

Abbreviations

ACM: Association for Computing Machinery
Bi-LSTM: bi-directional long short-term memory
CNN: convolutional neural network
MAS: Microsoft Academic Search

PMC: PubMed Central
ReLu: rectified linear unit

Edited by CL Parra-Calderón; submitted 28.11.18; peer-reviewed by C Lin, L Zhang, J Brenas; comments to author 31.03.19; accepted 19.04.19; published 24.05.19.

Please cite as:
Feng X, Zhang H, Ren Y, Shang P, Zhu Y, Liang Y, Guan R, Xu D
The Deep Learning–Based Recommender System “Pubmender” for Choosing a Biomedical Publication Venue: Development and Validation Study
J Med Internet Res 2019;21(5):e12957
URL: http://www.jmir.org/2019/5/e12957/
doi:10.2196/12957
PMID:31127715

Review

Design Choices and Trade-Offs in Health Care Blockchain Implementations: Systematic Review

Odhran O'Donoghue[1]; Anuraag A Vazirani[1]; David Brindley[2], MEng, MSc, DPhil, FRSA; Edward Meinert[2,3], MA, MSc, MBA, MPA, PhD, EUR ING, CEng FBCS

[1]Medical School, Medical Sciences Division, University of Oxford, Oxford, United Kingdom

[2]Healthcare Translation Research Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom

[3]Global Digital Health Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Edward Meinert, MA, MSc, MBA, MPA, PhD, EUR ING, CEng FBCS
Healthcare Translation Research Group
Department of Paediatrics
University of Oxford
Level 2, Children's Hospital
John Radcliffe Hospital
Oxford, OX3 9DU
United Kingdom
Phone: 44 7824446808
Email: edward.meinert@paediatrics.ox.ac.uk

Abstract

Background: A blockchain is a list of records that uses cryptography to make stored data immutable; their use has recently been proposed for electronic medical record (EMR) systems. This paper details a systematic review of trade-offs in blockchain technologies that are relevant to EMRs. Trade-offs are defined as “a compromise between two desirable but incompatible features.”

Objective: This review’s primary research question was: “What are the trade-offs involved in different blockchain designs that are relevant to the creation of blockchain-based electronic medical records systems?”

Methods: Seven databases were systematically searched for relevant articles using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Papers published from January 1, 2017 to June 15, 2018 were selected. Quality assessments of papers were performed using the Risk Of Bias In Non-randomized Studies—of Interventions (ROBINS-I) tool and the Critical Assessment Skills Programme (CASP) tool. Database searches identified 2885 articles, of which 15 were ultimately included for analysis.

Results: A total of 17 trade-offs were identified impacting the design, development, and implementation of blockchain systems; these trade-offs are organized into themes, including business, application, data, and technology architecture.

Conclusions: The key findings concluded the following: (1) multiple trade-offs can be managed adaptively to improve EMR utility; (2) multiple trade-offs involve improving the security of blockchain systems at the cost of other features, meaning EMR efficacy highly depends on data protection standards; and (3) multiple trade-offs result in improved blockchain scalability. Consideration of these trade-offs will be important to the specific environment in which electronic medical records are being developed. This review also uses its findings to suggest useful design choices for a hypothetical National Health Service blockchain.

International Registered Report Identifier (IRRID): RR2-10.2196/10994

(J Med Internet Res 2019;21(5):e12426) doi:10.2196/12426

KEYWORDS

blockchain; interoperability; distributed ledger technology; scalability; health information exchange

Introduction

Background

Blockchains contain *blocks* of data ordered chronologically. Each block is linked to previous blocks via cryptography. If previous blocks are edited, these cryptographic linkers will no longer match, making blockchains resistant to tampering. Blockchains originated in 1991 [1], but the first major adoption was 2008's Bitcoin [2]. For Bitcoin's creator, the use case for blockchains is to create data ledgers that can be trusted without centralized systems: updates to its blockchain have to be agreed on by its users through consensus processes. Bitcoin's use of blockchain works as described in the following sections.

Block Creation

Transactions between individuals are recorded. These data are *encrypted* via hash functions [3], which convert data into alphanumeric strings known as hashes (see Figure 1). The data inputted cannot be decrypted from these hashes. Each transaction is hashed and a summary hash, known as a *Merkle root*, is generated for each block of data [4] (see Figure 2).

Figure 1. A: hash functions convert data into fixed-length strings (hashes). B: similar but unidentical data have very different hashes. C: identical data have identical hashes, allowing for verification. D: all lengths of data produce fixed-length hashes.

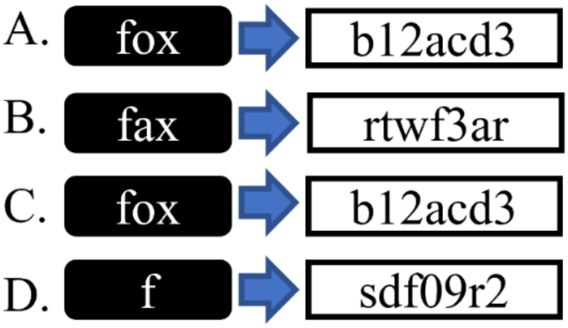
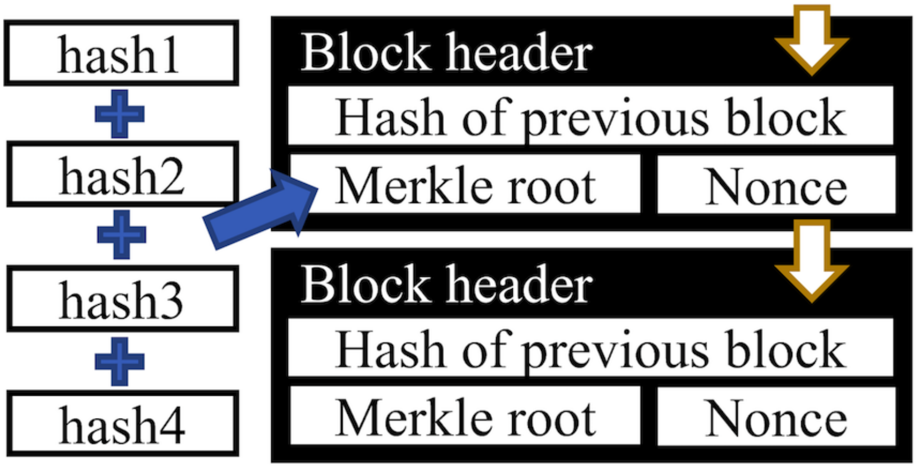


Figure 2. The hash of the previous block, the Merkle root, and the Nonce are combined into a single hash. This is included in the next block header, linking the two blocks.



Adding a New Block to the Blockchain

To add new blocks, a majority of users must want the same new block (ie, consensus). However, malicious users could create multiple accounts to gain control. To avoid this, Bitcoin uses proof of work (PoW); this requires the users' processing power for consensus, which will not increase for a user owning multiple accounts. PoW works when the network randomly generates a *target hash*. To add a block to the chain, users must find a number that, when added to the Merkle root and hashed, equals the target hash. This is known as *mining*.

Mining is computationally difficult but finding the right number (ie, a nonce) is rewarded with bitcoins. If miners use their combined computational power to mine the same block and share the rewards after finding the nonce, they are more likely to make money than if mining alone. This encourages consensus.

A new block is formed on the chain when a correct nonce is found. The hashes of the previous block are combined into the Merkle root of the new block, cryptographically linking the two blocks (see Figure 1).

Blockchain Security

Were someone to try to maliciously edit a transaction, they would have to mine a new nonce. This is because a different transaction value needs a different nonce to get the target hash. Furthermore, edits would need to be made to subsequent blocks to ensure chain coherence. For malicious users to find nonces faster than the rate at which nonces are found on the *true* chain is not feasible without obtaining majority control of a network: this is known as a 51% attack [5].

The Use of Blockchain as an Electronic Medical Record

As described in the 2019 Topol Review [6] and the 2019 National Health Service (NHS) Long-Term Plan [7], there is a drive to adopt electronic medical records (EMRs) throughout the NHS. The reports posit that such systems are needed for personalized long-term care and for obtaining the large-scale datasets necessary for predictive health modelling. Current strategic objectives are to operate multiple EMRs that are interoperable through standards laid out in the Local Health and Care Record Exemplars [8]. Blockchains have been proposed as an EMR platform and could be considered by the NHS as they may have some advantages over classical EMRs: distributed ledgers require no manual reconciliation of data between different providers [9]. This is a common issue with current EMRs, as patients can encounter many health care providers. Secondly, the ledger includes an audit trail of all changes. This helps to ensure EMR integrity and prevent data falsification [9]. Distributed systems would also allow interoperability to shift from a provider-driven model to a patient-driven model [10], in which patients are empowered to add to and exchange their own health data. An example of a blockchain EMR is MedRec [11], which is built on the Ethereum blockchain. Ethereum supports *smart contracts*, which are codes that self-execute when certain criteria are met [12]. MedRec only stores metadata on the blockchain and uses smart contracts to let certified users access full EMR data stored off-chain. This improves scalability, but a trade-off is that EMR data may be harder to audit.

Trade-Offs in Blockchain Design

While the principle of cryptographically linked blocks underpins a blockchain, many design choices can be made [13]. However, these choices may result in trade-offs [14]: trade-offs are defined as “a compromise between two desirable but incompatible features.” An example of where a trade-off has been made in the design of a blockchain-based EMR is in the creation of version 2.0 of MedRec, where the amount of information stored on the MedRec blockchain was reduced to improve scalability. Understanding the potential trade-offs in blockchain design is important, as it better informs the development of blockchain-based EMRs. Historic failings in the adoption of UK EMRs [15] indicate that new EMR technology should be investigated thoroughly before adoption. Therefore, how design trade-offs affect the suitability of a blockchain for EMR use should be explored. This review aims to improve the understanding of NHS clinicians and policy makers interested in adopting a blockchain EMR and to inform blockchain developers of design trade-offs relevant to EMRs.

Methods

Formulation of Research Question

This review’s research question was derived from Meinert et al’s review protocol [16]. In comparison to Meinert et al’s protocol, this review’s research question has three modifications: (1) to prevent the scope of the review from becoming too broad, the research question only considers blockchain design and does not consider blockchain implementation strategies and frameworks; (2) the review’s research question specifically focuses on trade-offs in blockchain design, as a research deficit in this area has been identified [14]; and (3) the review’s research question does not limit research domains to only privacy, efficiency, interoperability, and scalability.

Research Question

What are the trade-offs involved in different blockchain designs that are relevant to the creation of blockchain-based electronic medical records systems? This research question is based on the following definitions, which are sorted into Population, Intervention, Comparison, and Outcome (PICO) criteria (see Table 1).

Search Strategy and Inclusion Criteria

The selection process of this review is demonstrated using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (see Figure 3). PubMed, MEDLINE, Embase, Web of Science, and Scopus databases were searched. Papers published from January 1, 2017 to June 15, 2018 were selected. Search string creation (see Multimedia Appendix 1) was aided by the Nuffield Orthopaedic Centre Library. Inclusion criteria for database searches were based on PICO criteria (see Table 1). “Blockchain” (*Population*) was used in all search strings. This was combined with the term “trade-off” (*Outcome*) and variants thereof. “Blockchain” was then combined with “design” (*Intervention*) and variants thereof. “Blockchain” was also combined with “Electronic medical record” (*Comparison*) and variants thereof. Variants were found, in part, with Medical Subject Headings (MeSH) terms. Screening criteria based on key phrases were used to remove papers (see Table 2). Abstracts of remaining papers were screened to determine suitability. The PICO criteria formed the paper inclusion criteria. If abstracts contained PICO criteria or did not contain sufficient information, articles were read in full. A total of 90 articles were read in full to determine suitability. Following peer review from Institute of Electrical and Electronics Engineers (IEEE) Access, a search of Springer Link and IEEE databases was requested to find more nonmedical perspectives on blockchain design. Papers were searched that were published between January 1, 2017 and December 5, 2018. In this second search, 158 articles were read in full to determine suitability. Due to time limits, only papers written before the initial June 15, 2018 deadline were included in the qualitative synthesis. A total of 2885 papers were identified. After screening and duplicate removal, 248 articles remained, all of which were assessed for eligibility by analyzing their full text. A total of 233 articles were removed: the reasons for removal are outlined in Figure 3.

Table 1. PICO (Population, Intervention, Comparison, and Outcome) criteria, terms, and definitions.

PICO criterion	Term	Definition
Population	Blockchain	A growing list of records, called blocks, which are linked using cryptography
Intervention	Design	A decision about object function with a specific purpose in mind
Comparison	Electronic medical record	A systematized digital collection and storage of patient data
Outcome	Trade-off	A compromise between two desirable but incompatible features

Figure 3. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the search strategy.

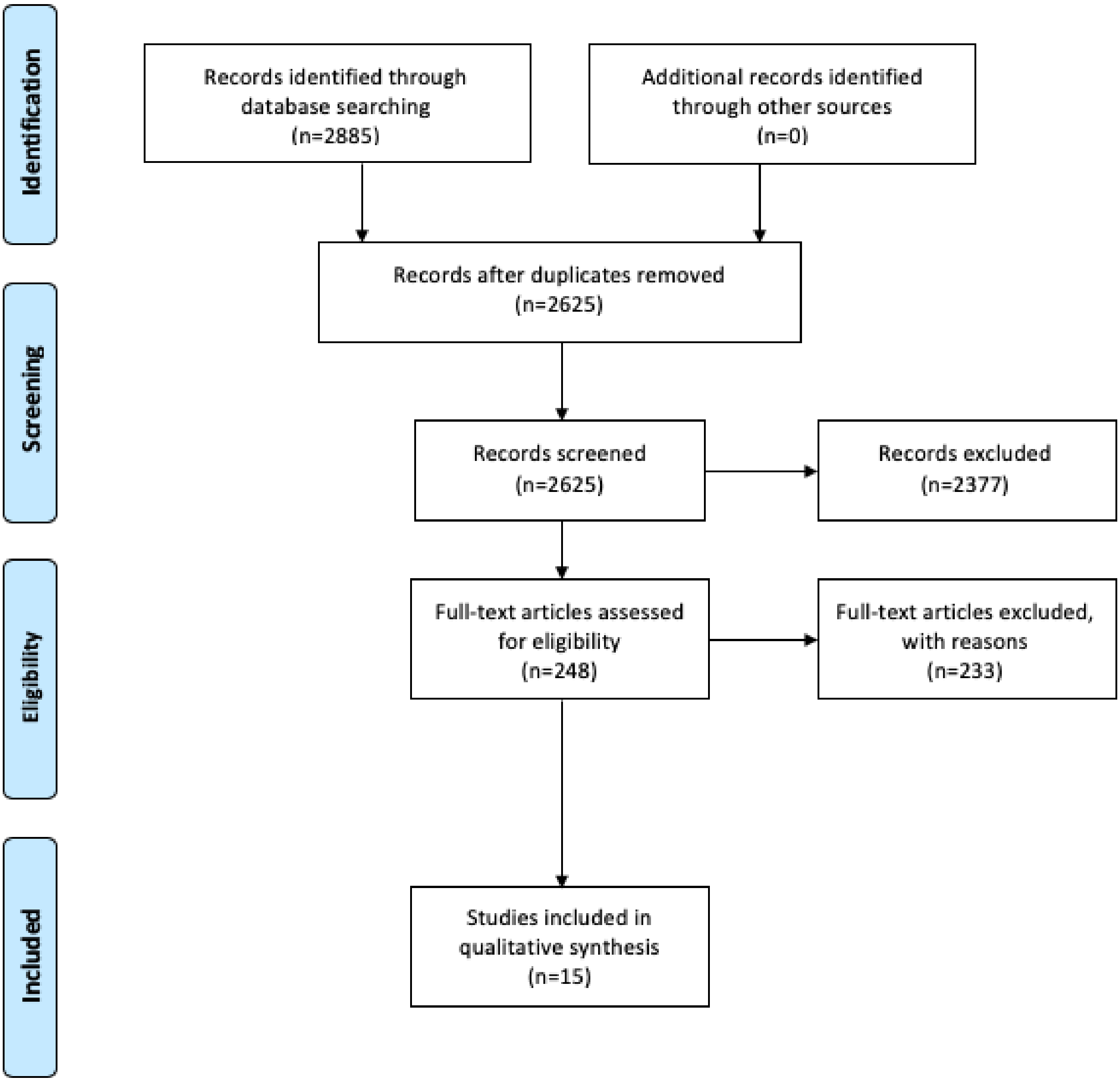


Table 2. Screening criteria used to remove papers.

Criterion number	Screening criteria	Justification	Removed in pre-peer-review search	Removed in post-peer-review search
1	Papers not containing abstracts	The papers cannot be analyzed at the abstract-screening stage and were therefore excluded.	75	0
2	Papers from 2013-2016	As an emerging technology, ideas about blockchain's possibilities and limitations are often changing. Older papers were not included, as they may contain information that does not reflect the current state of blockchain technologies. Papers before 2017 were chosen to be excluded, as 2017 was a year characterized by a significant change in blockchain valuation and regulation [26].	328	52
3	Full conference proceedings	The relevant full papers in the conference proceedings should have been identified as individual papers in database searches and included in the 1808 papers.	63	0
4	Duplicate titles	These articles contain repeated information.	125	2
5	Does not contain "block" in abstract	Articles that do not mention "block" in their abstract were deemed unlikely to be focused discussions of blockchain technologies and were therefore not included.	419	89
6	Does not contain "design" in abstract	As a PICO[a] term, a focus on blockchain design was essential to the papers. No alternative MeSH[b] terms for "design" were identified, so an absence of "design" in abstracts was used to filter out papers.	590	486
7	Title contains "Bitcoin" and not "block"	Papers that mentioned "Bitcoin" in their title and not "blockchain" were deemed to be too focused on the cryptocurrency to provide useful information concerning principles relevant to blockchain design.	12	6
8	Title contains "IoT" or "Internet of things"	Many papers contained a reference to IoT[c] devices. The abstracts of these papers contained little-to-no relevant information about blockchains. This exclusion criterion was thus used to remove papers.	15	24
9	Title contains "finance"	It was a concern that papers relating to financial elements of blockchains would be too focused on the economic aspects of cryptocurrency. As the intent of this research is to examine the technical, not economical, aspects of blockchains, these papers were removed.	15	2

[a] PICO: Population, Intervention, Comparison, and Outcome.

[b] MeSH: Medical Subject Headings.

[c] IoT: internet of things.

Assessment of Methodological Quality

The Risk Of Bias In Non-randomized Studies—of Interventions (ROBINS-I) framework was used to assess bias in the 15 remaining papers [17] (see Multimedia Appendix 2). For eight papers lacking experimental data, ROBINS-I was deemed unsuitable, and a qualitative medical data assessment tool developed by the Critical Assessment Skills Programme (CASP) was used instead [18] (see Multimedia Appendix 3). ROBINS-I identified a low risk of bias in six papers [19-24] and a serious risk of bias in one paper [25]: this paper was removed. Thus, trade-offs in 14 papers are discussed.

Data Synthesis

Only narrative data synthesis was appropriate due to study heterogeneity. To structure this, an enterprise architecture framework (ie, a set of principles used to guide enterprise design, planning, and implementation) was used. The Open Group Architecture Framework (TOGAF) [27] was chosen, as the framework is in use by software industry leaders [28]. TOGAF divides enterprise architecture into four domains: Technology (ie, the hardware underpinning programs), Data (ie, choices made in data formatting), Application (ie, software applications and procedures), and Business (ie, decisions about the business structure that relate to enterprise architecture development).

Results

Technology Architecture Trade-Offs

Current Implementation Versus Future Proofing

A critical technology architecture trade-off was current implementation versus future proofing [29]. The advent of quantum computing may compromise current blockchain cryptography. Quantum-resistant cryptography is being developed—lattice problems are being considered for this role

[30]—but research is still ongoing. This is a trade-off for hospital systems to consider: to invest in blockchain systems designed for today's computers or to wait until quantum-resistant cryptography is developed. Such a choice will depend on a health care provider's ability to update its system. If the provider is large and decentralized, it may struggle to simultaneously switch over to a new blockchain; initially investing in a postquantum blockchain may be more sensible, even if it delays implementation.

Stale Block Rate Versus Network Delay

Another trade-off is stale block rate versus network delay [23]. At any time, a miner may successfully find a nonce and broadcast it across the network so that miners can begin working on the next block. Delays in the network result in miners wasting computational power by continuing to mine a block that has already been solved, which is known as a *stale block*. Simulations in a study by Chen et al [23] show that the stale block rate correlates with network delay. Reducing delay involves increasing network infrastructure provision. Thus, decisions about the network infrastructure provided for EMRs will impact blockchain performance. An alternative detailed below is to increase the block window, but this also has trade-offs.

Data Architecture Trade-Offs

Block Window: Computation Versus Communication

One data architecture trade-off is block window: computation versus communication [31]. Block window is a measure of how many blocks are created per minute. Smaller windows allow for faster transaction confirmation but more network bandwidth is required, as a larger number of blocks is broadcast through the network. While larger windows will reduce the stale block rate, this would increase transaction confirmation times, which is undesirable in EMR medical emergency scenarios.

Data On-Chain Versus Off-Chain

Another data architecture trade-off is data on-chain versus off-chain [32]. Smaller data packets on-chain result in smaller storage costs but less data are available on-chain to audit. With a blockchain EMR, storing all patient data on-chain is likely infeasible: the International Data Corporation estimates that medical data will total 2314 exabytes by 2020 [33]. Duplication of data at such scales would be costly for health care providers. Systems that store patient data on-chain also require very strict security. Storing only metadata (ie, access permissions and edit history) is likely to be more suitable. If patient data were to be placed on-chain, unprocessed files should be avoided; raw genomic data, for instance, can be in excess of .1 TB per genome [34].

Block Size Versus Throughput

Block size versus throughput [20] is another data architecture trade-off. A blockchain network's maximum transaction throughput is calculated as maximum block size (ie, how many transactions can be stored in a block) divided by block window. Thus, a large block size is desired. If the maximum block size is too small, there is also a risk that many transactions will not be processed, as blocks will not have enough space to store all the transaction requests being made on the network. However, making the maximum block size too large will inadvertently reduce network throughput, as it will take longer to transmit blocks across a network for mining and validation. Both ends of this trade-off can delay access to EMRs in emergency situations. To maximize the throughput in a network, simulations should be run with different maximum block sizes, as per the study by Xin et al [20], to identify an optimum size. An adaptive block size may also be considered: at times of the day where there are high volumes of transactions, block size can be increased to prevent rejection of transaction requests. Outside of this period, block size can be reduced to increase network throughput. An adaptive system like this is relevant to medicine: not all types of professionals are active 24/7, which may result in times of the day where transaction requests become less frequent.

Read and Write Performance Versus Scalability

An additional data architecture trade-off is read and write performance versus scalability [19,22]. Blockchains have different options for holding edit data: data can be held in random access memory (RAM) or nonvolatile memory. RAM is generally faster to access than nonvolatile memory but has less space. Often on blockchains, recent edits are stored in RAM while remaining data are stored on a hard disk. Reducing the number of recent edits stored in RAM reduces read and write performance, delaying updates to the blockchain. However, it allows more data to be stored on nonvolatile memory, making systems more scalable. Given the likely growth in the size of health care data [33], ensuring scalability of blockchain EMRs should be a priority.

Method of Information Storage: Scalability Versus Immutability

One final data architecture trade-off is method of information storage: scalability versus immutability [35]. There are different ways to store data on blockchains. Three possibilities to do this for EMRs are as follows: (1) Mirroring: information is converted into hashes, and the hashes are stored on-chain; (2) Digital records: unhashed human-readable information is stored directly on-chain; and (3) Tokenization: information is treated as a token with value, and tokens are exchanged like currency. Mirroring EMRs (eg, those in use by the Estonian government [36]) are space efficient (ie, full EMRs are converted into single hash values) and can show when records stored off-chain have been tampered with, as hashes will no longer match. However, mirroring does not protect the original records through decentralized storage. Digital records guarantee immutability but are less space-efficient than mirroring. Tokenization also permits immutability, but it assumes that data is principally transactional in nature. This may not be true of many situations relevant to EMRs.

Application Architecture Trade-Offs

Expressibility in Blockchain Language

One application architecture trade-off is expressibility in blockchain language [19,22]. Bitcoin's scripting language has low functionality, making it difficult to exploit. Ethereum instead uses a Turing-complete scripting language (ie, it can compute anything that is computable given enough resources). While this increases versatility, flawed code can be submitted on Ethereum. An example was Ethereum's Decentralized Autonomous Organization (DAO): designed to operate as an investment fund, an exploit was discovered that allowed hackers to steal US $50,000,000 from the DAO [37]. A blockchain EMR should consider using a low-functionality language to avoid security being compromised by faulty code. However, the language should ideally be expressive enough to permit new types of hospital data to be integrated and to allow for smart contracts to be established.

Errors on the Blockchain: Filtering and Expiration

Another application architecture trade-off includes errors on the blockchain: filtering and expiration [38]. Instead of limiting script functionality to stop bugs, automated auditing could be used to identify bugs. However, automated auditing cannot guarantee that bugs will not end up on the chain and could increase network latency. A fail-safe mechanism is to give scripts expiration dates (ie, they will stop working after a certain amount of time). However, this could create confusion in EMR systems, with access to medical data being lost at critical moments.

Confirmation Blocks: Confidence Versus Speed

An additional application architecture trade-off includes confirmation blocks: confidence versus speed [21]. Honest and malicious miners may at the same time be working on different blocks to add to the blockchain. The probability of malicious miners creating a long *branch* consisting of multiple blocks is typically low. For this reason, blockchains often include *confirmation blocks* as a safety feature: a fixed number of confirmed blocks must proceed with a block in the chain before its contained transactions are performed. This causes latency between when a block is mined and when a transaction is performed. If more confirmation blocks are required, it reduces the chance of malicious transactions being processed but increases transaction latency. In a medical emergency, this delay could be critical. This trade-off can be adaptively managed; certain requests could require fewer confirmation blocks, allowing for emergency access to essential information.

Zero-Knowledge Proof: Security Versus Scalability

Zero-knowledge proof: security versus scalability [39] is another application architecture trade-off. Zerocoin is a blockchain that uses zero-knowledge proof (ZKP). ZKP hides a transaction's origin, destination, and content, but still allows transfers to be immutable. For EMRs, ZKP permits enhanced privacy; patients would be able to confirm EMR details without revealing that their EMR was queried on the blockchain. This potentially allows for sensitive exchanges to occur on public blockchains. While auditability is a concern for ZKP, Naganuma et al [39] describe a mechanism that permits designated auditors to audit information from ZKP transactions while preventing others from obtaining it. ZKP's disadvantage is its computational intensity: ZKP requires multiple rounds of communication between the sender and the receiver, with the certainty of truth increasing with every round of communication. This increases network delay and reduces scalability.

Blockchain Filtering: Safety Versus Flexibility

Another application architecture trade-off is blockchain filtering: safety versus flexibility [24]. Misconduct occurs on blockchains: Bitcoin contains links to pornography on its immutable ledger [40] and INTERPOL notes that chain updates may be used to inject malware [41]. Avoiding misconduct is essential on a blockchain EMR. One solution is to set a size limit on content added to the chain, as malware and pornography typically require large data packets. However, this could severely limit EMR functionality. An alternative is a filter system, but this could be circumvented and false positives may also occur, preventing the submission of medical data. While using human auditors can ensure accuracy, it could create a human bottleneck and could be used to attack the blockchain. To prevent misconduct, a combination of measures is likely to be the best solution.

Consensus: Byzantine Fault Tolerance Versus Non-Byzantine Fault Tolerance

Consensus: Byzantine Fault Tolerance (BFT) versus non-BFT [42] is another application architecture trade-off. A network of actors making a consensus decision is considered *Byzantine Fault Tolerant* if it can achieve consensus when some dishonest actors are present. Two of the properties that define BFT are *Agreement* (ie, all honest actors choose the same block) and *Termination* (ie, all honest actors eventually choose a block). The internet is an asynchronous network (ie, there is no guarantee that a message is delivered in a known time period). In asynchronous networks, BFT is mathematically proven to be impossible [43]: agreement and termination cannot be simultaneously guaranteed. Consensus protocols that require agreement between all honest actors may never terminate. In networks with only a few nodes, *weak* synchrony can be assumed and algorithms such as Practical Byzantine Fault Tolerance (PBFT) [44] allow consensus in these systems. However, PBFT fails in large networks as it requires heavy internode communication. Typical blockchain consensus algorithms (eg, PoW and proof of stake) instead use randomization systems (eg, mining a random nonce) to probabilistically approximate BFT. Such systems guarantee that transactions are processed, but transactions submitted by dishonest actors can be validated. On a blockchain EMR, this could result in data being obtained by malicious actors.

Consensus: Agreement Versus Termination

An additional application architecture trade-off is consensus: agreement versus termination [42]. For systems using probabilistic consensus, a trade-off can be made between agreement and termination; applications can have a *time-out* period, after which they consider a transaction to be valid even if full blockchain agreement has not yet been reached. Extending the time-out period delays transaction validation but results in fewer invalid transactions being processed.

Transparency Versus Privacy

A final application architecture trade-off is transparency versus privacy [45]. Bhaskaran et al make the case that blockchains involve trade-offs between preserving privacy and promoting transparency. While privacy is paramount in EMRs, there are situations where transparency should be considered. An example of this trade-off would be in verification: if a patient verifies their patient status on the blockchain with a provider, such verification could be shared with other health care providers to avoid a patient repeating the process. This transparency can improve efficiency, but information linking the patient to the original provider sacrifices privacy. While Bhaskaran et al propose a solution to this specific problem that makes use of smart contracts [45], there are more general considerations to be made about when transparency that benefits providers may create potential privacy issues for patients.

Business Architecture Trade-Offs

Control Versus Points of Failure

One business architecture trade-off is control versus points of failure [29]. Blockchain consensus can be determined by different actors. Options include the following: public (ie, consensus is determined by all network nodes), private (ie, a single organization manages blockchain updates), and consortium (ie, blockchain updates are determined by preselected validator nodes). In these systems, there is a trade-off between control and points of failure; in a private blockchain, it is easy to change settings, but a single point of failure means that if the organization is compromised, so is the blockchain. At the other extreme, a public blockchain has consensus verified by the entire network, but changing network properties requires all users to migrate to a new system. Consortium systems are an intermediate: these may be the most practical for EMRs, as there are multiple trusted authorities and businesses that could act as these preselected validator nodes (eg, hospital groups and health insurance companies).

Fraction of Consortium: Time Versus Trust

Another business architecture trade-off is fraction of consortium: time versus trust [29]. Alhadhrami et al suggest a hypothetical model where a fraction of validator nodes in a consortium blockchain must sign a transaction for it to be accepted. For an EMR consortium blockchain, choices should be made about the number of nodes that are needed to sign a transaction for it to be validated. More nodes verifying a transaction increases trustworthiness but results in longer verification times. In emergency medical situations, this delay could be critical. If the organizations involved in the consortium are seen to have high security, the fraction of consortium required can be kept low. Furthermore, the fraction of consortium required can be adaptively managed based on how time-critical the patient data are and the nature of the data being requested.

Discussion

Principal Findings

At the time of writing, this was the first manuscript identified in literature searches that systematically reviews trade-offs in blockchain design that are relevant to EMRs. While evidence was chosen based on relevance to EMRs, the trade-offs discovered are relevant to blockchain design in other disciplines and industries. Trade-offs were demonstrated with mathematical proofs, real-world examples, and simulations of blockchain networks. Some trade-offs were not directly evidenced, but these trade-offs were logical and comparable to real-world examples. While the list of trade-offs discussed is likely to be nonexhaustive due to study limitations, the trade-offs discovered can improve understanding of blockchain systems for medical professionals, while also providing useful design information for organizations looking to develop blockchain EMRs.

Limitations

Limitations have been divided into three types, as discussed below.

Unresolvable Issues

Firstly, non-English studies were excluded from the review. Secondly, as very few papers discuss the same trade-offs, there is little interpaper support for findings. Thirdly, while most papers contained logical proofs, not all papers contained simulation data. Thus, not all mathematical relationships in trade-offs are known.

Issues With Tools

Firstly, a lack of accepted MeSH terms for “design” and “trade-off” meant that alternative keyword terms were chosen nonsystematically. This may have caused exclusion of relevant literature. This is particularly pertinent given that not having “design” in the abstract was used to remove 1076 papers from the literature search. Subsequent reviews should aim to filter through these 1076 papers to see if other pertinent terms can be used to filter papers more effectively. Secondly, While the TOGAF was practical for exploring blockchains, it was not designed for this task and was designed in a preblockchain era. A more suitable framework specifically designed for blockchain analysis may be in existence but was not found for this study. Finally, for most papers, ROBINS-I was an inappropriate tool, as no experimental data was present. Papers instead made claims using logical and mathematical proofs. Furthermore, ROBINS-I was a tool developed for assessing bias in trials involving biological organisms subject to variation. ROBINS-I test questions on topics such as time-varying confounding and participant adherence were not appropriate for assessing bias in computer simulations. Questions relating to deviations from intended interventions were entirely irrelevant, due to the instantaneous nature of simulations. The CASP qualitative tool was used in instances where ROBINS-I was completely unsuited, but the tool is significantly less comprehensive than ROBINS-I.

Methodological Issues

Firstly, the choice to research trade-offs in blockchain design rather than potential solutions to trade-offs may have resulted in some irrelevant trade-offs being discussed in the study. Secondly, four papers that may have contained relevant information were excluded due to the level of technicality being too complex. This may have led to relevant trade-offs being excluded. Thirdly, only papers written after December 2016 were included in order to remove papers with outdated information. However, this cutoff year was relatively arbitrary and may have eliminated potentially useful pre-2017 studies. Fourthly, in paper filtering, exclusion criteria 7, 8, and 9 (see Table 2) were relatively arbitrary. They did not remove significant numbers of papers from the search and may have removed useful papers that could have been added to the final review. This is particularly true of criterion 9: papers with financial information may have discussed trade-offs at the business level of enterprise architecture in blockchain design. Finally, due to time constraints, following peer review, only articles in the second database search published before the initial cutoff date (ie, June 15, 2018) were selected to be added to the qualitative synthesis. Identified articles published after the cutoff date should be addressed in a subsequent systematic review, along with an analysis of their potential biases.

Conclusions

There were three key findings. Firstly, multiple trade-offs can be adaptively managed. These trade-offs include block size, the number of confirmation blocks required for a transaction, and the fraction of consortium required to process transactions in consortium blockchains. On a blockchain, multiple rule sets for exchanging data can be used. This will allow for adaptive management and maximization of the utility of blockchain EMRs.

Secondly, multiple trade-offs involve improving blockchain systems at the cost of security. Given this, data protection standards will be an important factor in determining the effectiveness of a blockchain EMR system. Given these trade-offs, research should be conducted on the potential effects of current data protection laws on blockchain EMRs. For instance, it may be necessary for further legislative development to allow for the effective development of NHS blockchain EMRs.

Finally, scalability can be encountered at one end of a trade-off. Blockchain scalability can be improved by moving data off-chain, using hash mirroring instead of digital records, and not using ZKP, among other methods. More infrastructure investment would be required to improve scalability otherwise. Given this, it is important that providers wishing to implement blockchain EMR systems understand the current and future scale of their institution. This will allow for the appropriate scalability trade-offs to be made.

A Model National Health Service Blockchain

This report would make the following specific recommendations for a hypothetical NHS blockchain:

1. Blockchains should not interfere with the Local Health and Care Record programs. Instead, blockchains can be used to manage long character (LCHR) metadata and access permissions for off-chain LCHR EMRs.
2. Given that there are multiple trusted institutions that form part of the NHS structure, a consortium blockchain managed by multiple NHS institutions should be considered. Such a system could potentially make use of PBFT.
3. The blockchain should have an adaptive block size, potentially implemented by machine learning.
4. A blockchain should be built with a scripting language that is specific to NHS needs, one that is difficult to exploit and minimizes the risk of insertion of potentially hazardous data like malware onto the blockchain. Expiration dates on scripts should also be mandated to prevent exploits being permanent.
5. A small number of confirmation blocks should be required for read-only access to patient basic emergency medical data (eg, allergies). More conformation blocks should be required for full access and edit permissions.

Recommendations for Future Research

Future studies should aim to quantify the mathematical relationships between all identified trade-offs in blockchain design. If this is done comprehensively, this will allow for the exact parameters of a blockchain to be purposely selected when designing a blockchain EMR for a specific provider. Identified trade-offs should also ideally be assessed with data from real-world blockchain EMRs. This will allow the impacts of the trade-offs discussed in this paper to be assessed. In addition, the impact of data protection legislation on blockchain systems should be investigated. The security requirements of health care providers may mean that large trade-offs must be made. The impacts of such trade-offs must be fully understood before a health care provider adopts a blockchain EMR. Given study limitations, further literature searches should be performed to find more examples of blockchain design trade-offs. Discovering more trade-offs that can be adaptively managed and discovering how the scalability of blockchains can be increased through trade-offs will help maximize the utility of blockchain-based EMRs.

Acknowledgments

We acknowledge our colleagues in the Department of Computer Science, Department of Engineering, Department of Paediatrics, and the Oxford Academic Health Science Centre for sharing their insights and expertise that assisted in initiating the idea of the project. We would also like to acknowledge Abrar Alturkistani for her help in formatting the paper. This project was supported by the Sir David Cooksey Research Fellowship, the Strategies for Engineered Negligible Senescence (SENS) Research Foundation, and the Oxford Academic Health Science Centre.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Exact search strings used in databases.

[PNG File, 234KB - jmir_v21i5e12426_app1.png]

Multimedia Appendix 2

Review of paper quality using the Risk Of Bias In Non-randomized Studies—of Interventions (ROBINS-I) checklist.

[PNG File, 114KB - jmir_v21i5e12426_app2.PNG]

Multimedia Appendix 3

Critical Assessment Skills Programme (CASP) analysis of papers.

[PNG File, 164KB - jmir_v21i5e12426_app3.PNG]

References

1. Haber S, Stornetta WS. How to time-stamp a digital document. J Cryptol 1991 Jan;3(2):99-111. [doi: 10.1007/BF00196791]
2. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. Saint Kitts and Nevis: Bitcoin Project; 2008. URL: https://bitcoin.org/bitcoin.pdf? [accessed 2019-04-22] [WebCite Cache ID 77pJYR5oV]
3. Preneel B. Analysis and Design of Cryptographic Hash Functions [doctoral dissertation]. Leuven, Belgium: Katholieke Universiteit te Leuven; 2003 Feb. URL: https://www.cosic.esat.kuleuven.be/publications/thesis-2.pdf [accessed 2019-04-22] [WebCite Cache ID 77pKHGBSN]
4. Franco P. Understanding Bitcoin: Cryptography, Engineering and Economics. Chichester, UK: John Wiley & Sons; 2015.
5. Bradbury D. The problem with Bitcoin. Comput Fraud Secur 2013 Nov;2013(11):5-8. [doi: 10.1016/S1361-3723(13)70101-5]
6. The Topol Review: Preparing the Healthcare Workforce to Deliver the Digital Future. Leeds, UK: Health Education England; 2019 Feb. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2019-04-17] [WebCite Cache ID 77hNJi2iG]
7. NHS long term plan. London, UK: National Health Service URL: https://www.longtermplan.nhs.uk/ [accessed 2019-04-17] [WebCite Cache ID 77hNi0Y6G]
8. Farenden F, Singh I. Local Health and Care Record Exemplars: A Summary. London, UK: NHS England; 2018 May. URL: https://www.england.nhs.uk/wp-content/uploads/2018/05/local-health-and-care-record-exemplars-summary.pdf [accessed 2019-04-17] [WebCite Cache ID 77pLYXwZ7]
9. Halamka JD, Lippman A, Ekblaw A. Harvard Business Review. 2017 Mar 03. The potential for blockchain to transform electronic health records URL: https://hbr.org/2017/03/the-potential-for-blockchain-to-transform-electronic-health-records [accessed 2019-04-22] [WebCite Cache ID 77pLtP5v6]
10. Gordon WJ, Catalini C. Blockchain technology for healthcare: Facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J 2018 Jan;16:224-230. [doi: 10.1016/j.csbj.2018.06.003]
11. Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: Using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Open and Big Data (OBD). New York, NY: IEEE; 2016 Presented at: 2nd International Conference on Open and Big Data (OBD); August 22-24, 2016; Vienna, Austria p. 25-30. [doi: 10.1109/OBD.2016.11]
12. Wood G. Ethereum: A Secure Decentralised Generalised Transaction Ledger. 2014 Apr. URL: https://gavwood.com/paper.pdf [accessed 2019-04-22] [WebCite Cache ID 77pMX29QX]
13. Zheng Z, Xie S, Dai H, Chen X, Wang H. An overview of blockchain technology: Architecture, consensus, and future trends. In: Proceedings of the IEEE International Congress on Big Data (BigData Congress). New York, NY: IEEE; 2017 Presented at: IEEE International Congress on Big Data (BigData Congress); June 25-30, 2017; Los Angeles, California p. 557-564. [doi: 10.1109/BigDataCongress.2017.85]
14. Risius M, Spohrer K. A blockchain research framework. Bus Inf Syst Eng 2017 Dec;59(6):385-409. [doi: 10.1007/s12599-017-0506-0]
15. Campion-Awwad O, Hayton A, Smith L, Vuaran M. The National Programme for IT in the NHS: A Case History. Cambridge, UK: University of Cambridge; 2014 Feb. URL: https://www.cl.cam.ac.uk/~rja14/Papers/npfit-mpp-2014-case-history.pdf [accessed 2019-04-22] [WebCite Cache ID 77pNEXm5Q]
16. Meinert E, Alturkistani A, Foley K, Osama T, Car J, Majeed A, et al. Blockchain implementation in health care: Protocol for a systematic review. JMIR Res Protoc 2019 Feb 08;8(2):e10994 [FREE Full text] [doi: 10.2196/10994] [Medline: 30735146]
17. Sterne J, Hernán MA, Reeves B, Savović J, Berkman N, 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]
18. CASP Qualitative Checklist. Oxford, UK: Critical Appraisal Skills Programme (CASP) URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf [accessed 2019-04-22] [WebCite Cache ID 77pNv7wqm]

19. Dinh T, Wang J, Chen G, Liu R, Ooi B, Tan K. Blockbench: A framework for analyzing private blockchains. In: Proceedings of the 2017 ACM International Conference on Management of Data. New York, NY: ACM; 2017 May 09 Presented at: 2017 ACM International Conference on Management of Data; May 14-19, 2017; Chicago, IL p. 1085-1100. [doi: 10.1145/3035918.3064033]
20. Xin W, Zhang T, Hu C, Tang C, Liu C, Chen Z. On scaling and accelerating decentralized private blockchains. In: Proceedings of BigDataSecurity 2017: The 3rd IEEE International Conference on Big Data Security on Cloud. New York, NY: IEEE; 2017 Presented at: BigDataSecurity 2017: The 3rd IEEE International Conference on Big Data Security on Cloud; May 26-28, 2017; Beijing, China p. 267-271. [doi: 10.1109/BigDataSecurity.2017.25]
21. Yasaweerasinghelage R, Staples M, Weber I. Predicting latency of blockchain-based systems using architectural modelling and simulation. In: Proceedings of the 2017 IEEE International Conference on Software Architecture (ICSA). 2017 Presented at: 2017 IEEE International Conference on Software Architecture (ICSA); April 3-7, 2017; Gothenburg, Sweden p. 253-256. [doi: 10.1109/ICSA.2017.22]
22. Dinh TTA, Liu R, Zhang M, Chen G, Ooi BC, Wang J. Untangling blockchain: A data processing view of blockchain systems. IEEE Trans Knowl Data Eng 2018 Jul 01;30(7):1366-1385. [doi: 10.1109/TKDE.2017.2781227]
23. Chen C, Qi Z, Liu Y, Lei K. Using virtualization for blockchain testing. In: Proceedings of the International Conference on Smart Computing and Communication (SmartCom 2017). Cham, Switzerland: Springer International Publishing; 2018 Presented at: International Conference on Smart Computing and Communication (SmartCom 2017); December 10-12, 2017; Shenzhen, China p. 289-299. [doi: 10.1007/978-3-319-73830-7_29]
24. Matzutt R, Henze M, Ziegeldorf JH, Hiller J, Wehrle K. Thwarting unwanted blockchain content insertion. In: Proceedings of the 2018 IEEE International Conference on Cloud Engineering (IC2E). New York, NY: IEEE; 2018 Presented at: 2018 IEEE International Conference on Cloud Engineering (IC2E); April 17-20, 2018; Orlando, FL p. 364-370. [doi: 10.1109/IC2E.2018.00070]
25. Xu X, Weber I, Staples M, Zhu L, Bosch J, Bass L, et al. A taxonomy of blockchain-based systems for architecture design. In: Proceedings of the 2017 IEEE International Conference on Software Architecture (ICSA). New York, NY: IEEE; 2017 Presented at: 2017 IEEE International Conference on Software Architecture (ICSA); April 3-7, 2017; Gothenburg, Sweden p. 243-252. [doi: 10.1109/ICSA.2017.33]
26. BitClave. Medium. 2017 Dec 26. 10 major events and trends in the blockchain industry in 2017 URL: https://medium.com/bitclave/10-major-events-and-trends-in-the-blockchain-industry-in-2017-90b17701fb81 [accessed 2019-04-17] [WebCite Cache ID 77haSZtNT]
27. The Open Group. The TOGAF Standard URL: https://www.opengroup.org/togaf [accessed 2019-04-22] [WebCite Cache ID 77pR9OWRy]
28. The Open Group. URL: https://www.opengroup.org/ [accessed 2019-04-17] [WebCite Cache ID 77hd6x1CB]
29. Alhadhrami Z, Alghfeli S, Alghfeli M, Abedlla JA, Shuaib K. Introducing blockchains for healthcare. In: Proceedings of the 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA). New York, NY: IEEE; 2017 Presented at: 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA); November 21-23, 2017; Ras Al Khaimah, United Arab Emirates p. 1-4. [doi: 10.1109/ICECTA.2017.8252043]
30. Perlner RA, Cooper DA. Quantum resistant public key cryptography: A survey. In: Proceedings of the 8th Symposium on Identity and Trust on the Internet. New York, NY: ACM; 2009 Presented at: 8th Symposium on Identity and Trust on the Internet; April 14-16, 2009; Gaithersburg, MD p. 85-93.
31. Tsai W, Bai X, Yu L. Design issues in permissioned blockchains for trusted computing. In: Proceedings of the 2017 IEEE Symposium on Service-Oriented System Engineering (SOSE). New York, NY: IEEE; 2017 Presented at: 2017 IEEE Symposium on Service-Oriented System Engineering (SOSE); April 6-9, 2017; San Francisco, CA p. 153-159. [doi: 10.1109/SOSE.2017.32]
32. Zhang P, Schmidt DC, White J, Lenz G. Blockchain technology use cases in healthcare. In: Raj P, Deka GC, editors. Advances in Computers, Volume 111. Blockchain Technology: Platforms, Tools and Use Cases. Cambridge, MA: Academic Press; 2018:1-42.
33. The International Data Corporation. The Digital Universe Driving Data Growth in Healthcare: Challenges & Opportunities for IT. Hopkinton, MA: Dell EMC; 2018. URL: https://www.emc.com/analyst-report/digital-universe-healthcare-vertical-report-ar.pdf [accessed 2019-04-17] [WebCite Cache ID 77hdVdri9]
34. He K, Ge D, He M. Big data analytics for genomic medicine. Int J Mol Sci 2017 Feb 15;18(2):412 [FREE Full text] [doi: 10.3390/ijms18020412] [Medline: 28212287]
35. Lemieux VL. A typology of blockchain recordkeeping solutions and some reflections on their implications for the future of archival preservation. In: Proceedings of the 2017 IEEE International Conference on Big Data (Big Data). New York, NY: IEEE; 2017 Presented at: 2017 IEEE International Conference on Big Data (Big Data); December 11-14, 2017; Boston, MA p. 2271-2278. [doi: 10.1109/BigData.2017.8258180]
36. Ruubel M. Guardtime. 2016 Feb 12. Estonian eHealth authority partners with Guardtime to accelerate transparency and auditability in health care URL: https://guardtime.com/blog/estonian-ehealth-partners-guardtime-blockchain-based-transparency [accessed 2019-04-17] [WebCite Cache ID 77hdgdwkv]

37. Leising M. Bloomberg. 2017 Jun 13. The ether thief URL: https://www.bloomberg.com/tosv2.html?vid=&uuid=e7efecf0-611c-11e9-8453-1fe20c0c76b0&url=L2ZlYXR1cmVzLzIwMTctdGhlLWV0aGVyLXRoaWVmLw== [accessed 2019-04-17] [WebCite Cache ID 77he3EBnZ]
38. Moubarak J, Filiol E, Chamoun M. On blockchain security and relevant attacks. In: Proceedings of the 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM). New York, NY: IEEE; 2018 Presented at: 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM); April 18-20, 2018; Jounieh, Lebanon p. 1-6. [doi: 10.1109/MENACOMM.2018.8371010]
39. Naganuma K, Yoshino M, Sato H, Suzuki T. Auditable Zerocoin. In: Proceedings of the 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). New York, NY: IEEE; 2017 Presented at: 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW); April 26-28, 2017; Paris, France p. 59-63. [doi: 10.1109/EuroSPW.2017.51]
40. Gibbs S. The Guardian. 2018 Mar 20. Child abuse imagery found within bitcoin's blockchain URL: https://www.theguardian.com/technology/2018/mar/20/child-abuse-imagery-bitcoin-blockchain-illegal-content [accessed 2019-04-17] [WebCite Cache ID 77heD05cV]
41. INTERPOL. 2015 Mar 26. INTERPOL cyber research identifies malware threat to virtual currencies URL: https://www.interpol.int/News-and-Events/News/2015/INTERPOL-cyber-research-identifies-malware-threat-to-virtual-currencies [accessed 2019-04-17] [WebCite Cache ID 77pTy6K7K]
42. Gramoli V. From blockchain consensus back to Byzantine consensus. Future Gener Comput Syst 2017 Sep 21:1. [doi: 10.1016/j.future.2017.09.023]
43. Fischer MJ, Lynch NA, Paterson MS. Impossibility of distributed consensus with one faulty process. J ACM 1985 Apr;32(2):374-382 [FREE Full text] [doi: 10.1145/3149.214121]
44. Castro M, Liskov B. Practical Byzantine fault tolerance. In: Proceedings of the Third Symposium on Operating Systems Design and Implementation (OSDI '99). 1999 Presented at: Third Symposium on Operating Systems Design and Implementation (OSDI '99); February 22-25, 1999; New Orleans, LA p. 173-186.
45. Bhaskaran K, Ilfrich P, Liffman D, Vecchiola C, Jayachandran P, Kumar A, et al. Double-blind consent-driven data sharing on blockchain. In: Proceedings of the 2018 IEEE International Conference on Cloud Engineering (IC2E). New York, NY: IEEE; 2018 Presented at: 2018 IEEE International Conference on Cloud Engineering (IC2E); April 17-20, 2018; Orlando, FL p. 385-391. [doi: 10.1109/IC2E.2018.00073]

Abbreviations

BFT: Byzantine Fault Tolerance
CASP: Critical Assessment Skills Programme
DAO: Decentralized Autonomous Organization
EMR: electronic medical record
IEEE: Institute of Electrical and Electronics Engineers
IoT: internet of things
LCHR: long character
MeSH: Medical Subject Headings
NHS: National Health Service
PBFT: Practical Byzantine Fault Tolerance
PICO: Population, Intervention, Comparison, and Outcome
PoW: proof of work
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAM: random access memory
ROBINS-I: Risk Of Bias In Non-randomized Studies—of Interventions
SENS: Strategies for Engineered Negligible Senescence
TOGAF: The Open Group Architecture Framework
ZKP: zero-knowledge proof

Edited by G Eysenbach; submitted 05.10.18; peer-reviewed by P Morita, E Chukwu, M Hölbl, J Liu, C Reis; comments to author 18.02.19; revised version received 23.03.19; accepted 30.03.19; published 10.05.19.

Please cite as:
O'Donoghue O, Vazirani AA, Brindley D, Meinert E
Design Choices and Trade-Offs in Health Care Blockchain Implementations: Systematic Review
J Med Internet Res 2019;21(5):e12426
URL: https://www.jmir.org/2019/5/e12426/
doi:10.2196/12426
PMID:31094344

Short Paper

Perspectives of English, Chinese, and Spanish-Speaking Safety-Net Patients on Clinician Computer Use: Qualitative Analysis

Elaine C Khoong[1], MD, MS; Roy Cherian[1,2], MHS; George Y Matta[3], BA; Courtney R Lyles[1,2], PhD; Dean Schillinger[1,2], MD; Neda Ratanawongsa[1,2], MD, MPH

[1]Division of General Internal Medicine, Department of Medicine at Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, CA, United States

[2]UCSF Center for Vulnerable Populations at Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, CA, United States

[3]School of Medicine, Boston University, Boston, MA, United States

Corresponding Author:
Elaine C Khoong, MD, MS
Division of General Internal Medicine
Department of Medicine at Zuckerberg San Francisco General Hospital
University of California San Francisco
1001 Potrero Avenue
Building 10, Ward 13
San Francisco, CA, 94110
United States
Phone: 1 4152063188
Fax: 1 4152065586
Email: elaine.khoong@ucsf.edu

Abstract

Background: Safety-net systems serve patients with limited health literacy and limited English proficiency (LEP) who face communication barriers. However, little is known about how diverse safety-net patients feel about increasing clinician electronic health record (EHR) use.

Objective: The aim of this study was to better understand how safety-net patients, including those with LEP, view clinician EHR use.

Methods: We conducted focus groups in English, Spanish, and Cantonese (N=37) to elicit patient perspectives on how clinicians use EHRs during clinic visits. Using a grounded theory approach, we coded transcripts to identify key themes.

Results: Across multiple language groups, participants accepted multitasking and silent clinician EHR use if focused on their care. However, participants desired more screen share and eye contact, especially when demonstrating physical concerns. All participants, including LEP participants, wanted clinicians to include them in EHR use.

Conclusions: Linguistically diverse patients accept the value of EHR use during outpatient visits but desire more eye contact, verbal warnings before EHR use, and screen-sharing. Safety-net health systems should support clinicians in completing EHR-related tasks during the visit using patient-centered strategies for all patients.

(J Med Internet Res 2019;21(5):e13131) doi:10.2196/13131

KEYWORDS

vulnerable populations; electronic health records; attitude to computers; physician-patient relations; communication barriers

Introduction

Electronic health record (EHR) system use during outpatient visits affects patient-provider communication, clinician workload, and clinician well-being [1-4]. Enabled by federal incentives, American safety-net clinics, publicly funded facilities providing care for socioeconomically disadvantaged populations, have experienced tremendous growth in EHR implementation. EHRs affect the care experience of safety-net patients, exacerbating communication barriers related to limited health literacy or limited English proficiency (LEP) [5]. However, delaying EHR use until after an office visit may increase stress in a workforce at risk for burnout or medical errors (from inaccurate charting or forgetting to place orders) [6-9].

Clinicians engage with EHRs by multitasking while talking with patients or using them silently, potentially diverting their attention from patients [10]. Little is known about how linguistically diverse safety-net patients feel about clinician EHR use during clinic visits. Prior work has suggested that LEP populations may have different perspectives, experiencing higher amounts of EHR use but perceiving greater benefits to that use [11]. This short study aims to further explore how safety-net patients, including those with LEP, feel about clinician EHR use.

Methods

Setting and Recruitment

We conducted focus groups with ethnically and linguistically diverse patients to elicit perspectives on clinician EHR use during outpatient visits. Through posted flyers and in-person recruitment at 7 primary and specialty care clinic waiting rooms (family medicine, adult primary care, obstetrics or gynecology, diabetes, anticoagulation, cardiology, and gastroenterology or hepatology), we identified eligible patients: (1) English, Cantonese, or Spanish-speaking, (2) adults, and (3) receiving primary or specialty care from an urban safety-net hospital. This safety-net system cares for a population that is 16% African-American, 23% Asian, and 37% Latinx and insured predominantly by Medicaid (58%) or Medicare (22%) [12]. A prior study found nearly 50% of patients have inadequate health literacy [13]. There are 24/7 interpretation services via phone interpreters with video or in-person interpreters available during business hours. This system utilizes an Office of National Coordinator (ONC) certified EHR [14]. We collected sociodemographic characteristics by telephone, including use of validated questions to assess English proficiency and health literacy [15,16]. We provided participants US $35 for participation.

Data Sources and Collection

The focus group guide (Multimedia Appendix 1) was created by the study team using an iterative process based on analysis of prior studies. We used consensus to ensure questions would be accurately translated before inclusion in the final focus group guide, which was translated into Chinese and Spanish by fluent, bilingual team members. It included questions on perceptions about overall communication, clinician communication, EHR use during visits, multitasking and silent EHR use, and preferences for EHR use. We used videos of reenacted examples [10] to demonstrate silent versus multitasking EHR use to help participants distinguish between these types of use. In Cantonese, there is no single word to convey the idea of *multi-tasking* so we felt demonstration of these concepts through videos was necessary.

We conducted 6 (3 English, 2 Cantonese, 1 Spanish) in-person 90-min focus groups in patients’ preferred languages from November 2017 to January 2018. A total of 2 Spanish focus groups were planned, but 1 focus group became a one-on-one interview when all other participants did not attend the focus group. The Spanish and English focus groups were conducted by a bilingual research team member; the Cantonese focus groups were conducted by a different bilingual team member. We acquired verbal and written consent from participants before participation. The focus groups were audio-recorded then transcribed into English for analysis. The research team contained Spanish and Cantonese speakers who could consult the audio files if participants’ meaning was unclear in English transcripts.

Data Analysis

Using a grounded theory approach [17,18], we (EK, GM, NR) independently coded the same 3 transcripts (1 from each language), then through consensus developed a codebook (Multimedia Appendix 2) that encompassed themes represented in these transcripts. One team member (GM) then applied this codebook to all transcripts using ATLAS.ti 7.0 (ATLAS.ti Scientific Software Development GmbH); no additional themes were identified during this process, suggesting primary thematic saturation [19]. We triangulated results with primary care clinicians through interactive presentations and determined findings most likely to impact clinical practice.

This study was approved by the University of California San Francisco Institutional Review Board.

Results

Participant Characteristics

There were 37 participants (Table 1). Of the 37, 11 were Cantonese-speaking and 5 Spanish-speaking. Mean age was 54, and 57% (21/37) were women. Nearly half (17/37, 46%) reported limited health literacy, 41% (15/37) reported infrequent personal computer use, and 41% (15/37) reported poor or fair health. Cantonese-speaking participants were more likely to report limited health literacy and feel their primary care providers (PCPs) do not know them well.

Table 1. Demographic characteristics of participants in focus groups (N=37).

Characteristic	All	English (n=21)	Cantonese (n=11)	Spanish (n=5)
Age (years), mean (SD)	54 (10.4)	55 (10.3)	57 (6.2)	46 (14.9)
Women, n (%)	21 (57)	11 (52)	7 (64)	3 (60)
Race, n (%)				
White	10 (27)	10 (48)	0	0
Black or African-American	6 (16)	6 (29)	0	0
Latinx	6 (16)	1 (5)	0	5 (100)
Asian or Pacific Islander	11 (30)	0	11 (100)	0
American Indian or Alaskan Native	1 (3)	1 (5)	0	0
More than one	2 (5)	2 (10)	0	0
Other	1 (3)	1 (5)	0	0
Limited English proficiency[a], n (%)	15 (41)	0	11 (100)	4 (80)[b]
Limited health literacy[c], n (%)	17 (46)	2 (10)	11 (100)	4 (80)[b]
Native language, limited health literacy[d], n (%)	—[e]	—	7 (64)	1 (20)[b]
Poor or fair health, n (%)	15 (41)	4 (19)	8 (73)	3 (60)[b]
Primary care provider knows me well, n (%)	17 (46)	12 (57)	2 (18)	3 (60)[b]
Uses computer never or less than monthly, n (%)	15 (41)	6 (29)	7 (64)	2 (40)[b]

[a] Participants who reported speaking English less than “well.”

[b] One participant declined to answer this question.

[c] Somewhat, a little bit, or not at all confident “filling out medical forms by yourself.”

[d] Somewhat, a little bit, or not at all confident “filling out medical forms by yourself” if in native language (Spanish or Chinese).

[e] Reported only for participants with limited English proficiency.

Perspectives on Electronic Health Record Use

Table 2 contains themes and representative quotes. Themes were consistent across language groups and classified into 2 categories: perspectives on EHR use and strategies for patient-centered EHR use.

Participants generally accepted EHR use, recognizing that its use assisted with care. This acceptance was conditional on the assumption that EHR use was focused on their care. Participants felt similarly about multitasking and silent EHR use, reporting that each type of use was appropriate during different parts of the visit.

Suggestions for Patent-Centered Electronic Health Record Use

Despite the general acceptance of silent, multitasking, and frequent EHR use, participants had suggestions for how clinicians could exhibit more patient-centered EHR use.

Participants uniformly desired more eye contact during EHR use. In particular, some felt computer use is inappropriate when patients are attempting to show physical concerns:

> *If I have a sore throat, don’t just put it in the computer—look!* [English-speaking participant]

All participants reported a desire for clinicians to communicate what they were doing in the computer. In particular, they asked clinicians to provide a warning (ie, signaling) before transitioning to silent EHR use.

As part of this desire for transparency about EHR use, participants—including LEP participants—desired screen sharing:

> *When the doctor is typing...can I look at it at the same time?* [Cantonese-speaking participant 1]
>
> *Like show a big TV screen.* [Cantonese-speaking participant 2]
>
> *Yes. When the doctor is typing, then I can see it.* [Cantonese-speaking participant 1]

Table 2. Key themes and example quotes.

Themes	Example quotes
Perspectives on EHR[a] use	
Patient-focused electronic health record use is acceptable	Spanish-speaking participant: "I think technology is important not just for keeping patient records but also for finding information. If a patient wants to know about a medication he's been prescribed, it's all right there" ; English-speaking participant: "It hasn't been an issue when she's—she's doing it [using the computer]. And, I'm still the only person in the room. So, I'm still getting 100% of our attention."
Silent and multitasking EHR use is expected and generally accepted	Cantonese-speaking participant: "When the doctor is ordering medications, he/she can stop talking to the patient, and concentrate on ordering the medication. When doctor is supposed to do one thing, then he/she should do that one thing" ; English-speaking participant: "I think that doctors have always had to multi-task. Throughout history, there's always a thousand instruments they're having to deal with, they had to deal with many different cases...they're always multi-tasking. Computer is just a tool."
Strategies for patient-centered EHR use	
More eye contact is desired	Cantonese-speaking participant: "If he (doctor) only looks at the medical record, not face the patient and only looks at the computer, then there is a distance between the doctor and patient" ; English-speaking participant: "Look at me when you're talking to me instead of looking at the screen and typing."
Limit computer use while patients show a physical concern to clinicians	Cantonese-speaking participant: "For some illness, you have to look at it to see it...For example, nails problem. You have to look at it to see it. If you only look at the computer...then you won't know it" ; English-speaking participant: "if a doctor was on the computer asking me 'hey how's your pain from 1 to 10'...I would want for her to… look closely at me...at my leg how I can move it, stuff like that...I wouldn't want my doctor being on the computer while that doctor was examining, giving me a physical..."
Communicate the purpose of computer use	English-speaking participant: "Just communicate. Communicate, communicate, communicate...'I got to take a minute and type this. I want to make sure it's right.' And then read it back or whatever; Cantonese-speaking participant: "Before the doctor orders medications, he/she should let you know: 'This is what is your situation, I am going to prescribe this medication for you.'...So are you worried? You're not worried." Spanish-speaking participant: "When he's on the computer he should explain what he's talking about."
Share the screen	Spanish-speaking participant: "...seeing the lab results. It's fantastic...I can look, too, and ask 'What about my anemia? What does that red line mean?' So, then she explains to me and tells me what we need to do..."; English-speaking participant: "I think the most important thing is just knowing that the patient would like to be a part of what's going on, on the computer."

[a]EHR: electronic health record.

Discussion

Principal Findings

In this short report, we found consensus among linguistically diverse safety-net patients on several themes regarding clinician EHR use. Previously, when this setting employed an EHR that did not meet ONC-certification requirements [14], we found that non-English-speaking patients reported more computer use but less concern that PCPs listened less carefully because of computer use; moreover, Asian patients had higher odds of reporting that computers helped PCPs remember patient concerns [11]. The findings of this study, conducted 3 to 4 years after implementation of a more comprehensive, certified EHR, suggest that safety-net patients across multiple languages experience frequent clinician EHR use but recognize its value to their care even if clinicians multitask or use EHRs silently during visits.

Despite acceptance of EHR use, participants provided suggestions for clinicians to improve the patient experience during EHR use. Safety-net patients—including limited health literate and LEP patients whom clinicians may not expect to read the EHR—wanted clinicians to be transparent about EHR use and even engage them in the process of EHR use. In triangulating our findings with clinicians, we found clinicians felt reassured that patients accepted exam room EHR use and surprised that LEP patients wanted to be included in what clinicians were doing on the computer.

Recommendations for Patient-Centered Electronic Health Record Use

These findings in safety-net patients augment existing recommendations for patient-centered EHR use [20-23]:

1. You can use EHRs during visits. Consider asking patients how they feel about EHR use, as participants in this study accepted EHR use if focused on their care, consistent with prior literature [4,24]. This may ease clinician concerns about in-room charting and reduce the burden of after-hours charting [3].
2. Tell and show patients, including LEP patients, what you are doing, and offer a warning before transitioning to silent EHR use. As patients frequently initiate conversations during silence, you should signal to patients if you need to focus temporarily to complete an EHR task safely [10,25-27].
3. Connect with patients by maximizing eye contact and limiting silent EHR use [22,28-30]. Cease computer use when discussing emotional concerns, as previously recommended [31,32], but also when patients are showing physical concerns on their body.

Study Limitations and Strengths

This study is limited by a small sample size within each language from a single setting, lack of information about patient-clinician language concordance, inability to report a response rate, and use of a single individual to code transcripts after developing themes through consensus. Strengths of this study are a diverse, safety-net population and recruitment from both outpatient primary care and specialty clinics.

Practice Implications

As educators develop communications curricula for patient-centered EHR use, these findings and other perspectives from diverse patients should inform the content [23,33,34]. Future efforts should investigate multilevel interventions to increase adoption of patient-centered EHR use strategies, including computer (EHR user interface or content), patient (activation or empowerment), environment (redesign or reposition equipment in rooms), and policy level (incentives) interventions.

Conclusions

Linguistically diverse safety-net patients accept the prevalence and utility of EHR use during outpatient visits, if focused on their care. However, there continues to be room for improvement for clinicians to adopt patient-centered strategies, including eye contact, signaling EHR use, and screen-sharing with safety-net patients.

Acknowledgments

The authors would like to thank Cynthia Cheung for assisting with recruitment and data collection from their Cantonese-speaking participants and colleagues for acting in video reenactments of types of EHR use. Research was supported by the Agency for Healthcare Research (AHRQ) and Quality Grants 1K08HS022561, P30HS023558, and R00HS022408 and National Institutes of Health’s (NIH) National Research Service Award (T32HP19025). DS is supported by the Health Delivery Systems Center for Diabetes Translational Research funded through the National Institute of Diabetes and Digestive and Kidney Diseases grant P30DK092924. The contents herein are solely the responsibility of the authors and do not necessarily represent the official views of AHRQ or the NIH. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of AHRQ or the NIH. Publication was made possible in part by support from the UCSF Open Access Publishing Fund.

Authors' Contributions

EK contributed in conceptualization, methodology, validation, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, and visualization. RC contributed in methodology, validation, formal analysis, data curation, writing—original draft, writing—review and editing, and visualization. GM contributed in conceptualization, methodology, investigation, resources, data curation, writing—review and editing, and project administration. CL contributed in conceptualization, methodology, writing—review and editing, and visualization. DS contributed in conceptualization, methodology, writing—review and editing, and visualization. NR contributed in conceptualization, methodology, validation, formal analysis, investigation, resources, writing—original draft, visualization, supervision, project administration, and funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Focus group interview guide.

[PDF File (Adobe PDF File), 957KB - jmir_v21i5e13131_app1.pdf]

Multimedia Appendix 2

Codebook.

[DOCX File, 17KB - jmir_v21i5e13131_app2.docx]

References

1. Alkureishi MA, Lee WW, Lyons M, Press VG, Imam S, Nkansah-Amankra A, et al. Impact of electronic medical record use on the patient–doctor relationship and communication: a systematic review. J Gen Intern Med 2016 May;31(5):548-560 [FREE Full text] [doi: 10.1007/s11606-015-3582-1] [Medline: 26786877]
2. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Sep 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]

3. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med 2017 Sep;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
4. Shachak A, Reis S. The impact of electronic medical records on patient-doctor communication during consultation: a narrative literature review. J Eval Clin Pract 2009 Aug;15(4):641-649. [doi: 10.1111/j.1365-2753.2008.01065.x] [Medline: 19522722]
5. Ratanawongsa N, Barton JL, Lyles CR, Wu M, Yelin EH, Martinez D, et al. Association between clinician computer use and communication with patients in safety-net clinics. JAMA Intern Med 2016 Jan;176(1):125-128 [FREE Full text] [doi: 10.1001/jamainternmed.2015.6186] [Medline: 26619393]
6. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004 Apr;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
7. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. Perspect Health Inf Manag 2013;10:1c [FREE Full text] [Medline: 24159271]
8. Yadav S, Kazanji N, Paudel S, Falatko J, Shoichet S, Maddens M, et al. Comparison of accuracy of physical examination findings in initial progress notes between paper charts and a newly implemented electronic health record. J Am Med Inform Assoc 2017 Dec;24(1):140-144. [doi: 10.1093/jamia/ocw067] [Medline: 27357831]
9. Friedberg MW, Reid RO, Timbie JW, Setodji C, Kofner A, Weidmer B, et al. Federally qualified health center clinicians and staff increasingly dissatisfied with workplace conditions. Health Aff (Millwood) 2017 Dec 1;36(8):1469-1475. [doi: 10.1377/hlthaff.2017.0205] [Medline: 28784740]
10. Ratanawongsa N, Matta GY, Lyles CR, Koenig CJ, Barton JL, Yu K, et al. . Multitasking and silent electronic health record use in ambulatory visits. JAMA Intern Med 2017 Sep 1;177(9):1382-1385. [doi: 10.1001/jamainternmed.2017.2668] [Medline: 28672379]
11. Ratanawongsa N, Barton JL, Schillinger D, Yelin EH, Hettema JE, Lum PJ. Ethnically diverse patients' perceptions of clinician computer use in a safety-net clinic. J Health Care Poor Underserved 2013 Nov;24(4):1542-1551 [FREE Full text] [doi: 10.1353/hpu.2013.0188] [Medline: 24185151]
12. San Francisco Department of Public Health. 2018. Annual Report 2016-2017 URL: https://www.sfdph.org/dph/files/reports/PolicyProcOfc/SFDPH-Annual-Report-2016-2017.pdf [accessed 2019-05-13] [WebCite Cache ID 78KfKYuSm]
13. Sarkar U, Schillinger D, López A, Sudore R. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. J Gen Intern Med 2011 Mar;26(3):265-271 [FREE Full text] [doi: 10.1007/s11606-010-1552-1] [Medline: 21057882]
14. Office of the National Coordinator for Health Information Technology (ONC). HealthIT.gov. 2019 Mar. About The ONC Health IT Certification Program URL: https://www.healthit.gov/topic/certification-ehrs/about-onc-health-it-certification-program [accessed 2019-05-08] [WebCite Cache ID 78DukU3G1]
15. Karliner LS, Napoles-Springer AM, Schillinger D, Bibbins-Domingo K, Pérez-Stable EJ. Identification of limited English proficient patients in clinical care. J Gen Intern Med 2008 Oct;23(10):1555-1560 [FREE Full text] [doi: 10.1007/s11606-008-0693-y] [Medline: 18618200]
16. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008 May;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]
17. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Newbury Park, CA: Sage Publications; 2014.
18. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
19. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. Qual Rep 2015;20(9):1408-1416 [FREE Full text]
20. Duke P, Frankel RM, Reis S. How to integrate the electronic health record and patient-centered communication into the medical visit: a skills-based approach. Teach Learn Med 2013;25(4):358-365. [doi: 10.1080/10401334.2013.827981] [Medline: 24112206]
21. Ventres W, Kooienga S, Marlin R. EHRs in the exam room: tips on patient-centered care. Fam Pract Manag 2006 Mar;13(3):45-47 [FREE Full text] [Medline: 16568596]
22. Frankel R. When It Comes to the Physician-Patient-Computer Relationship, the “Eyes” Have It. In: Papadakos P, Bertman S, editors. Distracted Doctoring. Cham, Switzerland: Springer International Publishing AG; 2017:87-99.
23. Alkureishi MA, Lee WW, Webb S, Arora V. Integrating patient-centered electronic health record communication training into resident onboarding: curriculum development and post-implementation survey among housestaff. JMIR Med Educ 2018 Jan 4;4(1):e1 [FREE Full text] [doi: 10.2196/mededu.8976] [Medline: 29301735]

24. Irani JS, Middleton JL, Marfatia R, Omana ET, D'Amico F. The use of electronic health records in the exam room and patient satisfaction: a systematic review. J Am Board Fam Med 2009;22(5):553-562 [FREE Full text] [doi: 10.3122/jabfm.2009.05.080259] [Medline: 19734402]
25. Ratanawongsa N, Matta GY, Bohsali FB, Chisolm MS. Reducing misses and near misses related to multitasking on the electronic health record: observational study and qualitative analysis. JMIR Hum Factors 2018 Feb 6;5(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.9371] [Medline: 29410388]
26. Matta GY, Khoong EC, Lyles CR, Schillinger D, Ratanawongsa N. Finding meaning in medication reconciliation using electronic health records: qualitative analysis in safety net primary and specialty care. JMIR Med Inform 2018 May 7;6(2):e10167 [FREE Full text] [doi: 10.2196/10167] [Medline: 29735477]
27. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med 2009 Mar;24(3):341-348 [FREE Full text] [doi: 10.1007/s11606-008-0892-6] [Medline: 19130148]
28. Street RL, Liu L, Farber NJ, Chen Y, Calvitti A, Zuest D, et al. Provider interaction with the electronic health record: the effects on patient-centered communication in medical encounters. Patient Educ Couns 2014 Sep;96(3):315-319 [FREE Full text] [doi: 10.1016/j.pec.2014.05.004] [Medline: 24882086]
29. Street R, Liu L, Farber N, Chen Y, Calvitti A, Weibel N, et al. Keystrokes, mouse clicks, and gazing at the computer: how physician interaction with the EHR affects patient participation. J Gen Intern Med 2018 Apr;33(4):423-428 [FREE Full text] [doi: 10.1007/s11606-017-4228-2] [Medline: 29188544]
30. Crampton NH, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. J Am Med Inform Assoc 2016 May;23(3):654-665. [doi: 10.1093/jamia/ocv178] [Medline: 26769911]
31. Rathert C, Mittler JN, Banerjee S, McDaniel J. Patient-centered communication in the era of electronic health records: what does the evidence say? Patient Educ Couns 2017 Jan;100(1):50-64. [doi: 10.1016/j.pec.2016.07.031] [Medline: 27477917]
32. Wolfe L, Chisolm MS, Bohsali F. Clinically excellent use of the electronic health record: review. JMIR Hum Factors 2018 Oct 5;5(4):e10426 [FREE Full text] [doi: 10.2196/10426] [Medline: 30291099]
33. Lee WW, Alkureishi ML, Isaacson JH, Mayer M, Frankel RM, London DA, et al. Impact of a brief faculty training to improve patient-centered communication while using electronic health records. Patient Educ Couns 2018 Dec;101(12):2156-2161. [doi: 10.1016/j.pec.2018.06.020] [Medline: 30007764]
34. Alkureishi M, Lee W, Lyons M, Wroblewski K, Farnan J, Arora V. Electronic-clinical evaluation exercise (e-CEX): a new patient-centered EHR use tool. Patient Educ Couns 2018 Dec;101(3):481-489. [doi: 10.1016/j.pec.2017.10.005] [Medline: 29042145]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
EHR: electronic health record
LEP: limited English proficiency
NIH: National Institutes of Health
ONC: Office of National Coordinator
PCP: primary care provider

Edited by G Eysenbach; submitted 13.12.18; peer-reviewed by A Shachak, M Alkureishi; comments to author 07.02.19; revised version received 02.04.19; accepted 24.04.19; published 22.05.19.

Please cite as:
Khoong EC, Cherian R, Matta GY, Lyles CR, Schillinger D, Ratanawongsa N
Perspectives of English, Chinese, and Spanish-Speaking Safety-Net Patients on Clinician Computer Use: Qualitative Analysis
J Med Internet Res 2019;21(5):e13131
URL: http://www.jmir.org/2019/5/e13131/
doi:10.2196/13131
PMID:31120020

Original Paper

Secure and Scalable mHealth Data Management Using Blockchain Combined With Client Hashchain: System Design and Validation

Tomomitsu Motohashi[1], MSc; Tomonobu Hirano[1], BSc; Kosuke Okumura[1], BSc; Makiko Kashiyama[1], MSc; Daisuke Ichikawa[1], MD, PhD; Taro Ueno[1], MD, PhD

SUSMED, Inc, Tokyo, Japan

Corresponding Author:
Taro Ueno, MD, PhD
SUSMED, Inc
Nihonbashi Life Science Bldg 2, 3-11-5, Honcho
Nihonbashi, Chuo-ku
Tokyo, 103-0023
Japan
Phone: 81 335273593
Email: t-ueno@umin.ac.jp

Abstract

Background: Blockchain is emerging as an innovative technology for secure data management in many areas, including medical practice. A distributed blockchain network is tolerant against network fault, and the registered data are resistant to tampering and revision. The technology has a high affinity with digital medicine like mobile health (mHealth) and provides reliability to the medical data without labor-intensive third-party contributions. On the other hand, the reliability of the medical data is not insured before registration to the blockchain network. Furthermore, there are issues with regard to how the clients' mobile devices should be dealt with and authenticated in the blockchain network in order to avoid impersonation.

Objective: The aim of the study was to design and validate an mHealth system that enables the compatibility of the security and scalability of the medical data using blockchain technology.

Methods: We designed an mHealth system that sends medical data to the blockchain network via relay servers. The architecture provides scalability and convenience of operation of the system. In order to ensure the reliability of the data from clients' mobile devices, hash values with chain structure (client hashchain) were calculated in the clients' devices and the results were registered on the blockchain network.

Results: The system was applied and deployed in mHealth for insomnia treatment. Clinical trials for mHealth were conducted with insomnia patients. Medical data of the recruited patients were successfully registered with the blockchain network via relay servers along with the hashchain calculated on the clients' mobile devices. The correctness of the data was validated by identifying illegal data, which were made by simulating fraudulent access.

Conclusions: Our proposed mHealth system, blockchain combined with client hashchain, ensures compatibility of security and scalability in the data management of mHealth medical practice.

Trial Registration: UMIN Clinical Trials Registry UMIN000032951; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000037564 (Archived by WebCite at http://www.webcitation.org/78HP5iFIw)

(J Med Internet Res 2019;21(5):e13385) doi:10.2196/13385

KEYWORDS

mobile health; electronic health records; blockchain; client hashchain; clinical trial

Introduction

Digital medicine, including the use of mHealth apps and internet of things (IoT) devices, has become popular in the everyday practice of medicine [1]. It has the potential to promote improved patient health outcomes, support care coordination, and improve communication with lower costs. While digital medicine has the potential for better practices to patients, we need to consider the security issues. Data tampering and impersonation are important security risks for digital medicine and clinical trials. Decision making in medical practice should be based on precise patient information. Data reliability is compromised if data tampering and impersonation are used to attack the system. External cyberattacks, including ransomware attacks, which result in compromised medical records, are huge threats against the health care sector [2,3]. Data breaches can lead to privacy violations, embarrassment, and social stigma, as well as to fraud and medical identity theft.

In addition to cybersecurity, data governance and authenticity are also important issues in the health care sector, especially in data management in clinical trials [4,5]. Recently, Web-based clinical trials have been conducted to streamline and improve the convenience of clinical trial participation [6]. Since the results of the clinical trials are the basis of the approval of medicine or medical devices by regulatory agencies, the reliability and transparency of the data obtained by the clinical trial must be maintained [7]. However, there are reports that 17% of clinical drug trials were fabricated [8-10]. The ability to easily trace data back to the original source is indispensable.

Blockchain technology has recently garnered attention as a means for transferring data between participating parties based on a “distributed ledger” model that affords a fully transparent and immutable record of data transactions [11]. A blockchain consists of a continuously growing list of transactional records organized into blocks that are replicated on the nodes of a peer-to-peer network. Valid transactions stored in a blockchain are digitally signed and timestamped by their sender, providing cryptographically irrefutable evidence of both provenance and existence of a record at a given point in time. The technology provides a verifiable and tamper-proof history of the data in the blockchain network. Bitcoin is the first implementation of blockchain as a digital asset in widespread use [12]. It eliminates the need for trusted third parties in financial transactions by providing a secure and verifiable history for every transaction in the system.

Beyond digital currency, researchers have started to focus on using blockchain technology for building cryptographic proof in many areas including medical sectors [13]. Blockchain has already been proposed for use in various health care settings, with potential applications in health supply chain management [14-16], insurance claims processing [17-19], medical record management [20,21], and data management in clinical trials [22-26]. Drug counterfeiting is a global problem with significant risks to consumers and the general public. Blockchain has the potential for tracking and tracing drug products and reagents, and counterfeit detection through information verification of supply chain participants. Blockchain technology can also be applied to managing insurance claim policies by the insurance providers and the patients. It can provide authorized access of data to researchers to analyze diseases. Blockchain-based models for electronic medical records have been proposed to enhance ownership of their medical data and data sharing between platforms [17,27,28]. Since the blockchain can be used to establish a permanent record agreed on by all participating parties, it has the potential to mitigate some of the threats to data validity, so that some researchers have proposed to support or even replace the traditional data infrastructure used in clinical trials with blockchain systems [22,23,29,30]. Our previous study also demonstrated an mHealth system for insomnia using a mobile phone app together with a blockchain storage platform and evaluated resistance against tampering of the data collected with mobile phones [24].

Although medical data registered in a blockchain network have proved tamper-resistant, the vulnerability of the medical data lies before registration to the blockchain network. Impersonation of client devices or fabrication of data outside of the blockchain network can impair the reliability of the medical data. In addition, if the blockchain network is open to the Internet, the network is vulnerable to attack and to the theft of medical records. Client devices, such as mobile phones used by patients, should not be dealt with as nodes of the blockchain network in order to preserve the confidentiality of the personal medical data and to reduce the operational cost for the management of the private key. The trade-off must be recognized and overcome by technological improvements.

This study aims to describe and validate an mHealth system using a client management architecture along with a blockchain network. To overcome the remaining problems described above, we designed the whole system with relay servers, which connect client devices to the blockchain network. In order to ensure reliability and to prevent impersonation or fabrication, a hashchain was calculated in client devices and sent to the blockchain network along with the medical data. We further verified the proposed system in the actual clinical trial of mHealth for insomnia patients and evaluated the resistance to various fraud attacks.

Methods

Clinical Trial and mHealth Records

The proposed system was applied and deployed into the mHealth app, which treats chronic insomnia based on cognitive behavioral therapy for insomnia (CBTi) [31]. After a favorable formal review by the Japanese Pharmaceuticals and Medical Devices Agency, research on the mHealth app was conducted by the digital therapeutics company, SUSMED, Inc. (Tokyo, Japan). Informed consent was obtained from the patients for publication of this study. The study has received ethical approval from the Ethics Committee and registered to clinical trial registry (UMIN000032951). All the methods were performed in accordance with the relevant guidelines and regulations.

The mHealth records collected from patients were divided into subjective and objective data. The subjective data, which include clinical indicators, sleep status, and the review of daytime activities, were collected through a self-administered questionnaire. The objective data, which include the results of the Psychomotor Vigilance Test [32], were evaluated by measuring the touch response using the function of mobile phones. For clinical indicators, Athens Insomnia Scale [33], Epworth Sleepiness Scale [34], and Quick Inventory of Depressive Symptomatology [35] were collected. For sleep status, time to go to bed, time to fall asleep, time to wake up, and time to get out of bed were recorded. Along with the medical information, timestamps of the app operation were collected. These clinical indicators were collected using a mobile phone app. All data were stored in the JavaScript Object Notation (JSON) format in the database.

mHealth Data Transfer Via Relay Servers to Blockchain Network

The collected data from the patients' devices were sent to the blockchain network via relay servers. We used three relay servers, and the app randomly selected two relay servers to send the data after the authentication of the client device. By deploying the relay proxy and setting the blockchain software development kits (SDKs) to write-only mode, the relay server sent the received data to the blockchain network. The authentication of the relay server was conducted with a single common authentication server. By configuring the Internet Protocol (IP) address restriction to the listed relay servers, the blockchain network, which contains the medical data, was protected against external attack. The blockchain network was made up of three organizations that contain two validating peers. Each account for nodes of the blockchain network and the relay servers were managed by independent departments in SUSMED, Inc. The strictness of the governance of the data management can be adjusted by individually managing the accounts of different stakeholders, such as pharmaceutical companies, contract research organizations, and regulatory agencies.

mHealth Data Registration in the Blockchain Network

We used Hyperledger Fabric v1.0 to operate the blockchain network because Hyperledger is an open-source blockchain platform and has become widely used [36,37].

The blockchain network was administered by a collection of organizations. Each organization had multiple nodes. In this study, the network had three organizations and each organization had two nodes. As the state database, CouchDB was used to store the JSON document [38].

The nodes executed an installed chaincode and returned hash values generated from the execution result. The secure hash algorithm SHA-256 was used to compute the hash values encoded into the blocks of the blockchain [39]. To execute the transaction, each node followed the consensus algorithm, which was called endorsement policy [11], although the previous version of Hyperledger Fabric used Practical Byzantine Fault Tolerance as a consensus algorithm [40,41].

The endorsement policy was set in units of organizations, and the flexible set was available according to the needs of the app. Each organization issued one signature. The node that validated the transactions in each organization was called the endorser. In this study, the validation of each transaction required more than two signatures from three organizations.

Under the endorsement policy, transactions were validated and accepted in the following processes:

1. Proposal: The transaction was sent from the client app to the endorsers in each organization.
2. Endorse: Each endorser verified that (1) the transaction proposal was well formed, (2) it had not been already submitted in the past, (3) the signature was valid, and (4) the client was properly authorized to perform the proposed operation, which was described in the chaincode. If the transaction was validated, chaincode was executed and the result with the signature was returned to the client.
3. Submit: The client verified that the number of signatures from the organizations satisfied the endorsement policy. If it was satisfied, the transaction was sent to the ordering service, which ordered the series of transactions in chronological order and created the block of the transactions.
4. Broadcast: The block was delivered to all nodes.
5. Commit: If each block was validated to fulfill the endorsement policy and was to be well formed, the block was appended to the chain in each node.

All data, including the client hashchain, were registered in the blockchain network via relay servers to secure the tamper-resistance of the data. In contrast, the secure string for the calculation of the hash value was preserved in the client device. At the end of the study, the secure string was sent to the blockchain network to verify the hashchain. The client hash values were calculated on the mobile phone based on medical data, the secure string, and the previous hash value using the SHA-256 algorithm [39].

Test Scenarios

There are issues regarding cybersecurity (ie, concerning external actors) and governance/authenticity (eg, internal actors like researchers) in medical data management. The tamper-resistance of the data registered in the blockchain network against external attack has been proven in previous studies. The reliability of the data against internal actors in the blockchain network can also be guaranteed by managing the accounts for each node by different stakeholders, such as pharmaceutical companies, contract research organizations, and regulatory agencies. Here, we evaluated how the data manipulation before registration to the blockchain network can be detected and distinguished by simulating the following malicious access. The artificial data were created for each scenario and tested if the fraudulent access were detected and the original data can be distinguished from the illegal data. Since the results of the manipulation of the data are deterministic due to collision-resistant hash functions of SHA-256 [42], we verified the result of a single manipulation in each scenario. Since the accounts for nodes of the blockchain network and the relay servers were managed by independent departments, both internal and external actors can be simulated by hacking each server.

1. Attack on the relay server: To simulate an artificial attack on the relay server (ie, outside actors) or misconduct by the owner of the relay server (ie, inside actors) during the clinical trial, one of the relay servers was hacked. The data sent from a client device can be modified before registration to the blockchain network by the malicious access. In this case, the secure string for the calculation of a hash value in the client device was not stolen. Hence, the client hash value was calculated with the previous hash value, modified medical data, and the incorrect string.
2. Attack on the authentication server: To simulate an artificial attack on the authentication server (ie, outside actors) or misconduct by the owner of the authentication server (ie, inside actors) during the clinical trial, the authentication key of an existing account on the authentication server was stolen and the data of the hacked account was uploaded by

multiple devices. In this case, the secure string for the calculation of a hash value in the client device was not stolen. Hence, the client hash value was calculated with the previous hash value, modified medical data, and the incorrect string.

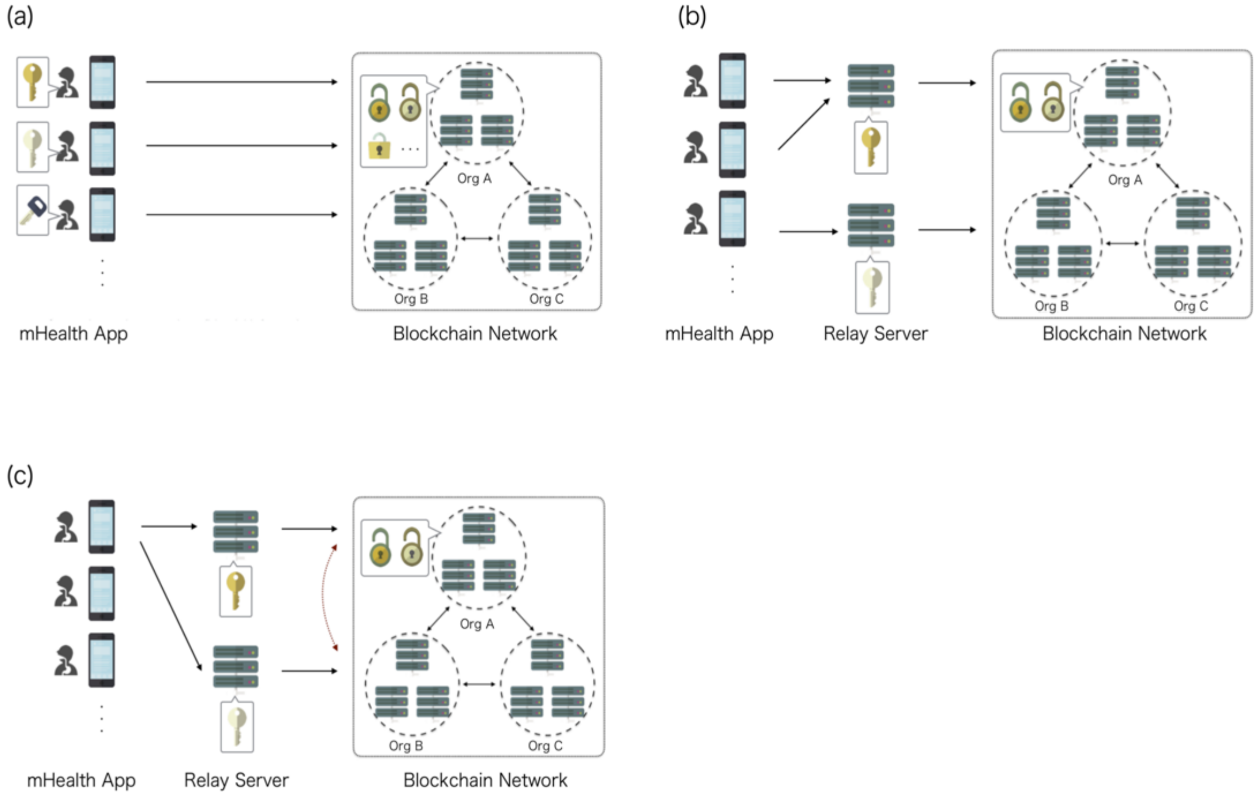
3. Attack on the client device: The secure string preserved in the client device was stolen by an attacker using a mobile malware root exploit. The authentication information to the relay servers was also stolen by the malicious infection. Hence, the client hash value was calculated by different devices with the previous hash value, modified medical data, and the correct secure string.

Results

Design of the Client for Blockchain Network in mHealth

In our previous study, the mHealth data obtained by a mobile phone were uploaded to the blockchain network. We then evaluated the robustness of the network and the tamper-resistance of the data in the blockchain network [24]. To advance further into the practical usage of blockchain in mHealth, it is necessary to design how the client devices, such as mobile phones, send medical data to the blockchain network. Client devices can send data to the blockchain network directly, but with this architecture, the mHealth app that patients installed to their mobile phone needs to have SDKs for blockchain. The blockchain network will also receive access from unspecified clients due to a lack of IP address restriction. In that situation, there are risks that unspecified clients can store medical information in the blockchain network using SDKs to read the data. The system is resistant against data tampering, but the operational costs increase since they must manage private keys for each client device (Figure 1). In order to overcome these obstacles, the usage of relay servers is one possible option. With the architecture using relay servers, the blockchain network can restrict the access by IP address selection and it is not necessary for the mHealth app to include SDKs for blockchain. With this architecture, blockchain networks were protected against unspecified access from the internet and the functions of SDKs in the relay server can be predetermined as write-only, resulting in the protection of the medical data stored in the blockchain network. On the other hand, there are risks that the hacking of the relay server will result in impersonation (Figure 1). To balance these trade-offs, we propose the following architecture of the client for the blockchain network in mHealth. The client devices, such as mobile phones, send their data to multiple relay servers and the data are compared between relay servers and verified. After verification, the relay server, which has permission to access by IP address restriction, sends the data to the blockchain network using write-only functions of SDKs. With this system, medical data stored in the blockchain network are protected against access from the internet and are resistant against the risks of server hacking (Figure 1).

Figure 1. mHealth system architecture, which sends the medical data to the blockchain network. Data sent to (a) the blockchain network without relay servers; client devices are dealt with as nodes of the blockchain network so that mHealth app needs to contain software development kits for blockchain; (b) the blockchain via a single relay server; public keys for each relay server should be managed; (c) the blockchain via multiple relay servers; public keys for each relay server should be managed. Data reliability can be verified by comparing the data to be registered (red line).



Authentication Between Clients and Relay Servers

For the usage of relay servers described above, it is indispensable to carefully design the authentication of client devices to send data to relay servers. It is possible that a single common server gives an authentication to client devices, but with this architecture, the system is vulnerable to server hacking. Malicious access to the single authentication server can result in impersonation (Figure 2). The alternative is to set an authentication server for each relay server. Although the risks of impersonation by server hacking can be reduced, the operational costs for authentication increase. In addition, it is not possible to verify the reliability of the original data if multiple authentication servers were maliciously accessed (Figure 2). In order to solve these problems, we implemented a single common authentication server with another method, using a hash value calculated on the client devices. As an initial setting, the client device generates and preserves a secure string. The client device calculates a hash value based on the medical data and the secure string, as well as the previous hash value using the SHA-256 hash algorithm. Thus, the hash value comprises the chain structure. The hash value was also registered in the blockchain network along with the medical data in order to guarantee tamper-resistance of the value, although the secure string was preserved in the client device. It is possible to verify the reliability of data using the secure string preserved in the client devices and to retrospectively reject impersonation after finishing the clinical trials. Even when a relay server was hacked by malicious access, we can verify the correct data based on the client hashchain and the secure string preserved in the client device. Since the hash value calculated in the client device makes up the chain structure, we called the technique “client hashchain” in contrast to blockchain (Figure 2). In the case of the device having been destroyed or disabled prior to the conclusion of the study, the secure string could be sent beforehand to the user’s personal storage, such as their email box.

Application of the Proposed System to mHealth and Data Management in a Clinical Trial

In order to validate the system proposed above, we have implemented the architecture into the mHealth app. The app was designed to treat insomnia patients based on CBTi and collects medical data using mobile phones. The app generates a secure string at login and stores it on the client device. The app also calculates a hash value based on the medical data, the previous hash value, and the secure string so that the hash values make up the chain structure. The medical data collected with the app, as well as the hash value, were sent to the blockchain network via relay servers. We used three relay servers and the app selected two relay servers at random to send the data to the blockchain network. The blockchain network comprised three organizations, which contain two validating peers.

With these systems, we conducted the clinical trial on the mHealth app for insomnia patients. Informed consent was obtained from the patients and the app account was provided by the medical doctor. mHealth data were collected with the mobile app and sent to the blockchain network via relay servers along with the client hash value (Figure 3). In the client devices, client hash values were calculated based on mHealth data, the previous client hash value, and the secure string stored on the client device. The client hash value constitutes the chain structure and proves the origin of the sequential data.

Both mHealth data and the client hash value were sent to the blockchain network to be registered in the ledger. The ledger is made up of the blockchain, sequenced records in blocks, and a state database. Each node of the blockchain maintains a copy of the ledger. If the transaction was validated under the endorsement policy, the chaincode was executed and the block of the transaction was appended in each node. The mHealth data and the client hash value were stored in CouchDB and the blockchain. The block includes a hash of the block’s transactions, as well as a hash of the prior block. In this way, it is not possible to tamper with the ledger data without breaking the hash links.

Although the block size is limited in blockchain, it is enough for our medical data since the clinical indicators are stored as JSON data. In addition, the transaction throughput in the Hyperledger Fabric platform that we used is 2250 transactions per second [27]. In contrast, in a permissionless network or public network, such as Bitcoin and Ethereum, it takes 600 seconds and 10 seconds respectively to write a transaction on the ledger [28]. Since the number of transactions in our mHealth system occur several times per day for every patient, the transaction performance of the blockchain will not be a bottleneck.

Figure 2. Authentication of client devices and relay servers. Client devices authenticated by (a) a single common authentication server; the system is vulnerable against server hacking; (b) multiple authentication servers for each relay servers; (c) a single common authentication server. In addition to authentication, client devices calculate a hash value based on data, secure string, and previous hash value, so the hash value consists of the chain structure (client hashchain).

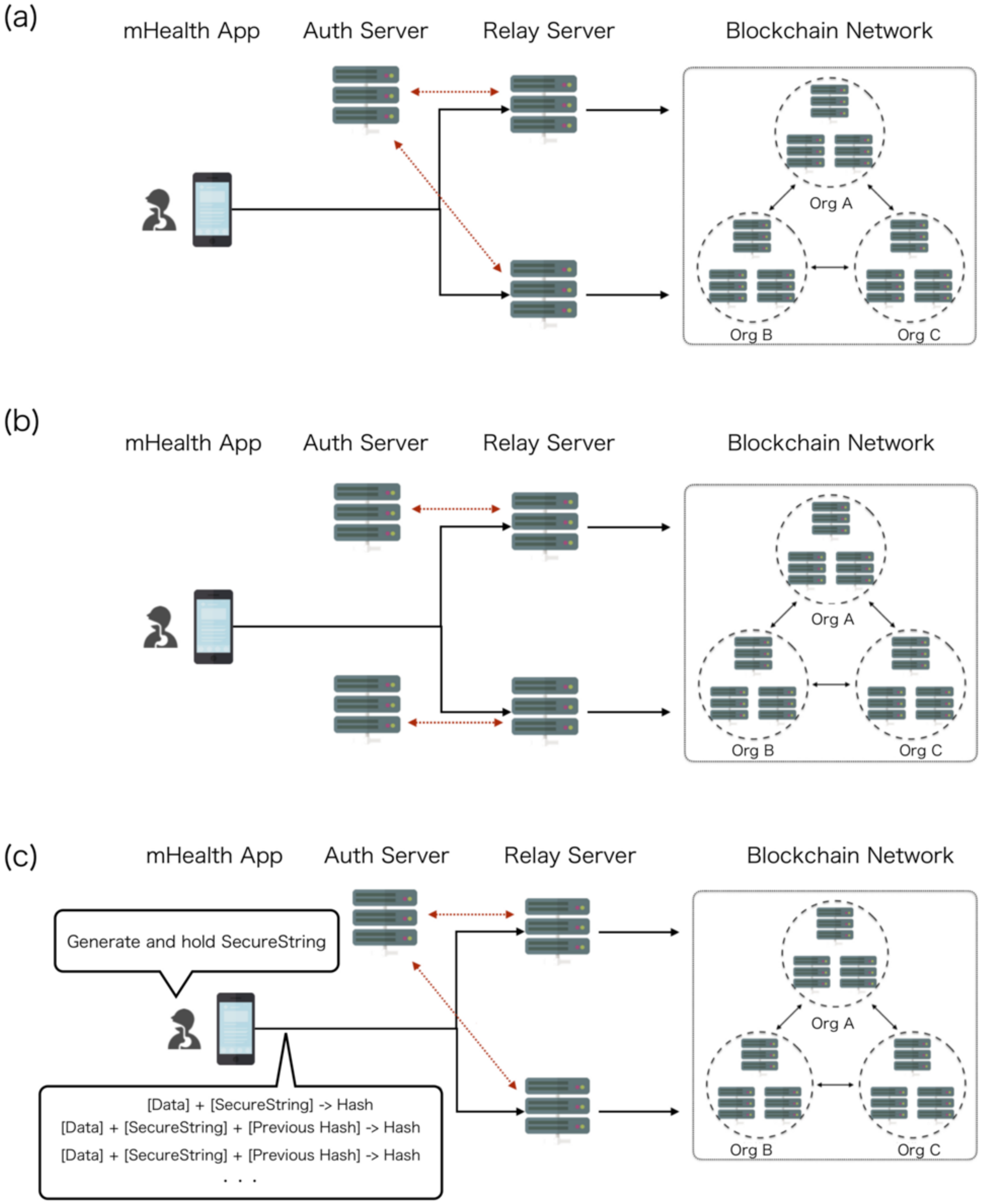
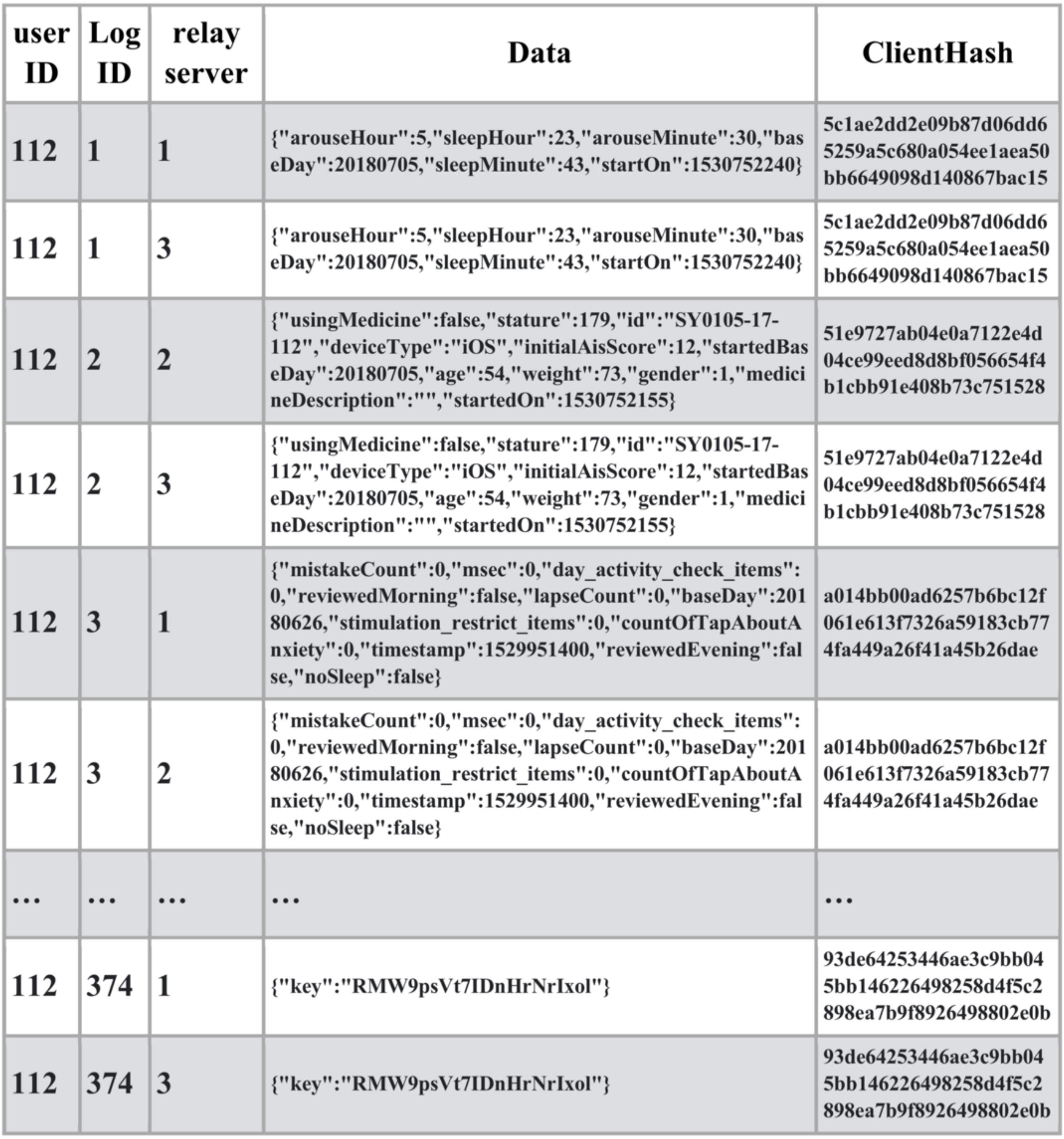


Figure 3. User data along with client hashchain registered to blockchain network.



user ID	Log ID	relay server	Data	ClientHash
112	1	1	{"arouseHour":5,"sleepHour":23,"arouseMinute":30,"baseDay":20180705,"sleepMinute":43,"startOn":1530752240}	5c1ae2dd2e09b87d06dd65259a5c680a054ee1aea50bb6649098d140867bac15
112	1	3	{"arouseHour":5,"sleepHour":23,"arouseMinute":30,"baseDay":20180705,"sleepMinute":43,"startOn":1530752240}	5c1ae2dd2e09b87d06dd65259a5c680a054ee1aea50bb6649098d140867bac15
112	2	2	{"usingMedicine":false,"stature":179,"id":"SY0105-17-112","deviceType":"iOS","initialAisScore":12,"startedBaseDay":20180705,"age":54,"weight":73,"gender":1,"medicineDescription":"","startedOn":1530752155}	51e9727ab04e0a7122e4d04ce99eed8d8bf056654f4b1cbb91e408b73c751528
112	2	3	{"usingMedicine":false,"stature":179,"id":"SY0105-17-112","deviceType":"iOS","initialAisScore":12,"startedBaseDay":20180705,"age":54,"weight":73,"gender":1,"medicineDescription":"","startedOn":1530752155}	51e9727ab04e0a7122e4d04ce99eed8d8bf056654f4b1cbb91e408b73c751528
112	3	1	{"mistakeCount":0,"msec":0,"day_activity_check_items":0,"reviewedMorning":false,"lapseCount":0,"baseDay":20180626,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1529951400,"reviewedEvening":false,"noSleep":false}	a014bb00ad6257b6bc12f061e613f7326a59183cb774fa449a26f41a45b26dae
112	3	2	{"mistakeCount":0,"msec":0,"day_activity_check_items":0,"reviewedMorning":false,"lapseCount":0,"baseDay":20180626,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1529951400,"reviewedEvening":false,"noSleep":false}	a014bb00ad6257b6bc12f061e613f7326a59183cb774fa449a26f41a45b26dae
…	…	…	…	…
112	374	1	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b
112	374	3	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b

Validation of the Resistance Against Attack on the Relay Server

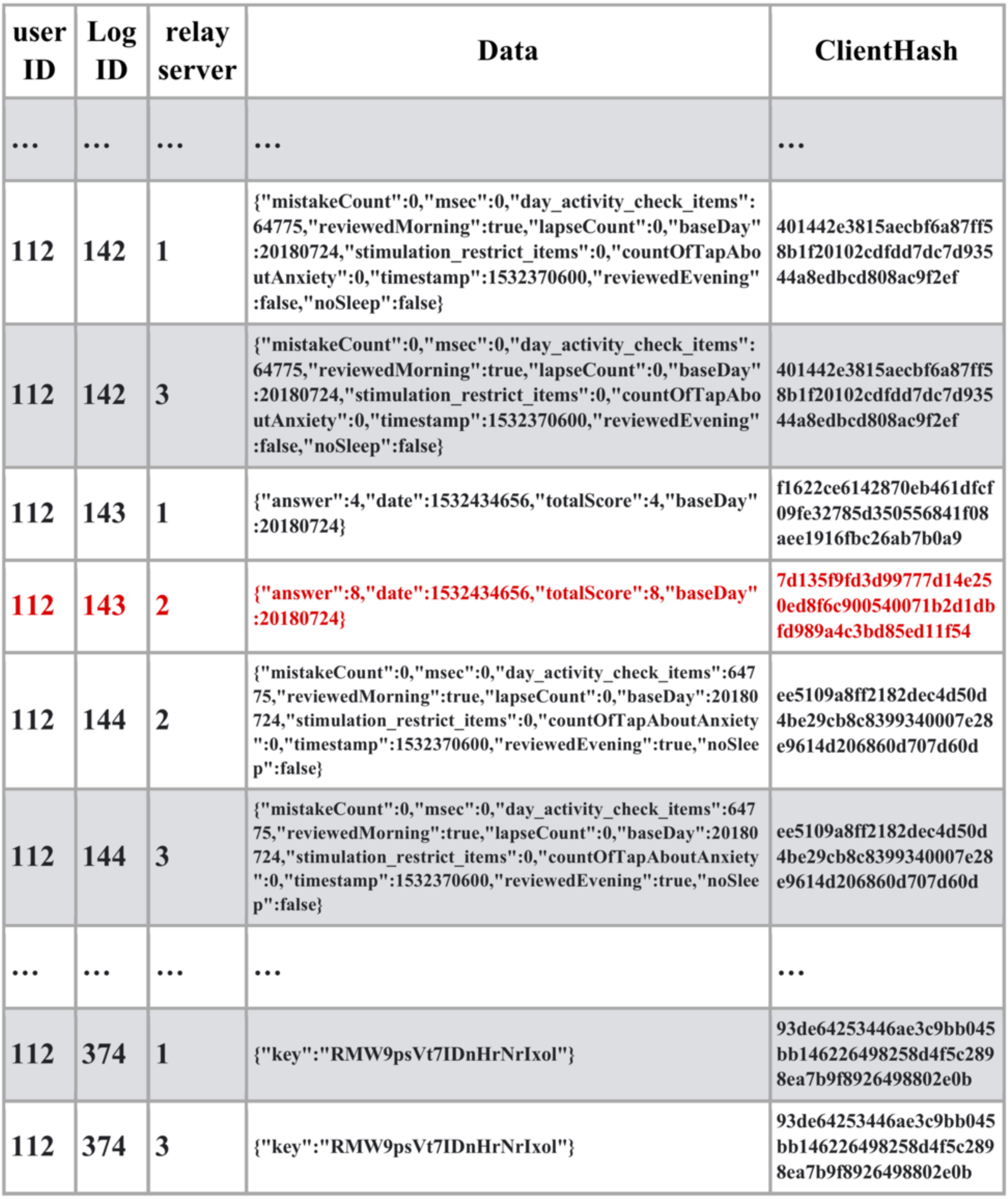
To investigate the resistance of our proposed system, we first simulated an artificial attack on the relay server (ie, outside actors) or misconduct by the owner of the relay server (ie, inside actors) and evaluated if the malicious access was detected and if the original data are distinguishable from the illegal data. As described above, the client device sends the data to the blockchain network via multiple relay servers. We used three relay servers and the app selected two relay servers at random to send the data to the blockchain network. If one of the relay servers is hacked, the attacker can modify the data sent from the client device prior to blockchain submission and steal the authorization token. In this case, the secure string used in the calculation of the client hashchain was not stolen by the attacker.

As shown in Figure 4, an attacker hacked relay server #2 and modified some medical data as well as the client hash value. In this case, the client hash value was calculated with the previous hash value, modified medical data, and the incorrect string. The access fraud can be detected by the mismatch between the data sent from other relay servers. It is also possible to distinguish and reject illegal data from original data automatically using the client hashchain. At the end of the clinical trial, the device sent the secure string to the blockchain to make it possible to validate the uploaded data retrospectively. Since all the medical data and the client hash values were stored in the blockchain network, it is not possible to rewrite the hash value based on the secure string, which was sent via the relay servers. Legal data can be guaranteed by combining the client hashchain, which rejects impersonation, with the blockchain, which provides tamper-proof history.

Validation of the Resistance Against Attack on the Authentication Server

We next simulated an artificial attack on the authentication server (ie, outside actors) or misconduct by the owner of the authentication server (ie, inside actors) during the clinical trial and evaluated whether the malicious access was detected if the original data are distinguishable from the illegal data. The authentication server possesses the authentication keys of every account. If the authentication key was stolen, the attacker can send illegal data from different devices. In this case, the secure string used in the calculation of the client hashchain was not stolen by the attacker.

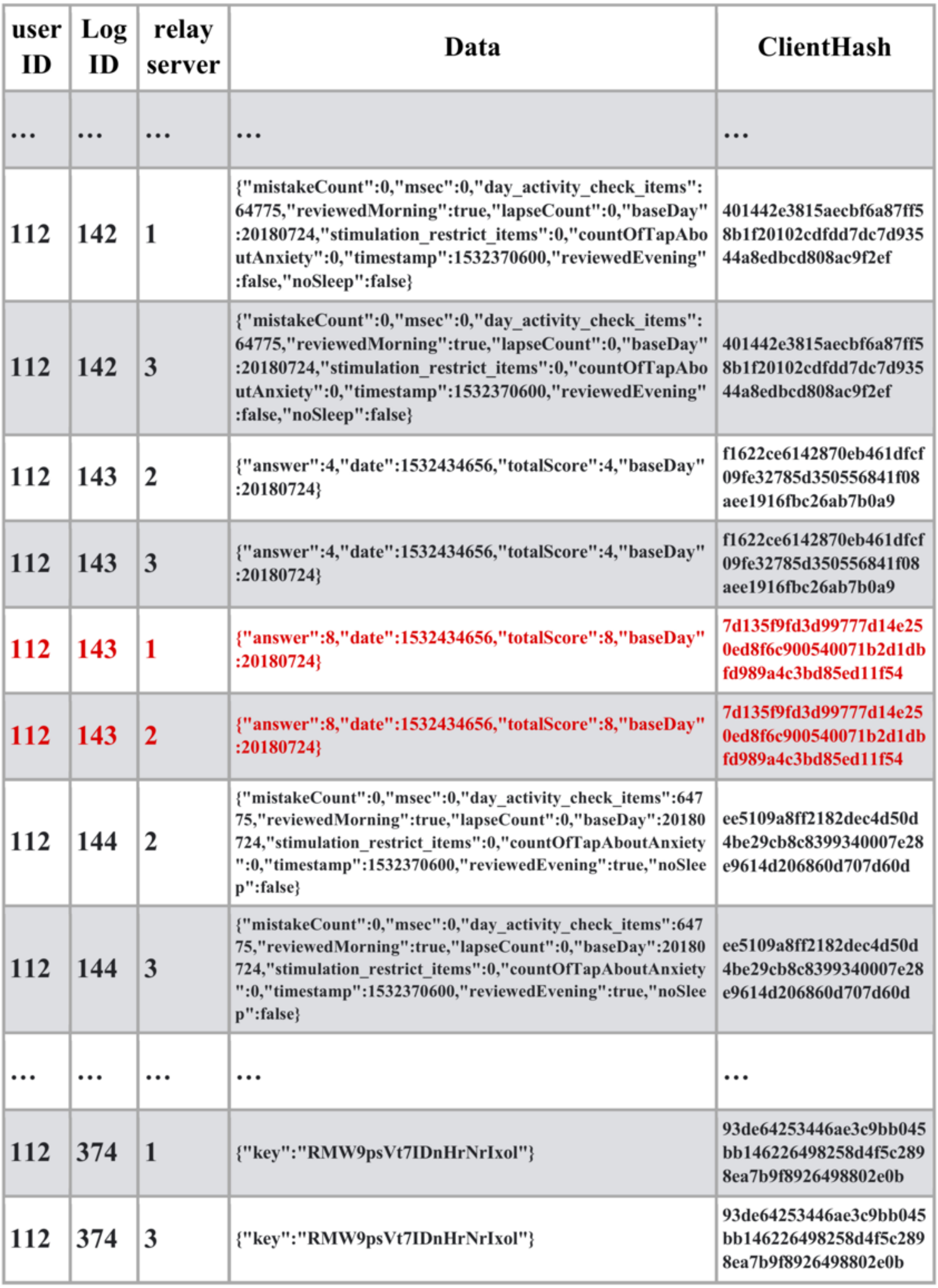
Figure 4. Relay server hacked and data modified by access fraud (client hash value in red).



user ID	Log ID	relay server	Data	ClientHash
…	…	…	…	…
112	142	1	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":false,"noSleep":false}	401442e3815aecbf6a87ff58b1f20102cdfdd7dc7d93544a8edbcd808ac9f2ef
112	142	3	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":false,"noSleep":false}	401442e3815aecbf6a87ff58b1f20102cdfdd7dc7d93544a8edbcd808ac9f2ef
112	143	1	{"answer":4,"date":1532434656,"totalScore":4,"baseDay":20180724}	f1622ce6142870eb461dfcf09fe32785d350556841f08aee1916fbc26ab7b0a9
112	143	2	{"answer":8,"date":1532434656,"totalScore":8,"baseDay":20180724}	7d135f9fd3d99777d14e250ed8f6c900540071b2d1dbfd989a4c3bd85ed11f54
112	144	2	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":true,"noSleep":false}	ee5109a8ff2182dec4d50d4be29cb8c8399340007e28e9614d206860d707d60d
112	144	3	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":true,"noSleep":false}	ee5109a8ff2182dec4d50d4be29cb8c8399340007e28e9614d206860d707d60d
…	…	…	…	…
112	374	1	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b
112	374	3	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b

As shown in Figure 5, using the stolen authentication key, multiple data with the same log ID were generated and sent to the blockchain network via relay servers by the attacker. Fraud detection can be achieved by identifying branched data. Furthermore, it is also possible to distinguish and reject illegal data from original data automatically using the client hashchain. At the end of the clinical trial, the device sent the secure string to the blockchain to enable the validation of the uploaded data retrospectively. Since all medical data and the client hash values were stored in the blockchain network, it is not possible to rewrite the hash value based on the secure string that was sent via relay servers.

Figure 5. Authentication server hacked and data generated from multiple devices using the authentication key (authentication key was stolen and the attacker created illegal data from different devices in red).



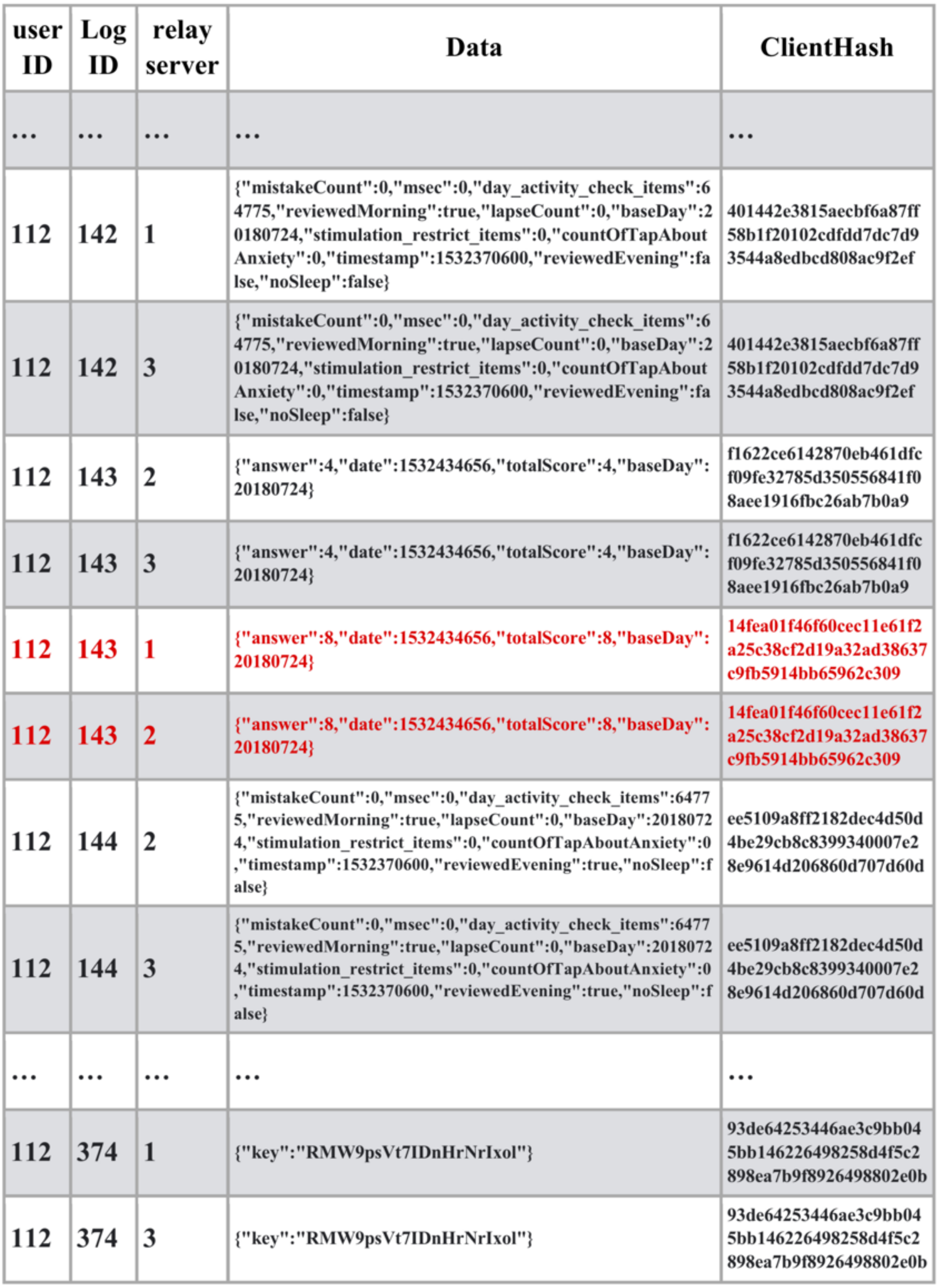
user ID	Log ID	relay server	Data	ClientHash
…	…	…	…	…
112	142	1	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":false,"noSleep":false}	401442e3815aecbf6a87ff58b1f20102cdfdd7dc7d93544a8edbcd808ac9f2ef
112	142	3	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":false,"noSleep":false}	401442e3815aecbf6a87ff58b1f20102cdfdd7dc7d93544a8edbcd808ac9f2ef
112	143	2	{"answer":4,"date":1532434656,"totalScore":4,"baseDay":20180724}	f1622ce6142870eb461dfcf09fe32785d350556841f08aee1916fbc26ab7b0a9
112	143	3	{"answer":4,"date":1532434656,"totalScore":4,"baseDay":20180724}	f1622ce6142870eb461dfcf09fe32785d350556841f08aee1916fbc26ab7b0a9
112	143	1	{"answer":8,"date":1532434656,"totalScore":8,"baseDay":20180724}	7d135f9fd3d99777d14e250ed8f6c900540071b2d1dbfd989a4c3bd85ed11f54
112	143	2	{"answer":8,"date":1532434656,"totalScore":8,"baseDay":20180724}	7d135f9fd3d99777d14e250ed8f6c900540071b2d1dbfd989a4c3bd85ed11f54
112	144	2	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":true,"noSleep":false}	ee5109a8ff2182dec4d50d4be29cb8c8399340007e28e9614d206860d707d60d
112	144	3	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":true,"noSleep":false}	ee5109a8ff2182dec4d50d4be29cb8c8399340007e28e9614d206860d707d60d
…	…	…	…	…
112	374	1	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b
112	374	3	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b

Validation of the Resistance Against Attack on the Client Device

To further investigate the resistance of our proposed system, we next simulated an artificial attack on the client device and evaluated if the malicious access was detected and if the original data are distinguishable from the illegal data. One of the most dangerous attacks is the malware root exploit, which enables the attacker to obtain the victim's private key. In this case, the authentication key as well as the secure string were stolen by the attacker, resulting in a more serious situation.

As shown in Figure 6, using the stolen authentication key, multiple data with the same log ID were generated and sent to the blockchain network. Fraud detection can be achieved by identifying the branched data. However, it is not possible to distinguish illegal data from original data automatically using the client hashchain because the attacker has stolen the secure string to calculate the client hash value. In this case, however, it is possible to judge which are the original data by checking the data in the patient's device offline, based on the fraud detection.

Figure 6. Client device washacked by malware root exploit and data generated from multiple devices using the authentication key (authentication key as well as the secure string for the client hash value were stolen by root exploit and the attacker created illegal data from different device in red).



user ID	Log ID	relay server	Data	ClientHash
…	…	…	…	…
112	142	1	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":false,"noSleep":false}	401442e3815aecbf6a87ff58b1f20102cdfdd7dc7d93544a8edbcd808ac9f2ef
112	142	3	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":false,"noSleep":false}	401442e3815aecbf6a87ff58b1f20102cdfdd7dc7d93544a8edbcd808ac9f2ef
112	143	2	{"answer":4,"date":1532434656,"totalScore":4,"baseDay":20180724}	f1622ce6142870eb461dfcf09fe32785d350556841f08aee1916fbc26ab7b0a9
112	143	3	{"answer":4,"date":1532434656,"totalScore":4,"baseDay":20180724}	f1622ce6142870eb461dfcf09fe32785d350556841f08aee1916fbc26ab7b0a9
112	143	1	{"answer":8,"date":1532434656,"totalScore":8,"baseDay":20180724}	14fea01f46f60cec11e61f2a25c38cf2d19a32ad38637c9fb5914bb65962c309
112	143	2	{"answer":8,"date":1532434656,"totalScore":8,"baseDay":20180724}	14fea01f46f60cec11e61f2a25c38cf2d19a32ad38637c9fb5914bb65962c309
112	144	2	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":true,"noSleep":false}	ee5109a8ff2182dec4d50d4be29cb8c8399340007e28e9614d206860d707d60d
112	144	3	{"mistakeCount":0,"msec":0,"day_activity_check_items":64775,"reviewedMorning":true,"lapseCount":0,"baseDay":20180724,"stimulation_restrict_items":0,"countOfTapAboutAnxiety":0,"timestamp":1532370600,"reviewedEvening":true,"noSleep":false}	ee5109a8ff2182dec4d50d4be29cb8c8399340007e28e9614d206860d707d60d
…	…	…	…	…
112	374	1	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b
112	374	3	{"key":"RMW9psVt7IDnHrNrIxol"}	93de64253446ae3c9bb045bb146226498258d4f5c2898ea7b9f8926498802e0b

Discussion

Principal Findings

In this study, we have developed a secure and scalable mHealth system using relay servers and blockchain combined with a client hashchain. Although blockchain technology provides tamper-resistance to medical data [24], the security was limited to the registered data and it cannot distinguish between original data and impersonated data. In addition, scalability will be compromised if the client devices were dealt with as a node of blockchain network. With our proposed system, we have shown that these problems can be resolved.

Bitcoin was the first implementation of blockchain as a digital asset in widespread use. Although bitcoin can be used as a platform for preventing data tampering, it is not appropriate since it is an open network and massive computing power is necessary for proof of work (PoW) to obtain consensus [43]. Private blockchain networks, such as Hyperledger Fabric, are more appropriate for the management of medical data since the node of stakeholders can be controlled. In addition, it is possible to process more transactions in a private blockchain without PoW. Using a private blockchain network, we used relay servers to send the data from authorized clients. With this architecture, it is not necessary to incorporate SDKs on client devices. To make the system robust against hacking of the relay server, the app sent data to the blockchain network via multiple relay servers. Even when one of the relay servers was hacked, we could detect access fraud and distinguish the original data from modified data. In our study, we used three relay servers and two were randomly selected to send the data. The robustness of the system against server attack can be increased if we use more relay servers, for instance, if three out of five relay servers are randomly selected to send the data.

To further clarify the origin of the data, we combined the client hashchain with the private blockchain. Hash values combined with the blockchain have been used as the metadata in a previous study for the management of rights for digital contents [44]. In contrast, we used hash values calculated on the client device to protect against impersonation and verify the origin of the data by chaining them. Hash values with chain structure (client hashchain) enable the identification of the original data sent from a specific client, which stores the secure string. In addition, in combination with the blockchain, the system also ensured tamper-resistance and the reliability of the hash values to prevent impersonation. We have shown that fraudulent data by compromised relay servers can be detected and distinguished from original data using the client hashchain. Even when the secure string used in calculating the client hash value was stolen by the attacker with root exploit [45], it is possible to detect the malformation in the branching of the chaincode. Based on the detection of the malformation, the researchers can ask the patients and check which are the original data. Therefore, the system is highly resistant against impersonation and tampering.

In this study, we designed the architecture for mHealth and verified the performance in a clinical study. Although mHealth is suitable for collecting medical data, such as patient reported outcomes [46], by changing the client device from patients' mobile phone to computers in medical institutions, the system can also be applied to clinical trials that use electronic data capture [47]. In addition, we could deploy smart contract, which is called chaincode in Hyperledger Fabric, to each node of the blockchain network to execute transactions. Since the smart contract may have the function to transform medical data into the determined format, it is possible to automatically complete the case report form of each patient if the app has access to additional medical data by deploying the smart contract for the clinical trial [25,48].

The system enables the verification of the accuracy of the medical data without confirmation by the third party, such as a contracted research organization, so that it is possible to reduce the cost of clinical trials as well as the possibility of human error. Thus, our system based on the blockchain technology combined with a client hashchain may enhance the development of drugs and medical devices.

Limitations

Further studies are needed to verify the scalability of the system for conducting multiple clinical trials simultaneously. In Hyperledger Fabric v1.0, it is possible to partition the network and define a communication channel using an ordering service, which enables multiple clinical trials to be conducted in the same system [49]. Although the transaction throughput in the Hyperledger Fabric platform that we used is much higher than public blockchain, one drawback of the private network described here is preventing 51% of attacks in networks that are composed of a limited number of nodes without public validation.

Although our system is resistant against impersonation and tampering, hacking of the client device is a great threat. Root exploit is a type of malware attackers use to modify the Android operating system kernel such that attackers are able to gain super-user privileges. When attackers gain root of the operating system kernel, they also gain access to full administrator privileges. Through this, attackers are able to install other malware types, such as botnets, worms, or Trojans into the system. Further studies like root exploit detection [45] may be beneficial for the improved security of the system.

Conclusion

In this study, we designed a secure and scalable mHealth system using blockchain. A client hashchain was combined with the blockchain network to protect against impersonation, enabling the usage of relay servers and reducing the complexity of authentication of client devices for mHealth. The system was validated in the clinical trial, and the resistance against various fraud attacks was evaluated.

Acknowledgments

This work was supported in part by the New Energy and Industrial Technology Development Organization and the Cabinet Office of Japan.

Authors' Contributions

TU designed the research; TH, KO, MK performed the research; TM, DI analyzed the data; and TM, TH, DI, and TU wrote the paper.

Conflicts of Interest

The authors are members of SUSMED, Inc.

References

1. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
2. Kierkegaard P. Electronic health record: Wiring Europe’s healthcare. Computer Law & Security Review 2011 Sep;27(5):503-515. [doi: 10.1016/j.clsr.2011.07.013]
3. Clarke R, Youngstein T. Cyberattack on Britain's National Health Service - A Wake-up Call for Modern Medicine. N Engl J Med 2017 Aug 03;377(5):409-411. [doi: 10.1056/NEJMp1706754] [Medline: 28591519]
4. Hao D, Patrick Y, Thomas D, Guannan G, Mingming K, Harlan K, et al. TrialChain: A Blockchain-Based Platform to Validate Data Integrity in Large, Biomedical Research Studies. arXiv 2018 Jul 10:03662 [FREE Full text]
5. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA 2003;289(4):454-465. [Medline: 12533125]
6. Orri M, Lipset CH, Jacobs BP, Costello AJ, Cummings SR. Web-based trial to evaluate the efficacy and safety of tolterodine ER 4 mg in participants with overactive bladder: REMOTE trial. Contemp Clin Trials 2014 Dec;38(2):190-197. [doi: 10.1016/j.cct.2014.04.009] [Medline: 24792229]
7. Pease AM, Krumholz HM, Downing NS, Aminawung JA, Shah ND, Ross JS. Postapproval studies of drugs initially approved by the FDA on the basis of limited evidence: systematic review. BMJ 2017 May 03;357:j1680 [FREE Full text] [doi: 10.1136/bmj.j1680] [Medline: 28468750]
8. Bhatt A. Quality of clinical trials: A moving target. Perspect Clin Res 2011 Oct;2(4):124-128 [FREE Full text] [doi: 10.4103/2229-3485.86880] [Medline: 22145122]
9. Sheehan JG. Fraud, conflict of interest, and other enforcement issues in clinical research. Cleve Clin J Med 2007 Mar;74 Suppl 2:S63-67; discussion S68. [Medline: 17471620]
10. Gupta A. Fraud and misconduct in clinical research: A concern. Perspect Clin Res 2013 Apr;4(2):144-147 [FREE Full text] [doi: 10.4103/2229-3485.111800] [Medline: 23833741]
11. Cai W, Wang Z, Ernst JB, Hong Z, Feng C, Leung VCM. Decentralized Applications: The Blockchain-Empowered Software System. IEEE Access 2018;6:53019-53033 [FREE Full text] [doi: 10.1109/ACCESS.2018.2870644]
12. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. URL: https://bitcoin.org/bitcoin.pdf [accessed 2019-05-05] [WebCite Cache ID 78A7OLdwe]
13. Kuo TT, Kim HE, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
14. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain Technology for Detecting Falsified and Substandard Drugs in Distribution: Pharmaceutical Supply Chain Intervention. JMIR Res Protoc 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: 10.2196/10163] [Medline: 30213780]
15. Mackey TK, Nayyar G. A review of existing and emerging digital technologies to combat the global trade in fake medicines. Expert Opin Drug Saf 2017 May;16(5):587-602. [doi: 10.1080/14740338.2017.1313227] [Medline: 28349715]
16. Tseng J, Liao Y, Chong B, Liao S. Governance on the Drug Supply Chain via Gcoin Blockchain. Int J Environ Res Public Health 2018 Dec 23;15(6):E1055 [FREE Full text] [doi: 10.3390/ijerph15061055] [Medline: 29882861]
17. Mertz L. (Block) Chain Reaction: A Blockchain Revolution Sweeps into Health Care, Offering the Possibility for a Much-Needed Data Solution. IEEE Pulse 2018;9(3):4-7. [doi: 10.1109/MPUL.2018.2814879] [Medline: 29757744]
18. Zhou L, Wang L, Sun Y. MIStore: a Blockchain-Based Medical Insurance Storage System. J Med Syst 2018 Jul 02;42(8):149 [FREE Full text] [doi: 10.1007/s10916-018-0996-4] [Medline: 29968202]
19. Kaur H, Alam MA, Jameel R, Mourya AK, Chang V. A Proposed Solution and Future Direction for Blockchain-Based Heterogeneous Medicare Data in Cloud Environment. J Med Syst 2018 Jul 10;42(8):156. [doi: 10.1007/s10916-018-1007-5] [Medline: 29987560]
20. Mense A, Athanasiadis L. Concept for Sharing Distributed Personal Health Records with Blockchains. Stud Health Technol Inform 2018;251:7-10. [Medline: 29968588]
21. Wang H, Song Y. Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain. J Med Syst 2018 Jul 05;42(8):152. [doi: 10.1007/s10916-018-0994-6] [Medline: 29974270]

22. Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials 2017 Dec 19;18(1):335 [FREE Full text] [doi: 10.1186/s13063-017-2035-z] [Medline: 28724395]
23. Maslove DM, Klein J, Brohman K, Martin P. Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
24. Ichikawa D, Kashiyama M, Ueno T. Tamper-Resistant Mobile Health Using Blockchain Technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
25. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Dec 22;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]
26. Zhuang Y, Sheets L, Shae Z, Tsai JJP, Shyu CR. Applying Blockchain Technology for Health Information Exchange and Persistent Monitoring for Clinical Trials. AMIA Annu Symp Proc 2018;2018:1167-1175 [FREE Full text] [Medline: 30815159]
27. Angraal S, Krumholz HM, Schulz WL. Blockchain Technology: Applications in Health Care. Circ Cardiovasc Qual Outcomes 2017 Dec;10(9):E003800. [doi: 10.1161/CIRCOUTCOMES.117.003800] [Medline: 28912202]
28. Roehrs A, da Costa CA, da Rosa Righi R. OmniPHR: A distributed architecture model to integrate personal health records. J Biomed Inform 2017 Dec;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
29. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
30. Zhang A, Lin X. Towards Secure and Privacy-Preserving Data Sharing in e-Health Systems via Consortium Blockchain. J Med Syst 2018 Jun 28;42(8):140. [doi: 10.1007/s10916-018-0995-5] [Medline: 29956061]
31. Jacobs GD, Pace-Schott EF, Stickgold R, Otto MW. Cognitive behavior therapy and pharmacotherapy for insomnia: a randomized controlled trial and direct comparison. Arch Intern Med 2004 Sep 27;164(17):1888-1896. [doi: 10.1001/archinte.164.17.1888] [Medline: 15451764]
32. Basner M, Dinges DF. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. Sleep 2011 May 01;34(5):581-591 [FREE Full text] [Medline: 21532951]
33. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res 2000 Jun;48(6):555-560. [Medline: 11033374]
34. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991 Dec;14(6):540-545. [Medline: 1798888]
35. Biggs MM, Shores-Wilson K, Rush AJ, Carmody TJ, Trivedi MH, Crismon ML, et al. A comparison of alternative assessments of depressive symptom severity: a pilot study. Psychiatry Res 2000 Nov 20;96(3):269-279. [Medline: 11084222]
36. Christian C. IBM research. Architecture of the Hyperledger Blockchain Fabric URL: https://www.zurich.ibm.com/dccl/papers/cachin_dccl.pdf [accessed 2019-05-06] [WebCite Cache ID 78A8lYPTB]
37. Hossein K. Lexology. The Blockchain Revolution: An Analysis of Regulation and Technology Related to Distributed Ledger Technologies URL: https://www.lexology.com/library/detail.aspx?g=5adaf528-b9ad-4575-8933-4cb033a2536d [accessed 2019-05-05] [WebCite Cache ID 78A91hLdN]
38. Abiteboul S. Web Data Management. London, UK: Cambridge University Press; 2012.
39. Shay G. Speeding Up SHA-1, SHA-256 and SHA-512 on the 2nd Generation Intel® Core™ Processors. In: Ninth International Conference on Information Technology - New Generations. 2012 Presented at: International Conference on Information Technology - New Generations; April 16-18, 2012; Las Vegas, NV URL: https://ieeexplore.ieee.org/document/6209073 [doi: 10.1109/ITNG.2012.62]
40. Castro M, Liskov B. Practical byzantine fault tolerance and proactive recovery. ACM Trans. Comput. Syst 2002;20(4):398-461 [FREE Full text] [doi: 10.1145/571637.571640]
41. Castro M, Liskov B. 3rd Symposium on Operating Systems Design and Implementation (OSDI) Proceedings. Berkeley, CA: USENIX Association; 1999.
42. Gilbert H, Handschuh H. Security Analysis of SHA-256 and Sisters. In: International Workshop on Selected Areas in Cryptography.: Springer; 2003 Presented at: International Workshop on Selected Areas in Cryptography; Aug. 14-15, 2003; Ottawa, ON, Canada p. 175-193 URL: https://link.springer.com/chapter/10.1007/978-3-540-24654-1_13
43. Michael C. Applied Innovation Review. 2016. BlockChain Technology: Beyond Bitcoin URL: http://scet.berkeley.edu/wp-content/uploads/AIR-2016-Blockchain.pdf [accessed 2019-05-06] [WebCite Cache ID 78AAoBhUS]
44. Shigeru F, Hiroki W, Atsushi N, Tomokazu Y, Akihito A, Junichi K. BRIGHT: A concept for a decentralized rights management system based on blockchain. In: IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin). 2015 Presented at: IEEE 5th International Conference on Consumer Electronics; Sept. 6-9, 2015; Berlin, Germany URL: https://ieeexplore.ieee.org/document/7391275 [doi: 10.1109/ICCE-Berlin.2015.7391275]
45. Firdaus A, Anuar NB, Razak MFA, Hashem IAT, Bachok S, Sangaiah AK. Root Exploit Detection and Features Optimization: Mobile Device and Blockchain Based Medical Data Management. J Med Syst 2018 May 04;42(6):112. [doi: 10.1007/s10916-018-0966-x] [Medline: 29728780]

46. Basch E. The missing voice of patients in drug-safety reporting. N Engl J Med 2010 Mar 11;362(10):865-869 [FREE Full text] [doi: 10.1056/NEJMp0911494] [Medline: 20220181]
47. Rorie DA, Flynn RWV, Grieve K, Doney A, Mackenzie I, MacDonald TM, et al. Electronic case report forms and electronic data capture within clinical trials and pharmacoepidemiology. Br J Clin Pharmacol 2017 Sep;83(9):1880-1895 [FREE Full text] [doi: 10.1111/bcp.13285] [Medline: 28276585]
48. Ene-Iordache B, Carminati S, Antiga L, Rubis N, Ruggenenti P, Remuzzi G, et al. Developing regulatory-compliant electronic case report forms for clinical trials: experience with the demand trial. J Am Med Inform Assoc 2009;16(3):404-408 [FREE Full text] [doi: 10.1197/jamia.M2787] [Medline: 19261946]
49. Fabrice B, Shai H, Tzipora H. Supporting Private Data on Hyperledger Fabric with Secure Multiparty Computation. In: 2018 IEEE International Conference on Cloud Engineering (IC2E). 2018 Presented at: 2018 IEEE International Conference on Cloud Engineering (IC2E); April 17-20, 2018; Orlando, FL URL: http://paperpile.com/b/1Cd2iA/77nZK [doi: 10.1109/IC2E.2018.00069]

Abbreviations

CBTi: cognitive behavioral therapy for insomnia
IoT: internet of things
IP: Internet Protocol
JSON: JavaScript Object Notation
PoW: proof of work
SDK: software development kit
SHA: secure hash algorithm

Edited by P Zhang, K Clauson; submitted 24.01.19; peer-reviewed by YC Lin, W Schulz, MS Aslam; comments to author 25.02.19; revised version received 23.03.19; accepted 27.04.19; published 16.05.19.

Please cite as:
Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T
Secure and Scalable mHealth Data Management Using Blockchain Combined With Client Hashchain: System Design and Validation
J Med Internet Res 2019;21(5):e13385
URL: http://www.jmir.org/2019/5/e13385/
doi:10.2196/13385
PMID:31099337

Original Paper

Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy

Amelia Fiske[1], BA, PhD; Peter Henningsen[2], PhD, MD; Alena Buyx[1], MPhil, PhD, MD

[1]Institute for History and Ethics of Medicine, Technical University of Munich School of Medicine, Technical University of Munich, Munich, Germany
[2]Department of Psychosomatic Medicine and Psychotherapy, Klinikum rechts der Isar at Technical University of Munich, Munich, Germany

Corresponding Author:
Amelia Fiske, BA, PhD
Institute for History and Ethics of Medicine
Technical University of Munich School of Medicine
Technical University of Munich
Ismaninger Straße 22
Munich, 81675
Germany
Phone: 49 8941404041
Email: a.fiske@tum.de

Abstract

Background: Research in embodied artificial intelligence (AI) has increasing clinical relevance for therapeutic applications in mental health services. With innovations ranging from ‘virtual psychotherapists’ to social robots in dementia care and autism disorder, to robots for sexual disorders, artificially intelligent virtual and robotic agents are increasingly taking on high-level therapeutic interventions that used to be offered exclusively by highly trained, skilled health professionals. In order to enable responsible clinical implementation, ethical and social implications of the increasing use of embodied AI in mental health need to be identified and addressed.

Objective: This paper assesses the ethical and social implications of translating embodied AI applications into mental health care across the fields of Psychiatry, Psychology and Psychotherapy. Building on this analysis, it develops a set of preliminary recommendations on how to address ethical and social challenges in current and future applications of embodied AI.

Methods: Based on a thematic literature search and established principles of medical ethics, an analysis of the ethical and social aspects of currently embodied AI applications was conducted across the fields of Psychiatry, Psychology, and Psychotherapy. To enable a comprehensive evaluation, the analysis was structured around the following three steps: assessment of potential benefits; analysis of overarching ethical issues and concerns; discussion of specific ethical and social issues of the interventions.

Results: From an ethical perspective, important benefits of embodied AI applications in mental health include new modes of treatment, opportunities to engage hard-to-reach populations, better patient response, and freeing up time for physicians. Overarching ethical issues and concerns include: harm prevention and various questions of data ethics; a lack of guidance on development of AI applications, their clinical integration and training of health professionals; ‘gaps’ in ethical and regulatory frameworks; the potential for misuse including using the technologies to replace established services, thereby potentially exacerbating existing health inequalities. Specific challenges identified and discussed in the application of embodied AI include: matters of risk-assessment, referrals, and supervision; the need to respect and protect patient autonomy; the role of non-human therapy; transparency in the use of algorithms; and specific concerns regarding long-term effects of these applications on understandings of illness and the human condition.

Conclusions: We argue that embodied AI is a promising approach across the field of mental health; however, further research is needed to address the broader ethical and societal concerns of these technologies to negotiate best research and medical practices in innovative mental health care. We conclude by indicating areas of future research and developing recommendations for high-priority areas in need of concrete ethical guidance.

(J Med Internet Res 2019;21(5):e13216) doi:10.2196/13216

KEYWORDS

artificial intelligence; robotics; ethics; psychiatry; psychology; psychotherapy; medicine

Introduction

Research in embodied artificial intelligence (AI) has increasing clinical relevance for therapeutic applications in mental health services, that is, in psychiatry, psychology, and psychotherapy. Innovations range from ‘virtual psychotherapists’ [1] to social robots in dementia care and autism disorder [2] and robots for sexual disorders [3]. Increasingly, artificially intelligent virtual and robotic agents are not only available for relatively low-level elements of mental health support, such as comfort or social interaction, but also perform high-level therapeutic interventions that used to be offered exclusively by highly trained, skilled health professionals such as psychotherapists [4]. Importantly, such ‘virtual’ or ‘robotic therapists’ include an artificially intelligent algorithm that responds independently of any expert human guidance to the client or patient through a virtually embodied presence, such as a face icon, or a physically embodied presence, such as a robotic interface. As such, these emerging applications are distinct from the many varieties of Web-based therapy, which usually involve either a human therapist, albeit remotely (telemedicine), or the patient herself, working independently with manuals, questionnaires, or other self-help materials [5].

Embodied AI applications in mental health care carry hopes of improving quality of care and controlling expenditure [6]. In addition, they also hold the promise of reaching underserved populations in need of mental health services and improving life opportunities for vulnerable groups. However, there is a persistent gap between current, rapid developments in AI mental health and the successful adoption of these tools into clinical environments by health professionals and patients. In addition, it has been demonstrated that the interventions are often designed without any explicit ethical considerations [7]. Furthermore, although studies often examine the effectiveness or ethical use of a single application, rarely do they consider the implications for the integration of AI across the field of mental health more broadly. In this paper, we argue that virtually and physically embodied artificially intelligent agents and applications have great potential in mental health care. However, their societal and ethical implications require further probing to identify pertinent concerns surrounding trust, privacy, and autonomy, as well as to anticipate concerns that may arise in the future. Identifying the broader ethical and societal implications of embodied AI is crucial for negotiating best research and medical practices in innovative mental health care. We conclude by indicating areas of future research and identifying points in need of ethical caution.

Overview: Existing Embodied Intelligent Applications

Although AI-enabled virtual and robot therapy has long been used across a number of medical fields [8-10], the integration of AI through the use of embodied agents is still at an early stage in mental health care; it is arguably the most recent addition to psychotherapeutic practice, supporting a host of emotional, cognitive, and social processes [11]. In what follows, we have sketched a range of applications with the aim of characterizing some of the embodied artificially intelligent innovations across the field of mental health. To maintain focus amid a broad and growing field, we have chosen to exclude from our analysis applications that are not intended to interact with patients, or that have no virtual presence or robotic interface; this includes AI-supported scanning and diagnostic tools. We have also excluded applications that may have a virtual or robotic interface but do not employ AI, such as telemedicine therapy (for further scholarship on this topic, please see [12-19]).

Virtually Embodied Artificially Intelligent Agents

AI-supported virtually embodied psychotherapeutic devices are currently developing at a rapid speed. For example, therapeutic apps such as Tess and other “chatbots” such as Sara, Wysa, and Woebot, which work over short message service text messaging, WhatsApp, or internet platforms, are being explored for addressing depression and anxiety. These applications come with interactive screen presences. Woebot and other programs engage with the patient like a virtual psychotherapist, with the aim of helping patients to recognize their emotions and thought patterns and to develop skills such as resilience or techniques for reducing anxiety. For example, using natural language processing, Tess is programed to flag expressions that indicate emotional distress. Often cited as a digital tool to reach underserved populations across the world that lack mental health services, the bots can explain to users the clinical terms for what they are experiencing—such as cognitive distortions—or provide concrete advice for recognizing and dealing with difficult situations [20]. Initial studies found that depression symptoms decreased with the use of Woebot more than groups who relied on electronic book resources [21], and another study found that Tess helped to reduce depression and anxiety among users [20].

A similar approach involves the use of avatars, such as the Avatar Project, for addressing persistent auditory hallucinations for patients with psychosis [22]. These usually involve computer-generated images of faces on computer screens or tablets that interact with a patient via intelligent algorithms. Avatars are also being explored in treatment of schizophrenia, for example, to improve medication adherence [23]. Similar to the Avatar Project, virtual reality–assisted therapy for schizophrenia often encourages patients to engage with the voices they hear through the use of an AI avatar. Initial studies found that the therapy could help in developing therapeutic targets [24] and also in particularly difficult cases of schizophrenia [25]. Another study found improvements in auditory visual hallucinations, symptoms of depression, and overall quality of life following therapy sessions for treatment-resistant schizophrenia patients [26]. ‘Avatar coaches’ have also been employed as part of an immersive virtual reality situation for treating the fear of heights [27] or as ‘virtual patients’ to provide medical students with lifelike interviewing practice [28]. Finally, avatars are also being implemented in risk prevention education, such as the Kognito program, which uses an avatar to help college students and faculty identify others at risk for suicide [29].

Artificially Intelligent Robot Therapy

In addition to these virtually embodied therapeutic applications, clinicians and scientists are exploring the translation of innovations at the intersection of AI and robotics into the clinic. For example, intelligent animal-like robots such as Paro, a fuzzy harp seal, are increasingly being used to help patients with dementia. Paro, along with the large furry eBear, is part of a class of ‘companion bots,’ engaging individuals as at-home health care assistants, responding to speech and movement with dynamic ‘dialog’, or seeking to help elderly, isolated, or depressed patients through companionship and interaction. Several studies have examined the role of such robots in reducing stress, loneliness, and agitation and in improving mood and social connections [30,31]. Thus far, the outcomes are promising [32,33].

AI robots also provide opportunities for different forms of engagement with children suffering from autism spectrum disorders (ASDs) [34]. Children with autism have been found to react positively to robots, even in cases where they have trouble interacting with others [35]. The Kaspar robot has demonstrated potential for integration in current education and therapy interventions [36] and is being investigated for the potential to improve social skills among children [37]. Similarly, RoboTherapy is an example of socially assistive robotics designed to help children with ASDs to develop social skills, and the robot Nao is designed to improve facial recognition and appropriate gaze response. The aim of such robotic interaction is to learn appropriate social skills (eg, imitation, taking turns, staying engaged, and empathy), with the hope that children can then apply the skills learned with the robot peer to their relationships with human peers. Initial studies are promising; individuals with ASDs performed better with their robot partners than human therapists, responded with social behaviors toward robots, and improved spontaneous language during therapy sessions [38]. However, the devices are still being developed and are not yet in wider therapeutic use.

AI-enabled robots are also being explored across a variety of other mental health areas including mood and anxiety disorders, children with disruptive behavior, and patients who may not have a specific diagnosis but who would benefit from assistance with mental health concerns [39]. Perhaps, most controversially, artificially intelligent robots have entered the field of human sexuality. Companies are now offering adult sex robots such as Roxxxy, which can speak, learn their human partners’ preferences, register touch, and provide a form of intimate companionship. Although the range of medical applications that sex robots can reportedly address remains debated, these include meeting the sexual needs of disabled and elderly individuals or as part of therapy for concerns such as erectile dysfunction, premature ejaculation, and anxiety surrounding sex [40]. Furthermore, some researchers have asked if sex robots could help to reduce sex crimes such as rape and assault or be used for treatment of paraphilia, such as pedophilia [3,41].

Methods

Based on a thematic literature analysis and established principles of medical ethics, an analysis of the ethical and social aspects of currently embodied AI applications was conducted across the fields of Psychiatry, Psychology, and Psychotherapy. To enable a comprehensive evaluation, the analysis was structured around the following three steps: assessment of potential benefits; analysis of overarching ethical issues and concerns; discussion of specific ethical and social issues of the interventions.

Results

Ethical and Social Implications and Concerns

The devices and applications described above have yet to be integrated into widespread clinical use. However, in view of the speed of research and development trajectories of these applications, it is reasonable to expect that therapeutic chatbots, avatars, socially assistive devices, and sex robots will soon translate into broader clinical applications in earnest. In some cases, initial ethical assessments are already available [42,43]; however, most of these studies focus on a single application. In general, for most of the applications we are discussing, large-scale rigorous research studies have not yet been conducted or are still in pilot stages [44-47]. Even in Web-based non-AI applications, evidence of patient acceptance and treatment outcomes in routine care is still limited and mixed [48,49]; there has so far been very little research on patient acceptance and contingent treatment outcomes of embodied AI applications in mental health fields. As with any medical innovation, the effects, impacts, and clinical utility of the applications can only be fully assessed once evidence has improved [50,51].

To enable responsible and responsive innovation and clinical translation into the field of mental health, further and more in-depth analysis of the ethical and social implications of embodied AI is necessary to flag areas of concern. Early identification of ethical issues can help researchers, designers, and developers consider these concerns in the design and construction of the next generation of AI agents and robots for mental health. In the following sections, we have provided an analysis of benefits, challenges, and risks of embodied AI in mental health from an ethical perspective. Beginning with a discussion of potential benefits, we have then turned to risks and challenges, followed by immediate concerns in clinical application and long-term effects.

Anticipated Benefits

All of the aforementioned intelligent applications promise significant benefits for the field of mental health, satisfying many aspects of the ethical principle of beneficence [52]. From a clinical point of view, the use of embodied AI applications holds the potential to open new avenues for intervention in places where there are still significant unmet health needs. AI interventions might be particularly well placed for detecting mental health concerns early on, for reaching high-risk groups such as veterans, or for those who are concerned about the social stigma associated with psychotherapy [53]. In some cases, patients may respond positively and productively to the fact that the counterpart is *not* a human therapist [54-56]. In one study, subjects overwhelmingly preferred the virtual agent over the human counterpart when being discharged from the hospital because they could self-direct the pace of information—something that is especially important for low-literacy patients [57]. Thus, in mental health services, nonhuman virtual or robotic applications might be preferable for some patients, reducing embarrassment when asking for specific information or services or feelings of shame when admitting noncompliance with a treatment plan. Embodied AI in mental health could also help to empower particular patient groups (such as those who are less familiar with the medical system), thereby helping to improve trust and openness between patients and the medical system. Another important advantage of AI applications is that many of them are low-threshold and self-administered, such that people who do not have an acute condition can elect services without going through the time-consuming process of being screened and admitted into the health care system.

Arguably, the greatest benefit of AI applications is structural, namely the potential to reach populations that are difficult to treat via traditional routes of provision. The provision of some mental health services, for example, through low-threshold, convenient therapeutic interventions via chatbots or avatars may be particularly beneficial for populations living in resource-poor settings. For those living in remote or rural locations or in settings where on-site mental health services are scarce, intelligent applications can increase geographical access and provide some minimal mental health care services where they are otherwise absent. The same may also be true for individuals living in higher income countries who do not have insurance or whose insurance does not cover therapy. Furthermore, it is likely that there are individuals who, for various reasons, do not respond to more traditional clinical services and might prefer low-threshold interventions that can be conducted in the privacy of their homes or on the go. For all of these patients, AI applications could complement existing services or constitute an entry point for pursuing more standard clinical interventions in the future.

In sum, embodied AI interventions may offer entirely new modes of treatment that are potentially more successful than traditional modalities either because they address hard-to-reach populations or because patients respond better to them. Whether, and for which conditions this is the case, requires further investigation. However, given that broadly speaking, conditions such as ASD and sexual dysfunction are increasing in incidence and patient populations with these and many other mental health conditions continue to have unmet health needs [58-62], increased exploration of embodied AI in these fields is promising.

Finally, there are also clear benefits of having a virtual or robotic therapist that is always accessible, has endless amounts of time and patience, never forgets what a patient has said, and does not judge [63,64], thus potentially offering a service that is highly reliable and particularly well-suited to certain patient populations. If integrated into a scaled provision of services, AI-enabled applications could provide support for mild cases of depression and other nonacute conditions [65], therefore helping health professionals to devote more time to the most severe cases. In view of overall increasing burden of illness in mental health and against a background of limited resources, these are important benefits to consider. However, it is likely that embodied AI may not be warmly received by all mental health care professionals, and some may even have serious misgivings about its use because of ethical or clinical concerns. Thus far, there has been no substantial review of the reception of AI across or within specific mental health fields, marking an area in need of further research.

Overarching Ethical Concerns

Harm Prevention and Data Ethics Issues

To satisfy the well-established ethical principle of nonmaleficence, more robust research is needed on embodied AI applications in mental health to prevent harm both within therapeutic encounters and in cases where robots could malfunction or operate in unpredictable ways. For instance, in interviewing respondents working with AI robotic technologies, Cresswell et al discuss an example of a woman who was stuck in an elevator with a robot and another who was run over by a robot [6]. Chatbots and avatars could also stop working or malfunction. Hence, it needs to be discussed if embodied AI devices—potentially including virtual agents and freely available mental health applications—should require the same kind of rigorous risk assessment and regulatory oversight that other medical devices are subject to before they are approved for clinical use.

Similar to other devices employed in medical settings, the use of any AI applications in mental health care requires careful consideration surrounding data security of devices that communicate personal health information, the ways that the data generated is used, and the potential for hacking and nonauthorized monitoring [66,67]. Clear standards are needed on issues surrounding confidentiality, information privacy, and secure management of data collected by intelligent virtual agents and assistive robots as well as their use for monitoring habits, movement, and other interactions [68,69]. Concerns around privacy may be amplified as the amount of data collected continues to expand; for example, we anticipate that applications that integrate video data would need to have specific privacy protections in place for the communication of sensitive information, or information pertaining to individuals other than the consenting patient.

Lack of Guidance on Development, Clinical Integration, and Training

With embodied AI being one of the newest and most rapidly changing areas of psychological and psychiatric research and treatment, existing legal and ethical frameworks are often not closely attuned to these changes. Rather than providing regulatory guidance, there is the risk that the ‘gaps’ between application and ethical frameworks would only be addressed once harm had already occurred [6]. Again, this is the case with many forms of emerging medical technologies; however, in view of the rapid pace of translation of embodied AI into practice in settings where traditional health technology assessment and medical oversight systems are not fully applicable—for example, through freely available therapy chatbots—this is an important concern. Although anticipating the ethical and legal questions that will emerge alongside future developments is difficult, active reflection on the ‘regulatory fit’ for embodied mental health AI is necessary. Initiatives for establishing guidelines are emerging, including the online collaboratively generated document “Moral Responsibility for Computing Artifacts: The Rules,” or the recent “An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations” [70]. However, thus far, no guidance exists that is specific to the field of mental health services; pointing to the need for the development of further recommendations to better guide advances in this area.

In addition to a lack of guidance on the development of these interventions for design, use, and regulatory questions, so far, there are also no frameworks available on how medical professionals can effectively engage with and train for increased use of embodied AI in the clinic; that is, although there is an increasing body of both academic and popular literature on how embodied AI can be integrated into clinical practice in mental health, there remains a lack of high-level guidance from professional bodies on the best use of AI in mental health services [15,71-73]. There are also no recommendations available on how to train and prepare young doctors for a mental health field in which such tools will increasingly be used by patients. Thus, further ethical guidelines are needed that are specific to assisting mental health professionals who will be supervising patients who have, or possibly will, engage with AI services.

Potential for Misuse to Reduce Service Provision

An ethically informed integration of AI should also consider questions of a just provision of mental health care [52]. There is the worry that the incorporation of embodied AI in mental health could be a justification for replacing established services, resulting in fewer available health resources or principally AI-driven services, thereby potentially exacerbating existing health inequalities. Many proponents insist that although informed by evidence-based psychotherapeutic approaches, chatbots, for instance, are not intended to replace therapists entirely. In some cases, forms of ‘blended’ care involving both in-person and virtual forms of therapy are being explored [74], which might also be appropriate for intelligent applications. Blended care models potentially offer the opportunity to draw on the strengths of both AI applications and in-person clinical supervision. However, whether or not it is appropriate to implement AI applications in mental health care depends in part on the availability of other resources in that area. As noted, in cases with limited mental health services, AI applications could provide a needed resource that is decidedly better than no services at all. However, at this point, AI mental health services are not a substitute or a stand-in for the kind of robust, multitiered mental health care available in high-resource health care systems. Appropriately considering the status quo of mental health resources in each context is thus highly relevant from an ethical perspective [75]. Otherwise, AI tools in mental health could be used as an excuse for reducing the provision of high-quality, multilayered care by trained mental health professionals in low-resource settings.

Discussion

Specific Challenges in Application

Risk-Assessment, Referrals, and Supervision

Considering the application of embodied AI tools in mental health practice, a host of specific challenges need to be kept in mind: mental health professionals have an ethical responsibility to inform other service providers as well as third parties or authorities if a patient indicates that they are a threat to themselves or to another individual. How this would work in artificially intelligent interventions, particularly when there is no supervision of the interaction between the AI agent and the patient by a qualified health professional, remains to be determined. It is unclear when, and how, assistive robots that patients have in their homes, or freely available virtual agents and chatbots, would effectively connect at-risk individuals with appropriate services, including hospitalization and other protections. This scenario is particularly relevant in the aforementioned situation of using AI mental health applications to extend access to rural, hard to reach, or uninsured populations. In these cases, some provision of service is arguably better than nothing. However, what should be done if, for example, a therapy bot detects through speech patterns that an individual is at higher risk for self-harm, yet appropriate referral services are not available in the area?

AI applications engaged in therapeutic relationships with clients will likely also need to be bound by similar ethical guidelines as those that bind mental health professionals. However, so far, how an AI duty of care or a code of practice on reporting harm should be operationalized is entirely unclear. An obvious suggestion would be to always mandate supervision through a qualified mental health clinician—when a human therapist evaluates a patient’s expression of self-harm, she also considers contextual information in her interpretation of the level of risk. Whether, and to what degree, robotic therapists are able to do this remains unclear. However, many AI applications are available outside established mental health settings; in addition, the capacity of computerized methods to identify and predict psychiatric illness are increasing [12,13], as are their skills of therapeutic interaction and communication. Thus, the question of whether supervision of embodied AI in health should always be provided, and how such a requirement could be successfully implemented, remains a subject for further debate.

Respecting and Protecting Patient Autonomy

Another concern for the application of embodied AI in mental health practice centers on enabling and respecting patient autonomy [52]. These are novel technologies that require assessment to guarantee that patients fully understand how the application or avatar works in order to ensure that a patient does not misunderstand or mistake the intelligent system for a human-driven application. For instance, it would be problematic if a patient were to assume that ‘at the other end’ of the chatbot there is a doctor communicating or reviewing her messages. Furthermore, obtaining consent for applications used outside of medical systems raises thorny concerns. For instance, an elderly person or a person with intellectual disabilities may not be able to understand what a robot is or what a robot does when it is installed at home to monitor the patient's activities, risking privacy infringement [42], manipulation, and even coercion if the conditions for informed consent are not satisfied. Such questions make consenting to surveillance, interaction, and data collection with the robot challenging matters. Distinctions could be drawn between interventions that are seen as helping and monitoring, as opposed to those that run the risk of manipulation and coercion; however, these lines are often blurred both in theory and in practice [76]. How AI applications should evaluate if a patient has fully understood the information provided when giving consent, and how to proceed in cases where it is not possible for individuals to provide consent, such as children, patients with dementia, those with intellectual disabilities, or those in acute phases of schizophrenia, needs to be addressed.

Another area of particular concern in relation to matters of promoting autonomy in the use of AI in mental health care is the engagement of vulnerable populations. People have been shown to be more compliant when a robot asks them to do something as compared with a person [9]. Although this could lead to better results when helping patients with autism or those needing to make difficult behavioral changes, the concern exists that people could be manipulated or coerced into doing things that they should not or that they have not fully thought through, either because of the novelty of the device or because of a lack of companions with whom to discuss alternatives. Some studies have made a distinction between a "suspension of disbelief" when anthropomorphizing a robot caregiver and deception *per se* [42,77], but this is a line that requires further investigation in practice.

Nonhuman Therapy?

In general, the question remains as to whether there are aspects of the therapeutic encounter that cannot be achieved through AI. Some therapeutic benefits may be difficult to anticipate, or highly specific to a particular individual's relationship to his or her therapist. One study found that embodied conversational agents had difficulty evaluating a user's emotional state in a real-time dialog and that the absence of a human therapist in Web-based mental health interventions for treating depression and anxiety had a negative effect on user adherence to the programs [78]. In the treatment of insomnia, some patients indicated that they missed having a human therapist [79]. Relatedly, because robots and artificially intelligent systems blur previously assumed boundaries between reality and fiction, this could have complex effects on patients. Similar to therapeutic relationships, there is the risk of transference of emotions, thoughts, and feelings to the robot. In particular, given that many of the target populations are vulnerable because of their illness, age, or living situation in a health care facility, there is the additional concern that patients would be vulnerable in their engagements with the robot because of their desire for company or to feel cared for [80]. Unlike with a therapist, however, there is no person on the other side of this transference. Whether robot therapists will ever be able to deal adequately with such transference remains to be seen. Further concerns are likely to emerge in practice; thus, embodied AI therapeutic aids need to also be evaluated carefully for unanticipated differences with standard therapy modalities.

Ethical Issues in Algorithms

It is necessary to note that AI mental health interventions work with algorithms, and algorithms come with ethical issues. It has been well-established that existing human biases can be built into algorithms, reinforcing existing forms of social inequality [81]. This raises the concern that AI-enabled mental health devices could also contain biases that have the potential to exclude or harm in unintended ways, such as data-driven sexist or racist bias or bias produced by competing goals or endpoints of devices [82,83]. Following other calls for transparency [84], the algorithms used in artificially intelligent applications for mental health purposes could be similarly open to scrutiny. This may require investing additional time in explaining to patients (and their families) what an algorithm is and how it works in relation to the therapy provided [85]. However, how to best do this, in particular with patients with compromised mental capacities, requires further consideration.

Concerns Regarding Long-Term Effects

Apart from these more immediate concerns, the implementation of embodied AI into mental health services also raises a number of broader questions regarding long-term impacts on patients, the mental health community, and society more widely. For instance, it has been noted that long-term use of AI interventions could lead to some patients or patient groups becoming overly attached to these applications. A study by Cresswell et al noted that robots that aim to alleviate loneliness or provide emotional comfort carry the risk that the patients they work with could become dependent on them [6]. More broadly, others have raised questions about ways that robots could contribute to changing social values surrounding care or situations in which caregiving is increasingly 'outsourced' to robotic aids. The impact of intelligent robots on relationships, both human-robot and human-human relationships, is an area that requires further probing, as do potential effects on identity, agency, and self-consciousness in individual patients. Specifically, research into the effectiveness of these applications needs to cover not only if the social skills of children with ASD are improved by working with robots but also their ability to apply these skills to relationships with other humans. Similarly, if a sex robot is provided therapeutically to an individual with paraphilia, the effects of this on the targeted behaviors with other humans also needs to be evaluated. The risk exists that if robotic interventions are not translatable to improving human interaction, that they merely remain a way of improving human relations with machines, or worse, an outlet that further limits human-to-human relationships. Similarly, engagement with embodied intelligent devices could also have important effects on the individual, such as on personal sense of identity or agency.

The integration of AI devices into our everyday lives and medical care is undoubtedly changing social expectations and practices of communication. There are essential differences between communicating with an AI device and communicating with another human. Anecdotal findings suggest that some users often speak to assistive devices such as Siri or Alexa in a curter or ruder manner than they would to a human [86]. Importantly, perceptions of the devices can vary by users: children often understand these devices differently than adults, sometimes attributing human characteristics to the device or believing that the device has a real individual inside [87]. Extrapolating from this example, it is clear that the ways that individuals interact with the AI applications in their lives can have implications for communication and social interaction. How this will evolve as more patients have the opportunity to interact with AI applications as part of their mental health care requires further empirical investigation to catch problematic trends early and correct for future development.

A related concern of objectification exists for some areas of AI applications, such as sex robots. The use of ‘sexbots’ has already been notably controversial, with scholars objecting that sexual dysfunction depends on a range of physical, psychological, and sociocultural factors that are profoundly relational and reciprocal. Rather than addressing issues of isolation associated with sexual dysfunction, robots might aggravate it or contribute to reductionist understandings of sexual violence [88]. It has been cautioned that the use of sex robots—also available in childlike models or programed with personalities such as “Frigid Farrah” to resist sexual advances—could instead increase the occurrence of sex crimes, normalize the production of social inequalities surrounding the male gaze [89], and contribute to unwanted sexual encounters. Furthermore, the creation of humanoid robots for use in sexual dysfunction raises concerns that it could reinforce or even legitimize the objectification of humans, in particular women and children [3,88]. As the use of AI in many therapeutic applications has not yet been validated in randomized controlled trials (RCTs), there is the risk that particular applications might make problems such as sexual violence worse. More broadly, embodied AI applications necessarily involve a relatively narrow understanding of illness. For instance, sex robots may help with some medical concerns but do not address other determinants of illness that would have to be taken into account from a bio-psycho-social understanding of mental health illness. Widespread AI use could thus exacerbate trends of reductionism in mental health.

Ideas around embodied AI are culturally and historically shaped. Whether providing motivational interviews in therapy [64], acting as embodied conversational agents for mental disorders [47], or working with populations with intellectual disabilities [90], discussion of embodied AI often turns to worries surrounding the limits of human control over technology. Conjuring images of the Terminator or other depictions of the nonhuman in science fiction or cinema, such tools can carry with them negative or scary associations that bring the issue of trust in medical practice into new light [6]. However, exposure to robotic devices, or living in places with positive or caring associations with robots, can influence the adoption of AI devices in different settings [91]. Initiatives that integrate embodied AI into health care practices need to be duly attuned to existing cultural understandings of the role of technology in social lives, and work to ensure that trust between patient and provider, or patient and the health care system, is not eroded.

Finally, AI agents for mental health raise fundamental questions about what it is to be human [6]. One of the principal contributions of science and technology studies scholarship has been to show how humans do not simply act upon objects but rather our relationships with objects also alters, transforms, and imposes limits upon human activity [92]. Interaction with embodied AI agents, just like interaction with other individuals or a therapist, alters behaviors and understandings of the world. Although social relationships are characterized by reciprocity, relationships with intelligent devices are neither mutual nor symmetric. In particular, some have raised the concern that interacting more with artificial agents may lead some individuals to engage less with other people around them or to develop forms of intimacy with intelligent robots [93], raising concerns specific to the use of robots with children or those with intellectual disabilities. As mentioned, people develop attachments to objects and have been shown to also develop attachments to simpler robotic systems such as AIBO. Thus, it is likely that as more intelligent and autonomous devices are developed, human relationships with them will become even more complicated [94].

Conclusions

In light of the demonstrated benefits and potential, such as expanding the reach of services to underserved populations or enhancing existing services provided by mental health professionals, embodied AI has emerged as an exciting and promising approach across the field of mental health. At present, the quality of research on embodied AI in psychiatry, psychology, and psychotherapy is varied, and there is a marked need for more robust studies including RCTs on the benefits and potential harms of current and future applications.

This is still an emerging field, and any analysis of ethical implications can only be preliminary at this point. However, a few conclusions and recommendations are warranted, based on the considerations presented in this paper:

1. It is necessary to develop clear guidance on whether (and which) embodied AI applications should be subject to standard health technology assessment and require regulatory approval. This should include a set of broader provisions for the use of AI services outside the supervision of a health care professional.
2. Professional associations in mental health should develop guidelines on the best use of AI in mental health services as well as recommendations on how to train and prepare young doctors for wide-spread use of embodied AI in mental health, including blended care models.
3. AI tools in mental health should be treated as an additional resource in mental health services. They should not be used as an excuse for reducing the provision of high-quality care by trained mental health professionals, and their effect on the availability and use of existing mental health care services will need to be assessed.
4. To satisfy duties of care and reporting of harm, ideally embodied AI should remain under the supervision of trained mental health professionals. Any applications offered outside of mental health care settings, such as apps and bots, should be required to demonstrate reliable pathways of risk-assessment and referral to appropriate services.
5. Embodied AI should be used transparently. Guidance on how to implement applications in a way that respects patient

autonomy needs to be developed, for example, regarding when and how consent is required and how to best deal with matters of vulnerability, manipulation, coercion, and privacy.

6. AI algorithms in mental health need to be scrutinized, for example, for bias. Ideally, health professionals should be trained in communicating to their patients the role of the algorithms used in different applications they might be using or consider using, and such algorithms should be open for public debate and shaping.
7. Increased use of embodied AI should be accompanied by research that investigates both direct and indirect effects on the therapeutic relationship, other human-human relationships, and effects on individual self-consciousness, agency, and identity. Long-term effects, ranging from health reductionism to increased objectification and impacts on our understandings of what it means to be human, need to be monitored.

Conflicts of Interest

None declared.

References

1. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health 2018 Apr 23;5(2):e32 [FREE Full text] [doi: 10.2196/mental.9423] [Medline: 29685865]
2. Góngora Alonso S, Hamrioui S, de la Torre Díez I, Motta Cruz E, López-Coronado M, Franco M. Social robots for people with aging and dementia: a systematic review of literature. Telemed J E Health 2018 Aug 23 Epub ahead of print(forthcoming). [doi: 10.1089/tmj.2018.0051] [Medline: 30136901]
3. Torjesen I. Sixty seconds on . . . sex with robots. Br Med J 2017 Dec 11;358:j3353. [doi: 10.1136/bmj.j3353] [Medline: 28698234]
4. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
5. Mehrotra S, Kumar S, Sudhir P, Rao GN, Thirthalli J, Gandotra A. Unguided mental health self-help apps: reflections on challenges through a clinician's lens. Indian J Psychol Med 2017;39(5):707-711 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_151_17] [Medline: 29200577]
6. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018 Jul 04;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
7. Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical design of intelligent assistive technologies for dementia: a descriptive review. Sci Eng Ethics 2018 Dec;24(4):1035-1055. [doi: 10.1007/s11948-017-9976-1] [Medline: 28940133]
8. Calderita LV, Manso LJ, Bustos P, Suárez-Mejías C, Fernández F, Bandera A. THERAPIST: towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. JMIR Rehabil Assist Technol 2014 Oct 07;1(1):e1 [FREE Full text] [doi: 10.2196/rehab.3151] [Medline: 28582242]
9. Broadbent E. Interactions with robots: the truths we reveal about ourselves. Annu Rev Psychol 2017 Jan 03;68:627-652. [doi: 10.1146/annurev-psych-010416-043958] [Medline: 27648986]
10. Liu C, Liu X, Wu F, Xie M, Feng Y, Hu C. Using artificial intelligence (Watson for Oncology) for treatment recommendations amongst Chinese patients with lung cancer: feasibility study. J Med Internet Res 2018 Sep 25;20(9):e11087 [FREE Full text] [doi: 10.2196/11087] [Medline: 30257820]
11. Eichenberg C, Küsel C. Deutsches Ärzteblatt. 2018. Roboter in der Psychotherapie: Intelligente artifizielle Systeme URL: https://www.aerzteblatt.de/pdf.asp?id=199391 [accessed 2019-04-24] [WebCite Cache ID 77rn5Im3V]
12. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015;1:15030 [FREE Full text] [doi: 10.1038/npjschz.2015.30] [Medline: 27336038]
13. LaFrance A. The Atlantic. 2015 Aug 26. Computers Can Predict Schizophrenia Based on How a Person Talks URL: https://www.theatlantic.com/technology/archive/2015/08/speech-analysis-schizophrenia-algorithm/402265/ [accessed 2019-04-22] [WebCite Cache ID 77pavCxgw]
14. Venturi D. Medium Corporation. 2017 Jan 26. How AI is revolutionizing mental health care URL: https://medium.com/@davidventuri/how-ai-is-revolutionizing-mental-health-care-a7cec436a1ce [accessed 2019-04-24] [WebCite Cache ID 77rnD5WjU]
15. Luxton D. Artificial Intelligence in Behavioral and Mental Health Care. Amsterdam, Boston: Academic Press; 2015:308.
16. Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B, et al. Predicting the risk of suicide by analyzing the text of clinical notes. PLoS One 2014;9(1):e85733 [FREE Full text] [doi: 10.1371/journal.pone.0085733] [Medline: 24489669]

17. Kuwamura K, Nishio S, Sato S. Can we talk through a robot as if face-to-face? Long-term fieldwork using teleoperated robot for seniors with Alzheimer's disease. Front Psychol 2016;7:1066 [FREE Full text] [doi: 10.3389/fpsyg.2016.01066] [Medline: 27486416]
18. Yamazaki R, Nishio S, Ishiguro H, Nørskov M, Ishiguro N, Balistreri G. Acceptability of a teleoperated android by senior citizens in Danish society. Int J of Soc Robotics 2014 Jul 15;6(3):429-442. [doi: 10.1007/s12369-014-0247-x]
19. Cavallo F, Esposito R, Limosani R, Manzi A, Bevilacqua R, Felici E, et al. Robotic services acceptance in smart environments with older adults: user satisfaction and acceptability study. J Med Internet Res 2018 Sep 21;20(9):e264 [FREE Full text] [doi: 10.2196/jmir.9460] [Medline: 30249588]
20. Sachan D. Self-help robots drive blues away. Lancet Psychiatry 2018 Jul;5(7):547. [doi: 10.1016/S2215-0366(18)30230-X]
21. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
22. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018 Jan;5(1):31-40. [doi: 10.1016/S2215-0366(17)30427-3]
23. Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, et al. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e18 [FREE Full text] [doi: 10.2196/mhealth.7030] [Medline: 28223265]
24. Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychother 2018 Nov;25(6):878-885. [doi: 10.1002/cpp.2322] [Medline: 30221431]
25. Dellazizzo L, Potvin S, Phraxayavong K, Lalonde P, Dumais A. Avatar therapy for persistent auditory verbal hallucinations in an ultra-resistant schizophrenia patient: a case report. Front Psychiatry 2018;9:131 [FREE Full text] [doi: 10.3389/fpsyt.2018.00131] [Medline: 29713292]
26. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: a pilot clinical trial. Schizophr Res 2018 Feb 24 Epub ahead of print(forthcoming). [doi: 10.1016/j.schres.2018.02.031] [Medline: 29486956]
27. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. Lancet Psychiatry 2018 Aug;5(8):625-632. [doi: 10.1016/S2215-0366(18)30226-8]
28. Ohio State University. 2015 Feb 25. Virtual patient: Avatar shows emotions as he talks to med students URL: https://www.sciencedaily.com/releases/2015/02/150225151639.htm [accessed 2019-04-22] [WebCite Cache ID 77pci9R3P]
29. Rein BA, McNeil DW, Hayes AR, Hawkins TA, Ng HM, Yura CA. Evaluation of an avatar-based training program to promote suicide prevention awareness in a college setting. J Am Coll Health 2018 Jul;66(5):401-411. [doi: 10.1080/07448481.2018.1432626] [Medline: 29461940]
30. Wada K, Shibata T. Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. IEEE Trans Robot 2007 Oct;23(5):972-980. [doi: 10.1109/TRO.2007.906261]
31. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: study protocol for a randomized controlled trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
32. Griffiths A. The Guardian. 2014 Jul 08. How Paro the robot seal is being used to help UK dementia patients URL: https://www.theguardian.com/society/2014/jul/08/paro-robot-seal-dementia-patients-nhs-japan [accessed 2019-04-22] [WebCite Cache ID 77pdxW9XI]
33. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. J Am Med Dir Assoc 2012 Feb;13(2):114-120.e1. [doi: 10.1016/j.jamda.2010.10.002] [Medline: 21450215]
34. Grossard C, Palestra G, Xavier J, Chetouani M, Grynszpan O, Cohen D. ICT and autism care: state of the art. Curr Opin Psychiatry 2018 Nov;31(6):474-483. [doi: 10.1097/YCO.0000000000000455] [Medline: 30256263]
35. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng 2012;14:275-294. [doi: 10.1146/annurev-bioeng-071811-150036] [Medline: 22577778]
36. Huijnen CA, Lexis MA, Jansens R, de Witte LP. How to implement robots in interventions for children with autism? A co-creation study involving people with autism, parents and professionals. J Autism Dev Disord 2017 Oct;47(10):3079-3096 [FREE Full text] [doi: 10.1007/s10803-017-3235-9] [Medline: 28685411]
37. Mengoni SE, Irvine K, Thakur D, Barton G, Dautenhahn K, Guldberg K, et al. Feasibility study of a randomised controlled trial to investigate the effectiveness of using a humanoid robot to improve the social skills of children with autism spectrum disorder (Kaspar RCT): a study protocol. BMJ Open 2017 Dec 22;7(6):e017376 [FREE Full text] [doi: 10.1136/bmjopen-2017-017376] [Medline: 28645986]
38. Pennisi P, Tonacci A, Tartarisco G, Billeci L, Ruta L, Gangemi S, et al. Autism and social robotics: a systematic review. Autism Res 2016 Feb;9(2):165-183. [doi: 10.1002/aur.1527] [Medline: 26483270]

39. Rabbitt SM, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded use. Clin Psychol Rev 2015 Feb;35:35-46. [doi: 10.1016/j.cpr.2014.07.001] [Medline: 25462112]
40. Sharkey N, van Wynsberghe A, Robbins S, Hancock E. Foundation for Responsible Robotics. 2017. Our Sexual Future with Robots URL: https://responsible-robotics-myxf6pn3xr.netdna-ssl.com/wp-content/uploads/2017/11/FRR-Consultation-Report-Our-Sexual-Future-with-robots-.pdf [accessed 2019-04-24] [WebCite Cache ID 77rnzVe17]
41. Torjesen I. Society must consider risks of sex robots, report warns. Br Med J 2017 Dec 05;358:j3267. [doi: 10.1136/bmj.j3267] [Medline: 28679604]
42. Vandemeulebroucke T, Dierckx de Casterlé B, Gastmans C. The use of care robots in aged care: a systematic review of argument-based ethics literature. Arch Gerontol Geriatr 2018 Jan;74:15-25. [doi: 10.1016/j.archger.2017.08.014] [Medline: 28926749]
43. Coeckelbergh M, Pop C, Simut R, Peca A, Pintea S, David D, et al. A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: ethical acceptability, trust, sociability, appearance, and attachment. Sci Eng Ethics 2015 Apr 17;22(1):47-65. [doi: 10.1007/s11948-015-9649-x]
44. Riek L. Robotics technology in mental health care. In: Artificial Intelligence in Behavioral Health and Mental Health Care. Cambridge, Massachusetts: Academic Press; 2015:185-203.
45. Piette JD, Krein SL, Striplin D, Marinec N, Kerns RD, Farris KB, et al. Patient-centered pain care using artificial intelligence and mobile health tools: protocol for a randomized study funded by the US Department of Veterans Affairs Health Services Research and Development Program. JMIR Res Protoc 2016 Apr 07;5(2):e53 [FREE Full text] [doi: 10.2196/resprot.4995] [Medline: 27056770]
46. Suganuma S, Sakamoto D, Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. JMIR Ment Health 2018 Jul 31;5(3):e10454 [FREE Full text] [doi: 10.2196/10454] [Medline: 30064969]
47. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 Dec 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
48. Dockweiler C, Kupitz A, Hornberg C. Acceptance of online-based therapy by patients with light to moderate depressive disorders. Gesundheitswesen 2018 Nov;80(11):1013-1022. [doi: 10.1055/s-0043-119287] [Medline: 29126341]
49. Rojas G, Guajardo V, Martínez P, Castro A, Fritsch R, Moessner M, et al. A remote collaborative care program for patients with depression living in rural areas: open-label trial. J Med Internet Res 2018 Dec 30;20(4):e158 [FREE Full text] [doi: 10.2196/jmir.8803] [Medline: 29712627]
50. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord 2012 Jan;6(1):249-262 [FREE Full text] [doi: 10.1016/j.rasd.2011.05.006] [Medline: 22125579]
51. Broekens J, Heerink M, Rosendal H. Assistive social robots in elderly care: a review. Gerontechnology 2009 Apr;8(2):94-103. [doi: 10.4017/gt.2009.08.02.002.00]
52. Beauchamp T, Childress J. Principles of Biomedical Ethics. New York: Oxford University Press; 2012.
53. Stix C. World Economic Forum. 2018 Mar 05. 3 ways AI could help our mental health URL: https://www.weforum.org/agenda/2018/03/3-ways-ai-could-could-be-used-in-mental-health/ [accessed 2019-04-24] [WebCite Cache ID 77roplA75]
54. Pugh A. The New Yorker. 2018 May 22. Automated Health Care Offers Freedom from Shame, But Is It What Patients Need? URL: https://www.newyorker.com/tech/annals-of-technology/automated-health-care-offers-freedom-from-shame-but-is-it-what-patients-need [accessed 2019-04-22] [WebCite Cache ID 77pfV94W6]
55. Costescu CA, Vanderborght B, David DO. Reversal learning task in children with autism spectrum disorder: a robot-based approach. J Autism Dev Disord 2015 Nov;45(11):3715-3725. [doi: 10.1007/s10803-014-2319-z] [Medline: 25479815]
56. Warren Z, Zheng Z, Das S, Young EM, Swanson A, Weitlauf A, et al. Brief report: development of a robotic intervention platform for young children with ASD. J Autism Dev Disord 2015 Dec;45(12):3870-3876 [FREE Full text] [doi: 10.1007/s10803-014-2334-0] [Medline: 25503680]
57. Bickmore TW, Pfeifer LM, Jack BW. Taking the Time to Care: Empowering Low Health Literacy Hospital Patients with Virtual Nurse Agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2009; New York p. 1265-1274.
58. Balon R. Burden of sexual dysfunction. J Sex Marital Ther 2017 Jan 02;43(1):49-55. [doi: 10.1080/0092623X.2015.1113597] [Medline: 26683616]
59. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveill Summ 2018 Dec 27;67(6):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6706a1] [Medline: 29701730]
60. McCabe MP, Sharlip ID, Lewis R, Atalla E, Balon R, Fisher AD, et al. Incidence and prevalence of sexual dysfunction in women and men: a consensus statement from the Fourth International Consultation on Sexual Medicine 2015. J Sex Med 2016 Feb;13(2):144-152. [doi: 10.1016/j.jsxm.2015.12.034] [Medline: 26953829]

61. Messias EL, Chen CY, Eaton WW. Epidemiology of schizophrenia: review of findings and myths. Psychiatr Clin North Am 2007 Sep;30(3):323-338 [FREE Full text] [doi: 10.1016/j.psc.2007.04.007] [Medline: 17720026]
62. Brown AS, Lau FS. Review of the Epidemiology of Schizophrenia. In: Pletnikov MV, Waddington JL, editors. A Handbook of Behavioral Neuroscience. Netherlands: Elsevier; 2016:17-30.
63. Gionet K. The Guardian. 2018 Apr 25. Meet Tess: the mental health chatbot that thinks like a therapist URL: https://www.theguardian.com/society/2018/apr/25/meet-tess-the-mental-health-chatbot-that-thinks-like-a-therapist [accessed 2019-04-22] [WebCite Cache ID 77pgIMS3h]
64. Galvão Gomes da Silva J, Kavanagh DJ, Belpaeme T, Taylor L, Beeson K, Andrade J. Experiences of a motivational interview delivered by a robot: qualitative study. J Med Internet Res 2018 Dec 03;20(5):e116 [FREE Full text] [doi: 10.2196/jmir.7737] [Medline: 29724701]
65. Schröder J, Berger T, Meyer B, Lutz W, Hautzinger M, Späth C, et al. Attitudes towards internet interventions among psychotherapists and individuals with mild to moderate depression symptoms. Cogn Ther Res 2017 Apr 22;41(5):745-756. [doi: 10.1007/s10608-017-9850-0]
66. Deutscher Ethikrat. 2017 Nov 30. Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung URL: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf [accessed 2019-04-24] [WebCite Cache ID 77rpJezPa]
67. Nuffield Council on Bioethics. 2014 Jul 01. The collection, linking and use of data in biomedical research and health care: ethical issues URL: http://nuffieldbioethics.org/wp-content/uploads/Biodata-a-guide-to-the-report-PDF.pdf [accessed 2019-04-24] [WebCite Cache ID 77rpMHI9y]
68. Feil-Seifer D, Mataric M. Socially assistive robotics. IEEE Robot Automat Mag 2011 Mar;18(1):24-31. [doi: 10.1109/MRA.2010.940150]
69. Körtner T. Ethical challenges in the use of social service robots for elderly people. Z Gerontol Geriatr 2016 Jun;49(4):303-307. [doi: 10.1007/s00391-016-1066-5] [Medline: 27220734]
70. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People - an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach (Dordr) 2018;28(4):689-707 [FREE Full text] [doi: 10.1007/s11023-018-9482-5] [Medline: 30930541]
71. Luxton DD. Recommendations for the ethical use and design of artificial intelligent care providers. Artif Intell Med 2014 Sep;62(1):1-10. [doi: 10.1016/j.artmed.2014.06.004] [Medline: 25059820]
72. Oliveira T, Novais P, Neves J. Development and implementation of clinical guidelines: an artificial intelligence perspective. Artif Intell Rev 2013 Mar 13;42(4):999-1027. [doi: 10.1007/s10462-013-9402-2]
73. Fulmer R. The Medical Futurist. 2018 May 24. 5 Reasons Why Artificial Intelligence Won't Replace Physicians URL: https://medicalfuturist.com/5-reasons-artificial-intelligence-wont-replace-physicians [accessed 2019-04-22] [WebCite Cache ID 77pgpQRMQ]
74. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
75. Ienca M, Jotterand F, Vică C, Elger B. Social and assistive robotics in dementia care: ethical recommendations for research and practice. Int J of Soc Robotics 2016 Jun 22;8(4):565-573. [doi: 10.1007/s12369-016-0366-7]
76. Anderson S. Stanf Encycl Philos Arch. 2017. Coercion URL: https://plato.stanford.edu/entries/coercion/ [accessed 2019-04-24] [WebCite Cache ID 77rpcS2YC]
77. Sharkey A, Sharkey N. Children, the elderly, and interactive robots. IEEE Robot Automat Mag 2011 Mar;18(1):32-38. [doi: 10.1109/MRA.2010.940151]
78. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. J Med Internet Res 2017 Dec 16;19(11):e383 [FREE Full text] [doi: 10.2196/jmir.7351] [Medline: 29146567]
79. Heim E, Rötger A, Lorenz N, Maercker A. Working alliance with an avatar: how far can we go with internet interventions? Internet Interv 2018 Mar;11:41-46 [FREE Full text] [doi: 10.1016/j.invent.2018.01.005] [Medline: 30135758]
80. Johnston A. KALW Local Public Radio. 2015 Aug 17. Robotic seals comfort dementia patients but raise ethical concerns URL: https://www.kalw.org/post/robotic-seals-comfort-dementia-patients-raise-ethical-concerns [accessed 2019-04-22] [WebCite Cache ID 77phfBLmf]
81. Tett G. Financial Times. 2018 Feb 09. When algorithms reinforce inequality URL: https://www.ft.com/content/fb583548-0b93-11e8-839d-41ca06376bf2 [accessed 2019-04-22]
82. Corea F. Machine Ethics and Artificial Moral Agents. In: Applied Artificial Intelligence: Where AI Can Be Used In Business. Switzerland: Springer; 2019:33-41.
83. Hammond K. TechCrunch. 2016. 5 unexpected sources of bias in artificial intelligence URL: https://techcrunch.com/2016/12/10/5-unexpected-sources-of-bias-in-artificial-intelligence/?guccounter=1&guce_referrer_us=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_cs=oDTg6AcZIiq57jAEze2jlw [accessed 2019-04-22]

84. Powles J. The New Yorker. 2017 Dec 20. New York City's Bold, Flawed Attempt to Make Algorithms Accountable URL: https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable [accessed 2019-04-22] [WebCite Cache ID 77pieWwCj]
85. Fiske A, Buyx A, Prainsack B. Health information counselors: a new profession for the age of big data. Acad Med 2019 Jan;94(1):37-41 [FREE Full text] [doi: 10.1097/ACM.0000000000002395] [Medline: 30095453]
86. Calfee I. Center for Mobile Communication Studies. 2017 Nov 16. Growing up with Alexa, Siri, and other A. I. technology URL: https://sites.bu.edu/cmcs/2017/11/16/growing-up-with-alexa-siri-and-other-a-i-technology/ [accessed 2019-04-22] [WebCite Cache ID 77piq3K2W]
87. Doucleff M, Aubrey A. National Public Radio. 2017 Oct 30. Alexa, Are You Safe For My Kids? URL: https://www.npr.org/sections/health-shots/2017/10/30/559863326/alexa-are-you-safe-for-my-kids [accessed 2019-04-22] [WebCite Cache ID 77pj2oDUT]
88. Facchin F, Barbara G, Cigoli V. Sex robots: the irreplaceable value of humanity. Br Med J 2017 Dec 15;358:j3790. [doi: 10.1136/bmj.j3790] [Medline: 28811272]
89. Scheutz M, Arnold T. Are we ready for sex robots? 2016 Presented at: 11th ACM/IEEE International Conference on Human-Robot Interaction; 2016; Christchurch, New Zealand p. 351-358.
90. Hall V, Conboy-Hill S, Taylor D. Using virtual reality to provide health care information to people with intellectual disabilities: acceptability, usability, and potential utility. J Med Internet Res 2011 Nov 14;13(4):e91 [FREE Full text] [doi: 10.2196/jmir.1917] [Medline: 22082765]
91. Misselhorn C, Pompe U, Stapleton M. Ethical considerations regarding the use of social robots in the fourth age. GeroPsych 2013 Jan;26(2):121-133. [doi: 10.1024/1662-9647/a000088]
92. Latour B, Woolgar S. Laboratory Life. Princeton: Princeton University Press; 1986.
93. Melson G, Kahn P, Beck A, Friedman B, Roberts T, Garrett E. Robots as dogs?: children's interactions with the robotic dog AIBO and a live australian shepherd. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems. 2005 Apr Presented at: CHI EA '05; April 02-07, 2005; Portland, OR, USA p. 1649-1652. [doi: 10.1145/1056808.1056988]
94. Dodig Crnkovic G, Çürüklü B. Robots: ethical by design. Ethics Inf Technol 2011 Aug 24;14(1):61-71. [doi: 10.1007/s10676-011-9278-2]

Abbreviations

AI: artificial intelligence
ASD: autism spectrum disorder
RCT: randomized controlled trial

Edited by C Eichenberg; submitted 21.12.18; peer-reviewed by M Beutel, D Johnson; comments to author 03.02.19; revised version received 21.02.19; accepted 26.02.19; published 09.05.19.

Please cite as:
Fiske A, Henningsen P, Buyx A
Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy
J Med Internet Res 2019;21(5):e13216
URL: https://www.jmir.org/2019/5/e13216/
doi:10.2196/13216
PMID:31094356

Review

Digital Education of Health Professionals on the Management of Domestic Violence: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration

Ushashree Divakar[1], MSc; Nuraini Nazeha[1], BSc; Pawel Posadzki[1], PhD; Krister Jarbrink[1], PhD; Ram Bajpai[1], PhD; Andy Hau Yan Ho[1,2,3], PhD, EdD; James Campbell[4], MSc; Gene Feder[5], BSc, MBBS, MD, FRCGP; Josip Car[1], MD, PhD, FRCPE, FFPH

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]Psychology Programme, School of Social Sciences, Nanyang Technological University Singapore, Singapore, Singapore
[3]Palliative Care Centre for Excellence in Research and Education, Singapore, Singapore
[4]Department of Health Workforce, World Health Organization, Geneva, Switzerland
[5]School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Josip Car, MD, PhD, FRCPE, FFPH
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road
Singapore,
Singapore
Phone: 65 6340 2480
Email: josip.car@ntu.edu.sg

Abstract

Background: The World Health Organization states that 35% of women experience domestic violence at least once during their lifetimes. However, approximately 80% of health professionals have never received any training on management of this major public health concern.

Objective: The objective of this study was to evaluate the effectiveness of health professions digital education on domestic violence compared to that of traditional ways or no intervention.

Methods: Seven electronic databases were searched for randomized controlled trials from January 1990 to August 2017. The Cochrane Handbook guideline was followed, and studies reporting the use of digital education interventions to educate health professionals on domestic violence management were included.

Results: Six studies with 631 participants met our inclusion criteria. Meta-analysis of 5 studies showed that as compared to control conditions, digital education may improve knowledge (510 participants and 5 studies; standardized mean difference [SMD] 0.67, 95% CI 0.38-0.95; I^2=59%; low certainty evidence), attitudes (339 participants and 3 studies; SMD 0.67, 95% CI 0.25-1.09; I^2=68%; low certainty evidence), and self-efficacy (174 participants and 3 studies; SMD 0.47, 95% CI 0.16-0.77; I^2=0%; moderate certainty evidence).

Conclusions: Evidence of the effectiveness of digital education on health professionals' understanding of domestic violence is promising. However, the certainty of the evidence is predominantly low and merits further research. Given the opportunity of scaled transformative digital education, both further research and implementation within an evaluative context should be prioritized.

(J Med Internet Res 2019;21(5):e13868) doi:10.2196/13868

KEYWORDS

systematic reviews; evidence-based; health workforce; domestic violence

Introduction

Domestic violence (also referred to as family violence) is a complex public health problem [1] that places a notable burden on the health care system [2]. The World Health Organization (WHO) defined domestic violence, an aggressive and oppressive form of interpersonal violence, as a situation where an individual uses control tactics to emotionally, physically, sexually, or economically abuse a family member or past/current romantic partner [3]. Forms of control behavior can include but are not limited to psychological, physical, sexual, financial, and emotional abuse [4].

The WHO commissioned a multinational study on domestic violence with data collected from 10 countries, which showed that 13%-61% of women between the ages of 15-49 years had experienced physical abuse from their intimate partners at least once in their lifetime [3]. However, due to various reasons including shame, embarrassment, social stigma, and fear of and dependency on the abuser, survivors are often unwilling to reveal their difficulties to others [5,6]. Domestic violence can have both short- and long-term effects on the mental and physical well-being of the survivors. Injuries and physical ailments resulting from prolonged exposure to domestic violence include chronic neurological disorders, cardiovascular diseases, respiratory, intestinal and digestive conditions, reproductive disorders, physical injuries, and even death [7-10]. The less visible, but equally detrimental, impact of domestic violence includes psychological and emotional sufferings, anxiety, fear, depression, and posttraumatic stress [11]. Survivors often require treatment and care from a spectrum of health professionals, ranging from family physicians to physical therapists or clinical psychologists [5].

By facilitating early detection and treatment through a well-structured system of education and support, health professionals could play an important role in promoting greater awareness on domestic violence, identifying survivors of violence, and enabling survivor protection [12]. As health professionals are usually the first line of contact for survivors, they would need to undergo special training to identify, support, and treat domestic violence patients. This would be of particular importance to conservative societies where domestic violence is known to be severely underreported [13,14]. Moreover, the immense shortage of health professionals worldwide, especially in developing countries [15], compounded by the lack of training of health professionals, poses immense challenges in tackling the global domestic violence crisis [13,16].

Occasional training programs on domestic violence consisting of seminars and workshops often claim poor retention rates, as they are mostly time-consuming [17], require health professionals to travel to training locations, and are taught by academics who themselves may have had little exposure to people who have experienced domestic violence [18,19]. With the increasing use of information communication technologies in health professions' education, leveraging on digital education to provide domestic violence management education could help address the various challenges of training and manpower shortage while improving the cost-effectiveness of educational programs [20-22].

The term digital education refers to a range of teaching and learning strategies that utilize digital media and devices for training and as interaction tools [23]. Digital education can be further subclassified into various types according to delivery methods (online or offline), content, learning objectives, pedagogical approaches, and delivery settings [24]. The use of the internet to deliver content is referred to as the online mode of digital education, while the use of software or PowerPoint without the need for the internet to deliver content is referred to as the offline mode of digital education. With its scalability, flexibility, cost-effectiveness, and ability to overcome geographical and temporal constraints, digital education has the potential to provide more independent, customized, and accessible domestic violence training. Studies comparing digital education to traditional methods in various specialties including medical education and engineering have found digital education to be more efficient and effective in building knowledge [25,26].

Although a previous study reviewed domestic violence education among health professionals [19], digital education has been gaining popularity in health professions' curricula and hence its use in domestic violence education should be studied further. To the best of our knowledge, there are no systematic reviews evaluating the effectiveness of digital health interventions specifically for domestic violence training among health professionals. Hence, the objective of this study was to evaluate the effectiveness of health professions' digital education on domestic violence compared to that of traditional ways or no intervention.

Methods

Search Strategy

We followed the Cochrane Handbook guidelines for this review. A more detailed description of the methodology is provided in the paper by Car et al [27]. This review is part of a global evidence-synthesis initiative for digital health professions' education [28-39]. The search for the relevant trials was conducted across 7 databases: MEDLINE (Ovid), EMBASE (Elsevier), the Cochrane Central Register of Controlled Trials (Wiley), PsychINFO (Ovid), Educational Resource Information Centre (Ovid), Cumulative Index to Nursing and Allied Health Literature (EBSCO), and Web of Science Core Collection (Thomson Reuters). The detailed search strategy for MEDLINE is presented in the Multimedia Appendix 1. A manual search was conducted to identify any relevant articles from the reference lists of all included articles. A search was also conducted in the International Clinical Trials Registry Platform Search Portal and the metaRegister of Controlled Trials to identify unpublished trials, meeting abstracts, and doctoral theses from Jan 1990 to August 2017.

Eligibility Criteria

The inclusion criteria are presented in Textbox 1. We adopted a broad definition of domestic violence, encompassing all subcategories of domestic violence, to capture a wide range of studies on the topic.

Textbox 1. Inclusion criteria for studies.

Design

- Randomized controlled trials
- Cluster randomized controlled trials

Participants

- Preregistration undergraduates enrolled in health-related courses (including allied health, nursing, and rehabilitation specialization).
    - Preregistration undergraduate education or basic vocational training is defined as any type of study leading to a qualification that (1) is recognized by the relevant government or professional bodies of the country where the study was conducted and (2) entitles the qualification holder to apply for entry-level positions in the health care workforce or have direct contact with patients
- Postregistration health professionals undertaking Continued Medical Education and Continued Professional Development.
    - Postregistration is defined as any type of qualification that is recognized by the relevant government bodies and enables the holder to gain entry into or continue to work in the health care workforce in a more independent or senior role, excluding traditional/complementary medicine practitioners
    - Continued Medical Education is defined as “educational activities which serve to maintain, develop, or increase the knowledge, skills, and professional performance and relationships that a health professional uses to provide services for patients, the public, or the profession” [41]
    - Continued Professional Development is defined as “a range of learning activities through which health and care professionals maintain and develop throughout their career to ensure that they retain their capacity to practice safely, effectively and legally within their evolving scope of practice” [42]

Interventions/exposure

- Studies that use digital education interventions to train pre- and postregistration health professionals in domestic violence management
- Training is delivered via digital education alone (fully) or partially (ie, blended learning)

Comparator(s)/control

- Studies comparing digital education interventions with traditional methods of learning domestic violence management
- Studies comparing digital education interventions with control groups that do not receive any training on domestic violence management
- Studies comparing one type of digital education intervention to another

Outcomes

- Primary outcomes (assessed using validated or nonvalidated measurement tools):
    - Learners’ knowledge postintervention
    - Learners’ skills postintervention
    - Learners’ attitudes
    - Learners’ improvement of self-efficacy defined as improved efficiency toward domestic violence management
- Secondary outcomes (assessed using validated or nonvalidated measurement tools):
    - Learners’ satisfaction postintervention
    - Patient-related outcomes
    - Cost and cost-effectiveness of the intervention
    - Any adverse or unintended effects of digital education interventions

Timeline

- Publications from January 1990 through August 2017

We included randomized controlled trials (RCTs) and quasi-randomized trials reporting the use of digital education interventions (including blended learning, which is a combination of conventional learning and digital education) to educate health professionals on domestic violence management. RCTs with and without control groups that received traditional interventions delivered by either health professionals or university personnel were included. Studies targeting both practicing health professionals and students were included in this review. No language restrictions were imposed. All digital education interventions were included. Cross-over studies were excluded due to the high likelihood of carry-over effects [40]. Non-RCTs and studies not focusing on computer-based interventions and interventions delivered to individuals other than health professionals were also excluded.

Study Selection

The search results from all the databases were combined in a single Endnote X8 library (Clarivate Analytics, Philadelphia, PA), and all duplicate records were removed. Search filters were used to remove articles not related to digital education for health professionals. Two reviewers (UD and NN) then independently screened the titles, abstracts, and full-text articles to identify studies potentially meeting the inclusion criteria. Disagreements were resolved through discussion between the reviewers. Primary outcomes included knowledge, skills, attitudes, self-efficacy, and satisfaction with the education measured using any validated and nonvalidated instruments. Secondary outcome measures included patient outcomes (eg, feedback from domestic violence survivors seeking treatment), change in health professionals’ behavior (ie, health professionals’ confidence in and ease of identifying and treating domestic violence survivors), and economic impact of the intervention.

Data Extraction

All the relevant data including study characteristics, type of digital education intervention, participant demographics, data for outcome measures, and other publication details were extracted independently by UD and NN using a structured data extraction form. We contacted one study author (Short LM) [43] for missing information.

Risk of Bias Assessment and the Overall Quality of Evidence

UD and NN independently assessed the risk of bias using the Cochrane Collaboration’s risk of bias tool [44]. When it was unclear if a trial was of low or high risk, the field was coded as unclear risk of bias. The following domains were evaluated: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome data, selective outcome reporting, and other bias. The following GRADE (Grading of Recommendations Assessment, Development and Evaluation) criteria for evaluating the overall quality of evidence were used: limitations of studies (risk of bias), inconsistency (heterogeneity), indirectness, imprecision, and publication bias [45].

Data Synthesis

Postintervention mean and SDs were used. The baseline mean value was used to calculate the final posttest mean and SD in studies that presented change scores rather than the final mean. When the studies compared more than two groups, the results from the comparison of the least active control group and the most active intervention group were presented.

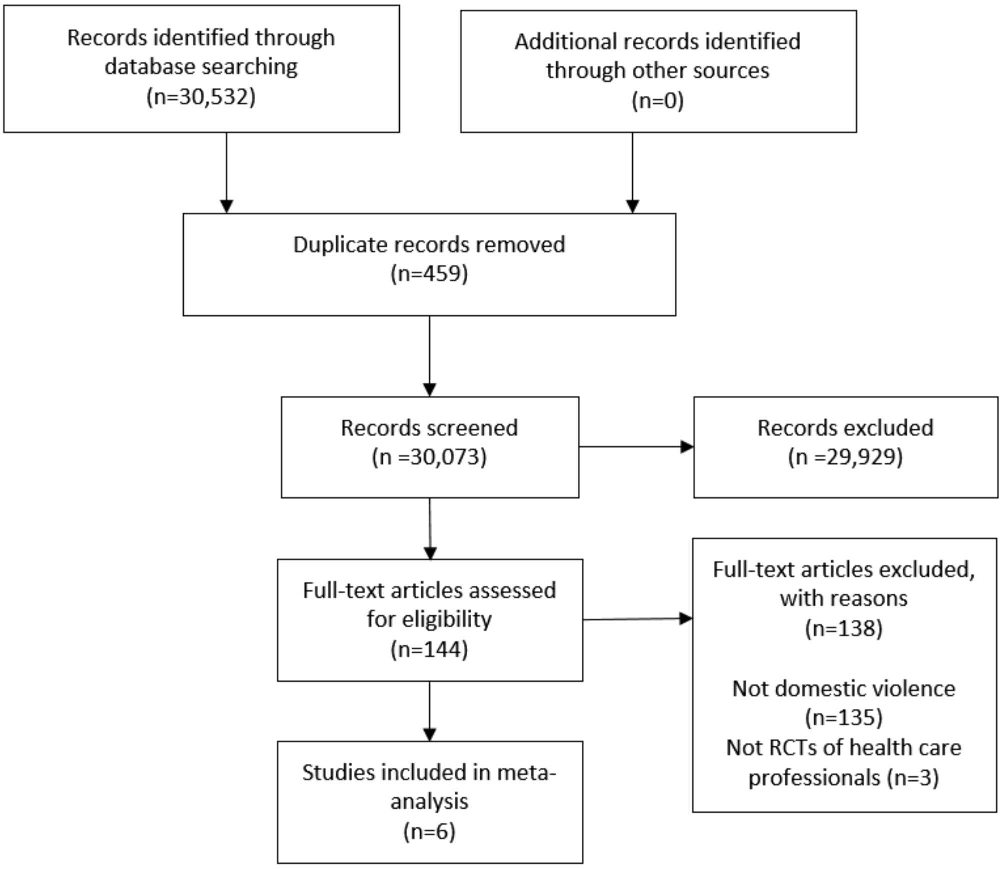
Statistical Analysis

We pooled the data using the random-effect model and calculated standardized mean differences (SMDs) with 95% CIs. Statistical heterogeneity across studies was assessed using the Cochran Q test and I^2 statistics (negligible: 0%-40%, moderate: 30%-60%, or substantial: 50%-90% heterogeneity) [44]. All statistical analyses were conducted using RevMan software (version 5.3; The Nordic Cochrane Centre, Copenhagen, Denmark).

Results

The searches generated a total of 30,073 references. Following abstract and title screening, 144 articles were found to be relevant to domestic violence and selected for full-text screening. Of those, six met our eligibility criteria (Figure 1).

Figure 1. Preferred Reported Items for Systematic Reviews and Meta-analyses chart summarizing the selection process. RCT: randomized controlled trial.



Study Characteristics

All the 6 included RCTs were published between 2000 and 2014 in high-income countries: 5 were from the United States [43,46-49] and 1 was from the Netherlands [50]. All the studies included were parallel RCTs. Three of the studies were conducted in a university setting, and the remaining three were conducted in community and hospital settings. Two of the studies [49,50] focused on child abuse; one, on intimate partner violence [43], and three, on domestic violence, in general [46-48]. In addition, three studies targeted dental professionals [46,48,49], two targeted physicians [43,47], and one targeted nurses [50].

A total of 631 participants were included in the six studies, of which 420 participants (66%) were dentists and dental students. Three studies used offline and three studies used online modes of delivering digital education intervention. The following primary outcomes were reported: knowledge [43,46-49], attitudes [43,46-48], self-efficacy [43,47,50], and skills [50]. Surveys, questionnaires, and checklist were used to measure these outcomes, of which only two instruments [43,47] were validated. The duration of the intervention varied between 15 minutes and 3 weeks. Tables 1 and 2 summarize the main characteristics of the included studies. The study by Shapiro [49], including second-year dental students, on recognizing child abuse was the only study comparing a digital education intervention with traditional lecture-based learning, whereas all the other studies compared digital education to no intervention.

Table 1. Characteristics of the included studies.

Study (year), country, setting	Characteristics of participants (preregistration/postregistration/mixed) and field of study (number of participants)	Intervention (duration)	Control	Results
Danley et al (2004), USA, university [46]	Mixed (dental students and dentists); dentistry (N=174)	Offline interactive multimedia tutorial on DV[a] designed to educate dentists to identify and respond to DV. Control group had no intervention. Assessment via questionnaires (15-25 min)	No intervention	Intervention demonstrated significantly improved attitudes and knowledge compared to the control group.
Harris et al (2002), USA, medical association [47]	Postregistration (physicians); primary care, emergency medicine, and orthopedics (N=121)	Online DV program designed to improve the confidence of practicing physicians in managing DV patients. Assessment via questionnaires (2 weeks to complete the program)	No intervention	Online education program on DV can improve physician confidence (measured by self-efficacy), attitudes, and self-reported knowledge in managing DV patients. In addition, 17.8% mean change in the self-efficacy domain score for the intervention group versus –0.6% change for the control group (*P*<.001) was observed. Self-reported user satisfaction with the program was high.
Hsieh et al (2006), USA, university and clinics [48]	Postregistration (dentists); dentistry (N=174)	Offline interactive multimedia tutorial on DV designed to educate dentists to identify and respond to DV. Assessment via questionnaires (15 min)	No intervention	The posttest comparison of the two groups was statistically significant (*P*=.01) in favor of the online training group.
Shapiro et al (2014), USA, university [49]	Preregistration (dental students); dentistry (N=72)	Online interactive training module to educate dental students on child abuse, assessed via questionnaires (3 weeks for reviewing the online module)	Traditional lecture-based session	In LG[b], 91.6% agreed or strongly agreed that the traditional lecture was a good way to learn the material.
Short et al (2006), USA, community practice [43]	Postregistration (community physicians); family medicine, pediatrics, obstetrics, and gynecology (N=52)	Online CME[c] program to educate HCPs[d] on IPV[e] program in a community practice setting assessed via self-administered, paper-based survey tool (minimum 4 hours)	No intervention	Online CME[f] survey program for physician readiness to manage intimate partner violence was successful in improving physicians' IPV knowledge, attitudes, and self-efficacy.
Smeekens et al (2011), The Netherlands, medical center [50]	Postregistration (nurses); emergency medicine (N=38)	Offline program designed to educate nurses to recognize child abuse in a simulated case, assessed via performance in simulated cases (minimum of 2 hours during a 2-week period)	No intervention	Nurses in the intervention group performed significantly better during the simulation than the control group and reported higher self-efficacy.

[a] DV: domestic violence.

[b] LG: lecture group

[c] CME: Continued Medical Education.

[d] HCP: health care professional.

[e] IPV: intimate partner violence.

[f] Continued Medical Education is defined as "educational activities which serve to maintain, develop, or increase the knowledge, skills, and professional performance and relationships that a health professional uses to provide services for patients, the public, or the profession" [48].

Table 2. Outcomes of the included studies.

Study and outcome measures	Intervention group score, mean (SD)	Control group score, mean (SD)
Danley et al [46]		
Knowledge	3.0 (0.76)	2.1 (0.78)
Attitude	4.6 (1.15)	3.9 (1.08)
Harris et al [47]		
Knowledge	3.3 (1.96)	2.5 (0.02)
Attitude	—[a]	—
Satisfaction	—	—
Self-efficacy	3.7 (1.20)	3.3 (0.04)
Hsieh et al [48]		
Knowledge	3.1 (2.29)	2.3 (0.18)
Attitude	5.5 (0.19)	4.8 (1.25)
Shapiro et al [49]		
Knowledge	80.5 (1.24)	76.1 (1.56)
Satisfaction	—	—
Short et al [43]		
Knowledge	28.4 (5.68)	25.8 (5.68)
Attitude	4.7 (1.00)	3.5 (1.00)
Self-efficacy	4.6 (1.15)	3.8 (1.15)
Smeekens et al [50]		
Skills	71 (18)	89 (19)
Self-efficacy	447 (98)	502 (96)

[a] Not available.

Effects of Interventions

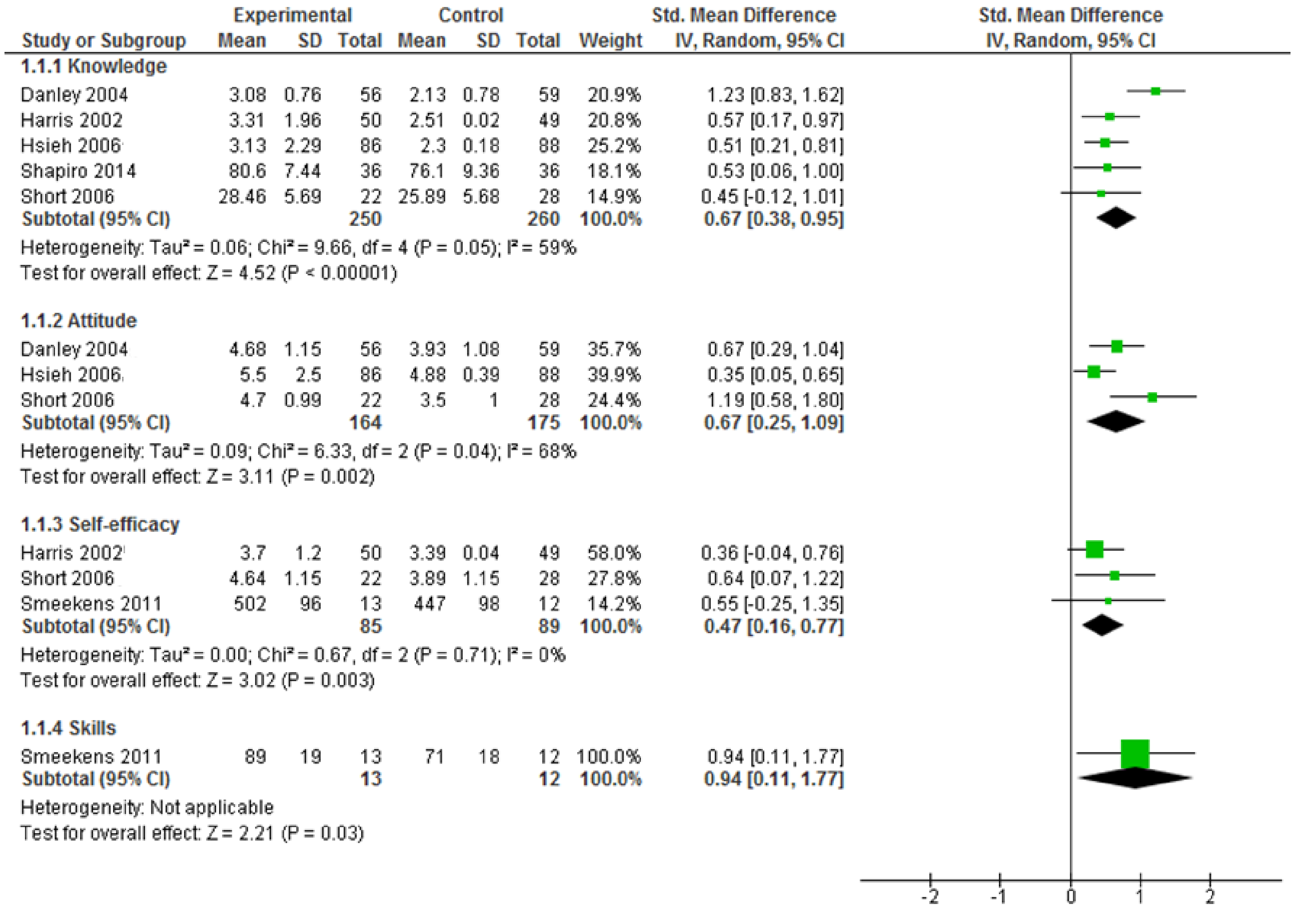
Meta-analysis of five studies [43,46-49] considered to be sufficiently homogeneous found that digital education (offline and online) may increase knowledge of domestic violence in dentists, physicians, and allied health professionals (510 participants; SMD 0.67, 95% CI 0.38-0.95; I^2=59%; low certainty evidence) compared with no intervention and traditional learning postintervention. There was evidence of moderate heterogeneity among the studies (ζ^2=0.06; χ^2_4=9.7; P=.05; I^2=59%; Figure 2).

Meta-analysis of three studies [43,46,48] found that compared to no intervention, digital education (offline and online) may increase postintervention attitude toward domestic violence management in dentists and physicians (339 participants; SMD 0.67, 95% CI 0.25-1.09; I^2=68%). There was a substantial level of heterogeneity among the studies (ζ^2=0.09; χ^2_2=6.3; P=.04; I^2=68%).

Meta-analysis of three studies [43,47,50] found that compared to no intervention, digital education (offline and online) may increase postintervention self-efficacy toward domestic violence management in physicians and nurses (174 participants; SMD 0.47, 95% CI 0.16-0.77; I^2=0%). The was no evidence of heterogeneity (ζ^2=0.00; χ^2_2=0.7; P=.71; I^2=0%).

One study [50] comparing change of score in skills found that digital education (offline program) may improve domestic violence skills in nurses (25 participants; SMD 0.94, 95% CI 0.11-1.77) compared to no intervention.

Figure 2. Forest plot comparing the experimental and control groups in terms of outcomes. IV: interval variable; random: random effect model; std: standardized.



Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.1.1 Knowledge								
Danley 2004	3.08	0.76	56	2.13	0.78	59	20.9%	1.23 [0.83, 1.62]
Harris 2002	3.31	1.96	50	2.51	0.02	49	20.8%	0.57 [0.17, 0.97]
Hsieh 2006	3.13	2.29	86	2.3	0.18	88	25.2%	0.51 [0.21, 0.81]
Shapiro 2014	80.6	7.44	36	76.1	9.36	36	18.1%	0.53 [0.06, 1.00]
Short 2006	28.46	5.69	22	25.89	5.68	28	14.9%	0.45 [-0.12, 1.01]
Subtotal (95% CI)			**250**			**260**	**100.0%**	**0.67 [0.38, 0.95]**
Heterogeneity: Tau² = 0.06; Chi² = 9.66, df = 4 (P = 0.05); I² = 59%								
Test for overall effect: Z = 4.52 (P < 0.00001)								
1.1.2 Attitude								
Danley 2004	4.68	1.15	56	3.93	1.08	59	35.7%	0.67 [0.29, 1.04]
Hsieh 2006	5.5	2.5	86	4.88	0.39	88	39.9%	0.35 [0.05, 0.65]
Short 2006	4.7	0.99	22	3.5	1	28	24.4%	1.19 [0.58, 1.80]
Subtotal (95% CI)			**164**			**175**	**100.0%**	**0.67 [0.25, 1.09]**
Heterogeneity: Tau² = 0.09; Chi² = 6.33, df = 2 (P = 0.04); I² = 68%								
Test for overall effect: Z = 3.11 (P = 0.002)								
1.1.3 Self-efficacy								
Harris 2002	3.7	1.2	50	3.39	0.04	49	58.0%	0.36 [-0.04, 0.76]
Short 2006	4.64	1.15	22	3.89	1.15	28	27.8%	0.64 [0.07, 1.22]
Smeekens 2011	502	96	13	447	98	12	14.2%	0.55 [-0.25, 1.35]
Subtotal (95% CI)			**85**			**89**	**100.0%**	**0.47 [0.16, 0.77]**
Heterogeneity: Tau² = 0.00; Chi² = 0.67, df = 2 (P = 0.71); I² = 0%								
Test for overall effect: Z = 3.02 (P = 0.003)								
1.1.4 Skills								
Smeekens 2011	89	19	13	71	18	12	100.0%	0.94 [0.11, 1.77]
Subtotal (95% CI)			**13**			**12**	**100.0%**	**0.94 [0.11, 1.77]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.21 (P = 0.03)								

Summary Risk of Bias

Of the six studies, four [43,46,48,50] were found to have an overall low risk of bias and the remaining two [47,49] had a high or an unclear risk of bias.

The random sequence generation method was reported in four [43,46,48,50] of the six studies. Blinding and protection against selective reporting was achieved through the nature of the intervention and the reporting of all the results in all the studies. Attrition and other biases were of low risk for five [43,46,48-50] of the six studies. One study had a high risk of attrition bias resulting from a high drop-out rate (42%). However, details of allocation concealment were not reported in any of the studies, and blinding of outcome assessment was attempted in only one study [50]. Similarly, the method for random sequence generation was not clearly stated in two studies [47,49]. At the individual-study level, of 56 domains, 14 (25%) were reported as unclear and one (2%) was reported as high risk (Figure 3). The summary of findings table shows the evidence to be of low to moderate quality as analyzed per the GRADE criteria (Table 3).

Figure 3. Risk of bias summary.

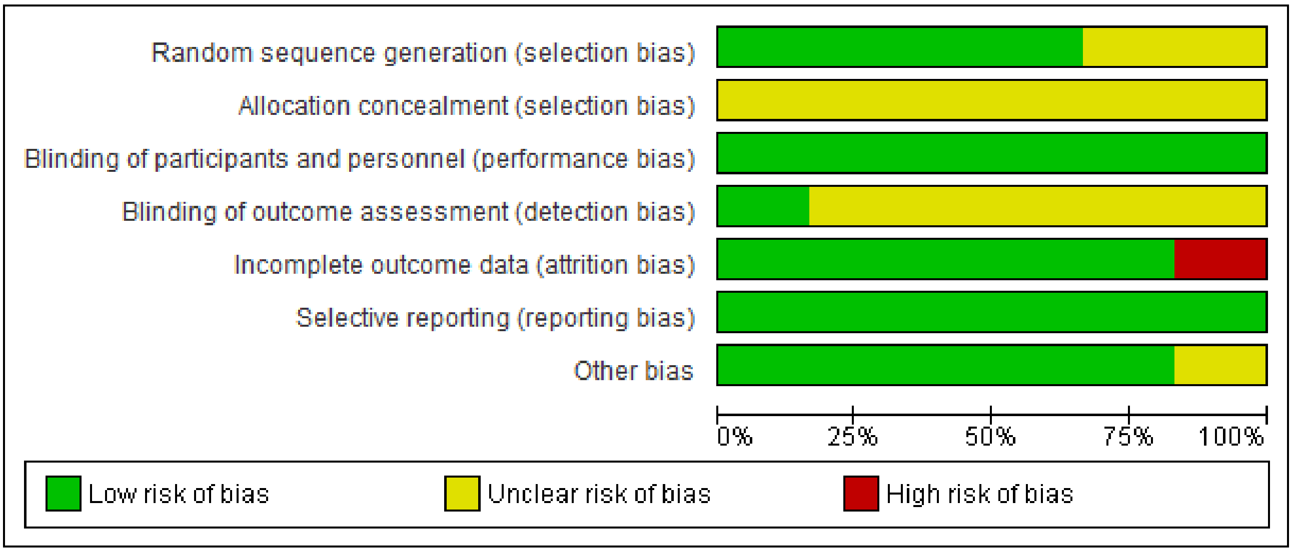


Table 3. Summary of findings table. Patient or population: health care professionals; Setting: university; Intervention: digital education; Comparison: traditional or no intervention.

Outcomes	Anticipated absolute effects[a] (95% CI)		Number of participants (number of RCTs[b])	Certainty of the evidence (GRADE[c])	Comments
	Assumed risk with controls	Corresponding risk with electronic learning			
Knowledge	The mean outcome score in the control groups was 21.79	The mean knowledge score in the intervention groups was 0.67 SD higher (0.38-0.95 higher)	510 (5)	Low[d,e,f]	None
Attitude	The mean outcome score in the control groups was 4.10	The mean attitude score in the intervention groups was 0.67 SD higher (0.25-1.09 higher)	339 (3)	Low[d,e,f]	The results of one study (121 participants) were not pooled due to incomplete data
Self-efficacy	The mean outcome score in the control groups was 151.43	The mean self-efficacy score in the intervention groups was 0.47 SD higher (0.16-0.77 higher)	174 (3)	Moderate[e,g]	None
Skills	The mean outcome score in the control groups was 71	The mean skill score in the intervention groups was 0.94 SD higher (0.11-1.77 higher)	25 (1)	Low[d,e,f]	None

[a] The risk in the intervention group (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

[b] RCT: randomized controlled trial.

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

[d] The heterogeneity was high with large variations in effects and the lack of overlap among CIs.

[e] Rated down by one level for study limitations. The risk of bias was unclear for allocation concealment in all studies.

[f] Low: Our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect.

[g] Moderate: We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Discussion

In this paper, we systematically reviewed and pooled data on the use of digital education for domestic violence management. We draw attention to the gap in digital education on domestic violence and the potential benefits of this educational strategy. Our findings provide preliminary evidence to show that using digital education to address socially sensitive issues such as domestic violence may improve certain educational outcomes in health professionals receiving the training. Although competencies, trustable professional activities, knowledge, skills, and attitudes do not automatically translate into change of practice, they are indispensable for improving patient outcomes.

All the studies included in this review were published after the year 2000, which is the period when digital education started becoming more prominent in health care professions' education [51] and more laws were implemented to tackle domestic violence [5,52,53].

Interestingly, we found that in all the studies, the intervention groups had improved knowledge, skills, attitudes, and self-efficacy, even though the studies employed different methodologies, sample sizes, sampling periods, settings, and types of domestic violence education. Additionally, the changes in primary outcomes were observed within short time periods of up to 2 weeks after the intervention, with only one study [43] measuring retention at the 12-month follow-up. Although the variability suggests that digital education has the versatility to reach a wide range of health professionals in different populations and settings, it underscores the potential of homogeneous short-term digital interventions in improving the quality of care that these professionals provide.

Risk of bias was mostly unclear for blinding of outcome assessment and allocation concealment, but it was mainly low for sequence generation. While the nature of the interventions does not allow blinding of participants, we believe it would not have had any effect on bias risk. We minimized biases by having two reviewers independently assess the articles for inclusion, complete data extraction, risk of bias, and use of the GRADE criteria. The overall quality of the evidence was low or moderate due to the risk of bias and inconsistency across the studies (Table 3).

This review has some important strengths including a strict adherence to the gold-standard Cochrane methods and use of validated, comprehensive, and reproducible searches across seven databases. Our review adds to previous research on domestic violence education for health professionals, as it focuses on the use of digital education, which is a growing area of research. Some weaknesses have to be kept in mind when interpreting the results of this systematic review. For instance, although our searches were comprehensive, we cannot be certain that all relevant trials were included.

However, the evidence evaluated has some limitations. First, only a few studies were published in this area, and they were all from high-income countries, making generalizability challenging. Although digital education may potentially serve as an effective and impactful solution to educating health professionals in domestic violence management, applicability, scalability, and implementation in low- and middle-income countries have to be studied further [54]. Second, we acknowledge that in certain countries such as the United Kingdom, social workers are the first "line of response" to domestic violence. Third, only two studies [47,49] measured and reported learners' satisfaction as one of the primary outcomes. This further highlights the need for uniform and validated outcomes and methods of measuring them to make conclusive judgements. Moreover, only two studies [43,47] used validated measurement instruments to measure outcomes, thereby making it challenging to compare the use of digital education between settings. Subsequently, the lack of data on retention rate, costs, or patient outcomes prevents policy makers from making informed decisions or assessing the transferability of digital education to other settings. Finally, none of the RCTs reported secondary outcomes such as patient outcomes, health professionals' behavior change, and economic impact. Hence, we are unable to assess how these outcomes changed with digital education.

Future studies should be designed to evaluate the effect of digital education on these outcomes. We further recommend that future studies consider including other professionals such as social carers, psychologists, counsellors, or teachers. Findings of this review suggest that digital education could contribute to developing the competencies that health professionals need to respond to complex psychosocial problems such as domestic violence. Therefore, future studies should focus on recording more practical outcomes of the trainings such as change in detection and referral rates [55]. This will help ensure a better understanding of the actual value of integrating digital education modules into pre- and postregistration as well as the continuing professional development curricula. In addition, although domestic violence is more accepted and prevalent in low- and middle-income countries, the education gap is wide and digital education is still at the developing stage in these countries [14,56]. Data should be collected beyond geographical regions with inclusion of the cost analysis to obtain a better understanding of the impact and feasibility of integrating electronic learning modules on domestic violence management into the medical curriculum in low- and middle-income countries. We further recommend that future studies be designed with larger, appropriately powered RCTs, in both developed and low- and middle-income countries alike in order to ensure better representation. Researchers could use methods such as the Consolidated Standards of Reporting Trials for Social and Psychological Interventions 2018 (CONSORT-SPI 2018) checklist as guidance to design and report future studies on digital education for domestic violence in order to ensure that the data collected are of high quality and representative [57].

We believe that while digital education could help increase identification of and support to patients experiencing domestic violence, research with study designs incorporating blended learning might hold the highest potential. Such designs would combine the best of digital education, such as smartphones, apps, emails, text messages, and virtual patients, with the best of traditional classroom practices such as personalized contact or feedback, meetings, and discussions.

In conclusion, we found some promising, predominantly low-quality evidence for the effectiveness of digital education on domestic violence. We also highlighted the need for further research evaluating and validating culturally tailored digital education interventions geared toward more holistic management of domestic violence.

Acknowledgments

This research was supported by the Nanyang Technological University eLearning grant. This review was conducted in collaboration with the Health Workforce Department of the World Health Organization. We would also like to thank Mr Carl Gornitzki, Ms GunBrit Knutssön, and Mr Klas Moberg from the University Library, Karolinska Institutet, Sweden, for developing the search strategy, and the peer reviewers for their comments.

Authors' Contributions

JoC and UD conceived the idea for the review. UD wrote the protocol; performed the searches; screened, extracted, analyzed, and synthesized the data; and wrote and revised the drafts. NN screened and extracted the data. RB analyzed the data. PP, NN, KJ, AHYH, GF, JoC, and JaC critically revised the drafts. All authors contributed to writing and reviewing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE (Ovid) search strategy.

[PDF File (Adobe PDF File), 69KB - jmir_v21i5e13868_app1.pdf]

References

1. Krantz G. Violence against women: a global public health issue!. J Epidemiol Community Health 2002 Apr;56(4):242-243 [FREE Full text] [doi: 10.1136/jech.56.4.242] [Medline: 11896128]
2. Hewins E, DiBella D, Mawla J. Domestic Violence and the Role of the Healthcare Provider: The Value of Educating on Assessment and Intervention Strategies. 2013. URL: https://www.nj.gov/dcf/women/archive/WhitePaper_DomesticViolence.pdf [accessed 2019-05-11] [WebCite Cache ID 78HqukOOU]
3. World Health Organization. Responding to Intimate Partner Violence and Sexual Violence Against Women: WHO Clinical and Policy Guidelines 2013. Geneva, Switzerland: WHO Press; 2013. URL: https://apps.who.int/iris/bitstream/handle/10665/85240/9789241548595_eng.pdf?sequence=1 [accessed 2019-05-20]
4. GOV.UK. 2018 Oct 05. Guidance - Domestic abuse: how to get help URL: https://www.gov.uk/guidance/domestic-abuse-how-to-get-help [accessed 2019-05-11] [WebCite Cache ID 78HrLgtRD]
5. Kaur R, Garg S. Addressing domestic violence against women: an unfinished agenda. Indian J Community Med 2008 Apr;33(2):73-76 [FREE Full text] [doi: 10.4103/0970-0218.40871] [Medline: 19967027]
6. World Health Organization. Understanding and addressing violence against women. Intimate partner violence URL: http://apps.who.int/iris/bitstream/handle/10665/77432/WHO_RHR_12.36_eng.pdf?sequence=1 [accessed 2018-10-01] [WebCite Cache ID 76UiiccJ9]
7. Florida State University. 2014. Conditions & Injuries Related to Domestic Violence-National Prevention Toolkit on Domestic Violence for Medical Professionals URL: https://dvmedtraining.csw.fsu.edu/wp-content/uploads/2014/01/Conditions-and-Injuries-2014.pdf [accessed 2019-05-09] [WebCite Cache ID 78EpIdzHO]
8. Davis R. Domestic violence ‐ related deaths. Jnl Aggress Conflict Peace Res 2010 Apr 14;2(2):44-52. [doi: 10.5042/jacpr.2010.0141]
9. Perona M, Benasayag R, Perelló A, Santos J, Zárate N, Zárate P, et al. Prevalence of functional gastrointestinal disorders in women who report domestic violence to the police. Clin Gastroenterol Hepatol 2005 May;3(5):436-441 [FREE Full text] [Medline: 15880312]
10. Stene LE, Jacobsen GW, Dyb G, Tverdal A, Schei B. Intimate partner violence and cardiovascular risk in women: a population-based cohort study. Journal of women's health 2013;22(3):250-258. [Medline: 23428282]
11. Trevillion K, Oram S, Feder G, Howard LM. Experiences of domestic violence and mental disorders: a systematic review and meta-analysis. PLoS One 2012;7(12):e51740. [doi: 10.1371/journal.pone.0051740] [Medline: 23300562]
12. Yut-Lin W, Othman S. Early detection and prevention of domestic violence using the Women Abuse Screening Tool (WAST) in primary health care clinics in Malaysia. Asia Pac J Public Health 2008;20(2):102-116. [doi: 10.1177/1010539507311899] [Medline: 19124304]
13. Usta J, Antoun J, Ambuel B, Khawaja M. Involving the health care system in domestic violence: what women want. Ann Fam Med 2012;10(3):213-220 [FREE Full text] [doi: 10.1370/afm.1336] [Medline: 22585885]
14. Sardinha L, Nájera Catalán HE. Attitudes towards domestic violence in 49 low-middle-income countries: A gendered analysis of prevalencecountry-level correlates. PLOS ONE 2018;13(10):e0206101 [FREE Full text] [Medline: 30379891]
15. Anyangwe SC, Mtonga C. Inequities in the global health workforce: the greatest impediment to health in sub-Saharan Africa. Int J Environ Res Public Health 2007 Jun;4(2):93-100 [FREE Full text] [Medline: 17617671]

16. Du Plat-Jones J. Domestic violence: the role of health professionals. Nurs Stand 2006;21(14-16):44-48. [doi: 10.7748/ns2006.12.21.14.44.c6392] [Medline: 17252869]
17. Arcoraci V, Squadrito F, Altavilla D, Bitto A, Minutoli L, Penna O, et al. Medical simulation in pharmacology learning and retention: A comparison study with traditional teaching in undergraduate medical students. Pharmacol Res Perspect 2019 Feb;7(1):e00449. [doi: 10.1002/prp2.449] [Medline: 30651989]
18. Congdon TW. A medical student's perspective on education about domestic violence. Acad Med 1997 Jan;72(1 Suppl):S7-S9. [Medline: 9008582]
19. Zaher E, Keogh K, Ratnapalan S. Effect of domestic violence trainingystematic review of randomized controlled trials. Canadian Family Physician 2014;60(7):618-624 [FREE Full text] [Medline: 25022633]
20. Runyan CW, Gunther-Mohr C, Orton S, Umble K, Martin SL, Coyne-Beasley T. PREVENT: a program of the National Training Initiative on Injury and Violence Prevention. Am J Prev Med 2005 Dec;29(5 Suppl 2):252-258. [doi: 10.1016/j.amepre.2005.08.013] [Medline: 16376727]
21. Zehry K, Halder N, Theodosiou L. E-Learning in medical education in the United Kingdom. Procedia - Social and Behavioral Sciences 2011;15:3163-3167 [FREE Full text] [doi: 10.1016/j.sbspro.2011.04.265]
22. Shefet D, Dascal-Weichhendler H, Rubin O, Pessach N, Itzik D, Benita S, et al. Domestic violence: a national simulation-based educational program to improve physicians' knowledge, skills and detection rates. Med Teach 2007 Jun;29(5):e133-e138 [FREE Full text] [doi: 10.1080/01421590701452780] [Medline: 17885965]
23. Sangrà A, Vlachopoulos D, Cabrera N. Building an inclusive definition of e-learning: An approach to the conceptual framework. IRRODL 2012 Apr 13;13(2):145-159 [FREE Full text] [doi: 10.19173/irrodl.v13i2.1161]
24. Paul P, Toon E, Hadadgar A, Jirwe M, Lim KTK, Semwal M, et al. PROSPERO CRD42016045571. 2016. Online- and local area network (LAN)-based eLearning interventions for medical doctors' education URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016045571 [accessed 2019-05-16] [WebCite Cache ID 75r2li0AE]
25. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med 2006 Mar;81(3):207-212. [Medline: 16501260]
26. Wong WK, Ng PK. An Empirical Study on E-Learning versus Traditional Learning among Electronics Engineering Students. American Journal of Applied Sciences 2016 Jun 01;13(6):836-844 [FREE Full text] [doi: 10.3844/ajassp.2016.836.844]
27. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital Education in Health Professions: The Need for Overarching Evidence Synthesis. J Med Internet Res 2019 Dec 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
28. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
29. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GCY, Tudor Car L. Mobile Digital Education for Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
30. Huang Z, Semwal M, Lee SY, Tee M, Ong W, Tan WS, et al. Digital Health Professions Education on Diabetes Management: Systematic Review by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 21;21(2):e12997 [FREE Full text] [doi: 10.2196/12997] [Medline: 30789348]
31. George PP, Zhabenko O, Kyaw BM, Antoniou P, Posadzki P, Saxena N, et al. Online Digital Education for Postregistration Training of Medical Doctors: Systematic Review by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 25;21(2):e13269 [FREE Full text] [doi: 10.2196/13269] [Medline: 30801252]
32. Lall P, Rees R, Law GCY, Dunleavy G, Cotič Z, Car J. Influences on the Implementation of Mobile Learning for Medical and Nursing Education: Qualitative Systematic Review by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 28;21(2):e12895 [FREE Full text] [doi: 10.2196/12895] [Medline: 30816847]
33. Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital Problem-Based Learning in Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 28;21(2):e12945 [FREE Full text] [doi: 10.2196/12945] [Medline: 30816846]
34. Tudor Car L, Kyaw BM, Atun R. The role of eLearning in health management and leadership capacity building in health system: a systematic review. Hum Resour Health 2018 Sep 03;16(44):1-9 [FREE Full text] [doi: 10.1186/s12960-018-0305-9] [Medline: 30176899]
35. Xu X, Posadzki PP, Lee GE, Car J, Smith HE. Digital Education for Health Professions in the Field of Dermatology: A Systematic Review by Digital Health Education Collaboration. Acta Derm Venereol 2018 Oct 15;99(2):133-138 [FREE Full text] [doi: 10.2340/00015555-3068] [Medline: 30320871]
36. Wahabi HA, Esmaeil SA, Bahkali KH, Titi MA, Amer YS, Fayed AA, et al. Medical Doctors' Offline Computer-Assisted Digital Education: Systematic Review by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 01;21(3):e12998. [doi: 10.2196/12998] [Medline: 30821689]
37. Semwal M, Whiting P, Bajpai R, Bajpai S, Kyaw BM, Tudor Car L. Digital Education for Health Professions on Smoking Cessation Management: Systematic Review by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 04;21(3):e13000 [FREE Full text] [doi: 10.2196/13000] [Medline: 30829576]

38. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
39. Kyaw BM, Posadzki P, Dunleavy G, Semwal M, Divakar U, Hervatis V, et al. Offline Digital Education for Medical Students: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Mar 25;21(3):e13165. [doi: 10.2196/13165] [Medline: 30907731]
40. Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. Int J Epidemiol 2002 Feb;31(1):140-149. [Medline: 11914310]
41. Accreditation Council for Continuing Medical Education. CME Content: Definition and Examples URL: https://www.accme.org/accreditation-rules/policies/cme-content-definition-and-examples [accessed 2019-05-10] [WebCite Cache ID 78FzQbEOD]
42. Davis N, Davis D, Bloch R. Continuing medical education: AMEE Education Guide No 35. Med Teach 2008;30(7):652-666. [doi: 10.1080/01421590802108323] [Medline: 18777424]
43. Short LM, Surprenant ZJ, Harris JMJ. A community-based trial of an online intimate partner violence CME program. Am J Prev Med 2006 Feb;30(2):181-185 [FREE Full text] [doi: 10.1016/j.amepre.2005.10.012] [Medline: 16459218]
44. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. 2011. URL: http://handbook-5-1.cochrane.org/ [accessed 2019-05-10] [WebCite Cache ID 78FzGlbYZ]
45. Guyatt G, Oxman A, 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 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
46. Danley D, Gansky S, Chow D, Gerbert B. Preparing dental students to recognize and respond to domestic violence. The Journal of the American Dental Association 2004 Jan;135(1):67-73. [doi: 10.14219/jada.archive.2004.0022]
47. Harris JMJ, Kutob RM, Surprenant ZJ, Maiuro RD, Delate TA. Can Internet-based education improve physician confidence in dealing with domestic violence? Fam Med 2002 Apr;34(4):287-292. [Medline: 12017143]
48. Hsieh NK, Herzig K, Gansky SA, Danley D, Gerbert B. Changing dentists' knowledge, attitudes and behavior regarding domestic violence through an interactive multimedia tutorial. J Am Dent Assoc 2006 May;137(5):596-603. [Medline: 16739538]
49. Shapiro MC, Anderson OR, Lal S. Assessment of a novel module for training dental students in child abuse recognition and reporting. J Dent Educ 2014 Aug;78(8):1167-1175 [FREE Full text] [Medline: 25086150]
50. Smeekens AE, Broekhuijsen-van Henten DM, Sittig JS, Russel IM, ten Cate OT, Turner NM, et al. Successful e-learning programme on the detection of child abuse in emergency departments: a randomised controlled trial. Arch Dis Child 2011 Apr;96(4):330-334. [doi: 10.1136/adc.2010.190801] [Medline: 21233084]
51. Masic I. E-learning as new method of medical education. Acta Inform Med 2008;16(2):102-117 [FREE Full text] [doi: 10.5455/aim.2008.16.102-117] [Medline: 24109154]
52. Neville FG. Preventing violence through changing social norms. In: Oxford Textbook Of Violence Prevention: Epidemiology, Evidence, And Policy. Oxford University Press: Oxford Textbooks In Public Health; 2015:239-244.
53. World Health Organization. World Health Organization. 2009. Violence prevention the evidence: Changing cultural and social norms that support violence URL: https://apps.who.int/iris/bitstream/handle/10665/44147/9789241598330_eng.pdf?sequence=1&isAllowed=y [accessed 2019-02-02]
54. World Health Organization. Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva, Switzerland: WHO; 2013. URL: https://tinyurl.com/yxjzygor [accessed 2019-02-02] [WebCite Cache ID 78G0cSgNH]
55. Feder G, Davies RA, Baird K, Dunne D, Eldridge S, Griffiths C, et al. Identification and Referral to Improve Safety (IRIS) of women experiencing domestic violence with a primary care training and support programme: a cluster randomised controlled trial. Lancet 2011 Nov 19;378(9805):1788-1795. [doi: 10.1016/S0140-6736(11)61179-3] [Medline: 22000683]
56. Passey D, Laferrière T, Ahmad MYA, Bhowmik M, Gross D, Price J, et al. Educational Digital Technologies in Developing Countries Challenge Third Party Providers. Journal of Educational Technology & Society 2016;19(3):121-133.
57. Grant S, Mayo-Wilson E, Montgomery P, Macdonald G, Michie S, Hopewell S, on behalf of the CONSORT-SPI Group. CONSORT-SPI 2018 Explanation and Elaboration: guidance for reporting social and psychological intervention trials. Trials 2018 Jul 31;19(1):406 [FREE Full text] [doi: 10.1186/s13063-018-2735-z] [Medline: 30060763]

Abbreviations

CME: Continued Medical Education
CONSORT-SPI: Consolidated Standards of Reporting Trials for Social and Psychological Interventions
DV: domestic violence
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HCP: health care professional
IPV: intimate partner violence

LG: Lecture Group
RCT: randomized controlled trial
SMD: standardized mean difference
WHO: World Health Organization

Edited by A Marusic; submitted 02.03.19; peer-reviewed by W Turner, T Franic, S Pranic, D Adeloye; comments to author 27.03.19; revised version received 14.04.19; accepted 01.05.19; published 23.05.19.

Please cite as:
Divakar U, Nazeha N, Posadzki P, Jarbrink K, Bajpai R, Ho AHY, Campbell J, Feder G, Car J
Digital Education of Health Professionals on the Management of Domestic Violence: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration
J Med Internet Res 2019;21(5):e13868
URL: http://www.jmir.org/2019/5/e13868/
doi:10.2196/13868
PMID:31124462

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org