Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 22 (2020), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Letters to the Editor

Corrigenda and Addendas

Review

How Behavior Change Strategies are Used to Design Digital Interventions to Improve Medication Adherence and Blood Pressure Among Patients With Hypertension: Systematic Review

Kobra Etminani[1], PhD; Arianna Tao Engström[1], MSc; Carina Göransson[2], PhD; Anita Sant'Anna[1], PhD; Sławomir Nowaczyk[1], PhD

[1]Center for Applied Intelligent Systems Research, Halmstad University, Halmstad, Sweden

[2]Center for Research on Welfare, Health and Sport, Halmstad University, Halmstad, Sweden

Corresponding Author:
Kobra Etminani, PhD
Center for Applied Intelligent Systems Research
Halmstad University
Kristian IV:s väg 3
Halmstad, 301 18
Sweden
Phone: 46 35167332
Email: kobra.etminani@hh.se

Abstract

Background: Information on how behavior change strategies have been used to design digital interventions (DIs) to improve blood pressure (BP) control or medication adherence (MA) for patients with hypertension is currently limited.

Objective: Hypertension is a major modifiable risk factor for cardiovascular diseases and can be controlled with appropriate medication. Many interventions that target MA to improve BP are increasingly using modern digital technologies. This systematic review was conducted to discover how DIs have been designed to improve MA and BP control among patients with hypertension in the recent 10 years. Results were mapped into a matrix of change objectives using the Intervention Mapping framework to guide future development of technologies to improve MA and BP control.

Methods: We included all the studies regarding DI development to improve MA or BP control for patients with hypertension published in PubMed from 2008 to 2018. All the DI components were mapped into a matrix of change objectives using the Intervention Mapping technique by eliciting the key determinant factors (from patient and health care team and system levels) and targeted patient behaviors.

Results: The analysis included 54 eligible studies. The determinants were considered at two levels: patient and health care team and system. The most commonly described determinants at the patient level were lack of education, lack of self-awareness, lack of self-efficacy, and forgetfulness. Clinical inertia and an inadequate health workforce were the most commonly targeted determinants at the health care team and system level. Taking medication, interactive patient-provider communication, self-measurement, and lifestyle management were the most cited patient behaviors at both levels. Most of the DIs did not include support from peers or family members, despite its reported effectiveness and the rate of social media penetration.

Conclusions: This review highlights the need to design a multifaceted DI that can be personalized according to patient behavior(s) that need to be changed to overcome the key determinant(s) of low adherence to medication or uncontrolled BP among patients with hypertension, considering different levels including patient and healthcare team and system involvement.

(J Med Internet Res 2020;22(4):e17201) doi:10.2196/17201

KEYWORDS

digital intervention; hypertension; medication adherence; behavior change; intervention mapping; matrix of change objective.

Introduction

Overview

There is increasing demand for the wide adoption of digital tools and interventions as the entire healthcare ecosystem struggles to deal with the biggest burden of the 21st century: chronic diseases. However, both theory and practice on how to create digital interventions (DIs) remain underdeveloped and under analyzed. Therefore, we performed a systematic literature review to identify the specific components previously used in DIs targeting medication adherence (MA) and/or blood pressure (BP) control in patients with hypertension. We present the results of this analysis within a matrix of change objectives (MoCO) using a framework called Intervention Mapping (IM). This framework allowed us to identify both the activities that are common across different interventions and the goals these activities were intended to achieve. The ultimate impact of health innovations depends on how research evidence and existing theories are used [1]. Thus, organizing the published studies in a framework that considers an existing common theory supports more effective intervention development in the future. At the same time, outlining existing research in such a map provides more profound insight into the strengths, focus, and weaknesses of DIs that have already been developed.

Background

Hypertension, or high BP, is a global public health issue. Approximately 40% of people worldwide are estimated to have hypertension [2]. According to the World Health Organization (WHO) Global Brief on Hypertension [2], 9.4 million deaths annually are attributed to complications of hypertension. Hypertension was identified as the third most important factor for disability-adjusted life years globally in 2002 [3] and as the leading risk factor for global burden of disease in 2010 [4]. Based on the Framingham Heart Study [5], a recent article estimated the direct cost associated with hypertension to be 51.3 million EUR across five European countries [6]. Hypertension is a major modifiable risk factor for cardiovascular diseases [4]. To put it in a different way, patients with controlled BP are 50% less likely to suffer a cardiovascular event compared to those with uncontrolled BP [7].

Medication Adherence and Blood Pressure Control

The WHO defines adherence to long-term therapy as “the extent to which a person's behavior—taking medication, following a diet, and/or executing lifestyle changes—corresponds to agreed recommendations from a health care provider” [8]. In this work, we focus specifically on adherence to a medication regimen to improve BP control.

Studies have repeatedly found that long-term adherence to hypertensive medication is low. A recent systematic review concluded that over 45% of hypertensive patients failed to comply with their medication regimen [9]. It has been estimated that increasing the adherence rate to antihypertensive therapy to 70% would reduce cardiovascular events by over 80,000 cases across the five European countries listed in the article based on the Framingham Heart Study [6].

Adherence to medications is a behavioral process driven by the interaction of many factors, which the WHO classifies into 5 categories: socioeconomic factors, factors associated with the health care team and system in place, disease-related factors, therapy-related factors, and patient-related factors [8]. Several patient-related factors, including lack of understanding of their disease, lack of involvement in the treatment decision-making process, and suboptimal medical literacy, contribute to medication nonadherence [10]. Horne and Weinman correlated MA with two main factors: patients' personal beliefs about the necessity of their prescribed medication and their concerns about taking it [11]. They measured these two factors using the Beliefs about Medicines Questionnaire and showed, in a later systematic review of 94 studies, that higher MA was consistently associated with stronger perceptions of necessity and weaker perceptions of concern [12].

Behavior Change Theories

A meta-analysis reviewing MA intervention for adults with hypertension concluded that the most promising intervention components were those linking adherence behavior with habits, giving adherence feedback to patients, self-monitoring of BP, using pill boxes and other special packaging, and motivational interviewing [13]. Although the literature lists many factors correlated with MA, the causal relationships with such factors must be explored. Behavioral psychology professionals have developed several theories that try to explain how different factors are linked to behavior. Commonly used theories in health-related interventions include the Capability, Opportunity, Motivation, Behavior model [14]; Health Belief Model [15]; theory of planned behavior [16]; and transtheoretical model [17].

Digital Interventions

Interventions to support people with hypertension have the potential to improve outcomes; however, delivery on a wide scale and at low cost is challenging [18]. The use of digital technologies considerably increases the cost-effectiveness of interventions by quickly reaching a large number of people and enabling automation and personalization of content and delivery. Information and Communication Technology solutions such as telephone or video counseling, recorded audio messages, informational websites, and text messages are commonly used in behavioral interventions [19]

A meta-analysis of the effect of text-based interventions concluded that mobile phone text messaging approximately doubles the odds of adequate MA [20]. A systematic review of internet-based interventions for MA similarly found that 11 of 13 studies reported a high or moderate effect on adherence [21].

There is ample evidence that well-designed interventions can improve MA and that digital technologies can facilitate the delivery of these interventions. However, one must navigate a convoluted set of choices to go from theoretical design to practical implementation. A key challenge is that the same theoretical component can be implemented in very different ways, depending on the technology of choice.

Study Purpose

The purpose of this study was to identify the specific DI components used to target MA or BP control for patients with hypertension and to map those components with respect to their intended change in behavior. Although previous reviews have looked at the effectiveness of DI [22-24] and the behavior change techniques included in interventions [25,26], the specific components of these interventions have not been mapped with respect to their intended effect on behavior and the determinant it is intended to address. We reviewed all studies published in the last decade that describe an intervention with at least one digital communication channel. Then, we identified the digital components and mapped them to the intended behavior and relevant determinant factors using the IM approach [27,28].

This approach allowed us to compare the implementation choices across different studies. The selection of the appropriate theories and components for a given intervention remains unguided. For effective (personalized) DI design in the future, we highlight the most commonly used components and the intended effect of each on patient behavior. This knowledge is essential to evaluate how the specific components have been mapped to behavior change strategies. Furthermore, it supports professionals unfamiliar with behavior change theory in understanding how different functionalities can be used to deliver behavior change components.

Methods

Article Identification and Screening

This review intended to answer the following question: "How are DIs designed to improve MA and BP control among patients with hypertension?" We broke this question into three topics: the behaviors, if any, that are targeted for change; the key determinants of the targeted behaviors; and the key components of DIs that effectively improve MA and BP control among patients with hypertension.

We performed a systematic search of the PubMed electronic database. We limited our search to articles published between January 1, 2008, and December 31, 2018, to capture the increasing use of smartphones in health interventions.

Selection Criteria

In our analysis, we included articles fulfilling all three of the following criteria: articles describing the design and/or evaluation of an intervention to improve BP control and/or MA, studies including the use of digital technologies in the intervention delivery to the patients, and studies with hypertension as an inclusion criterion.

Reviews and surveys were excluded, as were articles written in languages other than English. Studies with sample size smaller than 15 or intervention duration <1 month were also excluded. Since the goal was to analyze interventions, we combined all articles that described the same study.

Analysis and Categorization

Articles were evaluated by one reviewer and double checked by another reviewer. From each study, we extracted information pertaining to study design (ie, type of study, number of participants, duration of intervention, country), theoretical underpinnings, primary outcome measures regarding MA and BP, intervention delivery modalities, determinant factors, targeted behaviors, and components of the DI. Finally, three investigators in several rounds of discussions verified the results.

In this review, we focused only on the digital components of the intervention to compare how specific digital components were being used across studies. Given the heterogeneity of intervention designs across studies, different theoretical frameworks used, and different technologies employed, we decided to map all studies to a common framework for comparison and for further development of interventions. We selected a format inspired by MoCO as used in the IM framework [13,28].

Intervention Mapping and Matrix of Change Objectives

IM has been proposed as a 6-step protocol to guide the intervention design process [29]. The second step is to create the MoCO by matching performance objectives (sub-behaviors) with determinants (factors affecting a patient's conduct). It is important to note that IM can be applied regardless of the underlying theory because it relates to the design process in general.

In IM, matrices that combine performance objectives with their determinants are the basis for intervention development. MoCOs are intended to answer the question "What has to change in a specific determinant in order to bring about the behaviors that need to be changed, to reach performance objectives?" The matrix is created by intersecting the performance objectives with determinants of behavior and environmental conditions.

Performance objectives are statements of what a program participant will do or how an environmental condition will be improved. These performance objectives describe exactly what needs to be done at each environmental level by the at-risk population and the agents or policy makers to achieve health outcome improvements.

Determinants answer the question of "why?" The barriers and facilitators to implementation are considered as determinants. Determinants can be constructed through different methods. They can be theoretical constructs from health promotion theories. They can also be found through comprehensive reviews of empirical literature. Some planners would also investigate qualitative methods used with the targeted population, performed independently or sequentially with a quantitative study using a structured questionnaire with questions based on the results of the qualitative phase, shedding light on more hidden determinants.

Change objectives specify what needs to change in the determinants of behavior or environmental conditions to accomplish the performance objectives.

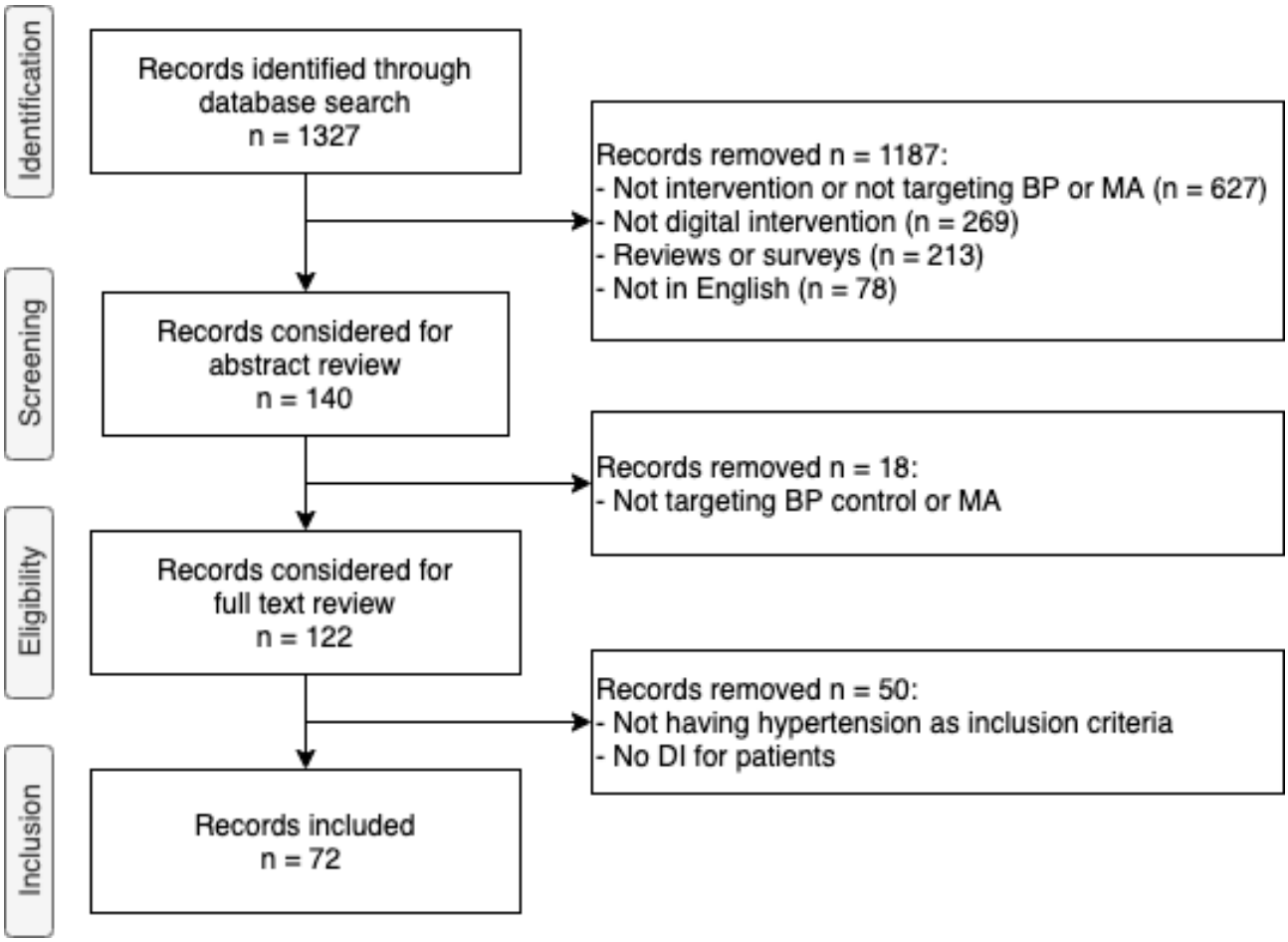
This framework allows us to organize and categorize existing work in a well-defined structure, offering two benefits. First, such a structure allows for a more in-depth comparison of existing studies in their targeted determinant, behaviors to

change, and the components of developed DIs. Second, it facilitates the creation of new DIs, by creating an understanding of the digital components that were previously used, the intended goals, and the areas that have received less or more attention. We created the MoCO by working backwards from each digital component of the DI delivered in each study. We had to identify which behaviors and determinants were addressed, based on the descriptions provided by the authors. After reading all the articles, we needed to understand each intervention component and extract all the DI components first. The second step, which was even more challenging, was to identify the goals behind each component. This rationale was rarely explicitly reported by the author(s) and required interpretation, sometimes requiring extensive discussion, by the reviewers. We identified the addressed determinant(s) and the targeted behavior(s) and then mapped that component to the relevant cell(s) in the matrix. The categories for targeted behavior and determinants were inductively generated from what we identified in the studies we reviewed, with initial inspiration from the Health Belief Model.

Results

The literature review resulted in 54 studies that met the research criteria (Figure 1). Because 8 studies published different aspects of their interventions in more than one paper, we mapped all the related articles into a single study for analysis.

Figure 1. Flowchart of study inclusion. BP: blood pressure; MA: medication adherence; DI: digital intervention.



Study Design

The intervention durations ranged from 1 month to 24 months. The most common duration was 12 months (16/54 studies), which seemed more likely to show benefits for improved MA or reduced BP as behavior change occurs over a period of time. Other reported durations were <6 months (11/54 studies), 6 months (11/54 studies), >12 months (7/54 studies), and 7-10 months (7/54 studies). The intervention duration was not explicitly mentioned in 2 of 54 the articles.

Of the 54 studies, 39 were randomized controlled trials, 5 were longitudinal studies, 5 were pilot/feasibility studies, 3 were quasi-experimental studies, and 1 each was an observational study and mixed-method study.

Across the 54 studies, 24 were conducted in North America, 14 in Europe, 8 in Asia, 4 in Latin America, 2 in Africa, and 2 in Australia.

The sample size varied from 19 to 4076 participants. The most common range was 100-499 participants (26/54 studies). The sample size in other studies ranged from 50 to 99 participants (9/54 studies), 500 to 999 participants (7/54 studies), 1000 to 4076 participants (7/54 studies), and 1 to 49 participants (5/54 studies). Almost all the studies had a multidisciplinary team delivering the interventions, ranging from nurses, physicians, pharmacists, social workers/community health workers, and family members.

Primary Outcome Measures: Medication Adherence and Blood Pressure

Among the 54 reviewed articles, 7 studies focused solely on MA, 18 studies on only BP, and 29 studies on both MA and BP. However, the studies approached those measures in quite different ways, as we explain in the subsequent two sections.

Several papers also included other primary outcome measures, depending on the specific goals of the intervention. Those typically varied between different studies. In brief, they included aspects of body composition (eg, BMI, weight, and waist circumference), scores for quality of life (eg, quality-adjusted life years and quality of life), and the Beliefs about Medicines Questionnaire to demonstrate how their intervention affected patient beliefs about medications. There were 8 studies that performed an economic/cost-effectiveness analysis or resource utilization of the intervention/care costs to demonstrate how important it is to invest more in treatment compliance to avoid emergency room admissions. Self-efficacy was compared using questionnaires in 2 of the 54 studies.

Measures of Medication Adherence

The papers included in this review used both subjective and objective approaches to determine whether patients took their medication as prescribed. Among the 54 studies reviewed, only 39 reported how MA was measured. In 3 studies, MA was not the primary outcome; instead, it was the secondary outcome [30,31] or used as a measure for medication adjustment [32]. In 8 papers, two separate measures were simultaneously used to determine MA, mainly to reconcile or verify with pharmacy records or pill counts.

Among the 39 studies that reported how MA was measured, self-reported values were used in 23 studies. Of these 23 studies, 12 used the Morisky Medication Adherence Scale (MMAS, MMAS-1, MMAS-4, or MMAS-8), 7 studies asked the participant to report whether they had taken the medication, and 1 study each used the 24-item Patient Medication Adherence Questionnaire, Medication Adherence Report Scale, Morisky-Green-Levine Measure of Patient Adherence to Medical Regimens, and Hill-Bone Blood Pressure Therapy Compliance Scale.

Objective MA criteria were used in 24 studies. None of the reviewed studies used direct measures; however, digital medicine (DM) was used in 2 pilot studies. DM is created by attaching an ingestible sensor to each pill that sends a signal to a wearable device when digested. With the help of precise recording of ingestion, the 2 studies were able to capture timing adherence (within ±1 hour around the prescribed dosing time). They were also the only studies that measured ingestion adherence. The most common MA measures were those based on secondary database analyses, such as pharmacy records or dispensing data (11 studies), where 6 studies used proportion of days covered, 1 study used pharmacy refill rate [33], and the remaining 4 studies either did not mention any specific equation or proposed their own. Electronic pill boxes (ePill) or medicine blisters were used in 7 studies, not always as a Medication Events Monitoring System but mainly to count the number of pills/doses taken by the patient. Finally, manual pill count was used in 4 studies.

Measures of Blood Pressure

BP was the most commonly reported primary outcome measure (47/54 studies). Some studies compared the mean change in BP, either self-measured or measured by a health care professional following specific guidelines, between several time points (eg, every month). Other studies reported the proportion of patients with adequately controlled BP (eg, clinic BP <140/90mmHg). There were a number of studies that only focused on systolic BP since most patients with hypertension who are <65 years old have systolic or a combination of systolic/diastolic hypertension and, for the majority, controlling systolic BP also results in control of diastolic BP [34,35].

To make better treatment decisions, some clinicians prefer to treat patients with hypertension based not only on the clinical measurements but also on additional measurements such as home BP monitoring. There were 21 studies in which patients were provided with home BP monitoring to reduce the white coat effect and allow for multiple and frequent readings, while others measured it during the patient visit following specific guidelines, such as the recommended measurement guidelines from the American Heart Association [36].

Intervention Results: Blood Pressure and Medication Adherence

We only report the DIs elicited from the reviewed studies. The intervention result is not solely based on DIs, since several of the studies considered a multifaceted intervention that included DI as part of the main intervention. Therefore, the reported intervention results in this section are based on all the components of the reviewed interventions and not only DI components. Since we only focused on BP and MA as the main outcomes, other outcomes are not reported here.

In total, 26 studies reported a significant reduction in BP, whereas 9 studies reported insignificant BP reduction in the DI group.

Regarding MA, 10 studies reported insignificant MA improvement, with 6 studies reporting a better MA rate among patients during the intervention. Another 6 studies reported significant BP reduction and MA improvement, and 2 studies reported insignificant changes in both BP and MA.

There were more studies reporting a significant BP reduction than reporting a significant improvement in MA. This might be because a greater number of studies focused on BP reduction. It might also be due to standardized BP measurements and BP as the final target of MA.

Theoretical Models and Constructs

Not all studies reported a theoretical underpinning for their DIs; 65% (35/54) of the studies did not explicitly mention the application of psychological theories, models, or principles to their interventions.

Of the 19 remaining studies, the cited theories, models, or principles were behavior change techniques (4/19 studies) targeting patients' psychological determinants [29], the Chronic

Care Model (4/19 studies) to improve care delivery by identifying essential components of the health care system [37], the Social Cognitive Theory (4/19 studies) to change patients' behavior using their own experiences and observations of others' actions [38,39], the Health Belief Model [40] (1/19 studies) that is based on patients' beliefs about the risks and perceptions of the potential benefits of the actions [41], and the Self-Determination Theory (2/19 studies) that highlights the importance of humans' evolved inner resources for personality development and behavioral self-regulation without external influence and interference [42]. Two studies used the principles of patient empowerment as their underpinning cognitive model, and 2 studies used the Common-Sense Model of Self-Regulation [43] that explicates the processes involved in the initiation and maintenance of behaviors for mitigating illness threats.

Matrix of Change Objectives Components

To map the studies into the MoCO framework, we built the matrix components inductively by reviewing the studies and then mapped them to the corresponding cell in the matrix.

Here, we present an example to illustrate the inductive analysis. The following process was performed for all the studies. All the terms used will be presented further.

Frias et al [44] investigated the effectiveness of DM, in which an ingestible sensor is attached to each medication that sends a signal to a wearable sensor whenever digested. Through the wearable sensor (ie, telemetric devices [TelDev]), physical activity data can also be collected, such as mean daily step count and duration of physical activity and rest. The researchers provided a mobile application for patients and a Web portal for physicians to visualize the DM data: MA, as indicated by ingestion, and physical activity. The mobile device app also prompted the patient to take their medication as scheduled. In brief, the following 3 DI components were used: DM data visualization through the mobile app for patients, DM data visualization through the Web portal for physicians, and medication reminders through the mobile app for patients.

The first component provided visualization of MA adherence and physical activity through a mobile app for patients with the help of TelDev and DM. Without this information, patients were not precisely aware of their medication intake history. Therefore, this component targets the "lack of self-awareness" determinant and "taking medication" behavior through "mobile health (mHealth)" and "DM." The same interpretation occurs for physical activity through the mobile app and the TelDev; hence, the crossover of the "lack of self-awareness" determinant and "lifestyle management" behavior is addressed through "TelDev" and "mHealth."

The second component provided DM data visualization to physicians through a Web portal. Thus, this component provides additional information about the patient's health status and MA. Regarding MA, the DI targets the "clinical inertia" determinant and the patient behavior of "taking medication" through "Web" and "DM." For physical activity data, the DI focuses on the "lifestyle management" behavior of patients and the same "clinical inertia" determinant through "Web" and "TelDev" at the health care team and system levels.

The last component provides mobile app–based patient reminders to take their medication. Thus, it targets the "forgetfulness" determinant for the patient behavior of "taking medication" through "mHealth."

Digital Intervention Delivery Modalities

We extracted all digital components of the interventions that were delivered in all 54 studies. We divided the delivery modalities into the following categories:

- Phone (23 studies): phone calls, including manual (19 studies), Interactive Voice Response (3 studies), and videoconferencing (3 studies)
- Web (26 studies): Web-based platforms
- SMS (13 studies)
- mHealth (16 studies): mHealth smartphone apps
- Email (16 studies)
- Electronic health records (EHR; 7 studies): EHR-based software
- Video (3 studies): non-Web-based educational multimedia content
- Com (2 studies): computer-based programs
- TelDev (8 studies): telemetric devices, including automatic BP monitoring devices or automatic weighing scales
- ePill (6 studies)
- DM (2 studies)

Almost all the studies used a mix of these categories to deliver their interventions. ePill and DM are means of measuring MA. We include them in this categorization for two reasons. First, they are a means of digitalization. With the help of these, one can measure MA without manually entering medication intake. Second, they can be considered an intervention factor, rather than just monitoring, since the act of being observed, especially at such high fidelity, is almost certain to influence a patient's behavior. Two key factors should generally be considered when discussing MA: monitoring and intervention [45]. Monitoring refers to the means that can reveal whether the patient has taken the medication as prescribed, while intervention refers to the tools that can be used to enhance MA or correct it once a mistake or drift is detected. The use of these electronic devices fulfills both criteria to some extent. The same reasons apply to including TelDev in this categorization.

Targeted Behaviors

The targeted patient behaviors to change through DI were extracted from 54 studies and then we inductively categorized them into the following:

- PB1: obtain the correct prescription when needed
- PB2: initiate/refill medication
- PB3: take/ingest medication
- PB4: interactive patient-provider communication
- PB5: self-measurement
- PB6: lifestyle management
- PB7: get support from peers or family members

Two-way communication between the patient and provider facilitates resolution of many issues regarding MA and BP control [46,47]. This communication is always happening when the patient visits the provider in person, but that is a relatively

rare occurrence. However, through digital communication channels, such as mHealth, the Web, SMS, phone, and email, two-way communication can occur when the patient needs to discuss something with the provider, but in a remote or even offline manner.

Key Determinants

Although we only included studies that delivered a DI for patients regarding BP control or MA improvement, some studies also simultaneously delivered a DI for the health care team or health care system. Therefore, the key determinants, mainly barriers, were investigated at these 3 levels. Then, we merged the health care team and system levels, resulting in two of the WHO categories: patient level and health care team and system level.

At the health care team and system level, the following determinants were extracted:

- HD1: limited software capability, including a limited appointment system [18,48] and lack of record keeping in EHR [30,49]
- HD2: lack of education, when the health care team had a lack of information about adherence guidelines and general health behavior recommendations or exhibited low adherence to the guidelines [50,51]
- HD3: clinical inertia (aka therapeutic inertia), defined as the failure of health care providers to initiate or intensify therapy according to the current guidelines when treatment goals are unmet [52] (eg, persistently elevated BP), resulting from a combination of health care system, provider, and patient factors [53] including patient factors of low MA [54]. Accordingly, experts recommend MA assessment prior to treatment intensification [55].
- HD4: inadequate health workforce, which occurs when the physician workload is high. Different studies discussed whether to shift some of the workload to pharmacists, community health workers, or nurses or make use of technology.

The latter two were the most targeted determinants.

Regarding the patient level, five determinants were extracted:

- PD1: low perceived risk to health, for which 2 studies aimed to create awareness of the disease risk factors [18,56]
- PD2: lack of education, which occurs when the targeted barrier is a general lack of health literacy or patient/family-based education regarding the different disease conditions
- PD3: lack of self-efficacy, which occurs when the patient barrier is related to a lack of motivation or inefficiency in decision making
- PD4: lack of self-awareness, which occurs when the patients are unaware of their health status recorded either in clinical data (eg, EHR) during visits to doctors or as self-measured data (eg, MA, BP, physical activity)
- PD5: forgetfulness, which occurs when the barrier is multiple medication management or a stressed and busy life, when patients are prone to forget to take their medication

The latter three were the most commonly targeted determinants for behavior change in patients.

Matrix of Change Objectives

In the resulting MoCO, we listed the delivered DIs in the selected studies with respect to the intended behaviors to change and the targeted key determinant to improve BP control or MA in patients.

Table 1 indicates the matrix for the patient level, and Table 2 illustrates the matrix for the health care team and system level. Of the 54 studies, 63% (34/54) of the studies involved both levels for the DI. We categorized the studies inside each cell based on the delivery modalities.

Table 1. Matrix of change objectives at the patient level, with the studies targeting medication adherence or blood pressure improvement in patients with hypertension categorized by digital intervention delivery modality.

Patient behaviour	PD1[a]	PD2[b]	PD3[c]	PD4[d]	PD5[e]
PB1[f]					
EHR[g]	—[h]	[49]	—	—	—
Email	—	[31]	—	—	—
PB2[i]					
EHR	—	[48,49]	[48]	—	—
mHealth[j]	—	—	—	—	[57]
SMS	—	—	—	—	[18]
Web	—	[48,58]	[48]	—	—
PB3[k]					
Com[l]	[59,60]	—	—	—	—
DM[m]	—	—	—	—	[44]
EHR	—	[49,61,62]	—	—	—
Email	—	—	[63]	—	[31,64]
ePill[n]	—	—	—	—	[64]
mHealth	[33,65,66]	[44,61,67,68]	[69]	—	[18,33,57,61,66-71]
Phone	[49,59-65,72-74]	[75]	[48,58,63,74,76,77]	—	[32,64,72,75,78]
SMS	[18]	[18]	[18,63,67,79]	[18]	[18,30,50,63,64,67,77,79-81]
Video	[74,82]	—	[74]	—	[78]
Web	—	[61,62]	[48,83,84]	—	[61,85]
PB4[o]					
Com	[59,60]	—	[60]	—	—
EHR	—	[49,86]	[48]	—	—
Email	—	[87]	[58,63]	—	[84]
mHealth	[65,69]	[57]	[70,88]	—	—
Phone	[49,59-65,73,74]	[87]	[48,58,60,63,74,76]	—	[32]
SMS	—	—	—	—	[18,30,81]
Web	[51]	[51,60,86,87]	[48,51,56,60,83]	—	[62]
PB5[p]					
EHR	—	[61]	—	—	—
Email	[60,89]	[31,87]	[60]	—	[31,64,68,71,84,87,88]
mHealth	[66]	[57,61,66-68,71,86,88,90]	[70,88]	—	[57,68,90]
Phone	[73]	[72,75,87]	[48]	—	[68,75,87]
SMS	—	—	[67]	—	[64,67,91,92]
TelDev[q]	—	[67,68,85,90,93]	[70,85]	—	[68]
Video	[82]	—	—	—	—
Web	[31,60,85]	[31,57,60,61,85,87,90,93]	[31,48,58,60,83,90]	[56]	[85]
PB6[r]					
Com	[59,60]	—	—	—	—

Patient behaviour	PD1[a]	PD2[b]	PD3[c]	PD4[d]	PD5[e]
EHR	—	[58,61]	—	—	—
Email	[60,89]	[93]	[31,58,60,94]	—	[31]
mHealth	[57,65,66,88,90]	[44,57,61]	[57,61,70,81]	—	[57,61,81]
Phone	[59-65,72,73,75,95]	[93]	[48,58,76,96]	—	[75,94]
SMS	[18,67]	[18]	[18,50,79,90]	[18]	[90]
TelDev	—	[44]	—	—	—
Web	[31,51,60,85,88]	[61]	[31,48,56,60,61,83,90]	[56]	[61]
PB7[s]					
mHealth	—	—	[61]	—	[68]
Phone	—	[72,75]	[75]	—	[75]
Web	[51]	[88]	[51,61,88]	—	—

[a]PD1: patient-level determinant 1, low perceived risk to health.
[b]PD2: patient-level determinant 2, lack of education.
[c]PD3: patient-level determinant 3, lack of self-efficacy.
[d]PD4: patient-level determinant 4, lack of self-awareness.
[e]PD5: patient-level determinant 5, forgetfulness.
[f]PB1: patient behavior 1, obtain the correct prescription when needed.
[g]EHR: electronic health record.
[h]No studies found.
[i]PB2: patient behavior 2, initiate/refill medication.
[j]mHealth: mobile health.
[k]PB3: patient behavior 3, take/ingest medication.
[l]Com: computer-based programs.
[m]DM: digital medicine.
[n]ePill: electronic pill boxes.
[o]PB4: patient behavior 4, interactive patient-provider communication.
[p]PB5: patient behavior 5, self-measurement.
[q]TelDev: telemetric device.
[r]PB6: patient behavior 6, lifestyle management.
[s]PB7: patient behavior 7, get support from peers or family members.

Table 2. Matrix of change objectives at the health care team and system level, with the studies targeting medication adherence and blood pressure improvement in patients with hypertension categorized by digital intervention delivery modality.

Patient behaviour	HD1[a]	HD2[b]	HD3[c]	HD4[d]
PB1[e]				
EHR[f]	[49]	—[g]	—	—
Email	—	—	[31]	—
mHealth[h]	[30]	—	[30]	[30]
Web	—	[50]	[70,87]	—
PB2[i]				
EHR	[49]	—	—	—
PB3[j]				
DM[k]	—	—	[44,91]	—
EHR	[49]	—	[97]	—
Email	—	—	[64,75]	[63,72]
ePill[l]	—	—	[64,76,77,97-99]	—
mHealth	—	—	[67,69]	—
Phone	—	—	[48,75]	[58,63]
Web	—	—	[31,44,48,69,70,77,91]	[72]
PB4[m]				
Com[n]	—	—	—	[59]
EHR	[49]	—	—	[86]
Email	—	—	—	[63]
Phone	[49]	—	[70]	[48,58,63]
SMS	[18]	—	—	—
Web	[48]	—	[90]	[48,56,86]
PB5[o]				
EHR	—	—	[86]	—
Email	—	—	[64,67,75]	[72]
mHealth	—	—	[57,67,88,90]	—
Phone	—	—	[48,72,75]	[67]
TelDev[p]	—	—	[64,67,90,91]	[93]
Web	—	—	[31,48,57,60,70,86,88,90-92]	[72,93]
PB6[q]				
Email	—	—	—	[72]
mHealth	[30]	—	[29,57]	[30]
Phone	—	—	[48]	[58]
TelDev	—	—	[44,91]	—
Web	—	[50]	[44,48,87,90,91]	[60,72]

[a]HD1: health care team and system-level determinant 1, limited software capabilities.

[b]HD2: health care team and system-level determinant 2, lack of education.

[c]HD3: health care team and system-level determinant 3, clinical inertia (aka therapeutic inertia).

[d]HD4: health care team and system-level determinant 4, inadequate health workforce.

[e]PB1: patient behavior 1, obtain the correct prescription when needed.
[f]EHR: electronic health record.
[g]No studies found.
[h]mHealth: mobile health.
[i]PB2: patient behavior 2, initiate/refill medication.
[j]PB3: patient behavior 3, take/ingest medication.
[k]DM: digital medicine.
[l]ePill: electronic pill boxes.
[m]PB4: patient behavior 4, interactive patient-provider communication.
[n]Com: computer-based programs.
[o]PB5: patient behavior 5, self-measurement.
[p]TelDev: telemetric device.
[q]PB6: patient behavior 6, lifestyle management.

Discussion

Principal Findings

To more effectively develop DIs, which leads to more personalized DIs, there is a need to review what has already been developed to identify the gaps. In this paper, we created a MoCO based on 54 reviewed studies to visualize the entire set of previously delivered DIs as a self-explainable map. Moreover, the map can be used to develop future DIs within this domain. With the help of the matrix, we identified a noticeable need to address the intent of the study regarding which determinants were targeted, what behaviors were targeted for change, and which components of DIs have been used. Tailored interventions addressing each patient's specific barriers to adherence successfully achieve improved MA on a larger scale [100].

Determinants targeted by the studies fell primarily into two levels: patient and health care team and system. This agrees with the results of a systematic review conducted by AlGhurair et al [101] that showed underrepresentation of condition, therapy, and socioeconomic barriers. Thus, there is still room for improvement in DI design by targeting factors from levels other than the patient level. However, according to a review conducted by Klasnja and Pratt [102], health care provider involvement is one of the five key intervention strategies that have been used in phone-based interventions. This also agrees with our inductive analysis, which showed that most of the DIs targeting the patient level (63%) also targeted the health care team and system level.

Using the IM technique can identify appropriate theories for creating DIs. The DIs that did not result in a significant improvement in MA and BP control tended to lack an underlying theory; in fact, 65% of the reviewed studies did not mention the use of theory, models, or principles in their intervention planning, which supports the results from a systematic review of mHealth intervention trials that concluded most interventions failed to incorporate behavioral theories [103]. In contrast, the 10 studies that designed their intervention based on a theoretical model were successful in achieving significant BP reduction or MA improvement or a better rate of BP control or MA.

Most Targeted Change Objectives

The most commonly targeted change objectives in the matrix were concentrated in the cross between the patient-level determinant of forgetfulness and the taking medication behavior (26/54 studies). One reason for this might be that reminders are one of the easiest and least expensive DIs that can address this change objective. The DIs were mainly delivered through SMS or mHealth, which are considered to attract attention sooner, compared to the Web or email. Moreover, it is less expensive than ePill, is less invasive than DM, and requires less effort from the health care team than phone calls.

The second most commonly targeted change objectives were also patient-level determinants, namely a lack of self-efficacy crossed with lifestyle management (22/54 studies) and a lack of education crossed with lifestyle management (22/54 studies). Lifestyle is an essential element in managing hypertension since optimal therapy involves consideration of the patient's diet, exercise, tobacco and alcohol use, compliance, and achievement of BP control [104]. The recent guideline for high BP management also provides new treatment recommendations including lifestyle changes as well as BP-lowering medications [105]. Furthermore, dietary risk factors are linked to poor knowledge of hypertension [106]. Evidence has shown that low self-efficacy is usually related to a low level of physical activity [107] and poor MA [101].

Clinical inertia crossed with self-measurement behavior was the third commonly cited change objective (20/54 studies). Clinical inertia crossed with the taking medication behavior was also found in many studies (17/54 studies). Clinical inertia and MA are most definitely intertwined [108], despite being two separate issues in managing hypertension. The results from this review support this fact. Previous studies tried to increase medication intake and self-measurement behaviors of patients to overcome clinical inertia. However, no study focused on the initiation or refill of medication crossed with the clinical inertia determinant. This might be due to several reasons, including patients are provided with medications, patients already have some medications, or the health care team can check medication disbursement via a system (eg, pharmacy pickups). Nevertheless, MA requires the initial purchase or refill of the medication; this could be a focus of future interventions.

Least Targeted Change Objectives

There were a number of determinants crossed with behaviors to change that were not as commonly addressed by the studies, leading to empty cells in the matrix. If empty cells belong to the health care team and system level, it implies that none of the reviewed studies developed a DI for both levels concurrently. One might find a study that is focused on DIs delivered only to the health care team; however, it was not of interest for this review and was not included. A clear example is many empty cells at the lack of education at the health care team and system level. Only one study [50] had delivered DI to patients and concurrently provided education through a Web portal for the health care team. However, there were some empty cells that were of no interest from a DI point of view. For example, no studies were in the initiation/refill medication behavior crossed with lack of education at the health care team and system level because providing education for the health care team to convince patients to buy or refill their medication does not require a DI or education.

Empty cells at the patient level warrant closer attention since the focus of the review was patient-level DIs. Most empty cells occurred for the get correct prescription when needed behavior. To improve BP control, physicians often adjust antihypertensive medication by increasing the dosage of drugs, switching BP-lowering drugs, or combining different classes of antihypertensive medications [109]. That is why this behavior was targeted at all health care team and system–level DIs but barely at the patient level. Only 2 studies targeted this behavior at the patient level, by increasing patient awareness about elevated BP levels and the need for medication change. However, future DIs can target some of the determinants from the patient level such as a lack of education.

Most of the included DIs did not involve support from family members, especially at the health care team and system level. Assistance and support from peers, family members, and friends can help enhance patients' optimism and self-esteem, ease the stress of being sick, calm depression caused by the disease, and improve sick-role behaviors, which can improve MA [110]. The limited social support indicates a gap that could be addressed in future personalized DIs.

Limitations

Systematic reviews, as key elements in evidence-based health care and research, should avoid selection bias. We identified all the relevant studies published from 2008 to 2018 in PubMed. Searching in one electronic database was possibly not the ideal option; however, the number of papers covering DIs delivered to patients with hypertension to improve BP control or MA was extensive. Further, as PubMed is a broad database targeting the area of interest, its use might reduce any selection bias.

To interpret the included articles, the reviewers rechecked and discussed the contents in the framework to gain consensus.

The distinction between the first four patient-level determinants was not always clear. Educational content can contain some information about the risks as well. Since we did not have access to the content of the DIs and the authors did not always describe them in detail, we categorized the DIs according to what was reported in the published articles.

Conclusions

There is an increasing demand for the wide adoption of digital tools and interventions for patients with hypertension to improve MA and BP control in the short term and quality of life in the long term. To illustrate the analytic results in a self-explainable map based on a common behavior change theory, we built a MoCO using the IM framework. This process highlighted a path for further research in DI design and development in pursuit of MA and BP control in patients with hypertension.

This review highlights the need to design a multi-faceted DI that can be personalized according to patient behavior(s) that need to be changed to overcome the key determinant(s) and considering different levels including patient and provider involvement.

Authors' Contributions

KE conducted the review process and led the creation of this manuscript. AS suggested the use of MoCO to structure the study. Search strategy was developed by KE, AS, and ATE. ATE conducted the search for studies and ran the first round of the review process. KE, AS, ATE, and CG reviewed the papers and ran all the other rounds and discussion groups. ATE wrote the initial draft. Revisions and subsequent drafts were organized and completed by KE, AS, CG, and SN.

Conflicts of Interest

None declared.

References

1. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
2. A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis. Geneva, Switzerland: World Health Organization; 2013.
3. Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ. Selected major risk factors and global and regional burden of disease. The Lancet 2002 Nov;360(9343):1347-1360. [doi: 10.1016/s0140-6736(02)11403-6]

4. Lim S, Vos T, Flaxman A, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2012 Dec;380(9859):2224-2260. [doi: 10.1016/S0140-6736(12)61766-8]
5. Levy RI. The Framingham Study: The Epidemiology of Atherosclerotic Disease. JAMA 1981 Feb 06;245(5):512. [doi: 10.1001/jama.1981.03310300064027]
6. Mennini FS, Marcellusi A, von der Schulenburg JMG, Gray A, Levy P, Sciattella P, et al. Cost of poor adherence to anti-hypertensive therapy in five European countries. Eur J Health Econ 2015 Jan 5;16(1):65-72. [doi: 10.1007/s10198-013-0554-4] [Medline: 24390212]
7. Kohlman-Trigoboff D. Hypertension management in patients with vascular disease: An update. J Vasc Nurs 2016 Sep;34(3):87-92. [doi: 10.1016/j.jvn.2016.01.002] [Medline: 27568315]
8. World Health Organization, UNAIDS. Adherence To Long-term Therapies: Evidence For Action. Geneva, Switzerland: World Health Organization; 2003.
9. Abegaz TM, Shehab A, Gebreyohannes EA, Bhagavathula AS, Elnour AA. Nonadherence to antihypertensive drugs: A systematic review and meta-analysis. Medicine (Baltimore) 2017 Jan;96(4):e5641 [FREE Full text] [doi: 10.1097/MD.0000000000005641] [Medline: 28121920]
10. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
11. Horne R, Weinman J. Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. Journal of Psychosomatic Research 1999 Dec;47(6):555-567. [doi: 10.1016/s0022-3999(99)00057-4]
12. Horne R, Chapman SCE, Parham R, Freemantle N, Forbes A, Cooper V. Understanding patients' adherence-related beliefs about medicines prescribed for long-term conditions: a meta-analytic review of the Necessity-Concerns Framework. PLoS One 2013 Dec 2;8(12):e80633 [FREE Full text] [doi: 10.1371/journal.pone.0080633] [Medline: 24312488]
13. Conn VS, Ruppar TM, Chase JD, Enriquez M, Cooper PS. Interventions to Improve Medication Adherence in Hypertensive Patients: Systematic Review and Meta-analysis. Curr Hypertens Rep 2015 Dec 11;17(12):94 [FREE Full text] [doi: 10.1007/s11906-015-0606-5] [Medline: 26560139]
14. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
15. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984 Jan 01;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
16. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
17. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Sep;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
18. Bobrow K, Brennan T, Springer D, Levitt NS, Rayner B, Namane M, et al. Efficacy of a text messaging (SMS) based intervention for adults with hypertension: protocol for the StAR (SMS Text-message Adherence suppoRt trial) randomised controlled trial. BMC Public Health 2014 Jan 11;14(1):28 [FREE Full text] [doi: 10.1186/1471-2458-14-28] [Medline: 24410738]
19. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018 May 07;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
20. 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 1;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
21. Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JCM. Effects of eHealth interventions on medication adherence: a systematic review of the literature. J Med Internet Res 2011 Dec 05;13(4):e103 [FREE Full text] [doi: 10.2196/jmir.1738] [Medline: 22138112]
22. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010 Mar;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
23. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mHealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
24. Granger BB, Bosworth HB. Medication adherence: emerging use of technology. Curr Opin Cardiol 2011 Jul;26(4):279-287 [FREE Full text] [doi: 10.1097/HCO.0b013e328347c150] [Medline: 21597368]
25. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

26. Kassavou A, Sutton S. Automated telecommunication interventions to promote adherence to cardio-metabolic medications: meta-analysis of effectiveness and meta-regression of behaviour change techniques. Health Psychol Rev 2017 Sep 12:1-18. [doi: 10.1080/17437199.2017.1365617] [Medline: 28805162]
27. Fernandez ME, Ten Hoor GA, van Lieshout S, Rodriguez SA, Beidas RS, Parcel G, et al. Implementation Mapping: Using Intervention Mapping to Develop Implementation Strategies. Front Public Health 2019 Jun 18;7:158 [FREE Full text] [doi: 10.3389/fpubh.2019.00158] [Medline: 31275915]
28. Eldredge LKB, Markham CM, Ruiter RAC, Fernandez ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco, CA: Jossey-Bass; 2016.
29. Jha D, Gupta P, Ajay VS, Jindal D, Perel P, Prieto-Merino D, et al. Protocol for the mWellcare trial: a multicentre, cluster randomised, 12-month, controlled trial to compare the effectiveness of mWellcare, an mHealth system for an integrated management of patients with hypertension and diabetes, versus enhanced usual care in India. BMJ Open 2017 Aug 11;7(8):e014851 [FREE Full text] [doi: 10.1136/bmjopen-2016-014851] [Medline: 28801393]
30. Band R, Morton K, Stuart B, Raftery J, Bradbury K, Yao GL, et al. Home and Online Management and Evaluation of Blood Pressure (HOME BP) digital intervention for self-management of uncontrolled, essential hypertension: a protocol for the randomised controlled HOME BP trial. BMJ Open 2016 Nov 07;6(11):e012684 [FREE Full text] [doi: 10.1136/bmjopen-2016-012684] [Medline: 27821598]
31. Bosworth HB, Olsen MK, McCant F, Stechuchak KM, Danus S, Crowley MJ, et al. Telemedicine cardiovascular risk reduction in veterans: The CITIES trial. Am Heart J 2018 May;199:122-129. [doi: 10.1016/j.ahj.2018.02.002] [Medline: 29754649]
32. Patel S, Jacobus-Kantor L, Marshall L, Ritchie C, Kaplinski M, Khurana PS, et al. Mobilizing your medications: an automated medication reminder application for mobile phones and hypertension medication adherence in a high-risk urban population. J Diabetes Sci Technol 2013 May 01;7(3):630-639 [FREE Full text] [doi: 10.1177/193229681300700307] [Medline: 23759395]
33. Appel LJ, Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Writing Group of the PREMIER Collaborative Research Group. Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. JAMA 2003 Apr 23;289(16):2083-2093. [doi: 10.1001/jama.289.16.2083] [Medline: 12709466]
34. Julius S, Nesbitt SD, Egan BM, Weber MA, Michelson EL, Kaciroti N, et al. Feasibility of Treating Prehypertension with an Angiotensin-Receptor Blocker. N Engl J Med 2006 Apr 20;354(16):1685-1697. [doi: 10.1056/nejmoa060838]
35. Pickering TG, Hall JE, Appel LJ, Falkner BE, Graves J, Hill MN, et al. Recommendations for Blood Pressure Measurement in Humans and Experimental Animals. Hypertension 2005 Jan;45(1):142-161. [doi: 10.1161/01.hyp.0000150859.47929.8e]
36. Bartholomew L, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
37. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4. [Medline: 10345255]
38. Locke EA. Self-efficacy: the exercise of control. Choice Reviews Online 1997 Nov 01;35(03):35-1826-35-1826. [doi: 10.5860/choice.35-1826]
39. Bandura A. Social cognitive theory. In: The International Encyclopedia of Communication. Wiley Online Library: Int Encycl Commun; 2008:1-60.
40. Rosenstock IM. Historical Origins of the Health Belief Model. Health Education Monographs 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
41. Glanz K, Rimer BK, Viswanath K. Health Behavior And Health Education: Theory, Research, And Practice. Wiley: Jossey-bass; 2020.
42. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
43. Leventhal H, Brissette I. The common-sense model of self-regulation of health and illness. In: Cameron L, Leventhal H, editors. The Self-Regulation of Health and Illness Behaviour. London: Routledge; 2003:42-65.
44. Frias J, Virdi N, Raja P, Kim Y, Savage G, Osterberg L. Effectiveness of Digital Medicines to Improve Clinical Outcomes in Patients with Uncontrolled Hypertension and Type 2 Diabetes: Prospective, Open-Label, Cluster-Randomized Pilot Clinical Trial. J Med Internet Res 2017 Jul 11;19(7):e246 [FREE Full text] [doi: 10.2196/jmir.7833] [Medline: 28698169]
45. Aldeer M, Javanmard M, Martin R. A Review of Medication Adherence Monitoring Technologies. ASI 2018 May 06;1(2):14. [doi: 10.3390/asi1020014]
46. Schoenthaler A, Chaplin WF, Allegrante JP, Fernandez S, Diaz-Gloster M, Tobin JN, et al. Provider communication effects medication adherence in hypertensive African Americans. Patient Educ Couns 2009 May;75(2):185-191 [FREE Full text] [doi: 10.1016/j.pec.2008.09.018] [Medline: 19013740]
47. Schoenthaler A, Allegrante JP, Chaplin W, Ogedegbe G. The effect of patient-provider communication on medication adherence in hypertensive black patients: does race concordance matter? Ann Behav Med 2012 Jun 20;43(3):372-382 [FREE Full text] [doi: 10.1007/s12160-011-9342-5] [Medline: 22270266]

48. Ralston JD, Cook AJ, Anderson ML, Catz SL, Fishman PA, Carlson J, et al. Home blood pressure monitoring, secure electronic messaging and medication intensification for improving hypertension control: a mediation analysis. Appl Clin Inform 2014;5(1):232-248 [FREE Full text] [doi: 10.4338/ACI-2013-10-RA-0079] [Medline: 24734136]
49. Persell SD, Eder M, Friesema E, Connor C, Rademaker A, French DD, et al. EHR-based medication support and nurse-led medication therapy management: rationale and design for a three-arm clinic randomized trial. J Am Heart Assoc 2013 Oct 24;2(5):e000311 [FREE Full text] [doi: 10.1161/JAHA.113.000311] [Medline: 24157649]
50. Mills KT, Dolan J, Bazzano LA, Chen J, He J, Krousel-Wood M, et al. Comprehensive Approach for Hypertension Control in Low-Income Populations: Rationale and Study Design for the Hypertension Control Program in Argentina. Am J Med Sci 2014 Aug;348(2):139-145. [doi: 10.1097/maj.0000000000000298]
51. Barbera M, Mangialasche F, Jongstra S, Guillemont J, Ngandu T, Beishuizen C, et al. Designing an Internet-Based Multidomain Intervention for the Prevention of Cardiovascular Disease and Cognitive Impairment in Older Adults: The HATICE Trial. JAD 2018 Feb 20;62(2):649-663. [doi: 10.3233/jad-170858]
52. Phillips LS, Branch WT, Cook CB, Doyle JP, El-Kebbi IM, Gallina DL, et al. Clinical inertia. Ann Intern Med 2001 Nov 06;135(9):825-834. [doi: 10.7326/0003-4819-135-9-200111060-00012] [Medline: 11694107]
53. Henriksen K, Battles JB, Marks ES, Lewin DI. Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology). US: Rockville (MD): Agency for Healthcare Research and Quality; 2005.
54. Huebschmann A, Mizrahi T, Soenksen A, Beaty B, Denberg T. Reducing clinical inertia in hypertension treatment: a pragmatic randomized controlled trial. J Clin Hypertens (Greenwich) 2012 May;14(5):322-329 [FREE Full text] [doi: 10.1111/j.1751-7176.2012.00607.x] [Medline: 22533659]
55. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Joint National Committee on Prevention, Detection, Evaluation,Treatment of High Blood Pressure. National Heart, Lung,Blood Institute, National High Blood Pressure Education Program Coordinating Committee. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension 2003 Dec;42(6):1206-1252 [FREE Full text] [doi: 10.1161/01.HYP.0000107251.49515.c2] [Medline: 14656957]
56. Goessens BMB, Visseren FLJ, de Nooijer J, van den Borne HW, Algra A, Wierdsma J, SMART Study Group. A pilot-study to identify the feasibility of an Internet-based coaching programme for changing the vascular risk profile of high-risk patients. Patient Educ Couns 2008 Oct;73(1):67-72. [doi: 10.1016/j.pec.2008.06.004] [Medline: 18639410]
57. Bengtsson U, Kjellgren K, Hallberg I, Lindwall M, Taft C. Improved Blood Pressure Control Using an Interactive Mobile Phone Support System. J Clin Hypertens (Greenwich) 2016 Feb;18(2):101-108 [FREE Full text] [doi: 10.1111/jch.12682] [Medline: 26456490]
58. Green BB, Ralston JD, Fishman PA, Catz SL, Cook A, Carlson J, et al. Electronic communications and home blood pressure monitoring (e-BP) study: design, delivery, and evaluation framework. Contemp Clin Trials 2008 May;29(3):376-395. [doi: 10.1016/j.cct.2007.09.005] [Medline: 17974502]
59. Reed SD, Li Y, Oddone EZ, Neary AM, Orr MM, Grubber JM, et al. Economic evaluation of home blood pressure monitoring with or without telephonic behavioral self-management in patients with hypertension. Am J Hypertens 2010 Feb 01;23(2):142-148 [FREE Full text] [doi: 10.1038/ajh.2009.215] [Medline: 19927132]
60. Salisbury C, O'Cathain A, Thomas C, Edwards L, Gaunt D, Dixon P, et al. Telehealth for patients at high risk of cardiovascular disease: pragmatic randomised controlled trial. BMJ 2016 Jun 01;353:i2647 [FREE Full text] [doi: 10.1136/bmj.i2647] [Medline: 27252245]
61. Redfern J, Usherwood T, Harris MF, Rodgers A, Hayman N, Panaretto K, et al. A randomised controlled trial of a consumer-focused e-health strategy for cardiovascular risk management in primary care: the Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) study protocol. BMJ Open 2014 Jan 31;4(2):e004523 [FREE Full text] [doi: 10.1136/bmjopen-2013-004523] [Medline: 24486732]
62. 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 08;17(10):e226 [FREE Full text] [doi: 10.2196/jmir.4872] [Medline: 26449757]
63. Choudhry NK, Isaac T, Lauffenburger JC, Gopalakrishnan C, Khan NF, Lee M, et al. Rationale and design of the Study of a Tele-pharmacy Intervention for Chronic diseases to Improve Treatment adherence (STIC2IT): A cluster-randomized pragmatic trial. Am Heart J 2016 Oct;180:90-97 [FREE Full text] [doi: 10.1016/j.ahj.2016.07.017] [Medline: 27659887]
64. McGillicuddy JW, Taber DJ, Mueller M, Patel S, Baliga PK, Chavin KD, et al. Sustainability of improvements in medication adherence through a mobile health intervention. Prog Transplant 2015 Sep;25(3):217-223. [doi: 10.7182/pit2015975] [Medline: 26308780]
65. Najafi Ghezeljeh T, Sharifian S, Nasr Isfahani M, Haghani H. Comparing the effects of education using telephone follow-up and smartphone-based social networking follow-up on self-management behaviors among patients with hypertension. Contemp Nurse 2018 Mar 05;54(4-5):362-373. [doi: 10.1080/10376178.2018.1441730] [Medline: 29451091]
66. Ashoorkhani M, Bozorgi A, Majdzadeh R, Hosseini H, Yoonessi A, Ramezankhani A, et al. Comparing the effectiveness of the BPMAP (Blood Pressure Management Application) and usual care in self-management of primary hypertension and adherence to treatment in patients aged 30-60 years: study protocol for a randomized controlled trial. Trials 2016 Oct 21;17(1):511 [FREE Full text] [doi: 10.1186/s13063-016-1638-0] [Medline: 27769282]

67. Ovbiagele B. Phone-based intervention under nurse guidance after stroke: concept for lowering blood pressure after stroke in Sub-Saharan Africa. J Stroke Cerebrovasc Dis 2015 Jan;24(1):1-9 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2014.08.011] [Medline: 25440360]
68. Morawski K, Ghazinouri R, Krumme A, McDonough J, Durfee E, Oley L, et al. Rationale and design of the Medication adherence Improvement Support App For Engagement-Blood Pressure (MedISAFE-BP) trial. Am Heart J 2017 Apr;186:40-47. [doi: 10.1016/j.ahj.2016.11.007] [Medline: 28454831]
69. Anglada-Martínez H, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Sequeira E, Aragunde V, et al. Feasibility and Preliminary Outcomes of a Web and Smartphone-Based Medication Self-Management Platform for Chronically Ill Patients. J Med Syst 2016 Apr 12;40(4):99. [doi: 10.1007/s10916-016-0456-y] [Medline: 26872781]
70. Xu L, Fang W, Zhu F, Zhang H, Liu K. A coordinated PCP-Cardiologist Telemedicine Model (PCTM) in China's community hypertension care: study protocol for a randomized controlled trial. Trials 2017 May 25;18(1):236 [FREE Full text] [doi: 10.1186/s13063-017-1970-z] [Medline: 28545514]
71. Albini F, Xiaoqiu Liu, Torlasco C, Soranna D, Faini A, Ciminaghi R, et al. An ICT and mobile health integrated approach to optimize patients' education on hypertension and its management by physicians: The Patients Optimal Strategy of Treatment(POST) pilot study. In: Conf Proc IEEE Eng Med Biol Soc. An ICT and mobile health integrated approach to optimize patients? education on hypertension and its management by physicians: The Patients Optimal Strategy of Treatment(POST) pilot study. Inth Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016 Aug Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 16-20 Aug. 2016; Orlando, FL, USA p. 517-520. [doi: 10.1109/EMBC.2016.7590753]
72. Piette JD, Marinec N, Janda K, Morgan E, Schantz K, Yujra ACA, et al. Structured Caregiver Feedback Enhances Engagement and Impact of Mobile Health Support: A Randomized Trial in a Lower-Middle-Income Country. Telemed J E Health 2016 Apr;22(4):261-268 [FREE Full text] [doi: 10.1089/tmj.2015.0099] [Medline: 26352854]
73. Lynch CP, Williams JS, J Ruggiero K, G Knapp R, Egede LE. Tablet-Aided BehavioraL intervention EffecT on Self-management skills (TABLETS) for Diabetes. Trials 2016 Mar 22;17(1):157 [FREE Full text] [doi: 10.1186/s13063-016-1243-2] [Medline: 27005766]
74. Williams A, Manias E, Walker R, Gorelik A. A multifactorial intervention to improve blood pressure control in co-existing diabetes and kidney disease: a feasibility randomized controlled trial. J Adv Nurs 2012 Nov;68(11):2515-2525. [doi: 10.1111/j.1365-2648.2012.05950.x] [Medline: 22335395]
75. Piette JD, Datwani H, Gaudioso S, Foster SM, Westphal J, Perry W, et al. Hypertension management using mobile technology and home blood pressure monitoring: results of a randomized trial in two low/middle-income countries. Telemed J E Health 2012 Oct;18(8):613-620 [FREE Full text] [doi: 10.1089/tmj.2011.0271] [Medline: 23061642]
76. Ogedegbe GO, Boutin-Foster C, Wells MT, Allegrante JP, Isen AM, Jobe JB, et al. A randomized controlled trial of positive-affect intervention and medication adherence in hypertensive African Americans. Arch Intern Med 2012 Feb 27;172(4):322-326 [FREE Full text] [doi: 10.1001/archinternmed.2011.1307] [Medline: 22269592]
77. Brath H, Morak J, Kästenbauer T, Modre-Osprian R, Strohner-Kästenbauer H, Schwarz M, et al. Mobile health (mHealth) based medication adherence measurement - a pilot trial using electronic blisters in diabetes patients. Br J Clin Pharmacol 2013 Sep 06;76 Suppl 1:47-55 [FREE Full text] [doi: 10.1111/bcp.12184] [Medline: 24007452]
78. Petry NM, Alessi SM, Byrne S, White WB. Reinforcing adherence to antihypertensive medications. J Clin Hypertens (Greenwich) 2015 Jan;17(1):33-38 [FREE Full text] [doi: 10.1111/jch.12441] [Medline: 25413264]
79. 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 21;19(12):1276-1284 [FREE Full text] [doi: 10.1111/jch.13098] [Medline: 28941056]
80. Haramiova Z, Stasko M, Hulin M, Tesar T, Kuzelova M, Morisky DM. The effectiveness of daily SMS reminders in pharmaceutical care of older adults on improving patients' adherence to antihypertensive medication (SPPA): study protocol for a randomized controlled trial. Trials 2017 Jul 18;18(1):334 [FREE Full text] [doi: 10.1186/s13063-017-2063-8] [Medline: 28720121]
81. Jindal D, Gupta P, Jha D, Ajay VS, Goenka S, Jacob P, et al. Development of mWellcare: an mHealth intervention for integrated management of hypertension and diabetes in low-resource settings. Glob Health Action 2018;11(1):1517930 [FREE Full text] [doi: 10.1080/16549716.2018.1517930] [Medline: 30253691]
82. Yeung DL, Alvarez KS, Quinones ME, Clark CA, Oliver GH, Alvarez CA, et al. Low-health literacy flashcards & mobile video reinforcement to improve medication adherence in patients on oral diabetes, heart failure, and hypertension medications. J Am Pharm Assoc (2003) 2017 Jan;57(1):30-37 [FREE Full text] [doi: 10.1016/j.japh.2016.08.012] [Medline: 27816544]
83. Glaser E, Richard C, Lussier M. The impact of a patient web communication intervention on reaching treatment suggested guidelines for chronic diseases: A randomized controlled trial. Patient Educ Couns 2017 Nov;100(11):2062-2070. [doi: 10.1016/j.pec.2017.05.022] [Medline: 28535926]
84. Thiboutot J, Sciamanna CN, Falkner B, Kephart DK, Stuckey HL, Adelman AM, et al. Effects of a web-based patient activation intervention to overcome clinical inertia on blood pressure control: cluster randomized controlled trial. J Med Internet Res 2013 Sep 04;15(9):e158 [FREE Full text] [doi: 10.2196/jmir.2298] [Medline: 24004475]

85. Watson AJ, Singh K, Myint- U, Grant RW, Jethwani K, Murachver E, et al. Evaluating a web-based self-management program for employees with hypertension and prehypertension: a randomized clinical trial. Am Heart J 2012 Oct;164(4):625-631. [doi: 10.1016/j.ahj.2012.06.013] [Medline: 23067923]
86. Istepanian R, Sungoor A, Earle K. Technical and compliance considerations for mobile health self-monitoring of glucose and blood pressure for patients with diabetes. In: Conf Proc IEEE Eng Med Biol Soc. 2009 Presented at: InAnnual International Conference of the IEEE Engineering in Medicine and Biology Society; 2009; Minneapolis, MN, USA p. 5130-5133. [doi: 10.1109/IEMBS.2009.5334580]
87. DeAlleaume L, Parnes B, Zittleman L, Sutter C, Chavez R, Bernstein J, et al. Success in the Achieving CARdiovascular Excellence in Colorado (A CARE) Home Blood Pressure Monitoring Program: A Report from the Shared Networks of Colorado Ambulatory Practices and Partners (SNOCAP). J Am Board Fam Med 2015 Sep 09;28(5):548-555 [FREE Full text] [doi: 10.3122/jabfm.2015.05.150024] [Medline: 26355126]
88. Kim JY, Wineinger NE, Steinhubl SR. The Influence of Wireless Self-Monitoring Program on the Relationship Between Patient Activation and Health Behaviors, Medication Adherence, and Blood Pressure Levels in Hypertensive Patients: A Substudy of a Randomized Controlled Trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
89. Devroey D, Senesael, Van De Vijver, Borgermans. Effectiveness of a quality improvement intervention targeting cardiovascular risk factors: are patients responsive to information and encouragement by mail or post? VHRM 2013 Feb:13. [doi: 10.2147/vhrm.s39919]
90. Yoo HJ, Park MS, Kim TN, Yang SJ, Cho GJ, Hwang TG, et al. A Ubiquitous Chronic Disease Care system using cellular phones and the internet. Diabet Med 2009 Jun;26(6):628-635. [doi: 10.1111/j.1464-5491.2009.02732.x] [Medline: 19538239]
91. DiCarlo LA, Weinstein RL, Morimoto CB, Savage GM, Moon GL, Au-Yeung K, et al. Patient-Centered Home Care Using Digital Medicine and Telemetric Data for Hypertension: Feasibility and Acceptability of Objective Ambulatory Assessment. J Clin Hypertens (Greenwich) 2016 Sep;18(9):901-906 [FREE Full text] [doi: 10.1111/jch.12787] [Medline: 26890041]
92. Lee P, Liu J, Hsieh M, Hao W, Tseng Y, Liu S, et al. Cloud-based BP system integrated with CPOE improves self-management of the hypertensive patients: A randomized controlled trial. Comput Methods Programs Biomed 2016 Aug;132:105-113. [doi: 10.1016/j.cmpb.2016.04.003] [Medline: 27282232]
93. Lee C, Chang B. Effect of Disease Improvement with Self-Measurement Compliance (Measurement Frequency Level) in SmartCare Hypertension Management Service. Telemed J E Health 2016 Mar;22(3):238-245. [doi: 10.1089/tmj.2015.0045] [Medline: 26252620]
94. Cicolini G, Simonetti V, Comparcini D, Celiberti I, Di Nicola M, Capasso LM, et al. Efficacy of a nurse-led email reminder program for cardiovascular prevention risk reduction in hypertensive patients: a randomized controlled trial. Int J Nurs Stud 2014 Jun;51(6):833-843. [doi: 10.1016/j.ijnurstu.2013.10.010] [Medline: 24225325]
95. Scala D, Menditto E, Caruso G, Monetti VM, Orlando V, Guerriero F, et al. Are you more concerned about or relieved by medicines? An explorative randomized study of the impact of telephone counseling by pharmacists on patients' beliefs regarding medicines and blood pressure control. Patient Educ Couns 2018 Apr;101(4):679-686. [doi: 10.1016/j.pec.2017.12.006] [Medline: 29249596]
96. Hyman DJ, Pavlik VN, Greisinger AJ, Chan W, Bayona J, Mansyur C, et al. Effect of a physician uncertainty reduction intervention on blood pressure in uncontrolled hypertensives--a cluster randomized trial. J Gen Intern Med 2012 Apr 27;27(4):413-419 [FREE Full text] [doi: 10.1007/s11606-011-1888-1] [Medline: 22033742]
97. Schoenthaler A, De La Calle F, Barrios-Barrios M, Garcia A, Pitaro M, Lum A, et al. A practice-based randomized controlled trial to improve medication adherence among Latinos with hypertension: study protocol for a randomized controlled trial. Trials 2015 Jul 02;16(1):290 [FREE Full text] [doi: 10.1186/s13063-015-0815-x] [Medline: 26134837]
98. van Onzenoort HAW, Verberk WJ, Kroon AA, Kessels AGH, Neef C, van der Kuy PM, et al. Electronic monitoring of adherence, treatment of hypertension, and blood pressure control. Am J Hypertens 2012 Jan 01;25(1):54-59. [doi: 10.1038/ajh.2011.153] [Medline: 21993365]
99. Kronish IM, Moise N, McGinn T, Quan Y, Chaplin W, Gallagher BD, et al. An Electronic Adherence Measurement Intervention to Reduce Clinical Inertia in the Treatment of Uncontrolled Hypertension: The MATCH Cluster Randomized Clinical Trial. J Gen Intern Med 2016 Nov 2;31(11):1294-1300 [FREE Full text] [doi: 10.1007/s11606-016-3757-4] [Medline: 27255750]
100. Zullig LL, Blalock DV, Dougherty S, Henderson R, Ha CC, Oakes MM, et al. The new landscape of medication adherence improvement: where population health science meets precision medicine. Patient Prefer Adherence 2018;12:1225-1230 [FREE Full text] [doi: 10.2147/PPA.S165404] [Medline: 30034226]
101. AlGhurair S, Hughes C, Simpson S, Guirguis L. A systematic review of patient self-reported barriers of adherence to antihypertensive medications using the world health organization multidimensional adherence model. J Clin Hypertens (Greenwich) 2012 Dec;14(12):877-886 [FREE Full text] [doi: 10.1111/j.1751-7176.2012.00699.x] [Medline: 23205755]
102. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]

103. 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]
104. Trilling J, Froom J. The urgent need to improve hypertension care. Arch Fam Med 2000;9(9):794-801. [doi: 10.1001/archfami.9.9.794] [Medline: 11031384]
105. Whelton P, Carey R, Aronow W, Casey D, Collins K, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
106. Chimberengwa PT, Naidoo M, cooperative inquiry group. Knowledge, attitudes and practices related to hypertension among residents of a disadvantaged rural community in southern Zimbabwe. PLoS One 2019 Jun 25;14(6):e0215500 [FREE Full text] [doi: 10.1371/journal.pone.0215500] [Medline: 31237883]
107. Prince SA, Reed JL, Nerenberg KA, Kristjansson EA, Hiremath S, Adamo KB, et al. Intrapersonal, social and physical environmental determinants of moderate-to-vigorous physical activity in working-age women: a systematic review protocol. Syst Rev 2014 Nov 04;3(1):132 [FREE Full text] [doi: 10.1186/2046-4053-3-132] [Medline: 25370736]
108. Engel-Nitz NM, Darkow T, Lau H. Antihypertensive medication changes and blood pressure goal achievement in a managed care population. J Hum Hypertens 2010 Oct 28;24(10):659-668. [doi: 10.1038/jhh.2010.2] [Medline: 20107489]
109. Václavík J, Vysočanová P, Seidlerová J, Zajíček P, Petrák O, Dlask J, et al. Reasons for Switching Antihypertensive Medication in General Practice. Medicine 2014;93(27):e168. [doi: 10.1097/md.0000000000000168]
110. Scheurer D, Choudhry N, Swanton KA, Matlin O, Shrank W. Association between different types of social support and medication adherence. Am J Manag Care 2012 Dec 01;18(12):e461-e467 [FREE Full text] [Medline: 23286676]

Abbreviations

BP: blood pressure
DI: digital intervention
DM: digital medicine
EHR: electronic health record
ePill: electronic pill boxes
IM: Intervention Mapping
MA: medication adherence
mHealth: mobile health
MoCO: matrix of change objectives
TelDev: telemetric devices
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by AR Razavi, S Madani; comments to author 07.01.20; accepted 26.01.20; published 09.04.20.

Please cite as:
Etminani K, Tao Engström A, Göransson C, Sant'Anna A, Nowaczyk S
How Behavior Change Strategies are Used to Design Digital Interventions to Improve Medication Adherence and Blood Pressure Among Patients With Hypertension: Systematic Review
J Med Internet Res 2020;22(4):e17201
URL: https://www.jmir.org/2020/4/e17201
doi:10.2196/17201
PMID:32271148

Review

Mobile Health for Perinatal Depression and Anxiety: Scoping Review

Neesha Hussain-Shamsy[1,2], BA, MHS; Amika Shah[1,2], BSc, MScCH; Simone N Vigod[1,3,4], MD; Juveria Zaheer[4,5], MD; Emily Seto[1,2], PEng, PhD

[1]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Centre for eHealth Global Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[3]Women's College Hospital and Women's College Research Institute, Toronto, ON, Canada

[4]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[5]Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Neesha Hussain-Shamsy, BA, MHS
Institute of Health Policy, Management, and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College Street 4th Floor
Toronto, ON, M5T 3M6
Canada
Phone: 1 416 978 4326
Email: neesha.hussainshamsy@mail.utoronto.ca

Abstract

Background: The perinatal period is a vulnerable time during which depression and anxiety commonly occur. If left untreated or undertreated, there may be significant adverse effects; therefore, access to rapid, effective treatment is essential. Treatments for mild-to-moderate symptoms according to a stepped-care approach involve psychoeducation, peer support, and psychological therapy, all of which have been shown to be efficaciously delivered through digital means. Women experience significant barriers to care because of system- and individual-level factors, such as cost, accessibility, and availability of childcare. The use of mobile phones is widespread in this population, and the delivery of mental health services via mobile phones has been suggested as a means of reducing barriers.

Objective: This study aimed to understand the extent, range, and nature of mobile health (mHealth) tools for prevention, screening, and treatment of perinatal depression and anxiety in order to identify gaps and inform opportunities for future work.

Methods: Using a scoping review framework, 4 databases were searched for terms related to mobile phones, perinatal period, and either depression or anxiety. A total of 477 unique records were retrieved, 81 of which were reviewed by full text. Peer-reviewed publications were included if they described the population as women pregnant or up to 1 year postpartum and a tool explicitly delivered via a mobile phone for preventing, screening, or treating depression or anxiety. Studies published in 2007 or earlier, not in English, or as case reports were excluded.

Results: A total of 26 publications describing 22 unique studies were included (77% published after 2017). mHealth apps were slightly more common than texting-based interventions (12/22, 54% vs 10/22, 45%). Most tools were for either depression (12/22, 54%) or anxiety and depression (9/22, 41%); 1 tool was for anxiety only (1/22, 4%). Interventions starting in pregnancy and continuing into the postpartum period were rare (2/22, 9%). Tools were for prevention (10/22, 45%), screening (6/22, 27%), and treatment (6/22, 27%). Interventions delivered included psychoeducation (16/22, 73%), peer support (4/22, 18%), and psychological therapy (4/22, 18%). Cost was measured in 14% (3/22) studies.

Conclusions: Future work in this growing area should incorporate active psychological treatment, address continuity of care across the perinatal period, and consider clinical sustainability to realize the potential of mHealth.

(J Med Internet Res 2020;22(4):e17011) doi:10.2196/17011

KEYWORDS

mental health; depression; anxiety; pregnancy; postpartum; smartphone; mobile phone; text message; mHealth

Introduction

Background

The perinatal period, often defined as the period between conception and up to 1 year after giving birth, is a vulnerable time during which depression and anxiety commonly occur. Depression and anxiety during this period of time affect 14.5% and 13% of women, respectively, and these are comorbid among 8.1% of new mothers [1-3]. If left untreated or undertreated, there are serious, adverse short- and long-term personal, social, and economic consequences for the mother, child, family unit, and society [4,5]. Access to prevention strategies, screening tools, and rapid and effective treatment is essential.

During the perinatal period, effective treatments may include pharmacological (ie, antidepressant) or psychological treatment, depending on the severity of illness and individual treatment preferences [6]. A stepped-care approach, reflecting high-intensity treatment as illness severity progresses, is recommended [7,8]. This involves a first step involving psychosocial interventions (eg, psychoeducation and peer support), followed by a second step to psychological therapies for illness of mild-to-moderate severity, as well as pharmacological and other biological treatments (ie, antidepressants) as a third step for more severe and/or lasting illness. The risks of antidepressant use in pregnancy are low (selective serotonin reuptake inhibitors are first-line drugs in pregnancy, but serotonin-norepinephrine reuptake inhibitors, bupropion, and mirtazapine can also be used in some circumstances) and must be weighed against the risks and benefits of all other treatment options (eg, no treatment or psychosocial, psychological, or pharmacological treatment alone or in combination) [9,10]. System-level barriers to accessing evidence-based care in this population include long wait lists for specialty psychiatric care and cost and availability of specialists [11]. This population also has unique practical barriers to accessing care, such as the cost and availability of childcare, transportation, and difficulty in scheduling appointments around an infant's often unpredictable needs [12,13].

Electronic health (eHealth) is a broad term used to describe the use of information and communication technologies for health, under which falls mobile health (mHealth) [14]. mHealth specifically relates to the use of mobile computing and communication technologies for health purposes, and it may include apps and text messaging–based (SMS) interventions as well as wearable devices [15]. In some cases, these can be developed and implemented at low cost [16]. These have the potential to help address barriers to accessing care, partly as they offer a greater level of portability and convenience for this population in a way that internet-based interventions designed for access on stationary devices (eg, desktop and laptop computers) cannot [17-19]. In high-income countries, mobile phone ownership is greater than 90%; smartphone (ie, a mobile phone that can access the internet and mobile applications) ownership is as high as 76%, with the highest rates of use among those under 35 years (in other words, who are also in their prime reproductive years) [20]. It is increasingly common to exclusively access the internet via a mobile device [21]. Rates of use and acceptance of mHealth interventions may differ based on a variety of individual factors including age, location (eg, developing country or not), and type of health outcome [22,23].

mHealth tools have the potential to be scalable, cost-effective, and to simultaneously benefit individual patients and the health care system [18,23]. To support mental health in the general population, mHealth tools are used for appointment and medication reminders, information and education, providing support, and self-monitoring. These have also been found to be effective at reducing symptoms when delivering psychotherapy and other behavioral interventions, although they have historically been used outside of the health care system [17,24-27]. Psychological treatments, such as cognitive behavioral therapy (CBT) and mindfulness, behavioral activation, and interpersonal therapy, are evidence-based treatments that can be efficaciously delivered in person and in digital formats, with emerging evidence to support their delivery via mHealth tools [28-30]. mHealth tools could tailor these interventions for perinatal populations to address the gaps in accessing mental health care. Publicly available apps to address postpartum depression were assessed to be of extremely low information quality [27]. The ubiquity of mobile phones, coupled with their potential to deliver effective psychosocial and psychological interventions, makes them a promising avenue through which to address the barriers to care in the delivery of mental health services among women with perinatal depression and anxiety.

Objective

Given the high rate of comorbidities and their similar treatment options, a comprehensive understanding of the current state of mHealth tools to support women with either perinatal depression or anxiety would be helpful to inform potential future work, but this understanding does not yet exist. The objective of this scoping review was to summarize the extent, range, and nature of academic literature about the use of mHealth tools for the prevention, screening, and treatment of perinatal depression and anxiety. Understanding the current state of the art through the results of this review will help identify the gaps in knowledge and practice, thereby informing opportunities for future work.

Methods

Study Design

A scoping review framework, defined by Arksey and O'Malley [31] and further developed by Levac et al [32], was used for this study. A scoping review was the most appropriate manner of assessing the breadth of the literature and identifying gaps and opportunities for further work in this field [33].

Databases and Search Terms

Four databases (Medical Literature Analysis and Retrieval System Online, EMBASE, PsycINFO, and Cumulative Index to Nursing and Allied Health Literature Plus) were searched for terms related to the following: (1) mobile phones, (2) the perinatal period (eg, pregnancy, postpartum, antenatal, and maternal), and (3) depression or anxiety (alone or in combination). Each database was individually searched by title,

abstract, keyword, and subject heading for relevant search terms. A librarian specializing in systematic and scoping review search strategies reviewed the search strings. Rayyan (Qatar Computing Research Institute, Hamad bin Khalifa University), a Web-based platform designed for data management for systematic reviews, was used to manage citations and filter results [34].

Inclusion and Exclusion Criteria

Studies were included if they met the following criteria: (1) defined the primary study population as women who were pregnant or had given birth in the past 12 months; (2) defined perinatal depression or anxiety as the outcome of interest for the intervention; (3) described a study (through a protocol or presentation of study results) related to a tool intentionally and primarily delivered through a mobile phone (including, but not limited to apps, SMS text messaging–based interventions) for the prevention, screening, or treatment of perinatal depression and/or anxiety; (4) were published in a peer-reviewed journal; and (5) intended to recruit (for study protocols) or reported a study sample of >1. Studies were excluded if they met the following criteria: (1) published in 2007 or earlier and (2) published in a language other than English (because of resource limitations in finding translations). The time restriction of articles published after 2007 or earlier reflects the upsurge in publications on mHealth care after this date, as a result of the introduction of the Apple iPhone [15]. Both quantitative and qualitative studies were eligible for inclusion provided all other criteria were met.

Article Selection and Abstraction

Hand searches were conducted by one reviewer (NH-S) who reviewed the reference lists of relevant scoping reviews, systematic reviews, and meta-analyses, as well as reference lists of all publications included in the study for any additional relevant articles.

Two stages of iterative screening for article selection were conducted [32]. In the first stage of screening after the removal of duplicate articles, two reviewers (NH-S and AS) independently assessed articles by title and abstract for potential inclusion in the study. Any articles that were clearly not relevant to this scoping review were excluded. In the second stage of screening, the same reviewers independently applied inclusion and exclusion criteria to all articles (including those found through hand searching) based on a full-text reading of each article. Both reviewers independently conducted data abstraction using an iterative process. Discrepancies in decisions at each stage were resolved through discussion and consensus. The initial search was conducted in February 2019, for a period spanning from 2007 to the search date. The search was updated to ensure capturing any new publications made between February 2019 and July 2019. These articles were selected using the same two-stage selection process described above.

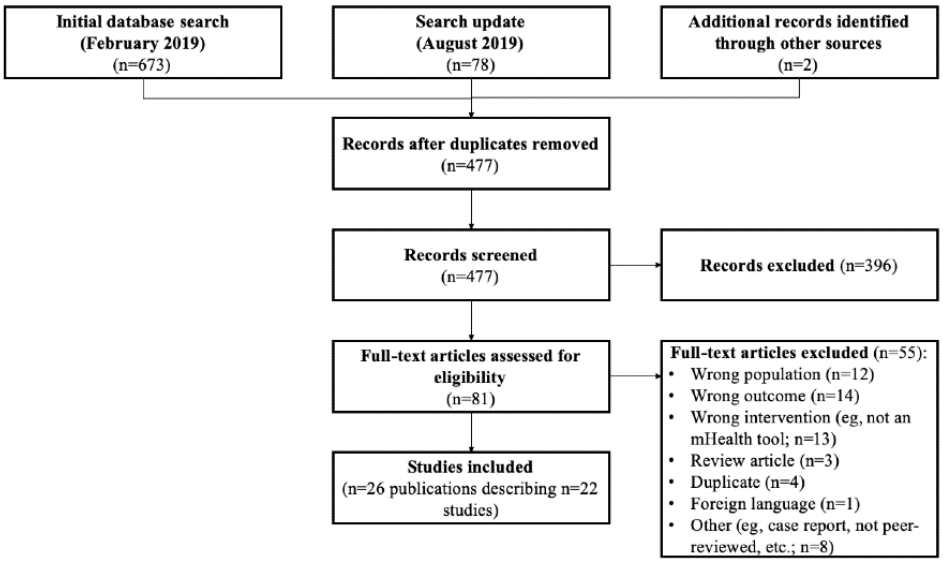
Results

Selected Articles

The initial search of four databases yielded 673 results. An additional 78 records were retrieved in the search update (Figure 1). Hand searching identified 2 additional records. After the removal of duplicate publications, 477 publications (427 from initial search, 48 from the search update, and 2 from hand searching) were reviewed. On the basis of title and abstract, 396 articles were excluded, including 13 systematic reviews whose reference lists were searched. Of the 81 publications reviewed by full text, 55 publications were excluded (wrong population, n=12; intervention targeting an outcome other than depression or anxiety, n=14; nonmobile phone–based intervention, n=13; not in English, n=1; review article, n=3; not peer reviewed, n=6; and duplicate, n=4). Hand searching of the reference lists of the reviews identified in full-text screening identified no additional relevant publications.

Overall, 26 publications (22 papers and 4 conference abstracts) describing 22 unique studies were included in this review. Of these publications, 5 were study protocols for which no results were available. Studies varied by the type of technology used (eg, SMS or app), purpose of the tool (eg, either prevention, screening, or treatment), illness (eg, depression or anxiety), and time point (eg, pregnancy and/or postpartum).

Figure 1. Preferred Reporting Items for Systematic Review and Meta-Analyses flowchart of study selection.



Population

Interventions were limited to be delivered either during pregnancy (7/22, 32%) [35-43] or in the postpartum period (10/22, 45%) [44-54]. One intervention (4%) was designed to be used during pregnancy or postpartum [55,56], and 2 interventions (9%) were designed to begin during pregnancy and continue into the postpartum period [57,58]. For an additional 2 studies (9%), the time period was unclear [59,60]. A total of 7 studies (32%) targeted their interventions to marginalized populations with potentially higher needs as compared with the general population. In all, 2 studies did so through direct approaches and targeted recruitment to low-income women and adolescent mothers [36,57]. A total of 4 studies did so indirectly by recruiting from centers that predominantly service marginalized populations [39,42,43,48,53,54]. One study indirectly included high-risk populations by specifically noting that women who had given birth to multiples or with an infant admitted to a neonatal intensive care unit (NICU) would not be excluded, as those mothers are at high risk for depression and anxiety [47]. Mothers with high-risk pregnancies or infants admitted into the NICU were often specifically excluded from other studies.

Intervention Purpose and Outcome

The purposes of the interventions in the included studies were the following: (1) prevention (10/22, 45%; Multimedia Appendix 1) [35-39,47,49,52,57,59], (2) screening (6/22, 27%; Multimedia Appendix 2) [40-44,50,55,58], and (3) treatment of perinatal mental illness (6/22, 27%; Multimedia Appendix 2) [45,46,48,51,53,60]. The majority of interventions focused on depression only (12/22, 54%) [36,37,40,41,44,46-48,50,51,53,54,58,60]; 1 study (1/22, 4%) focused only on anxiety [35], and 9 studies (9/22, 41%) focused on both illnesses [38,39,42,43,45,49,52,55,59]. Of the 10 studies in which outcomes were reported, 7 studies (70%) reported a positive impact on mental health [35,36,43,45,46,48,57]. Only 4 studies (4/22, 18%) measured long-term outcomes beyond a postintervention survey.

Intervention Strategy

Overall, there was a large variation in the types of strategies used by each mHealth tool to achieve its intended purpose, including psychoeducation (16/22, 73%), symptom monitoring (7/22, 32%), and/or peer support (4/22, 18%). A total of 4 studies (18%) used an active psychological therapy (CBT [2/22, 9%], mindfulness-based therapy [1/22, 4%], and attention bias modification training [1/22, 4%]). In all, studies used other strategies, including tracking movement (2/22, 9%) and encouraging exercise (1/22, 4%). Most studies involved a tool that incorporated multiple strategies (11/22, 50%); most frequently, it was psychoeducation in combination with another strategy (9/22, 41%). One preventative study delivered psychoeducation via SMS using text messages informed by CBT principles, but it was specifically noted that it did not deliver a complete CBT intervention; therefore, this study was counted as a psychoeducational intervention only [55,56].

A total of 6 studies (27%) included a component to help facilitate communication with a health care provider [42,43,50,53,54,58,60]. In all, 2 of these studies used passive communication, where the onus lay on the mother to choose to respond to an SMS text message to receive a phone call from a nurse [47,53,54]. A total of 4 studies used more directive communication with the health care provider, where an automatic alert to a health care provider was sent if symptoms were self-reported via the symptom-monitoring questionnaires delivered by app or SMS and were beyond an acceptable threshold [41,50,58,60].

Study Phase and Research Methodology

Publications described the various phases in the life cycle of such tools from design and development (5/22, 23%), feasibility and acceptability (16/22, 73%), to efficacy and effectiveness (10/22, 45%). Multiple studies had information (through study protocol or results) related to multiple phases of the intervention life cycle. No publications described implementation, scalability, and sustainability efforts. There are an increasing number of efficacy/effectiveness studies over time (Table 1). Most studies (13/22, 59%) solely described quantitative work, whereas only 1 study (1/22, 4%) described research that was solely qualitative in nature. A total of 7 studies (7/22, 32%) described the use of both quantitative and qualitative methods; 1 additional study (1/22, 4%) was explicitly described as being mixed method. Studies were typically designed as randomized controlled trials (RCTs; 9/22, 41%), quasi-experimental (6/22, 27%), or cohort studies (3/22, 13%). Overall, 4 studies (4/22, 18%) incorporated multiple study designs: a cross-sectional survey followed by a quasi-experimental design, a cohort study followed by the development of a machine learning–based algorithm, a longitudinal cohort study culminating in an RCT, and a study described as a qualitative design approach to the development of the tool and intervention.

In all, 3 studies (3/22, 13%) included some measures related to cost, although none provided any data to allow for a comparison between the mobile tool and standard care.

Table 1. Study phase over time, based on year of publication (N=22 studies).

Year[a]	Design and development, n (%)	Feasibility and acceptability, n (%)	Effectiveness and efficacy, n (%)
2008	N/A[b]	1 (4)	N/A
2009	N/A	N/A	N/A
2010	N/A	N/A	N/A
2011	N/A	N/A	N/A
2012	N/A	N/A	N/A
2013	1 (4)	1 (4)	1 (4)
2014	1 (4)	2 (9)	N/A
2015	1 (4)	1 (4)	N/A
2016	N/A	N/A	N/A
2017	N/A	2 (9)	1 (4)
2018	2 (9)	5 (23)	4 (18)
2019	N/A	4 (18)	4 (18)

[a]Note that multiple study phases are reported for some studies.

[b]Not applicable.

Setting

Of the 22 studies, 1 study (4%) was conducted in a lower-middle income economy (Kenya), 3 studies (13%) were conducted in upper-middle income economies (China, Iran, and Thailand), and the remaining studies (18/22, 81%) were conducted in high-income economies (Australia, n=5; Canada, n=2; Spain, n=1; the United Kingdom, n=1; and the United States, n=9), as per their World Bank income group classification [61]. Studies recruited women from a combination of hospitals (13/22, 59%), community-based settings (7/22, 32%), and/or through other means, such as social media or posters (4/22, 18%). One study (1/22, 4%) did not report this information. A total of 8 studies (8/22, 36%) specifically noted that recruitment occurred in an academic hospital or clinic.

Technology

There were more studies using apps (12/22, 54%) compared with SMS-based tools (10/22, 45%) over time, and there was an increase in the number of studies published after 2017 (17/22, 77%). Half of the apps (6/12, 50%) were developed for both iOS and Android devices or as a Web-app, and the other half of the app-based studies either built tools for a single platform (iOS or Android) or did not report this information. A total of 4 studies (4/22, 18%) were explicit in their use of a user-centered design process.

Discussion

Principal Findings

This scoping review found 22 studies related to the prevention, early detection or screening, or treatment of either depression or anxiety in the perinatal period using mobile phones. Interventions most frequently involved psychoeducation and focused on the prevention of illness. Results highlight that research into the use of mHealth tools to support perinatal mental health care is growing and that there are key areas of focus, as described below, for researchers to explore in order for the field to progress.

Results suggest that the potential for mHealth tools to improve access to stepped mental health care for women with either perinatal depression or anxiety is beginning to be realized. The expanded use of mHealth to facilitate delivery of mental health care is most appropriate for women experiencing mild-to-moderate symptoms of depression and anxiety in the perinatal period, which can be addressed using psychosocial and psychological treatments according to a stepped-care approach; the needs of the 11% of women with moderate-to-severe symptoms requiring more intensive clinical management fall beyond the scope of what mobile tools can provide on their own [62,63]. Interventions in this review, which used strategies such as psychoeducation and peer support, are appropriate for prevention and to treat the 53% of women with symptoms that are of mild severity [63]. Many women are already using publicly available apps to access informational support, despite evidence to suggest that the information provided in these apps is incomplete [40,64-66]. The development of mHealth tools in a manner that provides accurate, reliable information and the generation of evidence, as outlined in this review, to show they have the potential to efficaciously prevent or reduce mild symptoms of perinatal depression and/or anxiety is an important positive step.

A major gap highlighted by this review is that only a few tools engaged in the delivery of active psychological therapies (eg, CBT and mindfulness), despite evidence to support the efficacious Web-based delivery of these therapies for general mental health conditions, as well as the emerging evidence to support their delivery via mobile apps [29,67]. These types of strategies and interventions would be beneficial for the 36% of women with postpartum depression who experience symptoms of mild-to-moderate severity [63]. Internet-delivered psychotherapies have lower dropout rates, lower cost, and broader reach than in-person treatment, outcomes that in theory

should be maintained or enhanced by mobile delivery, given that it would also address additional practical barriers to care in this population [68]. For example, the use of an artificial intelligence chatbot to deliver a CBT intervention by SMS, included in this review, addresses many challenges in the delivery of perinatal mental health care by being accessible 24×7 to an unlimited number of patients, avoiding scheduling and transportation issues for users and at the same time having health care providers available to address issues beyond the scope of the chatbot's capacity (eg, talk of self-harm) [60]. In a separate study, when given the option of accessing a peer support group for postpartum women on the Web or through a mobile app, 93.7% of the users opted to use the mobile app, and usage data demonstrated that the majority of app use was between 6 pm and 8 am—a time of the day during which health care providers are not typically available [45]. Modern technological capacity provides the opportunity to develop and use mHealth tools that can address needs in a patient-centered manner, and multiple studies highlighted the interest among both health care professionals and women in the perinatal period to be able to deliver and access mental health support in this manner [69,70].

Challenges With Design, Evaluation, and Sustainability

Publications included in this review primarily described the design and development of an mHealth tool or assessed feasibility. The lack of scale and implementation studies suggests they have yet to be operationalized and scaled for widespread use, particularly in a scientifically rigorous manner. There may be several reasons for this. First, this is a growing area of research, noted in that the majority of studies were published after 2017 and in the increasing number of efficacy and effectiveness studies over time. It is possible that these efforts are underway but have not yet been described in the literature. Second, assessing efficacy and effectiveness are expensive and time-consuming endeavors, given that RCTs have long been considered the gold standard for determining the efficacy of clinical interventions. However, they are rigid and lengthy to conduct, creating the risk that the tool could be obsolete before scale and implementation can begin [71]. There are calls for the evaluation of mHealth tools, in general, and in psychiatry, in particular, to move toward a model of *naturalistic* evaluation through the use of embedded continuous data collection, in-app user surveys (ie, star rating systems for modules immediately after use), and qualitative interviews, thus enabling an evaluation that is more translational in nature [72,73]. Consideration should also be given to focusing on testing the principles of the intervention versus the technology that is used to allow tools to adapt, while remaining adherent to the treatment embedded within them [74]. This could facilitate rapid scale and implementation of mHealth tools.

The development of any tool with the potential to enhance clinical practice and patient experience should ideally incorporate factors that are likely to contribute to success and sustainability. For eHealth tools, user-level factors related to success are engagement with (eg, personal agency, motivation, and values) and the quality of the intervention itself [75]. At the system level, identified factors for success include those related to improvements in the quality of health care, including facilitating patient-provider communication and supporting patient-centered care, whereas factors related to failure are often relevant to cost, particularly the connection (or lack thereof) between quality of care provided by such tools and their system-wide cost savings [76]. Elements of patient-provider communication were used in less than one-third of the studies in this review, and only one study had a bidirectional chat feature, allowing for ongoing two-way communication with a provider [58]. In the same way that guided internet-based therapies are more effective and have higher adherence than self-directed internet-based interventions, attention should be paid to the best ways of integrating the expertise and capabilities of clinical providers into mHealth tools [77]. Publications included in this review rarely incorporated information on the cost or cost-effectiveness of the mobile tool in comparison with standard of care. Given that both communication and cost have been noted as key factors that influence sustainability, future work should incorporate these elements into the design of the tool and its evaluation.

Continuity of and Access to Care

Tools in this review address the need for mental health support during either pregnancy or the postpartum period, but these did not offer meaningful opportunities to provide continuity of care across the perinatal period, during which symptoms of depression or anxiety may ebb and flow [78]. Interventions housed in an obstetrics clinic may only be accessible to women currently accessing services, which typically end at 6 weeks postpartum, before the highest-risk period for the onset of symptoms of postpartum depression. Interventions based in clinics that specialize in the treatment of perinatal mental health concerns might be available to women throughout the perinatal period, but these are then limited to those who can be referred to and access these limited services. Most studies included in this review were based in urban, academic settings, which may have led to the passive exclusion of women receiving maternal care in the community (eg, from a midwife) or in rural settings, who may not otherwise have access to specialty perinatal mental health care that is more frequently available in academic centers. Any conclusions made about these tools are therefore potentially limited in their generalizability to urban, academic contexts, which may not reflect the perinatal population in general. Delivery of mental health care services via mobile tools has the potential to remove access barriers for hard-to-reach populations, particularly those outside of urban centers. Future work should address these possibilities by engaging in the recruitment of women with limited access to treatment, in-person or otherwise.

mHealth tools have the potential to address gaps in health care service delivery in low- and high-income economies alike; however, most studies included in this review were conducted in high-income countries. Further investigation is required to understand the nature of mHealth tools being developed and used in lower-middle income economies to address the mental health needs of women in the perinatal period who may have even more limited access to mental health support.

Just over half of the tools (12/22, 54%) included in this review were apps and most were available on both iOS and Android devices. No studies that used a single platform (ie, either iOS

or Android, not both) reported their reasoning for doing so, but this likely reflects the high cost of app development and of the smaller feasibility studies that are typically first required to justify and secure funding for large-scale studies, scaling, and wider implementation. Although mHealth has been widely touted as a way to address barriers to care in perinatal populations, restricting access to single platforms may impact equitable delivery of services. Overall, however, increasing smartphone ownership and availability of wireless internet access in public spaces suggest that mHealth tools have the potential to be delivered without restriction to those who have internet access at home.

Strengths and Limitations

Strengths of this review are its rigorous approach and inclusion of an updated search to capture newly published literature. However, this review did not assess interventions that were available through public app stores and that may not have undergone published scientific evaluation, even though these would also reflect both the current state of mobile tools available for women with perinatal depression or anxiety and the implementation of tools in the public domain. Because only an estimated 6% of the health-based apps have an associated scientific publication, such an endeavor is best suited to its own study [79]. Such a search was beyond the scope of this review, but it could be relevant for future work. This review also excluded tools that were designed to be accessed on stationary devices (ie, computers) but which could, in practice, be accessed on mobile devices. The exclusion of studies in other languages is an additional limitation that biases results to tools likely located in Western, and predominantly English-speaking, contexts.

Conclusions

Health systems worldwide lack the capacity to fulfill demand for appropriate mental health care services: only 1 in 10 women needing mental health treatment in the perinatal period receive it [80]. eHealth, in general, and mHealth, in particular, have the potential to address the significant barriers to care for women with perinatal depression and/or anxiety and mitigate the negative effects of untreated or undertreated mental illness during this time. This review recommends that future work should incorporate the use of active psychological treatment, address the need for continuity of care across the perinatal period, and include factors that affect long-term clinical sustainability. Results of this review fill an important gap by assisting stakeholders to understand the current state of evidence based mHealth tools for perinatal depression and anxiety. These results can be used to make informed decisions when determining how to develop and implement new or existing tools to fill gaps in knowledge. mHealth tools are part of the future of health care delivery and represent an exciting opportunity to evolve the ways in which psychiatric care is delivered, particularly to women during this vulnerable time.

Conflicts of Interest

SV receives royalties from UpToDate Inc for authorship of materials related to depression and pregnancy.

Multimedia Appendix 1
Overview of studies (n=10) related to mHealth tools for the prevention of perinatal depression and/or anxiety.
[DOCX File , 15 KB - jmir_v22i4e17011_app1.docx]

Multimedia Appendix 2
Overview of studies related to mHealth tools for the screening (n=6) or treatment (n=6) of perinatal depression and/or anxiety.
[DOCX File , 16 KB - jmir_v22i4e17011_app2.docx]

References

1. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005 Nov;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
2. Howard LM, Molyneaux E, Dennis C, Rochat T, Stein A, Milgrom J. Non-psychotic mental disorders in the perinatal period. Lancet 2014 Nov 15;384(9956):1775-1788. [doi: 10.1016/S0140-6736(14)61276-9] [Medline: 25455248]
3. Falah-Hassani K, Shiri R, Dennis C. The prevalence of antenatal and postnatal co-morbid anxiety and depression: a meta-analysis. Psychol Med 2017 Sep;47(12):2041-2053. [doi: 10.1017/S0033291717000617] [Medline: 28414017]
4. Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. Lancet 2014 Nov 15;384(9956):1800-1819. [doi: 10.1016/S0140-6736(14)61277-0] [Medline: 25455250]
5. Bauer A, Parsonage M, Knapp M, Iemmi V, Adelaja B. North West Clinical Senates. London: Centre for Mental Health; 2014. Cost of Perinatal Mental Health Problems URL: https://www.nwcscnsenate.nhs.uk/files/3914/7030/1256/Costs_of_perinatal_mh.pdf [accessed 2020-02-12]
6. Chabrol H, Teissedre F, Armitage J, Danel M, Walburg V. Acceptability of psychotherapy and antidepressants for postnatal depression among newly delivered mothers. J Reprod Infant Psychol 2004;22(1):5-12. [doi: 10.1080/02646830310001643094]
7. National Institute for Health and Care Excellence (NICE). London: National Institute for Health and Care Excellence (NICE); 2018. Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance URL: https://www.nice.org.uk/guidance/cg192 [accessed 2019-11-11]

8. MacQueen GM, Frey BN, Ismail Z, Jaworska N, Steiner M, Lieshout RJ, CANMAT Depression Work Group. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 6. Special Populations: Youth, Women, and the Elderly. Can J Psychiatry 2016 Sep;61(9):588-603 [FREE Full text] [doi: 10.1177/0706743716659276] [Medline: 27486149]
9. Fumeaux CJ, Harari MM, Weisskopf E, Eap CB, Epiney M, Vial Y, et al. Risk-benefit balance assessment of SSRI antidepressant use during pregnancy and lactation based on best available evidence - an update. Expert Opin Drug Saf 2019 Oct;18(10):949-963. [doi: 10.1080/14740338.2019.1658740] [Medline: 31430189]
10. Vigod SN, Wilson CA, Howard LM. Depression in pregnancy. Br Med J 2016 Mar 24;352:i1547. [doi: 10.1136/bmj.i1547] [Medline: 27013603]
11. Viveiros CJ, Darling EK. Barriers and facilitators of accessing perinatal mental health services: the perspectives of women receiving continuity of care midwifery. Midwifery 2018 Oct;65:8-15. [doi: 10.1016/j.midw.2018.06.018] [Medline: 30029084]
12. Rai M, Vigod SN, Hensel JM. Barriers to office-based mental health care and interest in e-communication with providers: a survey study. JMIR Ment Health 2016 Aug 1;3(3):e35 [FREE Full text] [doi: 10.2196/mental.6068] [Medline: 27480108]
13. Kopelman RC, Moel J, Mertens C, Stuart S, Arndt S, O'Hara MW. Barriers to care for antenatal depression. Psychiatr Serv 2008 Apr;59(4):429-432 [FREE Full text] [doi: 10.1176/ps.2008.59.4.429] [Medline: 18378843]
14. World Health Organization.: World Health Organization Global Observatory for eHealth URL: https://www.who.int/goe/en/ [accessed 2019-10-08]
15. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
16. Zhang M, Cheow E, Ho CS, Ng BY, Ho R, Cheok CC. Application of low-cost methodologies for mobile phone app development. JMIR Mhealth Uhealth 2014 Dec 9;2(4):e55 [FREE Full text] [doi: 10.2196/mhealth.3549] [Medline: 25491323]
17. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
18. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 5;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
19. Lee Y, Cho S. Technology-supported interventions for pregnant women: a systematic review. Comput Inform Nurs 2019 Oct;37(10):501-512. [doi: 10.1097/CIN.0000000000000535] [Medline: 31162146]
20. Silver L. Pew Research Centre. 2019 Feb 5. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-11-11]
21. Anderson M. Pew Research Centre. 2019 Jun 13. Mobile Technology and Home Broadband 2019 URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2019-11-11]
22. Do TT, Le MD, van Nguyen T, Tran BX, Le HT, Nguyen HD, et al. Receptiveness and preferences of health-related smartphone applications among Vietnamese youth and young adults. BMC Public Health 2018 Jun 19;18(1):764 [FREE Full text] [doi: 10.1186/s12889-018-5641-0] [Medline: 29921258]
23. Urrutia RP, Berger AA, Ivins AA, Beckham AJ, Thorp JM, Nicholson WK. Internet use and access among pregnant women via computer and mobile phone: implications for delivery of perinatal care. JMIR Mhealth Uhealth 2015 Mar 30;3(1):e25 [FREE Full text] [doi: 10.2196/mhealth.3347] [Medline: 25835744]
24. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
25. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
26. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov;39(6):785-804 [FREE Full text] [doi: 10.1177/0145445515595198] [Medline: 26187164]
27. Zhang MW, Ho RC, Loh A, Wing T, Wynne O, Chan SW, et al. Current status of postnatal depression smartphone applications available on application stores: an information quality analysis. BMJ Open 2017 Nov 14;7(11):e015655 [FREE Full text] [doi: 10.1136/bmjopen-2016-015655] [Medline: 29138195]
28. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
29. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]

30. Rathbone AL, 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 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
31. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
32. 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]
33. 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 9;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
34. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
35. Jareethum R, Titapant V, Chantra T, Sommai V, Chuenwattana P, Jirawan C. Satisfaction of healthy pregnant women receiving short message service via mobile phone for prenatal support: A randomized controlled trial. J Med Assoc Thai 2008 Apr;91(4):458-463. [Medline: 18556852]
36. Song H, May A, Vaidhyanathan V, Cramer EM, Owais RW, McRoy S. A two-way text-messaging system answering health questions for low-income pregnant women. Patient Educ Couns 2013 Aug;92(2):182-187. [doi: 10.1016/j.pec.2013.04.016] [Medline: 23711635]
37. Davis AM, Wambach KA, Nelson EL, Odar C, Lillis T, McKinley A, et al. Health behavior change in pregnant women: a two-phase study. Telemed J E Health 2014 Dec;20(12):1165-1169. [doi: 10.1089/tmj.2013.0374] [Medline: 25289706]
38. Dennis-Tiwary TA, Denefrio S, Gelber S. Salutary effects of an attention bias modification mobile application on biobehavioral measures of stress and anxiety during pregnancy. Biol Psychol 2017 Jul;127:148-156 [FREE Full text] [doi: 10.1016/j.biopsycho.2017.05.003] [Medline: 28478138]
39. Dalton JA, Rodger D, Wilmore M, Humphreys S, Skuse A, Roberts CT, et al. The Health-e Babies App for antenatal education: feasibility for socially disadvantaged women. PLoS One 2018;13(5):e0194337 [FREE Full text] [doi: 10.1371/journal.pone.0194337] [Medline: 29768407]
40. Doherty K, Barry M, Marcano-Belisario J, Arnaud B, Morrison C, Car J, et al. A mobile app for the self-report of psychological well-being during pregnancy (BrightSelf): qualitative design study. JMIR Ment Health 2018 Nov 27;5(4):e10007 [FREE Full text] [doi: 10.2196/10007] [Medline: 30482742]
41. Belisario JS, Doherty K, O'Donoghue J, Ramchandani P, Majeed A, Doherty G, et al. A bespoke mobile application for the longitudinal assessment of depression and mood during pregnancy: protocol of a feasibility study. BMJ Open 2017 May 29;7(5):e014469 [FREE Full text] [doi: 10.1136/bmjopen-2016-014469] [Medline: 28554914]
42. Faherty LJ, Hantsoo L, Appleby D, Sammel MD, Bennett IM, Wiebe DJ. Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy. J Am Med Inform Assoc 2017 Jul 1;24(4):746-753 [FREE Full text] [doi: 10.1093/jamia/ocx005] [Medline: 28339686]
43. Hantsoo L, Criniti S, Khan A, Moseley M, Kincler N, Faherty LJ, et al. A mobile application for monitoring and management of depressed mood in a vulnerable pregnant population. Psychiatr Serv 2018 Jan 1;69(1):104-107 [FREE Full text] [doi: 10.1176/appi.ps.201600582] [Medline: 29032705]
44. Jiménez-Serrano S, Tortajada S, García-Gómez JM. A mobile health application to predict postpartum depression based on machine learning. Telemed J E Health 2015 Jul;21(7):567-574. [doi: 10.1089/tmj.2014.0113] [Medline: 25734829]
45. Baumel A, Tinkelman A, Mathur N, Kane JM. Digital peer-support platform (7Cups) as an adjunct treatment for women with postpartum depression: feasibility, acceptability, and preliminary efficacy study. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e38 [FREE Full text] [doi: 10.2196/mhealth.9482] [Medline: 29439944]
46. Teychenne M, van der Pligt P, Abbott G, Brennan L, Olander EK. Feasibility and acceptability of a home-based physical activity program for postnatal women with depressive symptoms: A pilot study. Ment Health Phys Act 2018;14:82-89. [doi: 10.1016/j.mhpa.2018.02.007]
47. McCarter DE, Demidenko E, Hegel MT. Measuring outcomes of digital technology-assisted nursing postpartum: a randomized controlled trial. J Adv Nurs 2018 May 17:- [FREE Full text] [doi: 10.1111/jan.13716] [Medline: 29772609]
48. Niksalehi S, Taghadosi M, Mazhariazad F, Tashk M. The effectiveness of mobile phone text massaging support for mothers with postpartum depression: a clinical before and after study. J Family Med Prim Care 2018;7(5):1058-1062 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_120_17] [Medline: 30598957]
49. Fealy S, Chan S, Wynne O, Dowse E, Ebert L, Ho R, et al. The Support for New Mums Project: a protocol for a pilot randomized controlled trial designed to test a postnatal psychoeducation smartphone application. J Adv Nurs 2019 Jun;75(6):1347-1359. [doi: 10.1111/jan.13971] [Medline: 30740767]
50. Lawson A, Dalfen A, Murphy KE, Milligan N, Lancee W. Use of text messaging for postpartum depression screening and information provision. Psychiatr Serv 2019 May 1;70(5):389-395. [doi: 10.1176/appi.ps.201800269] [Medline: 30717643]
51. Sawyer A, Kaim A, Le H, McDonald D, Mittinty M, Lynch J, et al. The effectiveness of an app-based nurse-moderated program for new mothers with depression and parenting problems (eMums Plus): pragmatic randomized controlled trial. J Med Internet Res 2019 Jun 4;21(6):e13689 [FREE Full text] [doi: 10.2196/13689] [Medline: 31165715]

52. Sun M, Tang S, Chen J, Li Y, Bai W, Plummer V, et al. A study protocol of mobile phone app-based cognitive behaviour training for the prevention of postpartum depression among high-risk mothers. BMC Public Health 2019 Jun 7;19(1):710 [FREE Full text] [doi: 10.1186/s12889-019-6941-8] [Medline: 31174494]
53. Broom MA, Ladley AS, Rhyne EA, Halloran DR. Feasibility and perception of using text messages as an adjunct therapy for low-income, minority mothers with postpartum depression. JMIR Ment Health 2015;2(1):e4 [FREE Full text] [doi: 10.2196/mental.4074] [Medline: 26543910]
54. Rhyne EP, Borawski A. Text messaging as an adjunct treatment for urban mothers with postpartum depression. J Pediatr Healthc 2014;28(6):e49. [doi: 10.1016/j.pedhc.2014.08.003]
55. La Porte LM, Kim JJ, Adams MG, Zagorsky BM, Gibbons R, Silver RK. Feasibility of perinatal mood screening and text messaging on patients' personal smartphones. Arch Womens Ment Health 2019 Jun 15:-. [doi: 10.1007/s00737-019-00981-5] [Medline: 31203440]
56. Silver RK, Kim JJ, La Porte L, Zagorsky B, Desmarais E, Gibbons RD. Mobile phone-based perinatal depression screening. Obstet Gynecol 2017;129:187S. [doi: 10.1097/01.aog.0000514235.57160.0f]
57. Chyzzy B, Dennis C. 16. Mobile phone-based peer support in the prevention of postpartum depression among adolescent mothers: a pilot randomized controlled trial. J Adolesc Health 2019;64:S8-S9. [doi: 10.1016/j.jadohealth.2018.10.030]
58. Ricketts S, Nguyen TN, Narasimhan V. Screening for depression in pregnancy-there's an app for that!. Ann Fam Med 2019 May;17(3):275 [FREE Full text] [doi: 10.1370/afm.2388] [Medline: 31085534]
59. Biggs L, Forster D, Shafiei T, McLachlan H, Smith T, Wyett C, et al. IVY: InVestigating an online community of support for emotional health in pregnancy. Women Birth 2018;31:S37. [doi: 10.1016/j.wombi.2018.08.112]
60. Green EP, Pearson N, Rajasekharan S, Rauws M, Joerin A, Kwobah E, et al. Expanding access to depression treatment in Kenya through automated psychological support: protocol for a single-case experimental design pilot study. JMIR Res Protoc 2019 Apr 29;8(4):e11800 [FREE Full text] [doi: 10.2196/11800] [Medline: 31033448]
61. World Bank Data Help Desk.: The World Bank; 2019. World Bank Country and Lending Groups URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519 [accessed 2019-07-03]
62. Rebello TJ, Marques A, Gureje O, Pike KM. Innovative strategies for closing the mental health treatment gap globally. Curr Opin Psychiatry 2014 Jul;27(4):308-314. [doi: 10.1097/YCO.0000000000000068] [Medline: 24840160]
63. Postpartum Depression: Action Towards CausesTreatment (PACT) Consortium. Heterogeneity of postpartum depression: a latent class analysis. Lancet Psychiatry 2015 Jan;2(1):59-67 [FREE Full text] [doi: 10.1016/S2215-0366(14)00055-8] [Medline: 26359613]
64. Demirci JR, Cohen SM, Parker M, Holmes A, Bogen DL. Access, use, and preferences for technology-based perinatal and breastfeeding support among childbearing women. J Perinat Educ 2016;25(1):29-36 [FREE Full text] [doi: 10.1891/1058-1243.25.1.29] [Medline: 26848248]
65. O'Donnell BE, Lewkowitz AK, Vargas JE, Zlatnik MG. Examining pregnancy-specific smartphone applications: what are patients being told? J Perinatol 2016 Oct;36(10):802-807. [doi: 10.1038/jp.2016.77] [Medline: 27195980]
66. Connor K, Wambach K, Baird MB. Descriptive, qualitative study of women who use mobile health applications to obtain perinatal health information. J Obstet Gynecol Neonatal Nurs 2018 Nov;47(6):728-737. [doi: 10.1016/j.jogn.2018.04.138] [Medline: 29856963]
67. Gratzer D, Khalid-Khan F. Internet-delivered cognitive behavioural therapy in the treatment of psychiatric illness. Can Med Assoc J 2016 Mar 1;188(4):263-272 [FREE Full text] [doi: 10.1503/cmaj.150007] [Medline: 26527829]
68. Simon GE, Ludman EJ. It's time for disruptive innovation in psychotherapy. Lancet 2009 Aug 22;374(9690):594-595. [doi: 10.1016/S0140-6736(09)61415-X] [Medline: 19699995]
69. Osma J, Barrera AZ, Ramphos E. Are pregnant and postpartum women interested in health-related apps? Implications for the prevention of perinatal depression. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):412-415. [doi: 10.1089/cyber.2015.0549] [Medline: 27327069]
70. Sprenger M, Mettler T, Osma J. Health professionals' perspective on the promotion of e-mental health apps in the context of maternal depression. PLoS One 2017;12(7):e0180867 [FREE Full text] [doi: 10.1371/journal.pone.0180867] [Medline: 28704442]
71. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
72. 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]
73. Nicholas J, Boydell K, Christensen H. mHealth in psychiatry: time for methodological change. Evid Based Ment Health 2016 May;19(2):33-34. [doi: 10.1136/eb-2015-102278] [Medline: 27044849]
74. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018 Feb 21;3(1):e000676. [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]
75. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]

76. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
77. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
78. Newham JJ, Martin CR. Measuring fluctuations in maternal well-being and mood across pregnancy. J Reprod Infant Psychol 2013;31(5):531-540. [doi: 10.1080/02646838.2013.834040]
79. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
80. Kingston D, Austin M, Heaman M, McDonald S, Lasiuk G, Sword W, et al. Barriers and facilitators of mental health screening in pregnancy. J Affect Disord 2015 Nov 1;186:350-357. [doi: 10.1016/j.jad.2015.06.029] [Medline: 26281038]

Abbreviations

CBT: cognitive behavioral therapy
eHealth: electronic health
mHealth: mobile health
NICU: neonatal intensive care unit
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by I Mircheva, R Ho, C Fernández; comments to author 29.12.19; revised version received 16.01.20; accepted 24.01.20; published 13.04.20.

Please cite as:
Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E
Mobile Health for Perinatal Depression and Anxiety: Scoping Review
J Med Internet Res 2020;22(4):e17011
URL: http://www.jmir.org/2020/4/e17011/
doi:10.2196/17011
PMID:32281939

Review

Educating Patients by Providing Timely Information Using Smartphone and Tablet Apps: Systematic Review

Thomas Timmers[1,2], MSc; Loes Janssen[3], PhD; Rudolf B Kool[1], MD, PhD; Jan AM Kremer[1], MD, PhD

[1]IQ healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[2]Interactive Studios, Rosmalen, Netherlands

[3]Máxima Medical Center, Veldhoven, Netherlands

Corresponding Author:
Thomas Timmers, MSc
IQ healthcare
Radboud Institute for Health Sciences
Radboud University Medical Center
Geert Grooteplein Zuid 10
Nijmegen, 6525 GA
Netherlands
Phone: 31 73 644 6069
Email: thomas.timmers@radboudumc.nl

Abstract

Background: Patient education is a crucial element within health care. It is a known predictor for increased engagement in shared decision making, improved medication and treatment adherence, higher levels of satisfaction, and even better treatment outcomes. Unfortunately, often patients only remember a very limited amount of medical information. An important reason is that most patients are simply not capable of processing large amounts of new medical information in a short time. Apps for smartphones and tablets have the potential to actively educate patients by providing them with timely information through the use of push notifications.

Objective: The objective of this systematic review is to provide an overview of the effects of using smartphone and tablet apps to educate patients with timely education. Within this review, we focused on patients that receive their care in a hospital setting. We assessed the effects of the interventions on outcomes, such as patients' knowledge about their illness and treatment, adherence to treatment instructions and to medication usage, and satisfaction with the care they received.

Methods: A comprehensive search of MEDLINE (Medical Literature Analysis and Retrieval System Online), Embase, CINAHL (Cumulative Index to Nursing and Allied Health Literature), and Web of Science was conducted. Randomized controlled trials (RCTs) published between January 2015 and November 2019 were eligible for inclusion. Two reviewers independently searched and screened articles, assessed study quality and risk of bias, and extracted the data. Due to the heterogeneity of populations, interventions, and outcomes, a meta-analysis was not deemed appropriate. Instead, a narrative synthesis is presented.

Results: A total of 21 RCTs with 4106 participants were included. Compared to usual care, overall effectiveness of the interventions was demonstrated in 69% of the outcomes. Effectiveness increased to 82% when the intervention had a duration shorter than one month and increased to 78% when the intervention provided at least one push notification per week. The interventions showed the highest effects on satisfaction with information, adherence to treatment instructions and to medication usage, clinical outcomes, and knowledge.

Conclusions: This review demonstrates that educating patients with timely medical information through their smartphones or tablets improves their levels of knowledge, medication or treatment adherence, satisfaction, and clinical outcomes, as well as having a positive effect on health care economics. These effects are most pronounced in interventions with a short duration (ie, less than a month) and with a high frequency of messages to patients (ie, once per week or more). With the knowledge that patient education is a predictor for improved outcomes and the fact that patients have obvious difficulties processing large amounts of new medical information, we suggest incorporating the delivery of timely information through smartphone and tablet apps within current medical practices.

(J Med Internet Res 2020;22(4):e17342) doi:10.2196/17342

KEYWORDS

patient education; push notification; self-management; eHealth; timely information; timely education; smartphone; tablet computer; self-care; mobile phone

Introduction

Patient education is a crucial element within health care. Health care professionals provide patients with information about the origins of complaints, treatment options, prognosis, how to prepare for treatment, or how to manage one's health during the recovery phase. Health care professionals educate their patients because knowledge is a known predictor for increased engagement in shared decision making, improved medication and treatment adherence, higher levels of satisfaction, and better outcomes [1,2].

Unfortunately, patients often only remember a limited amount of the medical information they receive. Many different factors contribute to this. Some of these factors are related to the health care professional, such as using jargon or communicating in a passive way. Other factors are related to the patient, such as age, learning style, and stress [3]. Another important reason is the fact that most patients are simply not capable of processing large amounts of new medical information in a short amount of time [4].

During the last decade, smartphones, tablets, and apps have become commonplace in our society. These innovations offer many new opportunities within health care, such as optimizing the process of patient education. Apps, for example, allow patients to look at medical information as often as they like, at any place, and at any time. The information is comprehensive and different modes of information delivery and interaction are available. Furthermore, push notifications allow health care providers to actively educate patients with timely information, which, in this review, is defined as providing patients with small pieces of information at the time that these are actually relevant to them.

Although interventions like these appear to have much potential in allowing patients to better understand and to remember medical information, an overview of all available evidence on the effectiveness of these technologies has thus far not been published. The objective of this systematic review is to provide an overview of the effectiveness of educating patients by providing timely information using smartphone and tablet apps. With this systematic review, we focused on patients that receive care in a hospital setting rather than in primary care. We have chosen to do so since projects in primary care have already demonstrated effectiveness of electronic health (eHealth) apps, but these primarily focused on chronically ill patients from a population perspective and on telemonitoring services from an intervention perspective.

In this paper, we assess the effects that these interventions have on outcomes, such as patients' knowledge about their illness and treatment options, adherence to medication or instructions, and satisfaction with the information or the care they received.

Methods

Search Strategy and Data Sources

To identify relevant studies, we used a two-step strategy. First, we conducted a preliminary search in PubMed to identify key articles, relevant keywords, and Medical Subject Headings (MeSH) terms. The second step was to have the search strategy be peer reviewed by an information specialist from the Radboud academic medical center's medical library. Multimedia Appendix 1 shows the search strategy for the final search. We comprehensively searched the following databases: MEDLINE (Medical Literature Analysis and Retrieval System Online) (Ovid); Embase (Elsevier); CINAHL (Cumulative Index to Nursing and Allied Health Literature) (EBSCO); and Web of Science. Relevant systematic reviews were also assessed for eligible articles. In order to compare the effectiveness of interventions, we preferred to only include randomized controlled trials (RCTs). Since we were unsure about the number and quality of RCTs, our primary search also included cohort and quasi-experimental studies. After assessing the number and quality of RCTs, we decided to only include these in the review. Reporting was done in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [5].

Based on the results of our preliminary search, we deliberately limited our search to articles published between January 1, 2015, and November 1, 2019, as the interventions described before this period did not meet the eligibility criteria or could no longer be repeated since the technique was outdated or no longer available. We searched for papers in English and looked at reference lists of included studies to optimize our search.

Eligibility Criteria

RCTs were included if they met a number of eligibility criteria: (1) interventions had a focus on patient education through a smartphone or tablet app, used in a hospital setting; (2) interventions had to use push notifications to actively notify patients about newly available information in the app; and (3) the intervention had to be available for multiple days.

We excluded trials that focused solely on the acceptance or feasibility of technology, content or design of the intervention, availability in app stores, telemedicine (ie, remote care), websites or online platforms, or trials that only described the usage of an SMS. Furthermore, articles focusing on data collection, security, behavior or characteristics of patients, and health care professionals were excluded, as were study protocols. Studies were not excluded on the basis of sociodemographic characteristics of patients, such as age, gender, ethnicity, or any other related characteristic.

Data Selection, Extraction, and Management

The search results from different electronic databases were combined within a single Endnote library, version 8.2 (Clarivate Analytics), and duplicates were removed. Two reviewers (TT

and LJ) independently screened titles and abstracts to identify studies that potentially met the inclusion criteria. The full text of these articles was retrieved and read. Two review authors (TT and LJ) independently assessed these articles against the eligibility criteria and extracted the data from the included studies using a structured data extraction form. Disagreements were resolved through discussion and, if necessary, a third reviewer (RBK) was consulted. We extracted information about the patient population, outcomes, interventions, controls, results, and outcome measures.

Assessment of Risk of Bias

Two reviewers (TT and LJ) independently assessed the risk of bias of included RCTs using the Cochrane Collaboration's *risk of bias* tool [6]. Judgements concerning the risk of bias for each study were classified as high, some concerns, or low.

Data Synthesis

Included studies were insufficiently homogenous in terms of patient population, outcomes, and type of intervention. The decision not to perform a meta-analysis was made as a consensus by all authors. For any outcome that was investigated in three or more studies, we present a narrative synthesis of results. In order to compare the effects of the different interventions over the different studies, a standardized mean difference (SMD) is reported, including the 95% CI for the effect. SMD is reported only when results are normally distributed and mean and SD are available. The magnitude of the effect is interpreted according to Cohen's guidelines: small (SMD is 0.2 or lower), medium (SMD is between 0.2 and 0.8), or large (SMD is 0.8 or higher) [7].

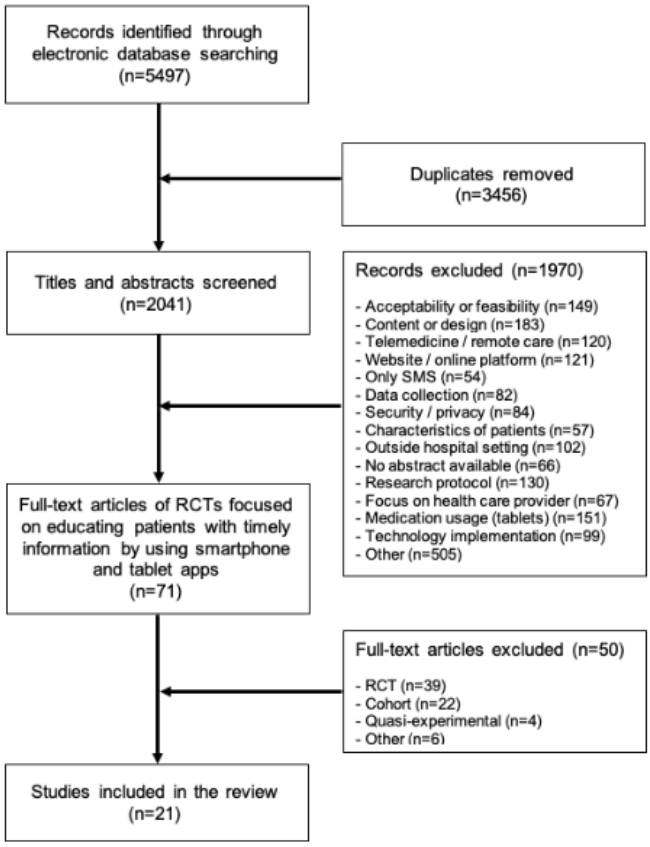
Furthermore, we created a narrative synthesis of overall results per outcome in relation to the duration of the intervention or the frequency with which messages were sent to the patient. Therefore, the duration of the intervention was subdivided into short (<1 month) and long (≥1 month). The frequency of messaging was subdivided into high (>1 message per week) or low (≤1 message per week). The relative effectiveness was calculated by dividing the total number of participants in studies that demonstrated an effect for the outcome by the total number of participants in studies linked to the outcome. Finally, a weighted overall effect was calculated summarizing all outcomes, specified for the duration of the intervention and the frequency of messages.

Results

Overview

Our searches yielded a total of 5497 articles from which 2041 unique articles were derived. After screening titles and abstracts, 1970 records were excluded. A total of 71 articles were assessed for eligibility by full-text screening. A total of 50 articles were excluded after full-text reading because of study type (ie, cohort, quasi-experimental, or other) or because the intervention used did not actually deliver timely education. In total, 21 RCTs were included in the review, including 4106 participants (see Figure 1). Sample sizes ranged from 34 participants [8] to 650 participants [9].

Figure 1. PRISMA (Preferred Reported Items for Systematic Reviews and Meta-Analyses) flowchart. RCT: randomized controlled trial.



Included Studies: Study Designs and Populations

Nine studies were conducted in Europe [10-18], four studies in North America [8,19-21], five studies in Asia [9,22-27], and one study in Africa [28]. In total, 4106 patients participated in the studies. Studies were divided over many different medical departments: gastroenterology [9,18,22,24,28], orthopedics [10,12,13], cardiology [17,20,25,26], oncology [21], surgery [11,19,23], urology [16], internal medicine [27], sports medicine [14], pulmonary disease [8], and neurology [15]. Six studies used a social media platform as the medium for the intervention [9,22-24,26,27]. Eight studies used apps that were already commercially available [10,12-14,16,17,19,20] and five studies used apps that were developed specifically for the study [8,18,21,25,28]. A total of five interventions that were used provided the possibility to interact with a health care provider [9,22,26-28].

Two studies included detailed information about the content and timing of notifications used in the intervention [10,17] and eight provided some details or images [9,12,13,15,21,22,25,28]. Regarding the phase of the treatment in which the study was conducted, seven studies focused on the period before the start of the treatment [9,12,18,21,22,24,28], 12 studies focused on the period after the start of the treatment [8,10,11,13-16,19,20,25-27], and one focused on both [23].

Details of the population, type of intervention, outcomes, and mean age of participants are presented in Table 1. The details of the interventions used, their duration, phase of the treatment, and frequency of notifying patients are presented in Table 2. An overview of all the measurement instruments used per study to assess these outcomes can be found in Multimedia Appendix 2.

Table 1. Details of the publications, interventions, outcomes, and populations.

Study	Year	Country	Department	Population (n)	Age (years), mean	Outcomes
Wang [22]	2019	China	Gastroenterology	Colonoscopy (392)	52	Bowel preparation adherence, quality of preparation, adenoma detection, and satisfaction
Timmers [10]	2019	Netherlands	Orthopedics	Knee replacement (212)	65	Pain, QoL[a], physical functioning, satisfaction, and health care consumption
Mata [19]	2019	Canada	Surgery	Colorectal surgery (97)	60	Recovery protocol adherence, length of stay, complications, and satisfaction
Li [23]	2019	China	Surgery	Pediatric day-care surgery (127)	4[b]	Quality of recovery, satisfaction, and time consumption during follow-up
Jeon [24]	2019	South Korea	Gastroenterology	Colonoscopy (281)	48	Bowel preparation adherence, quality of preparation, and adenoma detection
Van der Meij [11]	2018	Netherlands	Surgery	Abdominal surgery (344)	52	Return to work, first return to normal activity, physical functioning, QoL, and satisfaction
Timmers [12]	2018	Netherlands	Orthopedics	Knee replacement (213)	62	Knowledge, mobile device proficiency, treatment chosen, and satisfaction
Najafi Ghezeljeh [26]	2018	Iran	Cardiology	Hypertension (100)	65	Hypertension self-management
Hardt [13]	2018	Germany	Orthopedics	Knee replacement (60)	65	Range of motion, pain, and physical functioning
Alanzi [27]	2018	Saudi Arabia	Internal medicine	Diabetes mellitus (92)	41[c]	Knowledge and self-efficacy
Widmer [20]	2017	United States	Cardiology	Cardiac rehabilitation (80)	64	In-person hospital visits, clinical values, QoL, and mood
Asklund [16]	2017	Sweden	Urology	Stress urinary incontinence (123)	45	Symptom severity and QoL
Sharara [28]	2017	Lebanon	Gastroenterology	Colonoscopy (160)	53	Bowel preparation adherence, quality of preparation, and satisfaction
Perry [8]	2017	United States	Pulmonary disease	Asthma (34)	15	Asthma control and expiratory volume
Lee [21]	2017	United States	Oncology	Breast cancer (120)	52	Knowledge, readiness for mammography, and satisfaction
Lakshminarayana [15]	2017	United Kingdom	Neurology	Parkinson disease (158)	60	Medication adherence, QoL, quality of consultation, anxiety and depression, and beliefs about medication
Guo [25]	2017	China	Cardiology	Atrial fibrillation (209)	68	Knowledge, QoL, adherence, and satisfaction
Van Reijnen [14]	2017	Netherlands	Sports medicine	Ankle trauma (220)	38	Incidence of ankle sprains, residual pain, and ankle disability
Kang [9]	2016	China	Gastroenterology	Colonoscopy (650)	45	Bowel preparation adherence and compliance with instructions
Johnston [17]	2016	Sweden	Cardiology	Myocardial infarction (174)	57	Medication adherence, satisfaction, and QoL
Lorenzo-Zuniga [18]	2015	Spain	Gastroenterology	Colonoscopy (260)	50	Bowel preparation adherence and satisfaction

[a]QoL: quality of life.

[b]Age of the children who underwent surgery. In the study, their parents (age not mentioned) used the app and provided the data.

[c]Study only reports that 75% of the participants were 41 years or older.

Table 2. Details and duration of the interventions used, frequency of notifying patients, and treatment phase.

Study	Year	Country	Intervention and control	Duration	Notification frequency	Treatment phase[a]
Wang [22]	2019	China	Dietary preparation through the WeChat platform in the days before colonoscopy, as well as timing and usage of the bowel preparation solution; possibility to ask questions as well Control: Standard written information	3 days	Daily	Pre
Timmers [10]	2019	Netherlands	Day-to-day information and videos through an app on pain, wound care, physiotherapy exercises, medication usage, and self-care in the early postoperative phase after total knee replacement Control: Simplified version of the app with only basic information	28 days	Daily	Post
Mata [19]	2019	Canada	Recovery targets and educational information through an app on how to achieve them in the first days after surgery Control: Standard written instructions	2-4 days	Daily	Post
Li [23]	2019	China	Recovery education through the WeChat platform in the days before and after surgery Control: Telephone call by nursing staff	2-4 days	Daily	Pre/post
Jeon [24]	2019	South Korea	Self-management education through the WeChat platform in the days before colonoscopy by using videos Control: Standard written information	3 days	Daily	Pre
Van der Meij [11]	2018	Netherlands	Personalized eHealth[b] program through an app for patients undergoing abdominal surgery Control: Placebo website with standard recovery advice	3 months	Weekly	Post
Timmers [12]	2018	Netherlands	Subdivided and interactive information through an app in the week prior to the consultation with an orthopedic surgeon because of possible knee osteoarthritis Control: Standard information on website	7 days	Daily	Pre
Najafi Ghezeljeh [26]	2018	Iran	Self-management education through the Telegram platform in the weeks after hospitalization Control: Standard written information	6 weeks	Weekly	Post
Hardt [13]	2018	Germany	Postoperatively app-based, feedback-controlled, active muscle training program Control: Standard physiotherapy sessions	4 days	Daily	Post
Alanzi [27]	2018	Saudi Arabia	Diabetes mellitus education through the WhatsApp platform (eg, signs and symptoms, diet, and exercises) Control: Standard written information	8 weeks	Weekly	Post
Widmer [20]	2017	United States	Reporting of dietary and exercise habits through an app, as well as educational information on lifestyle during cardiac rehabilitation Control: Web-based platform	3 months	Occasionally	Post
Asklund [16]	2017	Sweden	Treatment program for pelvic floor muscles and information about stress urinary incontinence and lifestyle through an app Control: Standard written instructions	3 months	Daily	Post
Sharara [28]	2017	Lebanon	Dietary preparation through an app in the days before colonoscopy, as well as timing and usage of the bowel preparation solution Control: Standard written instructions	4 days	Daily	Pre

Study	Year	Country	Intervention and control	Duration	Notification frequency	Treatment phase[a]
Perry [8]	2017	United States	Education on medication usage and peak flow or asthma logging through an app Control: Standard written instructions	6 months	Occasionally	Post
Lee [21]	2017	United States	Personal, tailored multimedia messages through an app to prepare women for breast cancer screening Control: Standard written instructions	7 days	Daily	Pre
Lakshminarayana [15]	2017	United Kingdom	Reminding patients about medication usage, tracking of self-management skills, and educating patients about Parkinson disease through an app Control: Standard written instructions	4 months	Occasionally	Post
Guo [25]	2017	China	Educational program about atrial fibrillation and how to self-manage at home Control: Standard written instructions	3 months	Occasionally	Post
Van Reijnen [14]	2017	Netherlands	Neuromuscular training program through an app for athletes who suffered a sprained ankle Control: Standard written instructions	2 months	Occasionally	Post
Kang [9]	2016	China	Dietary preparation through the WeChat platform in the days before colonoscopy, as well as timing and usage of the bowel preparation solution; possibility to ask questions as well Control: Standard written instructions	4 days	Daily	Pre
Johnston [17]	2016	Sweden	Educational messages based on the data patients had registered about their medication usage Control: Simplified version of the app with only basic information	6 months	Weekly	Post
Lorenzo-Zuniga [18]	2015	Spain	Dietary preparation through an app in the days before colonoscopy, as well as timing and usage of the bowel preparation solution Control: Standard written instructions	4 days	Daily	Pre

[a]Pre: before the start of the treatment; post: after the start of the treatment.

[b]eHealth: electronic health.

Risk of Bias of Included Studies

All 21 included studies were assessed for risk of bias in the following domains: selection of the reported result, measurement of the outcome, missing outcome data, deviations from intended interventions, and randomization process. The levels of risk—low, some concerns, or high—per study, per domain are presented in Figure 2. An overview of the percentage of studies related to the level of risk and domain of bias is presented in Figure 3.

Figure 2. Level of risk of bias, per study, per domain.

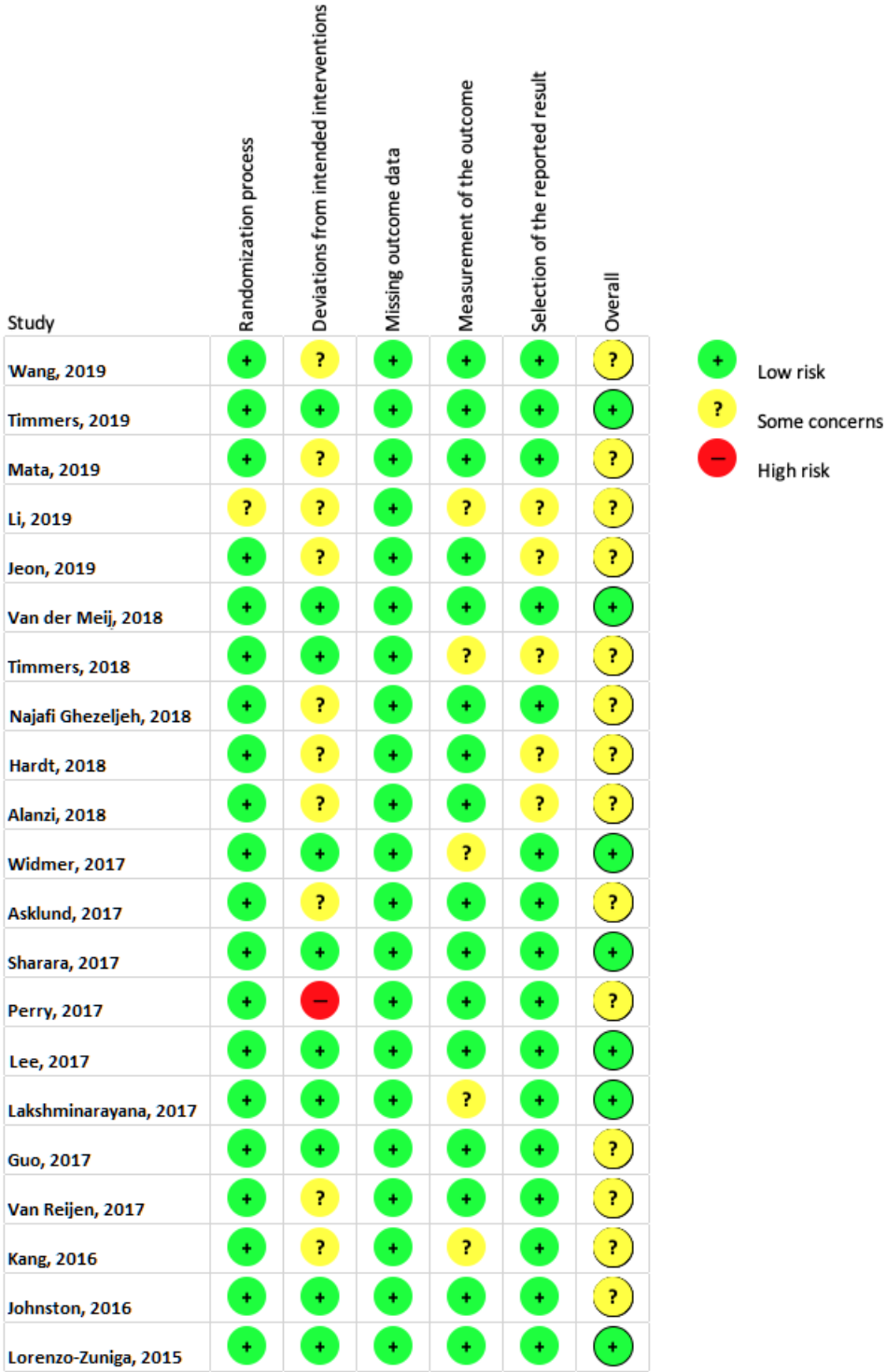
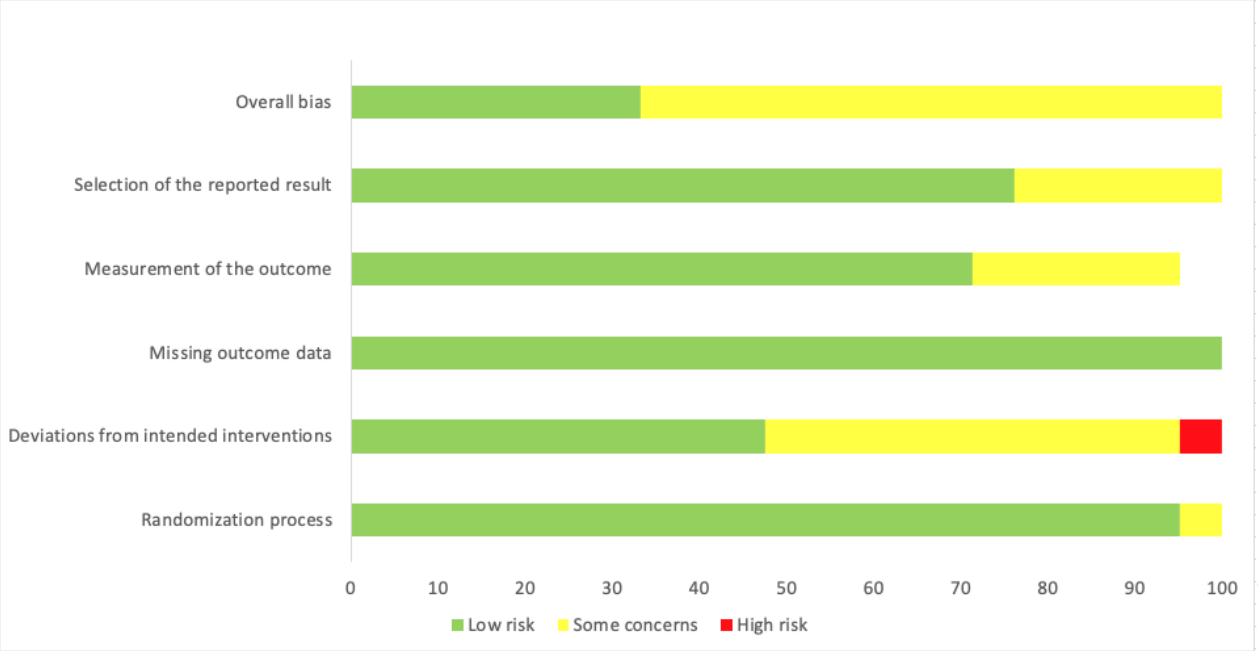


Figure 3. Risk of bias presented as the percentage of studies with low risk, some concerns, or high risk, scored for the different domains of bias.



Outcomes

Overview

Characteristics of the included studies are presented per outcome. Per study, the effect of the intervention on the outcome is described as *in favor of the intervention group*, *in favor of the control group*, or *no effect*.

Satisfaction

A total of 12 RCTs [10-12,15,17-19,21,22,25,28], in which 2466 patients participated, reported results related to satisfaction. Two main themes emerged from these studies: satisfaction with the information provided [10-12,17,19,21,28] and satisfaction with the overall care that was delivered [10-12,18,22,25] (see Table 3).

Regarding patients' satisfaction with the information, an effect in favor of the intervention group was demonstrated in eight out of 10 studies. Interventions included an app that was used to educate patients about the preparation for their colonoscopy [28], consultation with an orthopedic surgeon [12], postoperative self-management after knee replacement surgery [10], breast cancer screening [21], healthy lifestyle interventions in myocardial infarction patients [17], and return to normal activities after abdominal surgery [11]. One study, which focused on enhanced recovery education after colorectal surgery [19], showed no difference in terms of satisfaction between the intervention and control groups. SMD ranged from medium to large in five studies [10-12,17,21] and could not be calculated for the two other studies.

Regarding patients' satisfaction with the overall care they received, an effect in favor of the intervention group was demonstrated in four out of eight studies. These studies measured the patient-perceived level of involvement by the hospital after discharge [10], satisfactory bowel preparation [22], satisfaction with anticoagulation therapy [25], level of patient-centered care in Parkinson disease [15], and overall experience with the bowel preparation process [18]. Three other studies showed no differences between groups in patients' overall satisfaction with care related to abdominal surgery [11], patients' satisfaction related to the consultation with their orthopedic surgeon [12], and patients' overall satisfaction with the recovery process after pediatric surgery [23]. SMD ranged from small to large in six studies [10-12,15,18,22,25] and could not be calculated for the other study.

Table 3. Details about patients' satisfaction.

Satisfaction type and study	Population (n)	Description[a]	Effect[b]	SMD[c] (95% CI)
Satisfaction with information provided				
Van der Meij [11]	Abdominal surgery (344)	Personalized information on activity resumption	+	0.43 (0.22 to 0.65)
Lee [21]	Breast cancer (120)	Breast cancer screening instructions	+	0.55 (0.19 to 0.90)
Sharara [28]	Colonoscopy (160)	Bowel preparation	+	SMD could not be calculated[d]
Mata [19]	Colorectal surgery (97)	Postoperative adherence protocol	=	SMD could not be calculated[e]
Timmers [10]	Knee replacement (212)	Education on pain management, exercises, and self-care	+	0.97 (0.68 to 1.27)
Timmers [12]	Knee replacement (213)	Level of knowledge about treatment options	+	0.54 (0.26 to 0.82)
Timmers [12]	Knee replacement (213)	Preparation for medical consultation	+	0.70 (0.42 to 0.98)
Johnston [17]	Myocardial infarction (174)	Overall satisfaction with the app	+	0.56 (0.23 to 0.88)
Satisfaction with care received				
Van der Meij [11]	Abdominal surgery (344)	Overall satisfaction with care received	=	0.20 (–0.01 to 0.41)
Guo [25]	Atrial fibrillation (209)	Overall satisfaction with care received	+	0.58 (0.15 to 1.00)
Wang [22]	Colonoscopy (392)	Colonoscopy treatment itself	+	SMD could not be calculated[f]
Lorenzo-Zuniga [18]	Colonoscopy (260)	Overall satisfaction with care received	+	0.78 (0.52 to 1.04)
Timmers [10]	Knee replacement (212)	Hospital involvement during recovery	+	0.89 (0.60 to 1.19)
Timmers [12]	Knee replacement (213)	Medical consultation with orthopedic surgeon	=	0.29 (–0.02 to 0.58)
Lakshminarayana [15]	Parkinson disease (158)	Overall satisfaction with care received (Patient-Centered Outcomes Questionnaire for Parkinson's Disease)	+	0.35 (0.03 to 0.67)
Li [23]	Surgery (127)	Overall quality of recovery	=	0.20 (–0.15 to 0.55)

[a]All items were patient reported versus clinician reported.

[b]Effects were in favor of the intervention group (+) or there were no effects (=). No study had effects in favor of the control group (–).

[c]SMD: standardized mean difference.

[d]Outcome only measured in intervention group.

[e]No SD available (only average and *P* value).

[f]Nonnormal distributed data.

Adherence

A total of 11 RCTs [9,15,17-19,22,24-28], in which 2573 patients participated, reported results related to adherence. Two main themes emerged from these studies: adherence to treatment instructions [9,18,19,22,24,26-28] and adherence to medication usage [15,17,25] (see Table 4).

Regarding patients' adherence to treatment instructions, an effect in favor of the intervention group was demonstrated in five out of eight studies, focusing on patients' self-management in diabetes mellitus [27], hypertension [26], and adherence to purgative and dietary instructions for bowel preparation before their colonoscopy [9,22,24]. No differences between groups were reported in two other studies focusing on preparation for colonoscopy [18,28] and a postoperative recovery program after colorectal surgery [19]. SMD ranged from small to large in six studies [9,18,24,26-28], was negative in one study [19], and could not be calculated for the other study.

Regarding patients' adherence to their medication usage, an effect in favor of the intervention group was demonstrated in all three studies addressing this theme. These studies focused on drug adherence in Parkinson disease [15], anticoagulation adherence in patients who suffered from atrial fibrillation [25], or myocardial infarction [17]. With regard to the latter, patients in the intervention group reported lower missed medication doses. However, the same study also reported that there were no differences between groups in results related to the medication adherence questionnaire that was assessed. SMD ranged from small to medium in two studies [15,17] and could not be calculated for the other study.

Table 4. Details about patients' adherence.

Adherence type and study	Population (n)	Description[a]	Effect[b]	SMD[c] (95% CI)
Adherence to instructions				
Wang [22]	Colonoscopy (392)	Purgative and dietary instructions for bowel preparation (CR)	+	SMD could not be calculated[d]
Jeon [24]	Colonoscopy (281)	Purgative and dietary instructions for bowel preparation (PR)	+	SMD could not be calculated[e]
Jeon [24]	Colonoscopy (281)	Clinical Bowel Preparation score (CR)	+	0.28 (0.05 to 0.52)
Sharara [28]	Colonoscopy (160)	Purgative and dietary instructions for bowel preparation (PR)	=	SMD could not be calculated[d]
Sharara [28]	Colonoscopy (160)	Clinical Bowel Preparation score (CR)	=	0.12 (–0.19 to 0.43)
Kang [9]	Colonoscopy (650)	Purgative and dietary instructions for bowel preparation (CR)	+	0.51 (0.37 to 0.66)
Lorenzo-Zuniga [18]	Colonoscopy (260)	Purgative and dietary instructions for bowel preparation (CR)	=	0.16 (–0.08 to 0.42)
Mata [19]	Colorectal surgery (97)	Postoperative recovery elements (eg, mobilization) (PR)	=	–0.13 (–0.52 to 0.26)
Alanzi [27]	Diabetes mellitus (92)	Self-efficacy in diabetes mellitus	+	0.78 (0.36 to 1.21)
Najafi Ghezeljeh [26]	Hypertension (100)	Hypertension self-management (PR)	+	6.78 (5.34 to 8.21)
Adherence to medication				
Lakshminarayana [15]	Parkinson disease (158)	Parkinson disease drug adherence (PR)	+	0.37 (0.05 to 0.68)
Guo [25]	Atrial fibrillation (209)	Anticoagulation drug adherence (PR)	+	SMD could not be calculated[e]
Johnston [17]	Myocardial infarction (174)	Anticoagulation drug adherence (PR)	=	SMD could not be calculated[d]
Johnston [17]	Myocardial infarction (174)	Missed medication doses (PR)	+	0.14 (–0.16 to 0.46)

[a]Items were either clinician reported (CR) or patient reported (PR).

[b]Effects were in favor of the intervention group (+) or there were no effects (=). No study had effects in favor of the control group (–).

[c]SMD: standardized mean difference.

[d]No SD available (only average and *P* value).

[e]Nonnormal distributed data.

Quality of Life

Seven RCTs [10,11,15-17,20,25], in which 1300 patients participated, reported results related to quality of life (see Table 5). An effect in favor of the intervention group was demonstrated in four studies. These studies measured the effect of the intervention on quality of life at four weeks after knee replacement surgery [10], three months after starting a program for cardiac rehabilitation [20], three months after starting a program for pelvic floor muscle training [16], and three months after starting a program for enhanced self-management after atrial fibrillation [25]. Three studies did not report an effect in the intervention group at the following time points: 6 months after intermediate-grade abdominal surgery [11], 4 months after starting a self-management program in Parkinson disease [15], and 6 weeks after starting a support program on lifestyle changes and drug adherence in myocardial infarction patients [17]. SMD ranged from small to large in five studies [10,15,20] and could not be calculated for two studies [11,25].

Table 5. Details about patients' quality of life.

Study	Population (n)	Description[a]	Effect[b]	SMD[c] (95% CI)
Van der Meij [11]	Abdominal surgery (344)	After abdominal surgery	=	SMD could not be calculated[d]
Guo [25]	Atrial fibrillation (209)	After starting atrial fibrillation management program	+	SMD could not be calculated[d]
Widmer [20]	Cardiac rehabilitation (80)	After starting cardiac rehabilitation	+	3.30 (2.60 to 4.02)
Timmers [10]	Knee replacement (212)	After knee replacement surgery	+	0.44 (0.15 to 0.72)
Johnston [17]	Myocardial infarction (174)	After starting lifestyle and drug adherence support	=	0.33 (0.01 to 0.66)
Lakshminarayana [15]	Parkinson disease (158)	After starting self-management app	=	0.18 (–0.14 to 0.49)
Asklund [16]	Stress urinary incontinence (123)	After starting pelvic floor muscle training	+	0.81 (0.44 to 1.18)

[a]All items were patient reported versus clinician reported.

[b]Effects were in favor of the intervention group (+) or there were no effects (=). No study had effects in favor of the control group (–).

[c]SMD: standardized mean difference.

[d]No SD available (only average and *P* value).

Clinical Outcomes

A total of 11 RCTs [8,10,11,13-16,20,22,24,28], in which 1783 patients participated, reported results related to clinical outcomes. Three main themes emerged from these studies: physical functioning and pain [10,11,13,14], clinical values [20,22,24,28], and symptoms [8,15,16] (see Table 6).

Regarding physical functioning, an effect in favor of the intervention group was demonstrated in three out of four studies, albeit not on all outcomes. These results were related to physical functioning after abdominal surgery [11] and pain and knee function after knee replacement surgery [10,13]. No differences between groups were reported concerning pain and activities after abdominal surgery [11] or concerning knee function and physiotherapy assessment tests [13]. One study related to ankle function after sports-related trauma did not demonstrate a difference between groups either [14]. SMD was medium in one study [10] and could not be calculated for the other studies.

Regarding clinical values, an effect in favor of the intervention group was demonstrated in at least one of the outcomes of all four included studies. These effects were related to weight loss during cardiac rehabilitation [20] and adenoma detection during colonoscopy [22,24,28]. No differences between groups were found concerning cholesterol, glucose, and exercise capacity in cardiac rehabilitation [20]. SMD ranged from small to large in two studies [15,16] and could not be calculated for the other study.

Regarding symptoms, an effect in favor of the intervention group was demonstrated in one out of three studies. These results were related to a decrease in symptom severity after using an intervention to train pelvic floor muscles in women who suffer from stress-related urinary incontinence [16]. No differences between groups were reported in nonmotor symptoms related to Parkinson disease [15] and asthma [8]. SMD ranged from small to large within one study [20] and could not be calculated for the other studies.

Table 6. Details about clinical parameters.

Clinical parameters and study, population, and description[a]	Effect[b]	SMD[c] (95% CI)
Physical functioning and pain		
Van der Meij [11]		
Abdominal surgery (n=344)		
Physical function (PR)	+	SMD could not be calculated[d]
Physical activities (PR)	=	SMD could not be calculated[d]
Recovery (PR)	=	SMD could not be calculated[d]
Pain intensity (PR)	=	SMD could not be calculated[d]
Van Reijnen [14]		
Ankle trauma (n=220)		
Ankle function (PR)	=	SMD could not be calculated[e]
Hardt [13]		
Knee replacement (60)		
Knee range of motion (CR)	+	SMD could not be calculated[e]
Pain at rest (PR)	=	SMD could not be calculated[e]
Pain in motion (PR)	+	SMD could not be calculated[e]
Knee function (PR)	=	SMD could not be calculated[e]
Assessment tests (CR)	=	SMD could not be calculated[e]
Timmers [10]		
Knee replacement (n=212)		
Pain at rest (PR)	+	0.51 (0.23 to 0.79)
Pain during activity (PR)	+	0.49 (0.21 to 0.77)
Pain during the night (PR)	+	0.42 (0.14 to 0.71)
Knee function (PR)	+	0.47 (0.19 to 0.76)
Clinical values		
Widmer [20]		
Cardiac rehabilitation (n=80)		
Weight (CR)	+	0.80 (0.32 to 1.28)
Cholesterol (CR)	=	0.49 (–0.07 to 0.87)
Glucose (CR)	=	0.05 (–0.41 to 0.52)
Rehabilitation session attended (CR)	=	0.28 (–0.19 to 0.74)
Exercise capacity (VO_2 peak) (CR)	=	0.22 (–0.24 to 0.69)
Wang [22]		
Colonoscopy (n=392)		
Adenoma detection rate (1 adenoma detected) (CR)	=	SMD could not be calculated[d]
Adenoma detection rate (>1 adenoma detected) (CR)	+	SMD could not be calculated[d]
Jeon [24]		
Colonoscopy (n=281)		
Adenoma detection rate (overall) (CR)	+	SMD could not be calculated[e]
Sharara [28]		

Clinical parameters and study, population, and description[a]	Effect[b]	SMD[c] (95% CI)
Colonoscopy (n=160)		
Adenoma detection rate (overall) (CR)	+	SMD could not be calculated[d]
Symptoms		
Perry [8]		
Asthma (n=34)		
Asthma control rest (PR)	=	SMD could not be calculated[e]
Lakshminarayana [15]		
Parkinson disease (n=158)		
Range of nonmotor symptoms (PR)	=	0.16 (–0.16 to 0.48)
Asklund [16]		
Stress urinary incontinence (n=123)		
Symptom severity (PR)	+	0.95 (0.58 to 1.33)

[a]Items were either patient reported (PR) or clinician reported (CR).

[b]Effects were in favor of the intervention group (+) or there were no effects (=). No study had effects in favor of the control group (–).

[c]SMD: standardized mean difference.

[d]No SD available (only average and *P* value).

[e]Nonnormal distributed data.

Health Care Economics

Five RCTs [10,11,19,23], in which 860 patients participated, reported results related to health care economics (see Table 7). An effect in favor of the intervention group was demonstrated in three studies, concerning patients' contact with health care providers after total knee replacement surgery [10] and after pediatric day-care surgery [23], as well as after returning to work after abdominal surgery [11]. The other studies did not report an effect in favor of the intervention group for patients undergoing colorectal or abdominal surgery [11,19] or patients attending a cardiac rehabilitation program [20]. Regarding 30-day hospital readmissions, an effect in favor of the control group was demonstrated after colorectal surgery [19]. SMD ranged from small to large in two studies [19,23] and could not be calculated for the other studies.

Table 7. Details of health care economics of studies.

Study, Population (n), Description[a]	Effect[b]	SMD[c] (95% CI)
Van der Meij [11]		
Abdominal surgery (344)		
Postoperative complications (CR)	=	SMD could not be calculated[d]
Mean cost differences (CR)	=	SMD could not be calculated[d]
Return to work (PR)	+	SMD could not be calculated[d]
Return to 75% of normal activities (PR)	=	SMD could not be calculated[d]
Widmer [20]		
Cardiac rehabilitation (80)		
Emergency department visits (CR)	=	SMD could not be calculated[d]
Rehospitalization (CR)	=	SMD could not be calculated[d]
Emergency department visits plus rehospitalization (CR)	=	SMD could not be calculated[d]
Mata [19]		
Colorectal surgery (97)		
Length of stay (CR)	=	0.19 (–0.21 to 0.59)
Postoperative complications (CR)	=	SMD could not be calculated[d]
30-day reoperation (CR)	=	SMD could not be calculated[d]
30-day emergency department visits (CR)	=	SMD could not be calculated[d]
30-day hospital readmissions (CR)	–	SMD could not be calculated[d]
Timmers [10]		
Knee replacement (212)		
Contact with hospital, general practitioner, or home care organization during the 4 weeks after discharge (PR)	+	SMD could not be calculated[d]
Li [23]		
Surgery (127)		
Time consumed during follow-up (CR)	+	3.58 (3.02 to 4.14)

[a]Items were either clinician reported (CR) or patient reported (PR).

[b]Effects were in favor of the intervention group (+), in favor of the control group (–), or there were no effects (=).

[c]SMD: standardized mean difference.

[d]Nonnormal distributed data.

Knowledge

Four RCTs [10,21,25,27], in which 634 patients participated, reported results related to condition- or treatment-specific knowledge acquisition (see Table 8). An effect in favor of the intervention group was demonstrated in all four studies. All studies focused on disseminating disease-specific information, ranging from treatment options for patients with knee complaints due to osteoarthritis [12] to self-management in atrial fibrillation patients [25] or diabetes mellitus [27] and general knowledge about breast cancer and screening options [21]. SMD ranged from medium to large in three studies [12,21,27] and could not be calculated for one study.

Table 8. Details about disease-specific knowledge acquisition.

Study	Population (n)	Description[a]	Effect[b]	SMD[c] (95% CI)
Guo [25]	Atrial fibrillation (209)	Knowledge about atrial fibrillation	+	SMD could not be calculated[d]
Lee [21]	Breast cancer (120)	Knowledge about breast cancer and screening options	+	0.32 (–0.04 to 0.68)
Alanzi [27]	Diabetes mellitus (92)	Knowledge about diabetes mellitus and lifestyle	+	4.65 (3.87 to 5.44)
Timmers [12]	Knee replacement (213)	Actual knowledge about treatment options	+	1.27 (0.95 to 1.60)
Timmers [12]	Knee replacement (213)	Perceived knowledge about treatment options	+	0.87 (0.56 to 1.18)

[a]All items were patient reported versus clinician reported.

[b]Effects were in favor of the intervention group (+) for all studies, versus effects in favor of the control group (–) or no effects (=).

[c]SMD: standardized mean difference.

[d]Nonnormal distributed data.

Narrative Synthesis of Overall Results

Overall results demonstrate an average effectiveness of the intervention of 69% (see Table 9). Satisfaction with information, adherence to instructions and medication, clinical outcomes (eg, weight loss or adenoma detection), and knowledge acquisition showed the highest effects (>70%). When taking into account the duration of the intervention, a clear advantage in terms of effect is demonstrated by the interventions that have a duration of less than one month, compared to the interventions that take more than one month: 82% effectiveness versus 69%. A clear difference is noted in the comparison between the frequencies of messaging patients with information as well: an average effectiveness of 78% in the high-frequency group (more than once per week, on average) versus 64% in the low-frequency group (once per week, on average).

Table 9. Synthesis of results: average effectiveness per outcome.

Outcome	Dimension	Number of studies/ population members	Average effectiveness[a], %	Duration, %		Frequency, %	
				Short (<1 month)	Long (≥1 month)	High[b]	Low[c]
Satisfaction	Information provided	7/1320	93	88	100	88	100
Satisfaction	Overall care	8/1915	64	72	52	72	52
Adherence	Instructions	8/2032	75	72	100	75	N/A[d]
Adherence	Medication usage	3/541	84	N/A	84	50	100
Quality of life	Overall	7/1300	48	100	38	66	57
Clinical parameters	Physical functioning and pain	4/836	50	89	30	89	30
Clinical parameters	Clinical values	4/913	74	76	50	76	50
Clinical parameters	Symptoms	3/315	39	N/A	39	100	0
Health care economics	Overall	5/860	59	78	68	78	68
Knowledge	Overall	4/634	100	100	100	100	100
Average effect	N/A	N/A	69	82	69	78	64

[a]Average effectiveness is the weighted average of the population linked to an outcome and the part of the population with a positive effect on the outcome.

[b]High frequency is >1 message per week, on average.

[c]Low frequency is ≤1 message per week, on average.

[d]N/A: not applicable.

Discussion

Principal Findings

The objective of this systematic review was to evaluate the effectiveness of educating patients by providing timely information using smartphone and tablet apps. In particular, we focused on patients that had undergone treatment in a hospital. A total of 21 studies were identified, most with some concerns in terms of risk of bias. Included studies showed low levels of homogeneity in terms of populations and outcomes. Overall results demonstrate an average effectiveness of the interventions in 69% of the studies. Satisfaction with information, adherence to instructions and medication, improved clinical values (eg, weight loss or adenoma detection), and knowledge acquisition

showed the highest effects (>70%). An overall effect of 82% was observed in studies that lasted less than one month. Studies with a higher frequency of messaging (ie, more than once per week) were associated with an average effect of 78%. These results should not only be considered effective from a single outcome point-of-view, but should be, from a more holistic perspective, considered as important components required for effective patient self-management support as well [29].

Our results are in line with earlier reviews that focus on the effect of eHealth interventions on multiple outcomes in chronic health conditions [30,31]. A review by Schoeppe et al reported a positive effect in terms of prevention by focusing on lifestyle changes, such as diet, exercise, and sedentary behavior [32]. The average duration of the interventions in the Schoeppe et al review was 8 weeks, which is longer than the average duration of interventions in our review. However, this is probably due to the fact that the interventions in the Schoeppe et al review focused on behavioral changes related to lifestyle, whereas studies in our review sometimes lasted only 3 or 4 days, in which the aim is not to change one's lifestyle, but to optimize one's preparation for a one-time event such as a colonoscopy. The usage of frequent notifications has been recognized to encourage greater exposure to the intervention's content without deterring engagement [33].

Even though results seem to indicate that interventions of a short duration with a high frequency of notifications are beneficial to the patient, the low level of homogeneity across these studies makes it impossible to extract an optimal structure, duration, or frequency for messaging patients. Such a challenge has also been reported in a 2018 review on education via strategies and structures [34]. Unfortunately, only a few studies reported detailed information about the content that was provided to patients, its format (eg, text, photo, or video), and the actual timing of the content delivery. This information could have provided additional insights on what makes interventions successful or not.

Our results demonstrate the emerging character of this field of research: the 21 included studies were conducted in 10 different medical departments, covering 15 different types of treatments. Four medical specialties—cardiology, orthopedics, surgery, and gastroenterology—have had more than three studies included. Only interventions related to colonoscopy and knee replacement were studied more than once. The results regarding the number of studies that we excluded from this review also demonstrate that many studies still focus on feasibility, acceptance of technology, and the design and content of apps, rather than on the actual effect of this type of intervention.

Strengths and Limitations

To our knowledge, this review is the first to assess the effectiveness of educating patients in preparation of, during, or after their treatment in the hospital using an app for smartphones or tablets. This review adopted a detailed and comprehensive search strategy, followed by robust screening, data extraction, and risk-of-bias assessment, adhering to the PRISMA guidelines. A total of 21 studies were found eligible for inclusion, seven of them having a *low risk* level of bias and 14 of them having a level of bias with *some concerns* according to Cochrane's *risk of bias* assessment. The relatively large sample sizes allowed us to calculate SMDs and therewith enabled us to compare study outcomes. The observed high level of heterogeneity in terms of outcomes, population, and intervention characteristics, such as interaction models, commercial and noncommercial products, or social media platforms, made it inappropriate to perform a meta-analysis for any outcome.

In this review, we focused on the timely delivery of educational information to overcome patient-perceived information overload. The duration of the interventions within these studies ranged from 3 days to 6 months. In our opinion, this range is another indicator that this type of research is still at an early stage, in which the focus of the trial is really on the intervention itself instead of its long-term effects.

Implications for Practice

The results of our review demonstrate the effective application of smartphone and tablet apps to educate patients with timely information. The effects are visible within various outcomes and across various medical specialties. Medical practices could benefit from these effects by combining two already-existing resources: patient education materials and smartphones and tablets. Patient education is already available on hospital websites, brochures, and through the oral advice of health care professionals. Additionally, more and more patients, as well as their surrounding caretakers, possess a smartphone or tablet. By adding the concept of *timing* to existing educational materials, one could improve the likelihood that patients receive the right information at the right time. By using the push notification mechanism on smart devices, patients can also be actively made aware of newly available information related to their treatment. Medical practices may choose to either build an app themselves or use already-available commercial products or platforms, social media or otherwise. After the initial development of an app, little or no further adjustments to existing workflows are needed for successful implementation, which is regarded as a crucial factor for successful eHealth implementation [35]. Of course, some patients may require support during the initial downloading or configuring of the app, but when this effort is compared to the possible benefits in terms of improved outcomes, satisfaction, and health care consumption as described in this review, these efforts appear worthwhile.

Future Research

Delivering timely education to patients through an app for smartphones or tablets has the potential to contribute to the emerging field of patient education research, which may lead to a positive effect on numerous outcomes. Given the novelty of this area of research, more studies need to be performed in order to demonstrate the generalizability of the concept, as well as its long-term effects. In this review, we chose to include only RCTs, since this study design is currently considered to be the gold standard research design to assess the effectiveness of interventions. Yet, we believe it is legitimate to question whether this is the only appropriate study design, as eHealth innovations and research projects could be characterized by what we would like to refer to as "moving objects" and "moving targets." By moving objects, we refer to the interventions themselves, as

these may easily be adapted to the real-time needs of patients and health care providers by their inventors. By moving targets, we refer to outcomes that might not have been defined in the original research protocol but arose from the data and insights that were gathered during the study. Changing the intervention itself or adding outcomes during the course of a study is, however, often considered *not done*, as it could quickly lead to a high risk of bias and a lower overall quality of the research. As a consequence, many interventions might not be studied at all, because from a supplier's or producer's perspective, it feels unnatural not to be able to respond to these real-world demands "just because a study design won't allow you to." This challenge was also reported by two recent studies focusing on eHealth interventions in general [36] and, more specifically, in the field of psychiatry [37].

We suggest that other study designs, such as pragmatic RCTs, action research, or even real-world data, are considered to be eligible to demonstrate the effectiveness of these interventions. These designs more closely mimic a routine clinical setting from a health care provider's perspective (ie, no double blinding or placebo-controlled setting) and allow the interventions to be altered by the supplier during the course of the study if needed. This could lower some of the existing barriers and may convince more stakeholders to participate in eHealth research.

Conclusions

This review demonstrates that educating patients with timely medical information through their smartphones or tablets improves their levels of knowledge, medication or treatment adherence, satisfaction, and clinical outcomes, as well as having a positive effect on health care economics. These effects are most pronounced in interventions with a short duration (ie, less than a month) and with a high frequency of messaging patients (ie, once per week or more). With the knowledge that patient education is a predictor for improved outcomes and the fact that patients have obvious difficulties processing large amounts of new medical information, we suggest incorporating the delivery of timely information through smartphone and tablet apps within current medical practices.

Acknowledgments

The authors would like to thank Elmie Peters, medical librarian at Radboud University Medical Center (Radboudumc), for her input in shaping and optimizing the search strategy. Furthermore, the authors would like to thank Samuel Credgington for language editing.

Authors' Contributions

All authors were involved in the design of the protocol, development of the search strategy, and selection of data sources. TT and LJ independently screened articles and assessed the risk of bias of the included studies. TT drafted the manuscript. LJ, RBK, and JAMK critically revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE (Medical Literature Analysis and Retrieval System Online) search strategy.
[PDF File (Adobe PDF File), 43 KB - jmir_v22i4e17342_app1.pdf]

Multimedia Appendix 2
Overview of outcomes per study and instruments used to assess them.
[PDF File (Adobe PDF File), 122 KB - jmir_v22i4e17342_app2.pdf]

References

1. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
2. Dineen-Griffin S, Garcia-Cardenas V, Williams K, Benrimoj SI. Helping patients help themselves: A systematic review of self-management support strategies in primary health care practice. PLoS One 2019;14(8):e0220116 [FREE Full text] [doi: 10.1371/journal.pone.0220116] [Medline: 31369582]
3. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [doi: 10.1258/jrsm.96.5.219] [Medline: 12724430]
4. Khaleel I, Wimmer BC, Peterson GM, Zaidi ST, Roehrer E, Cummings E, et al. Health information overload among health consumers: A scoping review. Patient Educ Couns 2020 Jan;103(1):15-32. [doi: 10.1016/j.pec.2019.08.008] [Medline: 31451363]

5. 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]
6. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: A revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
7. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. New York, NY: Taylor & Francis Inc; 1988.
8. Perry TT, Marshall A, Berlinski A, Rettiganti M, Brown RH, Randle SM, et al. Smartphone-based vs paper-based asthma action plans for adolescents. Ann Allergy Asthma Immunol 2017 Mar;118(3):298-303. [doi: 10.1016/j.anai.2016.11.028] [Medline: 28111110]
9. Kang X, Zhao L, Leung F, Luo H, Wang L, Wu J, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016 Mar;14(3):429-435.e3. [doi: 10.1016/j.cgh.2015.09.038] [Medline: 26492848]
10. Timmers T, Janssen L, van der Weegen W, Das D, Marijnissen W, Hannink G, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: Randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323 [FREE Full text] [doi: 10.2196/15323] [Medline: 31638594]
11. van der Meij E, Anema JR, Leclercq WK, Bongers MY, Consten EC, Schraffordt Koops SE, et al. Personalised perioperative care by e-health after intermediate-grade abdominal surgery: A multicentre, single-blind, randomised, placebo-controlled trial. Lancet 2018 Jul 07;392(10141):51-59. [doi: 10.1016/S0140-6736(18)31113-9] [Medline: 29937195]
12. Timmers T, Janssen L, Pronk Y, van der Zwaard BC, Koëter S, van Oostveen D, et al. Assessing the efficacy of an educational smartphone or tablet app with subdivided and interactive content to increase patients' medical knowledge: Randomized controlled trial. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e10742 [FREE Full text] [doi: 10.2196/10742] [Medline: 30578185]
13. Hardt S, Schulz MR, Pfitzner T, Wassilew G, Horstmann H, Liodakis E, et al. Improved early outcome after TKA through an app-based active muscle training programme: A randomized-controlled trial. Knee Surg Sports Traumatol Arthrosc 2018 Nov;26(11):3429-3437. [doi: 10.1007/s00167-018-4918-2] [Medline: 29589050]
14. Van Reijen M, Vriend I, Zuidema V, van Mechelen W, Verhagen EA. The "Strengthen your ankle" program to prevent recurrent injuries: A randomized controlled trial aimed at long-term effectiveness. J Sci Med Sport 2017 Jun;20(6):549-554. [doi: 10.1016/j.jsams.2016.12.001] [Medline: 27988093]
15. Lakshminarayana R, Wang D, Burn D, Chaudhuri KR, Galtrey C, Guzman NV, et al. Using a smartphone-based self-management platform to support medication adherence and clinical consultation in Parkinson's disease. NPJ Parkinsons Dis 2017;3:2 [FREE Full text] [doi: 10.1038/s41531-016-0003-z] [Medline: 28649602]
16. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: A randomized controlled trial. Neurourol Urodyn 2017 Jun;36(5):1369-1376. [doi: 10.1002/nau.23116] [Medline: 27611958]
17. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
18. Lorenzo-Zúñiga V, Moreno de Vega V, Marín I, Barberá M, Boix J. Improving the quality of colonoscopy bowel preparation using a smart phone application: A randomized trial. Dig Endosc 2015 Jul;27(5):590-595. [doi: 10.1111/den.12467] [Medline: 25708251]
19. Mata J, Pecorelli N, Kaneva P, Moldoveanu D, Gosselin-Tardiff A, Alhashemi M, et al. A mobile device application (app) to improve adherence to an enhanced recovery program for colorectal surgery: A randomized controlled trial. Surg Endosc 2020 Feb;34(2):742-751. [doi: 10.1007/s00464-019-06823-w] [Medline: 31087175]
20. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: A randomized controlled trial. Am Heart J 2017 Jun;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
21. Lee H, Ghebre R, Le C, Jang YJ, Sharratt M, Yee D. Mobile phone multilevel and multimedia messaging intervention for breast cancer screening: Pilot randomized controlled trial. JMIR Mhealth Uhealth 2017 Nov 07;5(11):e154 [FREE Full text] [doi: 10.2196/mhealth.7091] [Medline: 29113961]
22. Wang S, Wang Q, Yao J, Zhao S, Wang L, Li Z, et al. Effect of WeChat and short message service on bowel preparation: An endoscopist-blinded, randomized controlled trial. Eur J Gastroenterol Hepatol 2019 Feb;31(2):170-177. [doi: 10.1097/MEG.0000000000001303] [Medline: 30418256]
23. Li C, Huang S, Su X, Zhang T, Jiang K. Monitoring of home recovery using the 317-nursing mobile application following day-case surgery in children: Perspectives from both nurses and patients. Medicine (Baltimore) 2019 Aug;98(31):e16639 [FREE Full text] [doi: 10.1097/MD.0000000000016639] [Medline: 31374033]
24. Jeon SC, Kim JH, Kim SJ, Kwon HJ, Choi YJ, Jung K, et al. Effect of sending educational video clips via smartphone mobile messenger on bowel preparation before colonoscopy. Clin Endosc 2019 Jan;52(1):53-58 [FREE Full text] [doi: 10.5946/ce.2018.072] [Medline: 30153724]

25. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GY. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: mAF App trial. Am J Med 2017 Dec;130(12):1388-1396.e6 [FREE Full text] [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
26. Najafi Ghezeljeh T, Sharifian S, Nasr Isfahani M, Haghani H. Comparing the effects of education using telephone follow-up and smartphone-based social networking follow-up on self-management behaviors among patients with hypertension. Contemp Nurse 2018;54(4-5):362-373. [doi: 10.1080/10376178.2018.1441730] [Medline: 29451091]
27. Alanzi T, Bah S, Alzahrani S, Alshammari S, Almunsef F. Evaluation of a mobile social networking application for improving diabetes type 2 knowledge: An intervention study using WhatsApp. J Comp Eff Res 2018 Sep;7(9):891-899. [doi: 10.2217/cer-2018-0028] [Medline: 29943647]
28. Sharara AI, Chalhoub JM, Beydoun M, Shayto RH, Chehab H, Harb AH, et al. A customized mobile application in colonoscopy preparation: A randomized controlled trial. Clin Transl Gastroenterol 2017 Jan 05;8(1):e211 [FREE Full text] [doi: 10.1038/ctg.2016.65] [Medline: 28055031]
29. Grady PA, Gough LL. Self-management: A comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
30. 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]
31. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: A systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
32. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
33. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. Correction: The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial. PLoS One 2018;13(5):e0198008 [FREE Full text] [doi: 10.1371/journal.pone.0198008] [Medline: 29787600]
34. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mHealth: Review of strategies and structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771 [FREE Full text] [doi: 10.2196/10771] [Medline: 30341042]
35. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: Systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
36. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
37. Tønning ML, Kessing LV, Bardram JE, Faurholt-Jepsen M. Methodological challenges in randomized controlled trials on smartphone-based treatment in psychiatry: Systematic review. J Med Internet Res 2019 Oct 27;21(10):e15362 [FREE Full text] [doi: 10.2196/15362] [Medline: 31663859]

Abbreviations

CINAHL: Cumulative Index to Nursing and Allied Health Literature
eHealth: electronic health
MEDLINE: Medical Literature Analysis and Retrieval System Online
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Radboudumc: Radboud University Medical Center
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by G Eysenbach; submitted 08.12.19; peer-reviewed by L Whitehead, D Reukers, MC Chang, G Dermody, C Mullens; comments to author 15.01.20; revised version received 21.02.20; accepted 22.03.20; published 13.04.20.

Please cite as:
Timmers T, Janssen L, Kool RB, Kremer JAM
Educating Patients by Providing Timely Information Using Smartphone and Tablet Apps: Systematic Review
J Med Internet Res 2020;22(4):e17342
URL: http://www.jmir.org/2020/4/e17342/
doi:10.2196/17342
PMID:32281936

Review

Improving the Theoretical Understanding Toward Patient-Driven Health Care Innovation Through Online Value Cocreation: Systematic Review

Atae Rezaei Aghdam[1], PhD; Jason Watson[1], PhD; Cynthia Cliff[2], PhD; Shah Jahan Miah[3], MSc, PhD

[1]School of Information Systems, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia

[2]Faculty of Health, Queensland University of Technology, Brisbane, Australia

[3]Business School, Victoria University, Melbourne, Australia

Corresponding Author:
Atae Rezaei Aghdam, PhD
School of Information Systems
Science and Engineering Faculty
Queensland University of Technology
2 George St
Brisbane, 4001
Australia
Phone: 61 431825500
Email: atae.rezaeiaghdam@hdr.qut.edu.au

Abstract

Background: Patient participation in the health care domain has surged dramatically through the availability of digital health platforms and online health communities (OHCs). Such patient-driven service innovation has both potential and challenges for health care organizations. Over the last 5 years, articles have surfaced that focus on value cocreation in health care services and the importance of engaging patients and other actors in service delivery. However, a theoretical understanding of how to use OHCs for this purpose is still underdeveloped within the health care service ecosystem.

Objective: This paper aimed to introduce a theoretical discussion for better understanding of the potential of OHCs for health care organizations, in particular, for patient empowerment.

Methods: This literature review study involved a comprehensive search using 12 electronic databases (EMBASE, PsycINFO, Web of Science, Scopus, ScienceDirect, Medical Literature Analysis and Retrieval System Online, PubMed, Elton B Stephens Co [academic], Cumulative Index of Nursing and Allied Health Literature, Accelerated Information Sharing for Law Enforcement, Association for Computing Machinery, and Google Scholar) from 2013 to 2019. A total of 1388 studies were identified from the database search. After removing duplicates and applying inclusion criteria, we thematically analyzed 56 articles using the Braun and Clarke thematic analysis approach.

Results: We identified a list of 5 salient themes: *communication extension, improved health literacy for patients and health care organizations, communication transparency with patients, informational and social support for patients, and patient empowerment in self-management*. The most frequent theme was communication extension, which covers 39% (22/56) of the literature. This theme reported that an extension of communication between patients, caregivers, and physicians and organizations led to new opportunities to create value with minimal time and cost restrictions. Improved health literacy and communication transparency with patients were the second and third most frequent themes, respectively, covering 26% (15/56) and 25% (14/56) of the literature, respectively. The frequency of these themes indicated that the use of OHCs to generate new knowledge from patients' interactions helped health care organizations to customize treatment plans and establish transparent and effective communication between health care organizations and patients. Furthermore, of the 56 studies, 13 (23%) and 10 (17%) studies contended the opportunity of using OHCs in terms of informational and emotional support and empowering patients in their self-management of diseases.

Conclusions: This review enables better understanding of the current state of the art of the online value cocreation and its potential for health care organizations. This study found that the opportunities for health care organizations through enhancement of patient participation and their cocreation of value in digital health platforms have been rapidly increasing. The identified gaps

and opportunities in this study would identify avenues for future directions in modernized and more effective value-oriented health care informatics research.

(J Med Internet Res 2020;22(4):e16324) doi:10.2196/16324

KEYWORDS

value cocreation; health care organizations; digital health platforms; online health communities; patient empowerment

Introduction

Background

The health care industry is under increasing pressure in terms of enhancing their service provision and quality with meeting the growing demands. This is because of population growth and the rise of chronic diseases [1] besides other factors. To respond to these pressures, the health care industry is continuously digitizing its service provisions to provide for more effective and cost-efficient care models, as well as self-care management for personalized health care [2]. As part of the digitization, health care organizations are establishing online health communities (OHCs) as part of their service offering [3] to cocreate value. An OHC refers to a group of people who interact with each other in online environments about similar health issues [4]. This is reflected in the health care literature with growing emphasis on health value cocreation and the benefits of consumer value cocreation in the health domain [5,6]. Historically, value creation was conceptualized as company centric with the value being provided by the company to the customer. More recently, cocreation is viewed as an appropriate customer-centric mechanism for health care organizations in which value is created *with* customers rather than *for* customers [7,8]. In the context of OHCs, the interaction between stakeholders within online platforms can create values that allow stakeholders to share their knowledge and experiences [9]. Through creating value by working with health care organizations, a patient can likely raise their feelings about the existential quality of life, improve the attainment of life goals, and support and reduce their psychological and physical distress [10]. OHCs can provide peer health knowledge, emotional support, and improve self-care for patients with chronic diseases, especially for lifestyle-related diseases such as cancer, obesity or type 2 diabetes [11]. For patients with chronic diseases, OHCs provide a set of anecdotal information [12,13], which helps patients increase their positive emotional experience and attitude toward chronic diseases, engaging them in the activities of the community [14]. Empowering patients improves their role in cocreating, co-designing, and co-delivering health services [15]. In addition, patient empowerment contributes to enhancing the quality of care and health outcomes [16,17]. For instance, OHC users with chronic diseases become more knowledgeable, feel better socially supported, and have improved behavioral and clinical outcomes compared with nonusers [18].

Over the last 5 years, the number of articles that focus on value cocreation in the health care services has increased, highlighting the significance of the collaboration and cocreation of value within the health care service ecosystem between the patients and health care providers [19,20]. In our last paper, we systematically reviewed the literature regarding the role of OHCs as facilitators of value cocreation in the health care service ecosystem within the last 5 years [21]. The findings showed that OHCs provide opportunities for members to cocreate value ubiquitously along with providing members with online informational and emotional support. However, the ability of health care organizations to engage patients in the health care service coproduction and value cocreation has been largely overlooked [22-24]. Due to the importance and apparent oversight of value cocreation for health care organizations, this paper sought to address the following research question: To what extent could online value cocreation add value for health care organizations?

To produce new insights into this research question, this study performed a descriptive literature review to investigate the potential of online value cocreation for health care organizations, identifying the current state of knowledge and the opportunities for health care organizations to engage in online value cocreation.

Value Cocreation in Health Care

Technological developments promoted a shift from a health care model dominated by professionals toward a patient-centered model in which patients and professionals collaborate to create a service that offers the most optimal health care solution [25]. In recent years, the health care domain has undergone a number of transformations because of the recent advances in technology [26]. In addition, significant priorities for service marketing research include the role of consumers in the cocreation of value within the service sector, the transformative potential of services, and the interface between consumer communities and organizations [27]. This is also reflected in health research with some research evidencing the need for health care service providers and physicians to understand the patient and their role in the provision of health information [26,28,29]. The positive impact of collaborations among patients, physicians, and other actors on health outcomes has justified the significance given to the health care domain in investigating the cocreation of value [30]. Value cocreation in health care is a framework that integrates quality enhancement efforts by health care community staff members with patients' engagement to promote innovation in creating value [31]. Value cocreation refers to the process through which health care providers collaboratively engage with customers to create value [32]. Organizations are increasingly offering value cocreation opportunities to create more value for both customers and themselves [32-34]. Hence, the provider should take a holistic view of service delivery and consider the important factors in the clinical encounter for empowering patients to assume an active participatory role. Information technology can nurture better health care along with cost reduction and develop service innovation [35,36]. Recently, online community research has gradually started to focus on

value cocreation and community outcomes [37]. OHCs provide opportunities for stakeholders such as patients, physicians, caregivers, and health care organizations to access and share health information as well as contribute to the value cocreation process, diminish geographical barriers, and provide informational, social, and emotional support [38,39].

Online Health Communities as Platforms for Value Cocreation

With the rapid growth of social media technology, OHCs provide opportunities for fostering cocreation among the different stakeholders in health care [10,40]. OHCs are a particular form of special interest community, centered on a shared interest in health conditions and diseases [41]. The primary objective of the most OHCs is to provide a platform for patients regarding interacting with each other to obtain emotional support and disease management and care [42]. Participation in OHCs leads to additional activities carried out by patients, which add value to the patient-provider interactions [43]. Research showed that the more effort patients put into value cocreation activities, the more likely they are to continue with the health care provider, to return to the health care provider when they need treatment in the future, and to recommend the provider to others [44]. In the context of service-dominant logic, which is focused on patients' contribution to the value creation, the customer is *always a cocreator of value* [8]. In essence, OHCs are changing the way patients treat and manage their health [45]. These communities facilitate self-management through health information exchange and disease experiences [46]. The prominent characteristics of online communities are strong social relationship among members, community-specific organizational structure and way of discussion, history sharing, community rituals, and common online meeting space [47]. All these characteristics support identity for the community, provide a long-lasting relationship between participants, and foster strong member commitment to community purposes [45]. Several studies have focused on the benefits of OHCs, including (a) availability of health-related information especially for people who live in remote areas, facilitating information and social support without the need for driving long distance for face-to-face support group [12,13]; (b) access to health-related information with minimum cost [14]; (c) accessibility to experience-based health information, sharing daily coping habits and user experiences with symptoms [23,48]; and (d) decrease in the feeling of loneliness, creating social interactions among patients with a stigmatized medical condition and avoiding asking for help outside of the online community [23,49]. By empowering patients in OHCs, it is possible to activate value cocreation path between them and health care organizations [50]. As such, by enabling patients to support each other in OHCs, organizations can also indirectly improve customers' ties with the product and with the organization [46]. Accordingly, OHCs offer cocreation of value opportunities among patients, physicians, and health care organizations to improve health care outcomes [51]. This process, in turn, fosters patients' access to health information [52].

Health Care Organizations and Online Value Cocreation

To date, many health care organizations are rapidly recognizing the importance of OHCs as a significant platform of complementary service to improve the total quality of health care services [53]. Although the advantages of using OHCs for health care providers are promising, health care organizations frequently discounted OHCs' information because of the lack of clinical training of contributors and perceived lower quality of online health information [54]. Health care organizations also experience barriers regarding privacy, confidentiality, reputation management, and the dissemination of inaccurate health information [55-57]. Many health care providers worry about broadcasting misinformation and its negative influence on patients' health decisions [54,58,59]. Some patients contended that the risk of misinformation in OHCs might be reduced by the participation of health care organization in the conversation. In return, health care organizations could receive valuable experience-based health information from patients, saving organizational resources [32]. The diverse needs of various patients prevent the setup of a single, one-size-fits-all community; rather, cognitive- and affective-related values in a community depend on who participate in the community, the foundation of their relationship, and their activities such as sharing experiences, assessing new ideas, and recommending alternative treatments. Therefore, the complexity of digital services, which involve different goals of interaction among different actors, demands a more granular view of value cocreation in online communities [60]. Although numerous businesses have started to harness the advance potential of online communities by utilizing them as an online environment for customer co-innovation and value cocreation, health care organizations are lagging behind [29,61]. Hence, health care organizations can extend a better understanding of various types of consumer value cocreation that is enabled by OHCs [10]. To address these gaps and to deeply understand the potential of online value cocreation for health care organizations, our main objective was as follows: to identify the salient themes of the current literature regarding the potential of online value cocreation for health care organizations.

Methods

Review Protocol

To identify the potential of online value cocreation for health care organizations, we needed to investigate the current literature. For doing so, a descriptive literature review is one of the suitable methods for providing a broad and comprehensive background about the current state of the art [62]. We employed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist to guide our systematic review of relevant peer-reviewed literature [63]. The main aim of the PRISMA is to assist authors in improving the reporting of systematic review and meta-analysis [63]. The following sections explain the article selection process and review protocol.

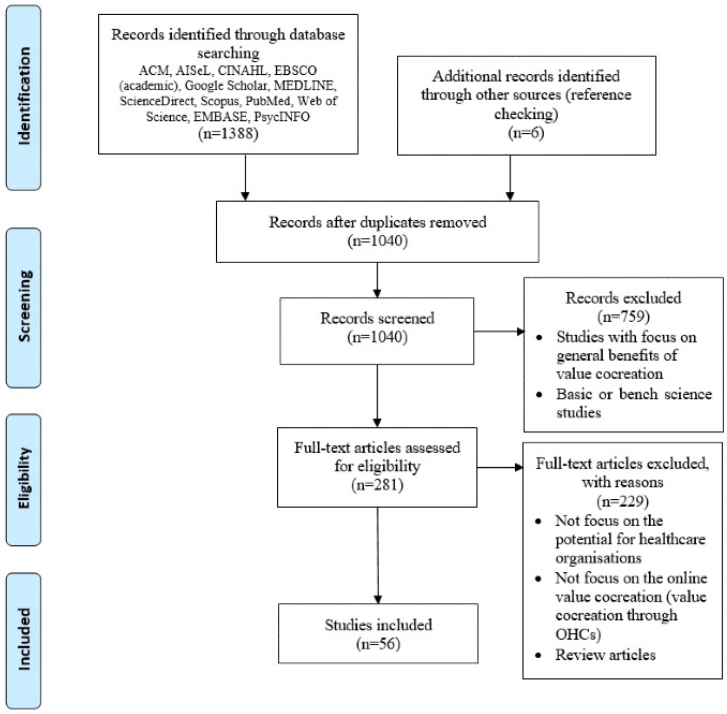
Search Method and Article Selection

The purpose of this study was to critically appraise extant work to answer the following question: to what extent could online value cocreation add value for health care organizations? To retrieve relevant articles, we reviewed several academic databases and derived database search terms. The search terms were derived based on keywords in the research question (online value cocreation and health care organizations). However, we did not merely focus on these particular keywords because some related papers discussed the same concepts using different terms. For instance, some articles discussed value cocreation activities in internet-based forums by various stakeholders without using specific terms such as *online value cocreation* or *health care ecosystem*. Hence, to ensure that relevant papers are not neglected, we broadened the search keywords and their synonyms to gather a comprehensive pool of related papers for this study (see Multimedia Appendix 1). Based on our research aim, a search string was defined using Boolean operators such as "AND" and "OR": ("online value co-creation" OR "value co-creation") AND ("healthcare organisation" OR "healthcare service providers") AND ("healthcare service ecosystem"). The variants of the search terms in Multimedia Appendix 1 were applied in different databases. In addition, we considered social science databases such as PsycINFO and EMBASE in terms of covering special studies in psychology and behavioral science. We also looked at controlled vocabularies such as Medical Subject Headings and thesaurus for a more complete search. In addition, three prominent researchers in the field of information systems (IS) and health informatics were asked to recommend any additional studies that met the inclusion criteria.

Eligibility Criteria

The searching process was complemented with articles identified from the reference lists as well as searching within the table of contents of selected journals. The outcome of phase 1 yielded 1040 papers after the removal of duplicates. Two scholars separately reviewed the abstracts of each paper and determined if the paper was relevant using the following inclusion criteria: (1) articles were written in English, (2) articles were published between 2013 and 2019, (3) articles focused on the role of health care organizations in the online value cocreation, and (4) articles employed quantitative or qualitative research studies that focused on the use of digital health platforms in value cocreation. During this process, the reference lists of the papers were also checked to identify other articles that are potentially eligible for inclusion. This returned 6 new papers. The two reviewers agreed with each other on the final pool of articles: 281 were identified to be eligible for inclusion in this review. Upon applying the inclusion criteria to the full-text papers, 56 articles were determined to be relevant. Figure 1 (adapted from Nili et al [63]) illustrates the article selection process, and Multimedia Appendix 1 denotes the number of relevant articles retrieved per database. In addition, a summary of the characteristics of the relevant papers is detailed in Multimedia Appendix 2.

Figure 1. Flow diagram of the literature search. ACM: Association for Computing Machinery; AISeL: Accelerated Information Sharing for Law Enforcement; CINAHL: Cumulative Index of Nursing and Allied Health Literature; EBSCO: Elton B Stephens Co; MEDLINE: Medical Literature Analysis and Retrieval System Online; OHC: online health community.



Data Analysis Approach

To answer the research question, we conducted a thematic analysis following the 6 steps of coding proposed by Braun and Clarke [64] to explore the major themes regarding the potential of online value cocreation for health care organizations. Thematic analysis is a technique that is commonly used to identify, analyze, and report patterns (themes) within data [64]. This technique is an inductive approach and involves coding all sections of findings, discussion, and conclusion of all selected

papers (N=56). Applying this method needs *careful reading and rereading of the data* [64] to identify the explicit and implicit meaning embedded within the text [65]. The 6 steps of thematic analysis process defined by Braun and Clarke [64] are collecting data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and writing the report (Figure 2). We used NVivo 12 (a qualitative data analysis software; QSR International, Melbourne, Australia) as a repository for storing the articles, and all coding was manually performed.

During the first step, we performed a preliminary analysis of the relevant articles and recorded our notes through *memo* and *annotation* features of NVivo and marked ideas for coding in the next step. In the second step, post familiarization with the data, we manually generated an initial list of codes. In total, 112 initial codes were generated. In the third step, after generating initial codes, we refocused the analysis process on the broader level of themes, rather than codes, sorting the different codes into potential themes and subthemes. This was facilitated by creating a thematic map by utilizing the *mind map* features of NVivo to visualize codes and themes. The thematic map revealed 10 themes and 41 miscellaneous codes. In the fourth step, we reviewed and refined themes. In the reviewing process, we reviewed all themes to make sure they followed a coherent pattern. During this phase, two themes collapsed into each other because of their common content. In the refining process, we recoded some additional data that have been missed in earlier stages, resulting in 11 themes. In the fifth step, defining and refining needs to be applied to our themes. In doing so, we named and defined themes [66]. We also performed a detailed analysis of each theme to ensure that the theme was relevant to the research question and that only minimal overlap existed between themes. Another significant factor in this stage is the naming of themes. Names of themes should be concise and punchy and should directly define what the theme is about [66]. In doing so, we named the themes that reflect the answer to our research question. The results of this stage revealed 5 themes (see Table 1). In the sixth step of the thematic analysis, we provided a concise, coherent, and logical report to summarize the themes as presented in the following sections.

Figure 2. Thematic analysis steps adapted from Braun and Clarke.

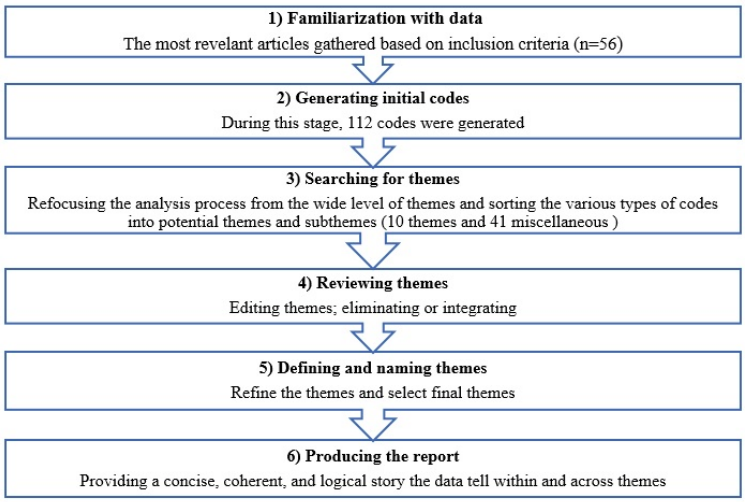


Table 1. Summary of the thematic analysis outcomes.

Theme	Description	Subthemes	Coverage of the final selection (N=56), n (%)
Communication extension	Digital health platforms such as OHCs[a] extend communication from the traditional power balance face-to-face consultation between patients and health care professionals to online interactions for facilitating dialogues between stakeholders.	• Resource integration • Easy access to health care information • Resource exchange with other stakeholders	22 (39)
Improved health literacy for patients and health care organizations	The use of digital platforms such as OHCs to generate new knowledge from patients' interactions and help health care organizations to customize treatment plans, offering some online advice especially for patients with a chronic disease.	• Easy-to-read and easy-to-understand health materials • Helping patients in the decision-making process • Co-learning	15 (26)
Communication transparency with patients	The use of digital platforms such as OHCs establishes transparent and effective communication between patient and patient, patient and physicians, and patient and health care organizations.	• Bridge builder • Facilitate communications • Effective interactions between patients and health care organizations • Improve mutual trust	14 (25)
Informational and social support for patients	The use of digital platforms such as OHCs by health care organizations provides informational and social support, which increases the quality of services and patient satisfaction.	• Positive comments to patients by health care organizations • Timely and appropriate responses to patients • Knowledge sharing and information exchange	13 (23)
Patient empowerment in self-management	The use of digital platforms such as OHCs to engage patients in the value cocreation process, assisting them in their self-management of diseases.	• Engaging members in OHCs • Encouraging members of OHCs • Positive patient-provider interactions • Self-management intervention for diseases	10 (17)

[a]OHCs: online health communities.

Trustworthiness

Trustworthiness is a key factor in a qualitative content analysis because text can have multiple meanings and diverse interpretation [67]. It is contended that if the analysis process provides adequate details, the validity of the research will be assured [68]. Accordingly, this research explicates the process of coding step by step to establish the trustworthiness of the study. In terms of testing the trustworthiness of the findings, we employed percent agreement as our method of intercoder reliability checking [69]. Percent agreement is a useful method of checking the reliability of the qualitative findings of less sensitive (eg, literature review) IS projects [69], where the original analyst and a second person compare the findings of their analyses for a sample of papers. The two scholars, experienced in qualitative research and thematic analysis, who checked different parts (paper selection and inclusion and exclusion criteria) placed particular emphasis on reviewing all codes, categories, and themes. They looked at the themes of the study and the way they have originated from categories and codes. During the first meeting, the percent agreement was 75%, and after the second meeting, that is, discussion on the essence of the themes, the general consensus was achieved and the overall results were 100%, making us confident about the reliability of our literature review.

Results

Themes

Through performing the thematic analysis, we identified 5 main themes: *communication extension*, *improve health literacy*, *communication transparency with patients*, *patient empowerment in self-management*, and *informational and social support for the patient*. Each theme consisted of several subthemes which are presented in Table 1. The following sections provide more detail into each of them.

As evidenced in Table 1, communication extension was the most explored theme (22/56, 39%) followed by improved health literacy (15/56, 26%), communication transparency with patients (14/56, 25%), informational and social support for patients (13/56, 23%), and patient empowerment in self-management (10/56, 17%).

Themes Explanation

Communication Extension

OHCs provide an opportunity for stakeholders to extend their communication from traditional approach, which comprises few face-to-face scheduled consultations, to effective interactions among the patient, health care professionals, and health care organizations within digital health platforms. In fact, OHCs have a potential to connect members who would never have met each other because of geographical distance. Communication extension assists health care organizations to empower patients to interact with health care professionals and organizations [66]. In this regard, OHCs can contribute to unleashing the provision of health care services and facilitate resource exchange and peer-to-peer social support. In essence, communication extension brings the patients into focus, and OHCs offer a mechanism for value cocreation among stakeholders with minimal time and cost restriction [17,19,46,66-75]. The vast majority of the literature emphasizes information accessing and creating, sharing, and recombination of resources [43,76]. OHCs assist patients to produce health information by sharing their health experiences and cocreate a value within these platforms [77]. Resource exchanges are the mutual actions taken by stakeholders in the health care service ecosystem to access, monitor, share, and integrate resources [75,78,79]. In fact, the communication extension provides health care organizations with the opportunity to look beyond the patient portal and consider how technology nurtures consumers' relationship, engagement, and contributions as well as the health care organizations to better understand the value cocreation process that occurs within digital health platforms [66,71].

Improved Health Literacy of Patients and Health Care Organizations

As indicated in the previous theme, health care organizations need to empower and engage patients in the value cocreation process. However, insufficient organizational health literacy impoverishes the ability of a health care organization to fulfill this task [80]. Health care organizations need to educate patients on their innovation that they are planning to offer to patients, and patients need to educate health care organizations in the context of their everyday use of innovation [71]. OHCs can be an effective tool for knowledge sharing and peer education, and it has the potential to impact patients' health literacy [46,71]. The new knowledge generated from those patient interactions in OHCs assists health care organizations to customize treatment plans, offering some online treatment options especially for patients with a chronic disease [71]. Some papers contend that by enhancing health literacy, it would be possible to decrease the use of medical services, improving value cocreation with stakeholders and, ultimately, improving the efficiency of the health care system [70,72,81,82]. One of the most relevant papers in this review contends that *knowledge should flow both ways*. It means that patients should be educated by health care organizations and vice versa [71]. In fact, interactions between patients and health care organizations in digital health platforms enable organizations to collect patients' health-related experiences and information. Therefore, health care organizations benefit from digital health platforms such as OHCs that improve the patient's ability to use digital platforms to find relevant health-related information and to apply the gained knowledge to address a health issue, which is known as *eHealth literacy* in the literature.

Communication Transparency With Patients

Communication transparency between patients and health care organizations is a viable method to provide health care organizations, health care professionals, and patients with new opportunities to cocreate value [83,84]. In this regard, OHCs can provide a proper, transparent, and effective communication between stakeholders [82]. The better health care organizations can effectively interact with patients in OHCs, the more online value will be created. Transparent communication and interaction between health care organizations and patients within OHCs are the fundamental building blocks for online value cocreation. One of the most important elements of having transparent and effective communication is trust. Trust can serve as a significant mechanism in reducing the uncertainty and complexity of exchange and enhancing the credibility of online health information [70]. The level of trust affects patients' behavior with health care organizations. The transparency of communication is critical for a successful and sustainable response to patients' health problems. This feature enables both health care organizations and patients to interact directly, allowing them to be actively involved with the value cocreation process [76].

Informational and Social Support for Patients

OHCs provide opportunities for stakeholders such as patients, physicians, caregivers, and health care organizations to access and share health information as well as contribute to the value cocreation process, diminish geographical barriers, and provide informational, social, and emotional support [38,39]. Value cocreation process occurs when organizations, stakeholders, and users integrate and renew each other's resources [85]. In this context, OHCs have immense potential to facilitate the process of value cocreation among actors in the health care ecosystem as well as provide an additional mechanism for obtaining informational and emotional support. Due to the nature of OHCs that provides access to information and coordinated social interaction, they constitute an alternative solution for patients' needs, such that they likely improve the well-being of individuals and society as a whole [86]. In fact, OHCs have developed as a part of the supplementary service of many health care organizations, where patients are directed to the health care organization's OHC to receive socioemotional support. OHCs are the major source of informational, social, and emotional support for people with health problems, and members of such communities interact with other stakeholders to seek, receive, and provide different types of supports (informational, social, and emotional) [4,70,82]. Users of OHCs participate and experience the various types of value cocreation via social support exchange within these types of communities [87]. Health care organizations can motivate patients in OHCs' activities by providing positive feedback to patients who are using digital health platforms. These types of feedback increase the patient's motivation and serve as a signal that the health care organization

approves and welcomes patients' activities in the health care domain [88].

Patient Empowerment in Self-Management

OHCs empower patients in the self-management process through the exchange of health-related information and patient experiences. In other words, OHCs can be used to actively engage and empower patients in their health care journey. Engagement in health care service innovation enables health care organizations to be more proactive, establishing and supporting effective approaches for online value cocreation. The empowerment of patients leads to improving the patients' role in value cocreation [80]. The patient's empowerment and engagement contribute to enhancing the quality of health care outcome, and health care organizations have started to recognize the significance of patient empowerment as a driver of patient-centered care. Self-management is where health care organizations can enhance patient participation, supporting patients in controlling their lives [82]. Helping patients with chronic diseases such as type 2 diabetes or asthma to self-manage their condition leads to enhanced quality of care and decreases cost and inappropriate use of health care resources [82].

Discussion

Principal Findings

The goal of this study was to improve the theoretical understanding toward patient-driven health care innovation and in this case, identifying the potential of online value cocreation for health care organizations. We conducted a descriptive review of the published papers on the potential of online value cocreation for health care organizations. After analyzing a large number of studies, we now, understand the significance of the digital health platforms in the value cocreation process for both health care organizations and patients by identifying the salient themes of the current literature. These themes were *communication extension*, *improved health literacy for patients and health care organizations*, *communication transparency with patients*, *informational and social support for patients*, and *patient empowerment in self-management*.

Numerous public and nonprofit health care organizations have started to embrace OHCs to support patients' requirements to locate others with similar health issues and share experiences at a peer-to-peer level [89]. Accordingly, we believe health care organizations perceive OHCs as a tool for extending the communication from the traditional power balanced between health care professionals and patients to online interaction among all stakeholders within the health care service ecosystem, empowering patients in their self-management of diseases by cocreating social and emotional value within OHCs, informationally and socially supporting patients, and establishing transparent communication with patients. Health care organizations might benefit from the interaction between members in OHCs. For example, by monitoring the user-generated content within the community, organizations can gain a deeper and better understanding of members' needs and finally co-innovate and coproduce health services with customers. This generated knowledge enhances members' and organizations' health literacy, helping patients in the decision-making process regarding health care services. In essence, productive collaboration with multiple stakeholders improves the resource, competencies, and capabilities, which fosters values in the cocreation process [85,90]. On the basis of our findings and in line with recommendations to the industry, we propose that health care organizations can prioritize the findings of this study (themes). By prioritizing the salient themes, health care organizations can leverage the potential of online value cocreation to improve their service quality and patient empowerment. According to the recent studies, "the organizations that have been shifting their strategies toward value-based care generally share certain advantages: financial stability, positive relationship with physicians, advanced information systems and (often) affiliation with health plan" [91]. Therefore, this study elucidates the potential that OHCs provide to health care organizations to engage in the value cocreation process. We employed the findings of this review in the real-world program for diabetes group education through conceptualizing OHCs and their potential for this cohort.

Strengths and Limitations

There are multiple strengths of this systematic review. It was conducted based on the PRISMA guidelines. It employed a rigorous and extensive search strategy to identify the most relevant outlets. Paper selection process based on the inclusion criteria, paper codding, and theme identification were conducted in duplicate by two members of the research team independently to ensure the accuracy of the findings. This study yielded beneficial findings that enabled us to synthesize and present the current state of the art of the potential of online value cocreation for health care organizations. Although this literature review sheds light on the potential of online value cocreation for health care organizations, some limitations of our review need to be considered. Our inclusion criteria limited our review results to only English-language articles and published, peer-reviewed literature between 2013 and 2019. Hence, these restrictions might have led to an exclusion of relevant literature.

Directions for Future Studies

This review helped us to explore some interesting directions for future work. Future studies can delve into providing a holistic view of the importance of health service coproduction and value cocreation in shaping a dynamic health care ecosystem [92-94]. There is also a lack of understanding on how interactions among stakeholders, especially at the meso and macro levels, contribute to the emergence of value cocreation [95]. Another avenue for future research is to investigate the perspective in which health care organizations are able to engage indirectly in online value cocreation. In fact, the *cocreation of value through engagement in health care warrants more detailed exploration* and highlights the need for more empirical analysis and data on this significant area in health care services [9]. Only a few studies explored the value cocreation at higher levels of health care service ecosystems such as the meso and macro levels. Accordingly, future research can examine the online value cocreation in the health care service ecosystem in the higher levels of the service ecosystem. Future research should be directed toward improving understanding of the engagement level of health care

organizations in value cocreation through OHCs. In particular, enhanced understanding of health care organizations' participation determinants in online value cocreation process and the factors that underpin this phenomenon is required. This may involve future studies with a long follow-up period. The following recommendations for future work might be useful for both health care researchers and organizations:

1. Investigate how health care organizations indirectly engage with online value cocreation process. Identify their challenges and policies for online activities in the virtual communities.
2. Enhance understanding of the level of engagement of health care organizations in online value cocreation and explore health care organizations' participation determinants in the online value cocreation process.
3. Develop strategies to boost ongoing engagement of health care consumers. This will empower patients in their self-management of chronic diseases.
4. Increase understanding of how health care organizations encourage health care consumers in the health care service coproduction and co-innovation through OHCs.
5. Implement security and data privacy rules considering health care organizations' perspectives on OHCs, trust-building measures, and challenges associated with the privacy in OHCs.

Conclusions

The findings of this study enrich our understanding of online value cocreation and its potential for health care organizations by providing a rich review of the literature in online value cocreation. The health care domain can be conceived as a cocreating service system based on the engagement of health care organizations and patients, caregivers, and health care professionals. In this regard, digital platform is one of the most prevalent sources of interaction and online value cocreation. Accordingly, this study aims to improve theoretical understanding toward a patient-driven innovation, such as OHCs, from the value cocreation lenses. Our findings reveal that to foster the implementation of an effective service ecosystem, health care organizations should be able to empower both patients and health care professionals to allow them to actively participate in value cocreation processes in digital platforms such as OHCs. We contend that the outcomes of our study can provide a bird's-eye view for health care organizations to leverage OHCs for improving their business intelligence along with patient empowerment. Our findings would be useful for health care organization policymakers on how digital health platforms such as OHCs facilitate the value cocreation for health care organizations. Moreover, the findings of this study would be able to guide health care organizations in choosing and implementing strategies and features in their online communities that lead to positive outcomes. Here, we argue that existing OHCs can assist researchers and health care organizations not only by identifying the benefits and potential but also by facilitating value cocreation in the health care service ecosystem.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The variant of search terms in different databases.
[DOCX File , 14 KB - jmir_v22i4e16324_app1.docx]

Multimedia Appendix 2
Profile characteristics of selected papers.
[DOCX File , 13 KB - jmir_v22i4e16324_app2.docx]

References

1. OECD. OECD iLibrary. Health Spending URL: https://www.oecd-ilibrary.org/social-issues-migration-health/health-spending/indicator/english_8643de7e-en [accessed 2018-09-14]
2. Caulfield BM, Donnelly SC. What is Connected Health and why will it change your practice? QJM Int J Med 2013;106(8):703-707. [doi: 10.1093/qjmed/hct114] [Medline: 23676416]
3. Nambisan P. Enabling consumer-driven service innovation in health care: the role of online Health Information Technologies (HIT). In: Nambisan S, editor. Information Technology and Product Development. Boston, MA: Springer; 2010:159-177.
4. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. J Med Internet Res 2017 Apr 24;19(4):e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]
5. Janamian T, Crossland L, Jackson CL. Embracing value co-creation in primary care services research: a framework for success. Med J Aust 2016 Apr 18;204(7 Suppl):S5-11. [doi: 10.5694/mja16.00112] [Medline: 27078794]
6. McColl-Kennedy JR. Health care in service science special issue of Australasian Marketing Journal. Australas Mark J 2014 Aug;22(3):165-167. [doi: 10.1016/j.ausmj.2014.08.012]
7. Prahalad CK, Ramaswamy V. Co‐creating unique value with customers. Strategy Leadersh 2004;32(3):4-9. [doi: 10.1108/10878570410699249]

8. Vargo SL, Lusch RF. The four service marketing myths: remnants of a goods-based, manufacturing model. J Serv Res 2004;6(4):324-335. [doi: 10.1177/1094670503262946]
9. Hardyman W, Daunt KL, Kitchener M. Value co-creation through patient engagement in health care: a micro-level approach and research agenda. Public Manag Rev 2015;17(1):90-107. [doi: 10.1080/14719037.2014.881539]
10. Loane SS, Webster CM. Consumer-to-consumer value within social networks. Mark Rev 2014;14(4):444-459. [doi: 10.1362/146934714x14185702841442]
11. Nambisan P, Nambisan S. Models of consumer value cocreation in health care. Health Care Manage Rev 2009;34(4):344-354. [doi: 10.1097/HMR.0b013e3181abd528] [Medline: 19858919]
12. Russo G, Tartaglione AM, Cavacece Y. Empowering patients to co-create a sustainable healthcare value. Sustain 2019;11(5):1315. [doi: 10.3390/su11051315]
13. Vilhauer RP. Perceived benefits of online support groups for women with metastatic breast cancer. Women Health 2009;49(5):381-404. [doi: 10.1080/03630240903238719] [Medline: 19851944]
14. Qiu B, Zhao K, Mitra P, Wu D, Caragea C, Yen J, et al. Get Online Support, Feel Better -- Sentiment Analysis and Dynamics in an Online Cancer Survivor Community. In: Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. 2011 Presented at: SocialCom'11; October 9-11, 2011; Boston. [doi: 10.1109/PASSAT/SocialCom.2011.127]
15. Hilliard ME, Sparling KM, Hitchcock J, Oser TK, Hood KK. The emerging diabetes online community. Curr Diabetes Rev 2015;11(4):261-272 [FREE Full text] [doi: 10.2174/1573399811666150421123448] [Medline: 25901500]
16. Palumbo R. The Bright Side and the Dark Side of Patient Empowerment: Co-creation and Co-destruction of Value in the Healthcare Environment. Switzerland: Springer; 2017.
17. Aung E, Donald M, Coll JR, Williams GM, Doi SA. Association between patient activation and patient-assessed quality of care in type 2 diabetes: results of a longitudinal study. Health Expect 2016 Apr;19(2):356-366 [FREE Full text] [doi: 10.1111/hex.12359] [Medline: 25773785]
18. Schmidt M, Eckardt R, Scholtz K, Neuner B, von Dossow-Hanfstingl V, Sehouli J, PERATECS Group. Patient empowerment improved perioperative quality of care in cancer patients aged ≥ 65 years - a randomized controlled trial. PLoS One 2015;10(9):e0137824 [FREE Full text] [doi: 10.1371/journal.pone.0137824] [Medline: 26378939]
19. Bronstein J. Is this OCD?: Exploring conditions of information poverty in online support groups dealing with obsessive compulsive disorder. Inform Res 2014;19(4):-.
20. Dodds SL. Semantic Scholar.: Massey University; 2015. Consumer Value and Value Co-Creation in Complementary and Alternative Medicine (CAM) Health Services: A Thesis Presented in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Marketing URL: https://pdfs.semanticscholar.org/aca0/a9d513c3e16905ace29081e01ced3d59b65f.pdf?_ga=2.239037376.1106761467.1581948404-1679671381.1567599385 [accessed 2020-02-17]
21. Aghdam AR, Watson J, Ziaimatin H. Online Value Co-Creation in the Healthcare Service Ecosystem: A Review. In: Proceedings of the 2018 Australasian Conference on Information Systems. 2018 Presented at: ACIS'18; Dec 3-5, 2018; Sydney.
22. Forouzandeh S, sheikhahmadi A, Rezaei Aghdam A, Xu S. New centrality measure for nodes based on user social status and behavior on Facebook. Int J of Web Inf Syst 2018;14(2):158-176 [FREE Full text] [doi: 10.1108/IJWIS-07-2017-0053]
23. Palumbo R. Contextualizing co-production of health care: a systematic literature review. Int J Public Sec Manag 2016;29(1):72-90. [doi: 10.1108/ijpsm-07-2015-0125]
24. Sharma S, Conduit J, Hill SR. Organisational capabilities for customer participation in health care service innovation. Australas Mark J 2014;22(3):179-188. [doi: 10.1016/j.ausmj.2014.08.002]
25. Willis C, Saul J, Bitz J, Pompu K, Best A, Jackson B. Improving organizational capacity to address health literacy in public health: a rapid realist review. Public Health 2014 Jun;128(6):515-524. [doi: 10.1016/j.puhe.2014.01.014] [Medline: 24923994]
26. Camacho N, Landsman V, Stremersch S. The connected patient. In: The Connected Customer: The Changing Nature of Consumer and Business Markets. Abingdon-on-Thames: Routledge Academic; 2010:107-139.
27. Osei-Frimpong K, Wilson A, Lemke F. Patient co-creation activities in healthcare service delivery at the micro level: the influence of online access to healthcare information. Technol Forecast Soc Change 2018 Jan;126:14-27. [doi: 10.1016/j.techfore.2016.04.009]
28. Ostrom AL, Parasuraman A, Bowen DE, Patrício L, Voss CA. Service research priorities in a rapidly changing context. J Serv Res 2015;18(2):127-159. [doi: 10.1177/1094670515576315]
29. Nambisan S, Nambisan P. How to profit from a better 'virtual customer environment'. MIT Sloan Manag Rev 2008 Mar;49(3):53-61 [FREE Full text]
30. Nambisan S. Designing virtual customer environments for new product development: toward a theory. Acad Manag Rev 2002;27(3):392-413. [doi: 10.5465/AMR.2002.7389914]
31. Holman H, Lorig K. Patients as partners in managing chronic disease. Partnership is a prerequisite for effective and efficient health care. Br Med J 2000 Feb 26;320(7234):526-527 [FREE Full text] [doi: 10.1136/bmj.320.7234.526] [Medline: 10688539]

32. Sabadosa KA, Batalden PB. The interdependent roles of patients, families and professionals in cystic fibrosis: a system for the coproduction of healthcare and its improvement. BMJ Qual Saf 2014 Apr;23(Suppl 1):i90-i94. [doi: 10.1136/bmjqs-2013-002782] [Medline: 24608555]
33. Vennik FD, Adams SA, Faber MJ, Putters K. Expert and experiential knowledge in the same place: patients' experiences with online communities connecting patients and health professionals. Patient Educ Couns 2014 May;95(2):265-270. [doi: 10.1016/j.pec.2014.02.003] [Medline: 24598314]
34. Durugbo C, Pawar K. Visualising the Co-Creation Process. In: Proceedings of the 2012 18th International ICE Conference on Engineering, Technology and Innovation. 2012 Presented at: ICE'12; June 18-20, 2012; Munich, Germany. [doi: 10.1109/ice.2012.6297663]
35. Payne AF, Storbacka K, Frow P. Managing the co-creation of value. J Acad Mark Sci 2007 Aug;36(1):83-96. [doi: 10.1007/s11747-007-0070-0]
36. Lusch RF, Nambisan S. Service innovation: a service-dominant logic perspective. Manag Inf Syst Q 2015;39(1):155-175. [doi: 10.25300/misq/2015/39.1.07]
37. Goh JM, Gao G, Agarwal R. The creation of social value: can an online health community reduce rural-urban health disparities? Manag Inf Syst Q 2016;40(1):247-263. [doi: 10.25300/MISQ/2016/40.1.11]
38. 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;35(9):629-639. [doi: 10.1002/mar.21111]
39. Yan Z, Wang T, Chen Y, Zhang H. Knowledge sharing in online health communities: a social exchange theory perspective. Inf Manag 2016;53(5):643-653. [doi: 10.1016/j.im.2016.02.001]
40. Amann J. Democratising healthcare: The role of eHealth technologies in driving patient participation. EMJ Innov 2017;1(1):40-46 [FREE Full text]
41. van Oerle S. Value Co-Creation in Online Health Communities: The Role of Participants' Posts, Network Position and Behavioral Patterns. Antwerp, Belgium: UAntwerp; 2016.
42. Nambisan P, Gustafson D, Pingree S, Hawkins R. Patients' sociability and usability experience in online health communities: impact on attitudes towards the healthcare organisation and its services. Int J Web Based Commun 2010;6(4):395-409. [doi: 10.1504/ijwbc.2010.035841]
43. Hartmann BJ, Wiertz C, Arnould EJ. Exploring consumptive moments of value-creating practice in online community. Psychol Mark 2015;32(3):319-340. [doi: 10.1002/mar.20782]
44. Sweeney JC, Danaher TS, McColl-Kennedy JR. Customer effort in value cocreation activities: improving quality of life and behavioral intentions of health care customers. J Serv Res 2015;18(3):318-335. [doi: 10.1177/1094670515572128]
45. Chen L, Straub DW. The Impact of Virtually Crowdsourced Social Support on Individual Health: Analyzing Big Datasets for Underlying Causalities. In: Proceedings of the Twenty-first Americas Conference on Information Systems. 2015 Presented at: AMCIS'15; August 13-15, 2015; Puerto Rico.
46. Willis E. The making of expert patients: the role of online health communities in arthritis self-management. J Health Psychol 2014 Dec;19(12):1613-1625. [doi: 10.1177/1359105313496446] [Medline: 23988679]
47. Stanoevska-Slabeva K, Schmid BF. A Typology of Online Communities and Community Supporting Platforms. In: Proceedings of the 34th Annual Hawaii International Conference on System Sciences. 2001 Presented at: HICSS'01; January 6, 2001; Maui, HI, USA. [doi: 10.1109/hicss.2001.927041]
48. Mazzoni D, Cicognani E. Sharing experiences and social support requests in an internet forum for patients with systemic lupus erythematosus. J Health Psychol 2014 May;19(5):689-696. [doi: 10.1177/1359105313477674] [Medline: 23479300]
49. Parrish ES, Cornelious L. Perceptions regarding online support, illness, information, and awareness of women with cancer. J Behav Stud Bus 2015 Feb;8:1 [FREE Full text]
50. Litchman ML, Rothwell E, Edelman LS. The diabetes online community: older adults supporting self-care through peer health. Patient Educ Couns 2018 Mar;101(3):518-523. [doi: 10.1016/j.pec.2017.08.023] [Medline: 28947360]
51. Ciasullo MV, Cosimato S, Palumbo R, Storlazzi A. Value co-creation in the health service ecosystems: the enabling role of institutional arrangements. Int Bus Res 2017;10(12):222. [doi: 10.5539/ibr.v10n12p222]
52. Serbanati LD, Ricci FL, Mercurio G, Vasilateanu A. Steps towards a digital health ecosystem. J Biomed Inform 2011 Aug;44(4):621-636 [FREE Full text] [doi: 10.1016/j.jbi.2011.02.011] [Medline: 21362497]
53. Amann J, Rubinelli S. Views of community managers on knowledge co-creation in online communities for people with disabilities: qualitative study. J Med Internet Res 2017 Oct 10;19(10):e320 [FREE Full text] [doi: 10.2196/jmir.7406] [Medline: 29017993]
54. Rupert DJ, Moultrie RR, Read JG, Amoozegar JB, Bornkessel AS, O'Donoghue AC, et al. Perceived healthcare provider reactions to patient and caregiver use of online health communities. Patient Educ Couns 2014 Sep;96(3):320-326. [doi: 10.1016/j.pec.2014.05.015] [Medline: 24923652]
55. 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]
56. Munson SA, Cavusoglu H, Frisch L, Fels S. Sociotechnical challenges and progress in using social media for health. J Med Internet Res 2013 Oct 22;15(10):e226 [FREE Full text] [doi: 10.2196/jmir.2792] [Medline: 24148206]

57. Moturu ST, Liu H. Quantifying the trustworthiness of social media content. Distrib Parallel Databases 2011;29(3):239-260. [doi: 10.1007/s10619-010-7077-0]
58. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct;26(10):1168-1174 [FREE Full text] [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
59. Abirami A, Askarunisa A. Sentiment analysis model to emphasize the impact of online reviews in healthcare industry. Online Inf Rev 2017;41(4):471-486. [doi: 10.1108/OIR-08-2015-0289]
60. Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. Br Med J 2006 Apr 22;332(7547):939-942 [FREE Full text] [doi: 10.1136/bmj.38753.524201.7C] [Medline: 16513686]
61. Chandler JD, Lusch RF. Service systems: a broadened framework and research agenda on value propositions, engagement, and service experience. J Serv Res 2015;18(1):6-22. [doi: 10.1177/1094670514537709]
62. Amann J, Zanini C, Rubinelli S. What online user innovation communities can teach us about capturing the experiences of patients living with chronic health conditions. A scoping review. PLoS One 2016;11(6):e0156175 [FREE Full text] [doi: 10.1371/journal.pone.0156175] [Medline: 27272912]
63. Paré G, Trudel M, Jaana M, Kitsiou S. Synthesizing information systems knowledge: a typology of literature reviews. Inf Manag 2015;52(2):183-199. [doi: 10.1016/j.im.2014.08.008]
64. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
65. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
66. Nili A, Tate M, Johnstone D. The process of solving problems with self-service technologies: a study from the user's perspective. Electron Commer Res 2018;19(2):373-407. [doi: 10.1007/s10660-018-9304-0]
67. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006;5(1):80-92. [doi: 10.1177/160940690600500107]
68. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
69. 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]
70. Rantala K, Karjaluoto H. Combining Digitization with Healthcare Service Processes: Value Co-creation Opportunities Through Standard Work. In: Proceedings of the 30th Bled eConference. 2017 Presented at: 8thInternational Research Conference Management Challenges in the 21stCentury,; June 18 - 21, 2017; Bled, Slovenia. [doi: 10.18690/978-961-286-043-1.33]
71. Hajli MN, Sims J, Featherman M, Love PE. Credibility of information in online communities. J Strateg Mark 2015;23(3):238-253. [doi: 10.1080/0965254X.2014.920904]
72. Nambisan S, Nambisan P. How should organizations promote equitable distribution of benefits from technological innovation in health care? AMA J Ethics 2017 Nov 1;19(11):1106-1115 [FREE Full text] [doi: 10.1001/journalofethics.2017.19.11.stas1-1711] [Medline: 29168682]
73. Pikkarainen M, Ervasti M, Hurmelinna-Laukkanen P, Nätti S. Orchestration roles to facilitate networked innovation in a healthcare ecosystem. Technol Innov Manag Rev 2017;7(9):30-43. [doi: 10.22215/timreview/1104]
74. Bornkessel A, Furberg R, Lefebvre RC. Social media: opportunities for quality improvement and lessons for providers-a networked model for patient-centered care through digital engagement. Curr Cardiol Rep 2014 Jul;16(7):504. [doi: 10.1007/s11886-014-0504-5] [Medline: 24947833]
75. Spanò R, Di Paola N, Bova M, Barbarino A. Value co-creation in healthcare: evidence from innovative therapeutic alternatives for hereditary angioedema. BMC Health Serv Res 2018 Jul 20;18(1):571 [FREE Full text] [doi: 10.1186/s12913-018-3389-y] [Medline: 30029666]
76. Beirão G, Patrício L, Fisk RP. Value cocreation in service ecosystems: investigating health care at the micro, meso, and macro levels. J Serv Manag 2017;28(2):227-249. [doi: 10.1108/josm-11-2015-0357]
77. Kuenne C, Agarwal R. Online Innovation Intermediaries as a Critical Bridge between Patients and Healthcare Organizations. In: Proceedings of the 12th International Conference on Wirtschaftsinformatik. 2015 Presented at: WI'15; March 4-6, 2015; Osnabrück, Germany.
78. Petrakaki DI. Producing Communal Health Through Self Care: The Emergence of Digital Patient Activism. In: Proceedings of the 25th European Conference on Information Systems. 2017 Presented at: ECIS'17; June 5-10, 2017; Portugal.
79. Agarwal N, Soh C, Yeow A. Value Co-creation in Service Ecosystems: A Member Perspective. In: Proceedings of the Thirty Seventh International Conference on Information Systems. 2016 Presented at: ICS'16; December 11-14, 2016; Dublin.
80. Zainuddin N, Tam L, McCosker A. Serving yourself: value self-creation in health care service. J Serv Mark 2016;30(6):586-600. [doi: 10.1108/jsm-02-2016-0075]

81. Palumbo R, Annarumma C, Musella M. Exploring the meaningfulness of healthcare organizations: a multiple case study. Int J Public Sec Manag 2017;30(5):503-518. [doi: 10.1108/ijpsm-10-2016-0174]
82. Tartaglione AM, Cavacece Y, Cassia F, Russo G. The excellence of patient-centered healthcare: investigating the links between empowerment, co-creation and satisfaction. TQM J 2018;30(2):153-167. [doi: 10.1108/tqm-11-2017-0138]
83. 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 25;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
84. Volland D, Eurich M. ICT-enabled Value Creation in Community Pharmacies: an Applied Design Science Research Approach. In: Proceedings of the Thirty Fifth International Conference on Information Systems. 2014 Presented at: ICIS'14; 11/12/2014; Auckland.
85. Cubaka VK, Schriver M, Cotton P, Nyirazinyoye L, Kallestrup P. Providers' perceptions of communication with patients in primary healthcare in Rwanda. PLoS One 2018;13(4):e0195269 [FREE Full text] [doi: 10.1371/journal.pone.0195269] [Medline: 29617429]
86. Ramaswamy V, Ozcan K. The Co-Creation Paradigm. Redwood City: Stanford Business Books; 2014.
87. Zhao J, Wang T, Fan X. Patient value co-creation in online health communities: social identity effects on customer knowledge contributions and membership continuance intentions in online health communities. J Serv Manag 2015;26(1):72-96. [doi: 10.1108/josm-12-2013-0344]
88. Loane SS, Webster CM, D'Alessandro S. Identifying consumer value co-created through social support within online health communities. J Macromark 2014;35(3):353-367. [doi: 10.1177/0276146714538055]
89. Chung JE. Patient-provider discussion of online health information: results from the 2007 Health Information National Trends Survey (HINTS). J Health Commun 2013;18(6):627-648. [doi: 10.1080/10810730.2012.743628] [Medline: 23590202]
90. Stewart Loane S, Webster CM. Social capital and consumer value co-created within an online health community. J Nonprofit Public Sec Mark 2017;29(3):317-345. [doi: 10.1080/10495142.2017.1326359]
91. Ramaswamy V, Chopra N. Building a culture of co-creation at Mahindra. Strateg Leader 2014;42(2):12-18. [doi: 10.1108/sl-01-2014-0005]
92. Kaiser LS, Lee TH. Harvard Business Review. 2015 Oct 8. Turning Value-Based Health Care into a Real Business Model URL: https://hbr.org/2015/10/turning-value-based-health-care-into-a-real-business-model [accessed 2020-02-17]
93. Annarumma C, Palumbo R. Contextualizing health literacy to health care organizations: exploratory insights. J Health Manag 2016;18(4):611-624. [doi: 10.1177/0972063416666348]
94. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Silva AP, Meira W, et al. Factors associated with weight change in online weight management communities: a case study in the LoseIt Reddit community. J Med Internet Res 2017 Jan 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]
95. Frow P, McColl-Kennedy JR, Payne A. Co-creation practices: their role in shaping a health care ecosystem. Ind Mark Manag 2016;56:24-39. [doi: 10.1016/j.indmarman.2016.03.007]

Abbreviations

IS: information systems
OHC: online health community
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by E Borycki; submitted 18.09.19; peer-reviewed by G Deckard, S Forouzandeh, P Naraei, N Sundar Rajan, P Balasooriya, N Hasan, S Xu; comments to author 30.11.19; revised version received 02.12.19; accepted 10.12.19; published 24.04.20.

Please cite as:
Rezaei Aghdam A, Watson J, Cliff C, Miah SJ
Improving the Theoretical Understanding Toward Patient-Driven Health Care Innovation Through Online Value Cocreation: Systematic Review
J Med Internet Res 2020;22(4):e16324
URL: http://www.jmir.org/2020/4/e16324/
doi:10.2196/16324
PMID:32329736

Review

Habit Strength, Medication Adherence, and Habit-Based Mobile Health Interventions Across Chronic Medical Conditions: Systematic Review

Sherif M Badawy[1,2*], MS, MD, MBBCh; Richa Shah[3*]; Usman Beg[4], BS; Mallorie B Heneghan[1,2], MD

[1]Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]Division of Hematology, Oncology and Stem Cell Transplant, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[3]Weinberg College of Arts and Sciences, Northwestern University, Evanston, IL, United States

[4]Midwestern University Arizona College of Osteopathic Medicine, Glendale, AZ, United States

[*]these authors contributed equally

Corresponding Author:
Sherif M Badawy, MS, MD, MBBCh
Department of Pediatrics
Northwestern University Feinberg School of Medicine
225 E Chicago Avenue, Box #30
Chicago, IL, 60654
United States
Phone: 1 3122274836
Email: sbadawy@luriechildrens.org

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2024/1/e63980

Abstract

Background: Unintentional medication nonadherence is common and has been associated with poor health outcomes and increased health care costs. Earlier research demonstrated a relationship between habit strength and medication adherence. Previous research also examined a habit's direct effect on adherence and how habit interacts with more conscious factors to influence or overrule them. However, the relationship between habit and adherence and the role of habit-based mobile health (mHealth) interventions remain unclear.

Objective: This review aimed to systematically evaluate the most recent evidence for habit strength, medication adherence, and habit-based mHealth interventions across chronic medical conditions.

Methods: A keyword search with combinations of the terms *habit*, *habit strength*, *habit index*, *medication adherence*, and *medication compliance* was conducted on the PubMed database. After duplicates were removed, two authors conducted independent abstract and full-text screening. The guidelines for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were followed when reporting evidence across the included and reviewed studies.

Results: Of the 687 records examined, 11 met the predefined inclusion criteria and were finalized for data extraction, grading, and synthesis. Most included studies (6/11, 55%) were cross-sectional and used a theoretical model (8/11, 73%). The majority of studies measured habit strength using the self-report habit index and self-report behavioral automaticity index (9/11, 82%). Habit strength was positively correlated with medication adherence in most studies (10/11, 91%). Habit mediated the effects of self-efficacy on medication adherence (1/11, 9%), and social norms moderated the effects of habit strength on medication adherence (1/11, 9%). Habit strength also moderated the effects of poor mental health symptoms and medication adherence (1/11, 9%). None of the included studies reported on using or proposing a habit-based mHealth behavioral intervention to promote medication adherence.

Conclusions: Habit strength was strongly correlated with medication adherence, and stronger habit was associated with higher medication adherence rates, regardless of the theoretical model and/or guiding framework. Habit-based interventions should be

used to increase medication adherence, and these interventions could leverage widely available mobile technology tools such as mobile apps or text messaging, and existing routines.

(J Med Internet Res 2020;22(4):e17883) doi:10.2196/17883

KEYWORDS

habit strength; medication adherence; habit index; medication compliance; mobile health; health; digital health; interventions; mobile phone

Introduction

Background

Medication adherence is defined as taking medication exactly as prescribed [1]; this includes taking the proper dose at the right time. Medication adherence comprises three components: initiation, implementation, and persistence [2]. Medication nonadherence can occur at any of these three stages because of the failure to initiate a new prescription, implement it as prescribed, or persist with treatment [2]. Medication adherence is not a dichotomous variable (ie, adherence vs nonadherence) [2] but is more of a continuum (ie, variable levels of adherence). Lower adherence and variations in adherence can lead to loss of drug effectiveness, toxicity, and drug resistance [3]. Only approximately 50% of medications are taken as recommended in different patient populations [4-6], including children with chronic conditions [7,8]. The costs of low medication adherence are both personal and economic. In the United States, this has been shown as a cycle where poor medication adherence leads to poor patient outcomes [9-15] and increased service utilization and health care costs [9,10,12,16], all of which are passed down to the patient, further affecting adherence [17]. The Institute for Healthcare Informatics identified US $500 billion in savings across 186 countries with the responsible use of medication and noted that about 8% of the global total health expenditure could be avoided by improving adherence to medication [2].

Habit is the context-dependent automatic completion of a behavior [18]. Medication adherence would be an example of such a behavior where patients may take the same number of pills in the same room at the same time of day. Therefore, high habit strength is the result of recurring contextual cues [19]. As habit is automatic, it works independently of, and can even override, conscious desires when strong enough [19]. There are 2 types of medication nonadherence: intentional and unintentional [20]. Forgetfulness is the number one cause of unintentional nonadherence [21]. As habit is independent of conscious cognitive processes, having high habit strength protects against forgetfulness. Earlier research demonstrated a relationship between the strength of habit and medication adherence. Previous research examined habit's direct effect on adherence and how habit interacts with more conscious factors to influence or overrule them [22]. However, the relationship between habit and adherence remains unclear.

Access to personal and mobile technology is ubiquitous [23-25], and there has been strong evidence to support the efficacy of digital or mobile health (mHealth) behavioral interventions, in particular text messaging and apps as tools to improve medication adherence [26-34]. These findings make mHealth interventions an appealing approach to optimize habit formation and medication adherence behavior in pediatric and adult patients with chronic health conditions [35]. However, the cost-effectiveness of these interventions remains unclear [36,37].

Objective

This review aimed to systematically evaluate the most recent evidence for habit strength, medication adherence, and habit-based mHealth interventions across chronic medical conditions.

Methods

Study Design

The guidelines for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were followed in the reporting of evidence across the studies reviewed herein [38]. The PRISMA checklist is included in Multimedia Appendix 1. To conduct this systematic review, a literature search was conducted on the PubMed database on June 25, 2019. Search terms were used in various combinations, including the following keywords: habit, habit strength, habit index, medication adherence, and medication compliance. For the first round of screening, 2 independent reviewers (RS and UB) conducted the keyword search and removed duplicates. Both the reviewers (RS and UB) then screened titles and abstracts independently for eligibility criteria and removed those that did not meet our inclusion criteria. Full texts were retrieved for the studies that were agreed on, and the 2 reviewers (RS and UB) completed full-text screening independently against our eligibility criteria. After conducting both screening steps, the results were compared, and any disagreements were settled by discussion with a third senior reviewer (SB).

Eligibility Criteria

Eligible studies were original research studies in English and included validated quantitative measures of habit strength and medication adherence that have been used in earlier published studies. Studies examining all ages, conditions, and countries were included. The included studies needed to evaluate habits specifically in the context of taking medication. This excluded lifestyle habits and general habit formation such as smoking, diet, and exercise. Studies that looked at adding medication to preexisting habits were also excluded. We excluded studies that evaluated habit strength and medication adherence solely from qualitative interviews without any validated measures.

Data Synthesis

A standardized form was used for data extraction. This form included the following categories: title, author, year of publication, country, number of participants, age, gender, study

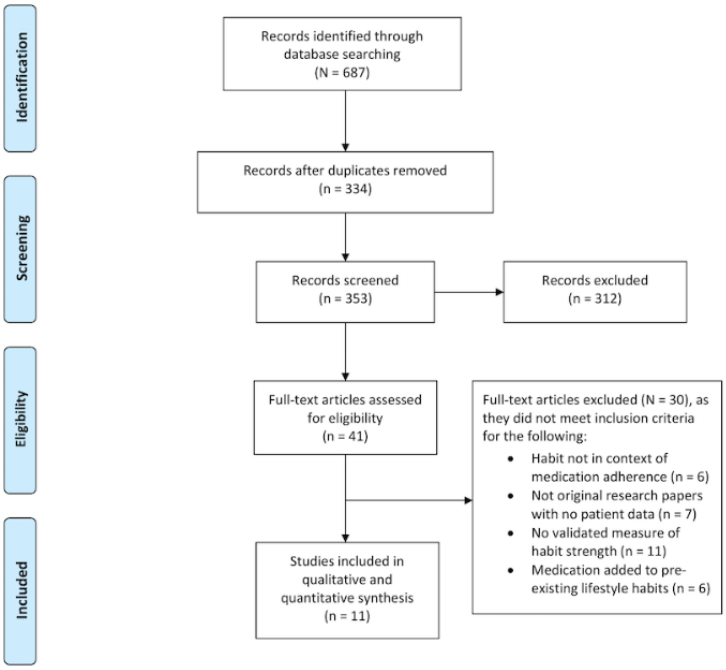
design, study approach, theoretical model, medical condition, habit strength instrument, adherence scale, measured habit strength, measured adherence rates, habit strength and medication adherence relationship (quantitative), main study conclusion, other study outcomes, and quality of the evidence. To assess the quality of the included studies, the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) criteria were used [39]. The GRADE approach evaluates a body of evidence by starting with a quality level based on the underlying methodology and then upgrading or downgrading the quality level based on various factors. Randomized trials or double-upgraded observational studies were rated as high. Downgraded randomized trials or upgraded observational studies were rated as moderate. Double-downgraded randomized trials or observational studies were rated as low. Triple-downgraded randomized trials and downgraded observational studies or case series/case reports were rated very low. Factors that downgrade the quality of evidence include limitations that suggest bias, indirectness of evidence, unexplained heterogeneity or inconsistency of results, imprecision of results, or a high probability of publication bias. Factors that improve the quality of evidence include a large magnitude of effect (ie, when all plausible confounding factors reduce a demonstrated effect or suggest a spurious effect when results show no effect) and dose-response gradient [38]. Data were analyzed and summarized qualitatively.

Results

Literature Search

Our literature search identified 687 studies for screening (title and abstract). Of these, 41 full-text studies were reviewed, and 11 studies [9-12,20,22,40-44] met all inclusion criteria. This process is outlined in the PRISMA flow chart (Figure 1).

Figure 1. Flow of studies through the review according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.



Description of Included Studies

Study Characteristics

Table 1 summarizes the study characteristics. The research from the included studies was conducted across a range of countries, including 1 study from the Netherlands [22], 2 from Canada [10,20], 2 from Ireland [40,41], 3 from the United Kingdom [11,43], and 3 from the United States [9,42,44]. All the included studies were published over the past 9 years, with the oldest published in 2011 [22] and the most recent in 2019 [11]. All studies included studies on habit strength and medication adherence in a specific chronic disease population including asthma [22], type 2 diabetes [10,20,42], hypertension [9,40,44], cystic fibrosis [11,12], and psoriasis [43], except for 1 study that looked at a population taking oral contraceptives [41]. All studies included adult subjects, but the participants' mean age ranged greatly from 22.41 to 69.86 years. The number of participants varied as well; the included studies ranged from 61 to 901 participants, with a mean of 331.82 participants and a median of 202 participants. The majority of the studies had a roughly equal ratio of male to female participants. One study observed a veteran population [44] with a mean age of 64.1 years, and only 14% women were included in the study [44]. Owing to the nature of the population, a study on oral contraceptives had a 100% female population [41]. Most studies (n=6) were cross-sectional [10,20,22,40,43], 4 were longitudinal studies [9,12,42,44], and 1 was a pilot randomized control trial [11]. In the included studies, medication adherence was measured using a combination of self-report questionnaires, remote monitoring using electronic pill bottles, and in-person interviews. The majority of studies (9/11, 82%) measured habit strength using the self-report habit index and self-report behavioral automaticity index [9-11,20,22,40-43].

Table 1. Summary of included studies that evaluated habit strength and medication adherence.

Source (country)	Health condition	Participants (N)	Age (years), mean (SD)	Sex (female), n (%)	Study design	Theoretical model	Study assessments	Quality of evidence[a]
Bolman et al [16] (Netherlands)	Asthma	139	31.5 (5.60)	99 (71)	Cross-sectional study	ASE[b] model	Mail-out survey, questionnaire	Low
Burns et al [14] (Canada)	Type 2 diabetes	790	64.05 (8.20)	387 (49)	Cross-sectional study	—[c]	Telephone interview, questionnaire	Moderate
Durand et al [19] (Ireland)	Hypertension	204	69.86 (10.69)	86 (42)	Cross-sectional study	CS-SRM[d]	Questionnaire	Low
Guenette et al [8] (Canada)	Type 2 diabetes	901	62.70 (9.10)	369 (41)	Cross-sectional study	TPB[e]	Questionnaire	Very low
Hoo et al [10] (United Kingdom)	Cystic fibrosis	123	25.00[f] (19-31)	52 (42)	Longitudinal study	Habit index measure	Electronic pill bottle	Very low
Hoo et al [9] (United Kingdom)	Cystic fibrosis	61	27.40 (21.70-37.10) - low adherence, 23.70 (18.40-32.00) - moderate adherence, and 26.10 (21.20-37.50) - high adherence[f]	28 (46)	Pilot randomized control trial	COM-B[g] model	Questionnaire, electronic pill bottle	Very low
Murphy et al [20] (Ireland)	Oral contraceptive pill	245	22.41 (4.78)	245 (100)	Cross-sectional study	—	Questionnaire	Very low
Phillips et al [7] (United States)	Hypertension	71	67.9 (12.28)	45 (63)	Longitudinal study	CS-SRM	Interview, MEMS[h]	Low
Phillips et al [21] (United States)	Type 2 diabetes	103	56.96 (12.94)	64 (62)	Longitudinal study	CS-SRM	Interview, electronic pill bottle, Fitbit, survey	Very low
Thorneloe et al [22] (United Kingdom)	Psoriasis	811	48.10 (13.10)	349 (43)	Cross-sectional cohort study	CS-SRM	Questionnaire	Moderate
Voils et al [23] (United States)	Hypertension	202	64.10 (11.00)	28 (14)	Longitudinal study	—	Survey	Very low

[a]Quality of evidence assessed using the Grades of Recommendation, Assessment, Development, and Evaluation criteria.

[b]ASE: attitude, social influence, and self-efficacy model.

[c]Missing data were not reported in the included studies.

[d]CS-SRM: common sense model of self-regulation.

[e]TPB: theory of planned behavior.

[f]Median age (years) is reported when the mean age was not provided in the included studies. IQR in parenthesis.

[g]COM-B: capability, opportunity, motivation, and behavior.

[h]MEMS: medication event monitoring system.

Description of Guiding Models

Different behavioral models exist to explain the process that occurs before a behavior takes place. In the context of this systematic review, the behavior being studied is medication adherence. Most studies (8/11, 73%) included theoretical models that comprised the guiding framework [9-12,22,40,42,43]. A variety of theoretical models were used by the included studies: the attitude, social influence, and self-efficacy model (ASE) [22]; the common sense model of self-regulation (CS-SRM) [9,40,42,43]; the theory of planned behavior (TPB) [10]; and the capability, opportunity, motivation, and behavior model (COM-B) [11].

The ASE is a behavioral explanatory model that takes a look at attitude, social influence, and self-efficacy as the predictors of intention and behavioral change [22]. Adapting the model to medication adherence, attitude is referred to as “the perceived pros and cons of taking medication,” social influence included “perceived norms and support of important others toward medication adherence and modeling which is the perceived behavior of others,” and self-efficacy was defined as “the

person's belief that they could adhere to medicine" [22]. Habit was observed as either a moderating or mediating factor in this model.

The CS-SRM proposes that an individual has a certain representation of an illness in their mind that guides how they respond to an illness threat [45]. The components of the illness representation are identity, causes, consequences, timeline, and controllability [40]. In the context of this systematic review, the response to illness threat would be medication adherence. According to the CS-SRM model, treatment-favorable beliefs lead to the initiation of behavior, and experiential feedback proves that those beliefs were correct (CS-SRM coherence), and the behavior was practiced until it became habit, leading to long-term medication adherence [9].

Fundamentally, the TPB states that attitudes, subjective norms, and perceived behavioral control create the intention to perform behaviors such as medication adherence [10]. When an individual is given sufficient control over their situation, intention will be turned into behavior when given the opportunity to do so [10]. The included study that used the TPB as a guiding model [10] examined habit as another factor that influences intention and the performance of medication adherence behavior.

The COM-B is a behavior system in which capability, opportunity, and motivation interact and lead to behavior [46]. The behavior itself, then, influences capability, opportunity, and motivation. Capability includes knowledge, cognitive ability, and physical skills to perform a behavior. Opportunity is defined as factors that lie outside the patient's environment that make taking medications possible or prompt them to do so according to the cultural milieu, including the access to medications and medical care as well as the perceptions related to disease stigma. In other words, opportunity includes any factor that is not in the hands of the individual. Motivation energizes and directs behavior. Goals, conscious decision-making, habitual processes, emotional responding, and analytical decision-making are all components of motivation. In the context of this systematic review, habit is part of the automatic motivation that energizes and directs the behavior of medication adherence.

Description of Habitat Strength Measures

Table 2 summarizes the habit strength and adherence measure as well as the main outcomes. The majority of studies (9/11, 82%) measured habit strength using the self-report habit index and self-report behavioral automaticity index [9-11,20,22,40-43]. The self-report behavioral automaticity index is a subset of the self-report habit index, and an example item from this index would be taking this medication is something I do automatically, which is rated on a scale of 1-5 from strongly disagree=1 to strongly agree=5 [21]. One study [9] modified this scale by adding 4 additional questions that evaluated the concept of habit strength more broadly and intuitively (ie, asking patients in different ways whether or not they have a habit of taking their medication vs asking them in different ways if they take their medication without conscious attention, without conscious awareness, etc). One of the studies [12] built and tested a new form of measuring habit strength, using the habit index scale as its guiding model. A total of 2 studies [12,44] used the multiplicative product of behavior frequency and context stability.

Table 2. Summary of habit strength, medication adherence measures, and outcomes in the included studies.

Source	Habit strength measure	Adherence scale and rates	Relationship between habit strength and adherence rates
Bolman et al [16]	• SRHI[a]	• MARS[b]	• Correlation r=0.61; P<.001
Burns et al [14]	• Self-report behavioral automaticity index	• *Did you ever forget to take your medication?* on a 5-point scale	• Depressive symptoms: beta=.08; P<.001; 95% CI 0.04 to 0.12 • Diabetes distress: beta=.09; P<.001; 95% CI 0.04 to 0.12 • Major depressive syndrome: beta=.07; P<.001; 95% CI 0.03 to 0.11
Durand et al [19]	• Self-report behavioral automaticity index	• Overall adherent range: 58.9%-79.7% • MARS: 36.7% nonadherent • MMAS[c]: 41.1% nonadherent • Prescription refill: 79.7% adherent • Urine assay • Total nonadherence, 2.1% • Partial nonadherence, 23.8%	• MARS: correlation r=0.36[d]; P<.001 • MMAS: correlation r=0.35[d]; P<.001 • Prescription refill: correlation r=0.08 • Urine assay: correlation r=−0.02 • Adherence composite: correlation r=0.36; P<.001 • Hierarchical regression analysis: beta=.44; P<.001; adjusted. R^2=0.22, ΔR^2=0.19; ; P<.001 • Unintentional adherence: beta=−.45; t_{203}=−7.04; P<.001 • Intentional adherence: beta=−.22; t_{203}=−3.08; P<.01
Guenette et al [8]	• SRHI • About 71% scoring *high* (at least 5/6) • Mea	• MMAS-8 modified French version • 45% high adherence • 40.7% medium adherence • 14.3% low adherence	• Adjusted OR[e] 1.65; 95% CI 1.35 to 2.03; P<.001
Hoo[a] et al [10]	• Multiplicative product of behavior frequency and context stability	• Electronic pill bottles • 47.30% median adherence • 4.9% low adherence • 80.5% variable adherence • 14.6% high adherence	• Overall cohort: R=0.40; 95% CI 0.36 to 0.44; beta=.30; 95% CI −1.04 to 1.65 • Adherence consistently low: R=0.24; 95% CI 0.04 to 0.44; beta=3.03; 95% CI −9.68 to 15.76 • Variable adherence: R=0.45; 95% CI 0.41 to 0.49; beta=.08; 95% CI −1.44 to 1.60 • Adherence consistently high: R=0.20; 95% CI 0.13 to 0.27; beta=.61; 95% CI −1.90 to 3.13

Hoo[b] et al [9]	• Self-report behavioral automaticity index	• Chipped nebulizer
		• 75.4% low adherence
		• 13.1% medium adherence
		• 11.5% high adherence
			• Median habit strength in different subgroups:
			  • Low adherence: 9.0, IQR 4.8-12.0
			  • Moderate adherence: 14.5, IQR 11.3-18.3
			  • High adherence: 18.0, IQR 14.0-20.0
			  • All significantly correlated with adherence levels, P<.001

Murphy et al [20]	• Self-report behavioral automaticity index • Mean habit strength per number OCP[f] missed per month	• MARS • Mean MARS score per number of OCP missed per month: • Never: 5.85 • Once: 7.49 • Twice or more: 10.12	• Correlation r=−0.24[g]; P<.001

Source	Habit strength measure	Adherence scale and rates	Relationship between habit strength and adherence rates

Phillips et al [7]	• Self-report habit index, with 4 additional questions	• MARS, MMAS, MEMS[h]
		• Mean adherence
		• MMAS=0.80
		• MEMS timing adherence=76%
		• MEMS dosing adherence=96%
			• Bivariate relationship (correlations):
			  • MARS: 0.37
			  • MMAS: 0.26
			  • MEMS dose frequency: 0.42
			  • MEMS dose timing: 0.49
			• Hierarchical regression analysis:
			  • MARS: ΔR^2=0.11; *P*<.01
			  • MMAS: ΔR^2=0.06; *P*=.04
			  • MEMS frequency: ΔR^2=0.17; *P*<.001
			  • MEMS timing: ΔR^2=0.27; *P*<.001
			• Unintentional nonadherence: beta=−.32; t_{66}=−2.55; *P*=.01
			• Intentional nonadherence: beta=−.23; t_{66}=−1.82; *P*<.07

Phillips et al [21]	• Self-report behavioral automaticity index
	• Mean medication-taking habit strength 3.75
		• MARS and MEMS
		• Mean adherence:
		• MARS=4.66
		• Self-reported intentional nonadherence=1.24
		• Self-reported unintentional nonadherence=1.76
		• MEMS % days adherent=76.19
		• MEMS % doses on time=60.68
			• Bivariate correlations:
			  • MARS: 0.40, *P*<.001
			  • Self-reported intentional nonadherence: −0.34; *P*<.001
			  • Self-reported unintentional nonadherence: −0.41; *P*<.001
			  • MEMS % days adherent: 0.37; *P*<.001
			  • MEMS % doses on time: 0.40; *P*<.001
			• MARS (with control variables): beta=0.15; β=.32; *P*<.001
			• MEMS (with control variables): beta=8.57; β=.32; *P*<.01

Thorneloe et al [22]	• Self-report habit index
	• Mean 41.5 for self-administered systemic therapy
		• MARS
		• Overall:
		• 22.4%, nonadherent
		• 12% intentional
		• 10.9% unintentional
		• Conventional:
		• 29.2% overall
		• 15.3% intentional
		• 14.5% unintentional
		• Biologic:
		• 16.4% overall
		• 9.1% intentional
		• 7.7% unintentional
			• Multivariable regression model:
			  • 0.94 overall nonadherence: 95% CI 0.91 to 0.97
			  • 0.95 intentional nonadherence: 95% CI 0.92 to 0.98
			  • 0.92 unintentional nonadherence: 95% CI 0.89 to 0.96

Voils et al [23]	• Product of frequency and mean of 5 situational consistency items	• Patient rating and MMAS-8 • 60% nonadherent	• Extent of nonadherence: correlation *r*=−0.39; *P*<.001

[a]SRHI: self-report habit index.

[b]MARS: medication adherence report scale.

[c]MMAS: Morisky medication adherence scale.

[d]Z-scores averaged. So greater MARS and MMAS scores represented greater nonadherence.

[e]OR: odds ratio.

[f]OCP: oral contraceptive pill.

[g]Negative because lower MARS score represents better adherence.

[h]MEMS: medication event monitoring system.

Description of Adherence Measures

The most common measures of adherence were the medication adherence report scale [9,22,40-43] and the Morisky medication adherence scale [9,10,40,44]. Other measures of adherence included medication event monitoring systems [7,9,21], medication possession ratio based on prescription refill data [19], single item self-report [14], and urine drug monitoring [14].

Study Methodological Quality

The quality of the included studies ranged from very low to moderate. Of the 11 included studies, 2 were of moderate quality [20,43], 3 were of low quality [9,22,40], and 6 were of very low quality [10-12,41,44]. Table 1 reports the quality of each included study.

Description of Study Outcomes

Habit Strength, Medication Adherence, and Mobile Health Interventions

Table 3 summarizes the study outcomes related to habit strength and medication adherence. Most studies showed a positive correlation between habit strength and medication adherence behavior, suggesting stronger habit formation with higher medication adherence rates [9-12,22,40-44]. Furthermore, compared with factors such as pill burden [40], illness coherence [40], treatment-related beliefs [9,40], and experiences with treatment-related efficacy [9], habit strength was the strongest predictor of adherence. Habit strength had the strongest association with medication adherence and medication event monitoring system dose timing among all the other adherence measures in 1 study [9]. However, habit strength was found to be equally correlated to dose timing and days taken in a later study [42].

Habit strength was also found to mediate the effects of self-efficacy on adherence [22]. The effect of self-efficacy on adherence disappeared once habit strength was added to the hierarchical multiple regression analysis model, and this relationship was confirmed with bootstrapping analysis [22]. Social norms moderated the relationship between habit strength and medication adherence; in weak habit, a supportive norm of taking medicine was positively related to adherence, and supportive norms were only positively correlated with adherence when habit strength score was low [22].

Even after adjusting for covariates, such as age and disease duration, habit strength moderated the association between poor mental health symptoms and medication adherence [20]. Interaction between habit strength and depressive symptoms was also observed. When habit strength was weak or average, depressive symptoms were negatively associated with adherence [20]. However, if habit was strong, no association was observed [20]. This same interaction was observed between diabetes distress and habit strength as well as between major depressive syndrome and habit strength [20].

Habit strength was more strongly associated with unintentional nonadherence than intentional nonadherence in 2 studies [9,40] but was equally predictive in another study [42]. None of the included studies reported on using or proposing a habit-based mHealth behavioral intervention to promote medication adherence.

Table 3. Summary of the main study findings.

Source	Study outcomes
Bolman et al [16]	• Higher habit strength is positively correlated with higher adherence. • Habit mediates the relationship between self-efficacy and medication adherence. • Social norms moderate the relationship between habit and adherence; in weak habit, a supportive norm of taking medicine was positively related to adherence, and in strong habit, supportive norm correlated with less adherence. • Perceiving few negative consequences of taking medicine was associated with better adherence. • Control variables of risk perception and asthma severity were positively correlated with adherence. • Female gender was positively correlated with adherence. • Control variable of internal locus of control negatively correlated with adherence. • From the central concepts, perceiving more pros, social support, higher self-efficacy, and stronger habit was associated with more adherence. • From the central concepts, habit strength and attitude pros had the strongest correlation with medication adherence. • Social norm and modeling were not significantly associated with adherence. • Social influence subscales were highly intercorrelated, as well as habit with risk perception, pros, social support, and self-efficacy. • After hierarchical multiple regression, habit strength proved to be significantly related to adherence. Of the control variables, only severity remained significant; of the ASE[a] concepts, only the cons remained significant.
Burns et al [14]	• Interaction between habit strength and depressive symptoms was observed. If habit strength was weak or average, depressive symptoms were negatively associated with adherence. However, if habit was strong, no association was observed. • Same significant interaction pattern was observed for diabetes distress and habit strength as well as major depressive syndrome and habit strength. • Habit strength moderates the association between poor mental health symptoms and medication adherence. • After adjusting for covariates, results remained significant.
Durand et al [19]	• Medication-taking habit strength was the strongest predictor of adherence (compared with pill burden, illness coherence, and treatment-related beliefs). • Habit strength explained 19% incremental variance in adherence beyond treatment-related beliefs. • Habit strength was more strongly associated with unintentional nonadherence than intentional. • Associations among adherence measures were weak to moderate, indicating that multiple measures are necessary to accurately assess adherence. • Neither treatment-related beliefs nor CSM[b] coherence predicted adherence, even for patients with weak habit strength. • Pill burden was not associated with habit strength or adherence. • There was no significant interaction between treatment-related beliefs, habit strength, and adherence.
Guenette et al [8]	• Strong habit was significantly associated with adherence. • Perceived behavioral control, older age, no perceived side effects, a longer period since T2D[c] diagnosis, and a lower number of NAID[d] daily doses were significantly associated with adherence. • Sex, level of education, and income are not associated with adherence. • Intention, insulin use, number and type of NIAD drugs prescribed, perceived cost of antidiabetes medications, and use of glucometer or weekly pill organizer were not associated with NIAD adherence. • Depressed mood, anxiety, and mental health were not associated with adherence. • Behavioral control was found to be significant, so the 26 underlying beliefs were analyzed, and 12 beliefs were found to be significant with adherence.
Hoo[a] et al [10]	• One unit increase in habit index was associated with a 0.3% increase in the subsequent week's adherence after controlling for current adherence. • Those with variable adherence displayed higher mean cross-correlation coefficients (0.45) compared with those with consistent adherence (0.20-0.40).
Hoo[b] et al [9]	• Higher adherers reported stronger habit compared with lower adherers. • A 1-unit increase in habit strength was associated with a 31% increase in odds of being in the next higher adherence category. • In a multiple ordinal regression model with both habit and concerns scores, only habit was associated with adherence. • Higher adherers had lower prior year intravenous use, tended to have higher %FEV[e] at baseline, and reported lower concerns.
Murphy et al [20]	• Stronger habit strength was associated with better adherence. • Those who never miss an OCP[f] reported significantly higher habit strength than those who miss 2 or more per month. • There was no difference between those who never miss an OCP and those who miss 1 OCP per month. • Having a fixed time of day to take the OCP was associated with better habit strength and adherence. • There is, however, no association between habit strength and taking OCP at different times of the day. • Having a fixed place to store the OCP was associated with habit strength but not adherence.

Source	Study outcomes
Phillips et al [7]	• Habit strength was the strongest predictor of medication adherence (compared with beliefs and experiences plus efficacy)—explains 6%-27% incremental variance in adherence to that explained by treatment-related beliefs. • Habit strength was more strongly related to unintentional medication nonadherence than intentional nonadherence. • Patients' CS-SRM[g] coherence was more strongly associated with intentional nonadherence than unintentional adherence. • Patients' treatment-related beliefs were not more strongly associated with intentional nonadherence than unintentional nonadherence. • Habit strength had the strongest association with MEMS[h] dose timing out of all the adherence measures. • The interaction between treatment-related beliefs and habit was not significant for any of the adherence measures. • Patients' beliefs and experiences did not predict overall adherence, even for weaker adherence. Patient experience, however, did predict intentional nonadherence.
Phillips et al [21]	• Habit strength consistently predicted incremental variance in measured outcomes, both self-reported and measured. • Correlations, between habit strength and % of the doses taken on time vs between habit strength and % of the days when medications were taken, were not significantly different. • Habit strength does not predict unintentional nonadherence better than intentional. • Habit strength is not relatively more important for predicting medication adherence than physical activity.
Thorneloe et al [22]	• Patients in the biological cohort were more likely to be male, have a younger age of onset of psoriasis, longer duration of disease, more likely to have a diagnosis of inflammatory arthritis, have lower quality of life scores at the start of therapy, have longer duration of systemic therapy, have stronger beliefs in the chronicity of their illness, stronger beliefs that systemic therapy is necessary, weaker concerns about therapy and medicine, greater coherence, and less symptoms of depression. • Patients using self-administered systemic therapy had strong habit strength. • Being on a conventional systemic therapy, having strong medication concerns, longer treatment duration, and younger age were factors associated with overall nonadherence. • Being on a conventional therapy and strong medication concerns were also significant for intentional nonadherence. • Being on a conventional systemic therapy, stronger perceptions of psoriasis being a chronic condition, younger age, and longer treatment duration were factors associated with unintentional nonadherence. • Group 1 membership (strongest medication concerns) was associated with intentional nonadherence, and weaker medication-taking routine or habit strength was associated with unintentional nonadherence.
Voils et al [23]	• Dual conceptualization (self-report with psychometric principles) of medication nonadherence has stronger validity and reliability than other forms that confound these 2 variables. • Extent of adherence was highly correlated with self-efficacy, where lower adherence levels were associated with lower self-efficacy. • In all, 3 items assessing the extent of nonadherence produced reliable scores. • Correlations between the extent and harm subscales with habit strength were above 0.3. • Correlations and comparison measures showed convergent and divergent validity. • Predictive validity was evidenced by correlations between extent and BP[i]. • Means of the reasons items were well below the scale midpoint, and several distributions were positively skewed and kurtosis. The Morisky scale did not measure a single underlying construct in this sample. • The Morisky score was not correlated with BP.

[a]ASE: attitude, social influence, and self-efficacy model.

[b]CSM: common sense model.

[c]T2D: type 2 diabetes.

[d]NAID: noninsulin antidiabetic drugs.

[e]FEV: forced expiratory volume.

[f]OCP: oral contraceptive pill.

[g]CS-SRM: common sense model of self-regulation.

[h]MEMS: medication event monitoring system.

[i]BP: blood pressure.

Other Adherence-Related Outcomes

In addition to habit strength, other factors significantly associated with increased medication adherence included perceived negative consequences of taking the medication [10,11,22,43], perceived behavioral control [10], and older age [10,43]. Longer treatment duration was found to be significantly associated with adherence in 2 different studies. One study [10] concluded that longer treatment duration led to higher adherence because a longer period led to the development of habit. In contrast, another study [43] concluded that longer treatment duration led to higher overall and unintentional nonadherence. Strong concerns were also associated with intentional nonadherence [43]. Stronger perceptions of having a chronic condition and younger age were also associated with unintentional nonadherence [43]. Disease severity also affected adherence behavior. For those taking asthma medications, asthma severity was positively correlated with medication

adherence [22]. Having a fixed time of the day for taking medications was associated with better habit strength and better adherence, but there was no association between habit strength and the time of the day medications were taken [41]. Having a fixed place to store the medication was associated with higher habit strength but not higher medication adherence [41].

In the included studies, factors found to not have a relationship with medication adherence were modeling [22], treatment-related beliefs [9,40], and pill burden [40]. The association of treatment-related beliefs was not stronger for intentional nonadherence than for unintentional nonadherence [9]. One study [40] found that common sense model–related coherence did not have a significant relationship with medication adherence; yet, another study [9] found that CS-SRM coherence was more strongly associated with intentional nonadherence than unintentional nonadherence. Demographics associated with adherence included sex [10,22], education level [10], and income [10]. Although social norms did moderate the relationship between habit strength and medication adherence, it did not have a significant relationship with medication adherence by itself [22]. One study [20] found that there was an interaction between habit strength and mental health symptoms; however, depressed mood, anxiety, and mental health by themselves were not associated with adherence [10]. Patient experiences did not predict overall nonadherence but did predict intentional nonadherence [9].

Discussion

Principal Findings

Our systematic review contributes to the literature on habit strength and medication adherence across chronic medication conditions. We found that habit strength was strongly correlated with medication adherence, with stronger habit being associated with higher medication adherence rates, regardless of the theoretical model and/or guiding framework. As the behavior becomes more automatic, there is less chance for an individual to forget to take their medicine. We also found that the effect of habit strength on adherence was also related to the individual's self-efficacy, social norms, and mental health symptoms. This has been explained in earlier studies investigating the dual-process theories, where an individual's behavior is a result of both deliberative/reflective processes and implicit/impulsive processes [47]. When conscious processes are strong, they might be able to overpower the automaticity of habits. It is also important to note that many of the social influence subscales, as well as habit, were highly intercorrelated in relation to risk perception, pros/cons of taking medication, social support, and self-efficacy [22].

In 1 study [40], it was determined that associations among adherence measures were weak to moderate. This indicates that multiple measures are necessary to accurately assess adherence, as was done in some of the included studies. In addition, self-report questionnaires run the risk of social desirability bias, so monitoring with electronic pill bottles or looking into prescription pharmacy refill records are other important and informative ways of measuring medication adherence.

Our findings in this review suggest that interventions to increase medication adherence could be more effective if they focused on developing a stronger habit among individuals. One way to build a stronger habit is by reminding the patient when it is time to take their medication [18], such as with pill bottle caps that light up when it is time to take the medication. Another approach to strengthen a patient's habit of taking medication is to leverage technology-based interventions and remind the patient to take their medication by sending a text message or alert on the patient's phone when it is time to take their dose. Interestingly, a recent review reported that a number of grants, funded by the US National Institutes of Health (2014-2018), were focused on developing and testing mHealth smartphone apps that were specifically designed to facilitate medication adherence behavior by reminding patients to take their scheduled medications, which could lead to the development of a habit [48]. Furthermore, a possible habit-based intervention to increase medication adherence is to incorporate the medication routine into existing lifestyle habits such as physical activity, mealtimes, or morning routines to develop a stronger habit, which should be explored further in future research. However, in this review, we focused our research question on the relationship between habit strength and adherence rates, solely in the context of medication-taking behavior. It is also important to note that most of the included studies in our review were of low quality, and the majority were observational studies, yet they are informative for the most recent evidence on habit strength and medication adherence.

Strengths

Our review has some strengths. In all, two authors independently completed the search process at each stage of the systematic review process, following established methodology guidelines (PRISMA). Some of the included studies used multiple forms of measure for medication adherence, making adherence assessment more accurate. Despite having no eligibility restrictions on the year of publication, all the included studies were published between 2011 and 2019, indicating an increasing interest in the topic of habit strength and medication adherence.

Limitations

It is important to note some of the limitations of this systematic review. Given that all the search results came from 1 database (ie, PubMed) during the literature search, it is possible that some relevant studies could have been missed during the process. However, most of the studies in other databases, such as PsycINFO, are also indexed in PubMed, and the chances of missing relevant studies are relatively less. Furthermore, many of the included studies used self-report questionnaires, an approach that has the inherent limitation of social desirability bias. In addition, most of the included studies were observational and cannot evaluate the direction or the cause-and-effect relationship between habit strength and medication adherence. Moreover, it is important to note that the range of countries represented in the included studies was limited and included only developed countries. This is important because lifestyle factors, prescribing practices, and social/cultural norms could be different in different countries, affecting both the development of habit strength and medication adherence behavior. Therefore, the inclusion of studies from only

developed countries limits the generalizability of this systematic review. Moreover, examining the relationship between habit strength and medication adherence should extend beyond developed countries where all the included studies were conducted. Developing countries have different clinical and research settings, and gaining insight from studies conducted there would be essential for future wide dissemination and implementation efforts of adherence-promoting behavioral interventions. Finally, this systematic review looked primarily at chronic health conditions, and further research should investigate the connection between habit strength and adherence behavior in nonchronic conditions. In addition, future research should also assess the longitudinal relationship between habit strength and medication adherence to better understand their cause-effect association, given that most of the included studies were cross-sectional.

Conclusions

In conclusion, stronger habit has been associated with higher medication adherence rates. This is consistent with published literature indicating that forgetfulness is the leading cause of unintentional medication nonadherence. All studies in the literature examined habit strength in the context of nonadherence. Future rigorous longitudinal studies are needed to examine the direction of the relationship between habit strength and medication adherence behavior.

Acknowledgments

We wish to acknowledge MMAS Research LLC for the use of ©Morisky Widget MMAS-8 software. Morisky Widget MMAS-8 software is protected by U.S. copyright and trademark laws. Permission for use of Morisky Widget MMAS-8 software copyright registration number TX 8-816-517 is required and was obtained for this research. A license agreement is available from MMAS Research LLC, 1501 Spring Hill Red, Petaluma CA, 94952 USA; strubow@morisky.org. We would like to express our gratitude to Steve Trubow for the generous training and the timely license provision.

This project was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (K23HL150232, principal investigator: Badawy). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.
[DOC File , 63 KB - jmir_v22i4e17883_app1.doc]

References

1. Vrijens B, de Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, Project Team. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
2. Vrijens B, Antoniou S, Burnier M, de la Sierra A, Volpe M. Current situation of medication adherence in hypertension. Front Pharmacol 2017;8:100 [FREE Full text] [doi: 10.3389/fphar.2017.00100] [Medline: 28298894]
3. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. Annu Rev Pharmacol Toxicol 2012;52:275-301. [doi: 10.1146/annurev-pharmtox-011711-113247] [Medline: 21942628]
4. Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2008 Apr 16(2):CD000011. [doi: 10.1002/14651858.CD000011.pub3] [Medline: 18425859]
5. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
6. Badawy SM, Thompson AA, Liem RI. Beliefs about hydroxyurea in youth with sickle cell disease. Hematol Oncol Stem Cell Ther 2018 Sep;11(3):142-148 [FREE Full text] [doi: 10.1016/j.hemonc.2018.01.001] [Medline: 29397333]
7. Gupta OT, Wiebe DJ, Pyatak EA, Beck AM. Improving medication adherence in the pediatric population using integrated care of companion animals. Patient Educ Couns 2018 Oct;101(10):1876-1878. [doi: 10.1016/j.pec.2018.05.015] [Medline: 29807672]
8. Badawy SM, Thompson AA, Penedo FJ, Lai J, Rychlik K, Liem RI. Barriers to hydroxyurea adherence and health-related quality of life in adolescents and young adults with sickle cell disease. Eur J Haematol 2017 Jun;98(6):608-614. [doi: 10.1111/ejh.12878] [Medline: 28306171]
9. Phillips LA, Leventhal H, Leventhal EA. Assessing theoretical predictors of long-term medication adherence: patients' treatment-related beliefs, experiential feedback and habit development. Psychol Health 2013;28(10):1135-1151. [doi: 10.1080/08870446.2013.793798] [Medline: 23627524]

10. Guénette L, Breton M, Guillaumie L, Lauzier S, Grégoire JP, Moisan J. Psychosocial factors associated with adherence to non-insulin antidiabetes treatments. J Diabetes Complications 2016 Mar;30(2):335-342. [doi: 10.1016/j.jdiacomp.2015.10.016] [Medline: 26615905]
11. Hoo ZH, Gardner B, Arden MA, Waterhouse S, Walters SJ, Campbell MJ, et al. Role of habit in treatment adherence among adults with cystic fibrosis. Thorax 2019 Feb;74(2):197-199. [doi: 10.1136/thoraxjnl-2017-211453] [Medline: 29886416]
12. Hoo ZH, Wildman MJ, Campbell MJ, Walters SJ, Gardner B. A pragmatic behavior-based habit index for adherence to nebulized treatments among adults with cystic fibrosis. Patient Prefer Adherence 2019;13:283-294 [FREE Full text] [doi: 10.2147/PPA.S186417] [Medline: 30863018]
13. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Health-related quality of life and adherence to hydroxyurea in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2017 Jun;64(6):e26369. [doi: 10.1002/pbc.26369] [Medline: 27896936]
14. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Adherence to hydroxyurea, health-related quality of life domains, and patients' perceptions of sickle cell disease and hydroxyurea: a cross-sectional study in adolescents and young adults. Health Qual Life Outcomes 2017 Jul 5;15(1):136 [FREE Full text] [doi: 10.1186/s12955-017-0713-x] [Medline: 28679417]
15. Badawy SM, Barrera L, Cai S, Thompson AA. Association between participants' characteristics, patient-reported outcomes, and clinical outcomes in youth with sickle cell disease. Biomed Res Int 2018;2018:8296139 [FREE Full text] [doi: 10.1155/2018/8296139] [Medline: 30105252]
16. Badawy SM, Thompson AA, Holl JL, Penedo FJ, Liem RI. Healthcare utilization and hydroxyurea adherence in youth with sickle cell disease. Pediatr Hematol Oncol 2018;35(5-6):297-308. [doi: 10.1080/08880018.2018.1505988] [Medline: 30636474]
17. Iuga AO, McGuire MJ. Adherence and health care costs. Risk Manag Healthc Policy 2014;7:35-44 [FREE Full text] [doi: 10.2147/RMHP.S19801] [Medline: 24591853]
18. Gardner B. A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev 2015;9(3):277-295 [FREE Full text] [doi: 10.1080/17437199.2013.876238] [Medline: 25207647]
19. Wood W, Rünger D. Psychology of habit. Annu Rev Psychol 2016;67:289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]
20. Burns RJ, Deschênes SS, Knäuper B, Schmitz N. Habit strength as a moderator of the association between symptoms of poor mental health and unintentional non-adherence to oral hypoglycemic medication in adults with type 2 diabetes. J Health Psychol 2019 Mar;24(3):321-326. [doi: 10.1177/1359105316674269] [Medline: 28810360]
21. Sawyer SM, Fardy HJ. Bridging the gap between doctors' and patients' expectations of asthma management. J Asthma 2003 Apr;40(2):131-138. [doi: 10.1081/jas-120017983] [Medline: 12765314]
22. Bolman C, Arwert TG, Völlink T. Adherence to prophylactic asthma medication: habit strength and cognitions. Heart Lung 2011;40(1):63-75. [doi: 10.1016/j.hrtlng.2010.02.003] [Medline: 20561874]
23. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
24. Andersen M, Jiang J. Pew Research Center. 2018. Teens, Social Media & Technology URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-01-17]
25. Pew Research Center. 2019. Mobile Fact Sheet URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-01-17]
26. 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]
27. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered ehealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
28. 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]
29. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mhealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
30. 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 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
31. 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]

32. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
33. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
34. Britto MT, Rohan JM, Dodds CM, Byczkowski TL. A randomized trial of user-controlled text messaging to improve asthma outcomes: a pilot study. Clin Pediatr (Phila) 2017 Dec;56(14):1336-1344. [doi: 10.1177/0009922816684857] [Medline: 28056541]
35. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
36. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mhealth? A systematic review of economic evaluations of mhealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
37. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
38. 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]
39. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
40. Durand H, Hayes P, Harhen B, Conneely A, Finn DP, Casey M, et al. Medication adherence for resistant hypertension: assessing theoretical predictors of adherence using direct and indirect adherence measures. Br J Health Psychol 2018 Nov;23(4):949-966. [doi: 10.1111/bjhp.12332] [Medline: 30014548]
41. Murphy J, Eustace N, Sarma KM, Molloy GJ. Habit strength and adherence to oral contraceptives: the role of time- and place-based cues. Int J Behav Med 2018 Aug;25(4):431-437. [doi: 10.1007/s12529-018-9729-9] [Medline: 29856008]
42. Phillips LA, Cohen J, Burns E, Abrams J, Renninger S. Self-management of chronic illness: the role of 'habit' versus reflective factors in exercise and medication adherence. J Behav Med 2016 Dec;39(6):1076-1091. [doi: 10.1007/s10865-016-9732-z] [Medline: 26980098]
43. Thorneloe RJ, Griffiths CE, Emsley R, Ashcroft DM, Cordingley L, British Association of Dermatologists Biologic Interventions Register, Psoriasis Stratification to Optimise Relevant Therapy Study Groups. Intentional and unintentional medication non-adherence in psoriasis: the role of patients' medication beliefs and habit strength. J Invest Dermatol 2018 Apr;138(4):785-794 [FREE Full text] [doi: 10.1016/j.jid.2017.11.015] [Medline: 29183731]
44. Voils CI, Maciejewski ML, Hoyle RH, Reeve BB, Gallagher P, Bryson CL, et al. Initial validation of a self-report measure of the extent of and reasons for medication nonadherence. Med Care 2012 Dec;50(12):1013-1019 [FREE Full text] [doi: 10.1097/MLR.0b013e318269e121] [Medline: 22922431]
45. Hagger MS, Koch S, Chatzisarantis NL, Orbell S. The common sense model of self-regulation: meta-analysis and test of a process model. Psychol Bull 2017 Nov;143(11):1117-1154. [doi: 10.1037/bul0000118] [Medline: 28805401]
46. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
47. Hagger MS. Non-conscious processes and dual-process theories in health psychology. Health Psychol Rev 2016 Dec;10(4):375-380. [doi: 10.1080/17437199.2016.1244647] [Medline: 27718880]
48. Hansen WB, Scheier LM. Specialized smartphone intervention apps: review of 2014 to 2018 NIH funded grants. JMIR Mhealth Uhealth 2019 Jul 29;7(7):e14655 [FREE Full text] [doi: 10.2196/14655] [Medline: 31359866]

Abbreviations

ASE: attitude, social influence, and self-efficacy model
COM-B: capability, opportunity, motivation, and behavior model
CS-SRM: common sense model of self-regulation
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TPB: theory of planned behavior

Edited by G Eysenbach; submitted 23.01.20; peer-reviewed by GE Iyawa, W Hansen, Y Virella Pérez; comments to author 13.02.20; revised version received 02.03.20; accepted 25.03.20; published 28.04.20.

Please cite as:
Badawy SM, Shah R, Beg U, Heneghan MB
Habit Strength, Medication Adherence, and Habit-Based Mobile Health Interventions Across Chronic Medical Conditions: Systematic Review
J Med Internet Res 2020;22(4):e17883
URL: http://www.jmir.org/2020/4/e17883/
doi:10.2196/17883
PMID:32343250

Review

Infodemiology and Infoveillance: Scoping Review

Amaryllis Mavragani[1], BSc, MSc

Department of Computing Science and Mathematics, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom

Corresponding Author:
Amaryllis Mavragani, BSc, MSc
Department of Computing Science and Mathematics
Faculty of Natural Sciences
University of Stirling
University Campus
Stirling, FK9 4LA
United Kingdom
Phone: 44 7523 782711
Email: amaryllis.mavragani1@stir.ac.uk

Abstract

Background: Web-based sources are increasingly employed in the analysis, detection, and forecasting of diseases and epidemics, and in predicting human behavior toward several health topics. This use of the internet has come to be known as infodemiology, a concept introduced by Gunther Eysenbach. Infodemiology and infoveillance studies use web-based data and have become an integral part of health informatics research over the past decade.

Objective: The aim of this paper is to provide a scoping review of the state-of-the-art in infodemiology along with the background and history of the concept, to identify sources and health categories and topics, to elaborate on the validity of the employed methods, and to discuss the gaps identified in current research.

Methods: The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed to extract the publications that fall under the umbrella of infodemiology and infoveillance from the JMIR, PubMed, and Scopus databases. A total of 338 documents were extracted for assessment.

Results: Of the 338 studies, the vast majority (n=282, 83.4%) were published with JMIR Publications. The Journal of Medical Internet Research features almost half of the publications (n=168, 49.7%), and JMIR Public Health and Surveillance has more than one-fifth of the examined studies (n=74, 21.9%). The interest in the subject has been increasing every year, with 2018 featuring more than one-fourth of the total publications (n=89, 26.3%), and the publications in 2017 and 2018 combined accounted for more than half (n=171, 50.6%) of the total number of publications in the last decade. The most popular source was Twitter with 45.0% (n=152), followed by Google with 24.6% (n=83), websites and platforms with 13.9% (n=47), blogs and forums with 10.1% (n=34), Facebook with 8.9% (n=30), and other search engines with 5.6% (n=19). As for the subjects examined, conditions and diseases with 17.2% (n=58) and epidemics and outbreaks with 15.7% (n=53) were the most popular categories identified in this review, followed by health care (n=39, 11.5%), drugs (n=40, 10.4%), and smoking and alcohol (n=29, 8.6%).

Conclusions: The field of infodemiology is becoming increasingly popular, employing innovative methods and approaches for health assessment. The use of web-based sources, which provide us with information that would not be accessible otherwise and tackles the issues arising from the time-consuming traditional methods, shows that infodemiology plays an important role in health informatics research.

(J Med Internet Res 2020;22(4):e16206) doi:10.2196/16206

KEYWORDS

big data; infodemiology; infoveillance; internet; review; web-based data

Introduction

Infodemiology (ie, information epidemiology) is a field in health informatics defined as "*the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy*" [1]. The first official mention of infodemiology according to a search in PubMed (ie, baring the term on the title) was by Gunther Eysenbach in 2002 [2]. However, infodemiology studies (ie, assessment of health-related topics using web-based data [3])

can be traced all the way back to 1996. Two more studies in the 2000s use the term (PubMed): one in 2004, where the quality of hospitals' websites was assessed [4], and one in 2006, showing that flu data from Google correlated with influenza cases [5].

The large corpus of publications in infodemiology were present after 2009, with the first complete presentation and assessment of the subject being found in the scoping review of Bernardo et al [6] published with JMIR Publications—the main publisher of infodemiology and infoveillance studies.

Social media and search queries are the most popular sources for retrieving information from web-based sources. The use of social media is constantly expanding [7], with more users and including more features. Search query data is also of significant value, as they take into account the revealed and not the stated users' preferences [8,9], but methodology should be designed with caution to ensure the validity of the results [10].

Popular social media data sources in infodemiology include Twitter [11-17], Facebook [18-22], and Instagram [23,24]. Queries from search engines are mostly retrieved by Google Trends [25-32], as well as Yandex [33-35], Baidu [36,37], Bing [38], Yahoo [39], and Daum [40,41]. Other popular sources include websites and platforms [42-45]; blogs, forums, and online communities [46-52]; and, what has received attention lately, mobile apps of certain health categories (eg, asthma [53] and heart failure self-care management [54]). Significant focus has been shown in combining two or more sources such as Facebook and Instagram [55], Facebook and Twitter Posts [56], US newspaper media and Facebook [57], and Google and Wikipedia [58].

The use of web-based sources offers an assessment of real-time information, whether it is from Twitter, Google, or other social media and search queries. For health data retrieved through traditional methods such as registries, questionnaires, or surveys, analysis and assessment can take time to perform. Thus, nowcasting using said methods is not trivial. However, web-based (real time) data has been shown to significantly contribute to the analysis and forecasting of certain diseases, outbreaks, and epidemics.

Therefore, this specific part of health informatics has benefitted from infodemiology. In particular, one of the most studied diseases is influenza, and several data sources have been employed to predict and assess flu-related topics [39,40,59-76]. Epidemics and infectious diseases that have been analyzed and assessed using infodemiology and infoveillance approaches include HIV/AIDS [77-79], measles [80-83], and the Zika virus [84-87].

Infodemiology topics have also been the subject of research for several reviews on various topics like curable sexually transmitted diseases (STDs) [88] and mental health disorders [89], and for data sources like search queries, social media [6], mobile phone apps [90], Twitter [91], and Google Trends [92].

Infodemiology has contributed to health assessment with the analysis of a range of topics. In specific, popular categories in the field of infodemiology and infoveillance include drugs [39,93,94], marijuana [95-97], depression and suicide [98-108], smoking and tobacco [109-116], electronic cigarettes (e-cigarettes) [117-126], and hookahs [127-130]. As far as chronic diseases are concerned, the assessment of diabetes [131-136] and multiple sclerosis [137,138] has benefitted from infodemiology and infoveillance studies. Other topics include breast cancer [139-142]; fitness and diet [143-146]; health care performance, evaluation, and dissemination [147,148]; and human papillomavirus (HPV) [149-154].

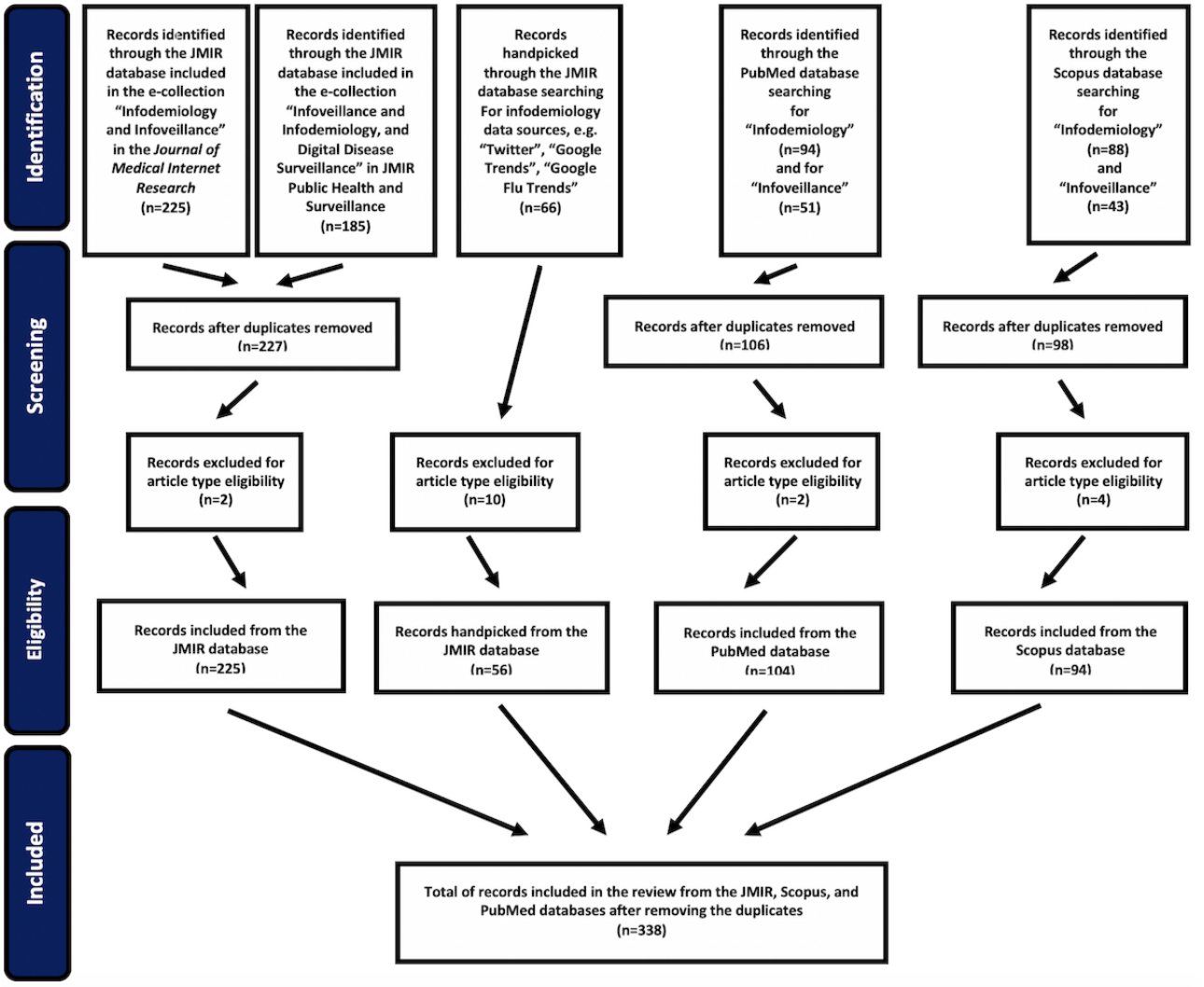
This review aims to update and expand the 2013 scoping review of Bernardo et al [6]. The authors of said review provided a well-structured outline of how infodemiology was employed in health informatics research up to that point, but as is evident, the large corpus of studies in the field have been published within the last couple of years. This update on the subject is important in identifying how infodemiology has contributed to health informatics over the past decade compared with traditional surveillance methods, the main web sources used, and the individual health categories and topics that have been explored. Apart from identifying the "metrics" of infodemiology studies (ie, numer of publications, thematic topics, journals and publishers, and data sources), this review aims to identify the advantages, disadvantages, and value and validity of infodemiology and infoveillance.

Methods

To select the publications in the fields of infodemiology and infoveillance, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Extension for Scoping Reviews [155,156] were followed. The procedure is depicted in the PRISMA flow diagram in Figure 1.

In the JMIR Publications database, all papers from the two relevant electronic collections (e-collections) were retrieved: 225 documents from the "Infodemiology and Infoveillance" [157] e-collection in the *Journal of Medical Internet Research* and 185 from the "Infodemiology, Infoveillance, and Digital Disease Surveillance" [158] e-collection in *JMIR Public Health and Surveillance*. After removing the duplicates and 2 documents for article type eligibility, 227 documents were extracted in total. Next, 66 documents were handpicked from the JMIR publications database based on searches of data sources (ie, "Twitter," "Google Trends," and "Google Flu Trends"). After 10 documents were excluded based on article type, a total of 56 documents were handpicked from the JMIR database.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram for selecting the publications from JMIR, PubMed, and Scopus.



Next, the term "infodemiology" was searched for in the PubMed database from January 1, 2009, to December 31, 2018, in the search field "Title-Abstract." The search returned 94 documents. The term "infoveillance" was then entered for the same period and in the same field, and the search returned 51 documents. The duplicates were 39 in total, and 2 documents were excluded based on article type; a total of 104 documents were extracted from the PubMed database. The terms "infodemiology" and "infoveillance" were then independently searched for in the Scopus database in the "Article title, Abstract, Keywords" field for the full years 2009-2018. The search returned 88 and 43 documents, respectively (ie, a total of 131). After removing 33 duplicates and 4 documents for article type criteria, a total of 94 documents were extracted from the Scopus database.

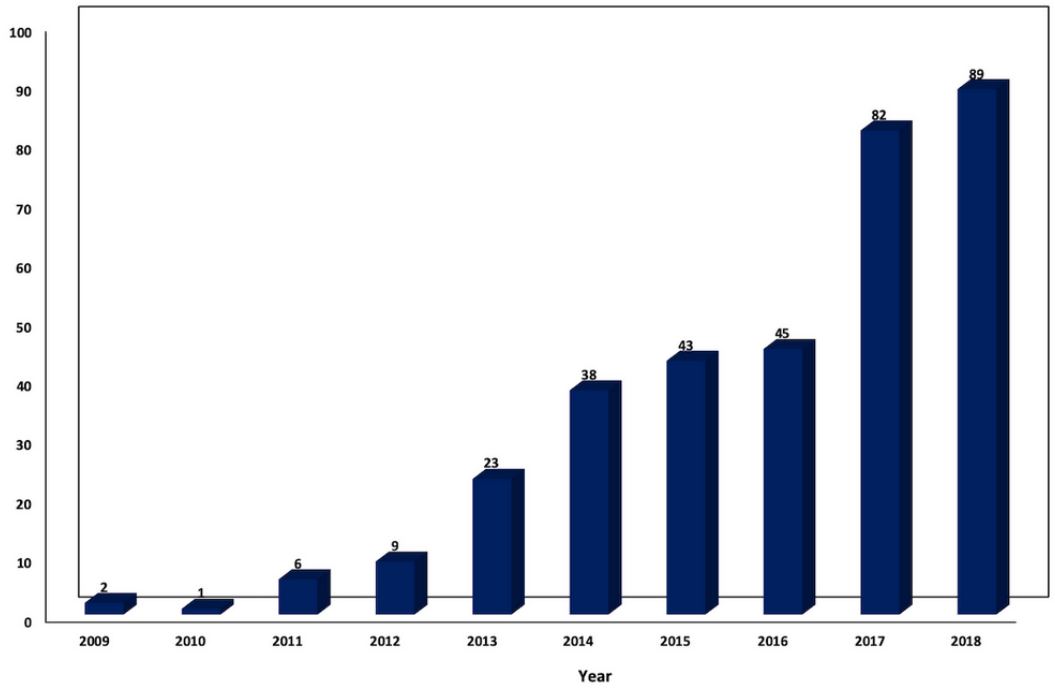
Overall, all studies that included the terms "infodemiology" or "infoveillance" in the "Title/Abstract" field in PubMed up to December 2018 and all studies including the terms "infodemiology" or "infoveillance" in the "Article title, Abstract, Keywords" field up to 2018 in Scopus were selected. For JMIR, all articles in the two relevant e-collections, as well as the articles derived by the individual data source searches, were included in this review. Articles were only excluded based on article type eligibility (eg, correction, erratum). After removing the duplicates from the JMIR, PubMed, and Scopus databases, the total extracted documents from all databases were 338.

Results

Table A1 in Multimedia Appendix 1 consists of the 338 selected publications included in this review, and shows the authors' names, publication year, the title, and the journal used for the analysis. Figure 2 depicts the number of publications by year from 2009 to 2018.

The number of publications in the subject is increasing every year, with 2018 featuring more than one-fourth of the 338 total publications (n=89, 26.3%), and the publications from 2017 and 2018 combined accounted for more than half (n=171, 50.6%) of the total number of publications in the past decade.

Figure 2. Number of publications in infodemiology and infoveillance by year (2009-2018).



The 338 extracted studies were published in 57 journals in total. The vast majority of the studies in infodemiology and infoveillance were published with JMIR Publications (n=282, 83.4%). Specifically, the *Journal of Medical Internet Research* features almost half of the publications (n=168, 49.7%), and *JMIR Public Health and Surveillance* features almost one-fourth of the examined studies (n=74, 21.9%).

Figure 3 consists of the numbers of publications per journal with >2 publications on the subject.

Journals that have published more than 1 paper in the subject included JMIR Mental Health (n=9), JMIR Research Protocols (n=7), JMIR Diabetes (n=4), JMIR Cancer (n=3), JMIR Medical Informatics (n=3), JMIR Pediatrics and Parenting (n=3), the Journal of Big Data (n=2), and the Interactive Journal of Medical Research (n=2).

Table A2 of Multimedia Appendix 1 consists of the 338 publications categorized by data source employed. Figure 4 depicts the popularity of the examined data sources in terms of number of publications (some publications employed more than one data source).

Figure 3. Number of publications in infodemiology and infoveillance by journal (2009-2018). IEEE: Institute of Electrical and Electronics Engineers; JPH: JMIR Public Health and Surveillance.

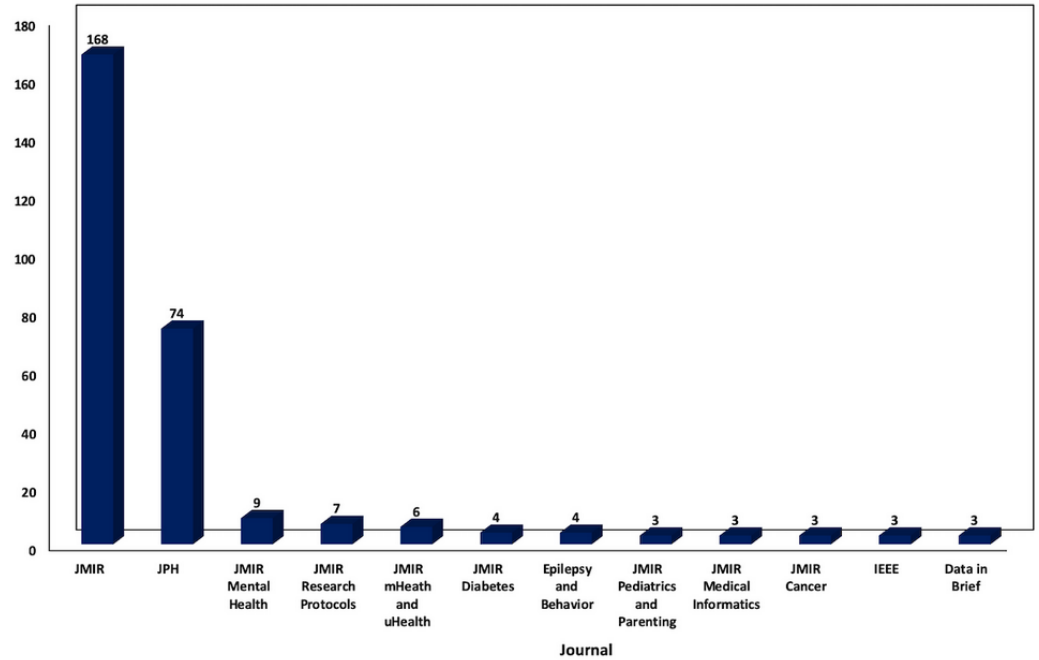
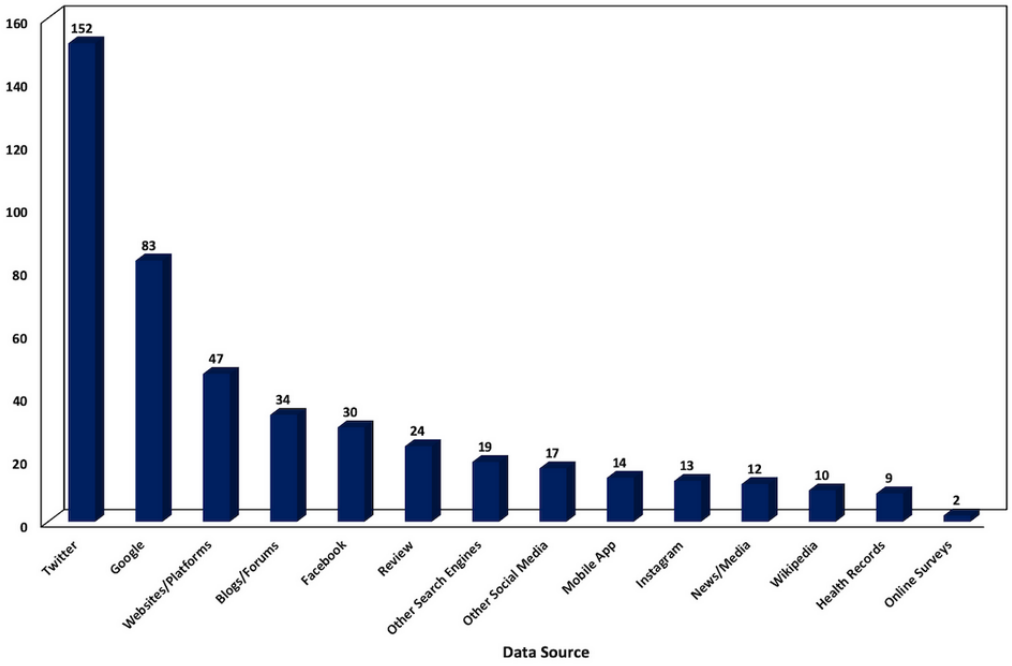


Figure 4. Number of publications by data source (2009-2018).



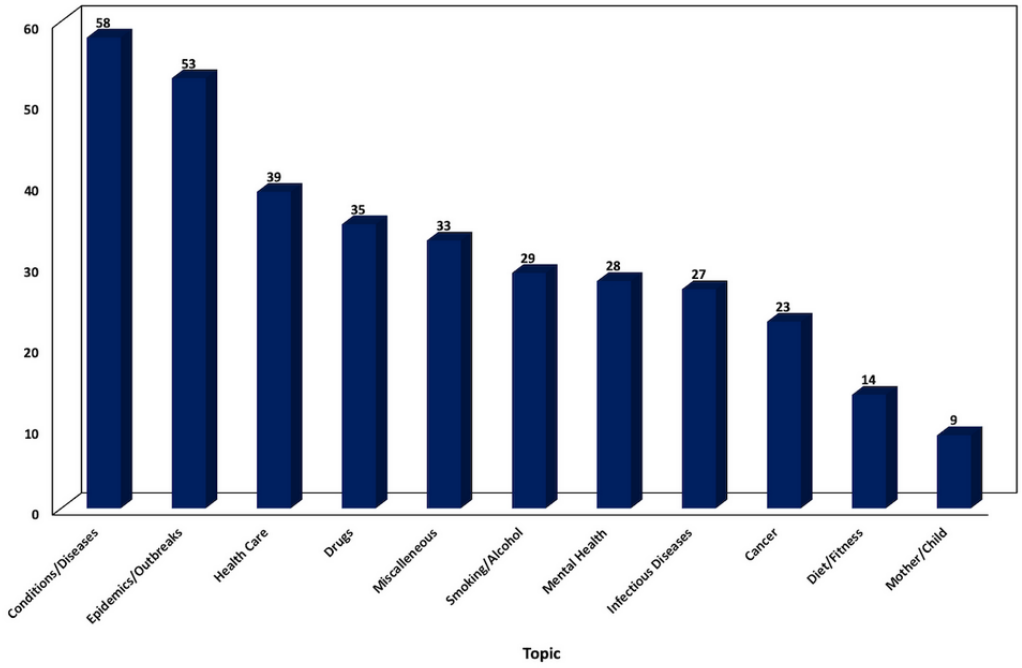
Of the 338 publications, the most popular source was Twitter with 45.0% (n=152) and is continuously gaining popularity. Google sources were in second place with 24.6% (n=83), followed by websites and platforms with 13.9% (n=47), blogs and forums with 10.1% (n=34), Facebook with 8.9% (n=30), and other search engines with 5.6% (n=19). The Google category consisted mainly of publications using Google Trends; although, the following Google tools have also been identified as main data sources in several publications: Google Flu Trends (n=6), Google Analytics (n=2), Google Insights (n=2), Google Correlate (n=1), Google Health (n=1), Google News (n=1), Google AdWords (n=1), Google Video (n=1), and Google Blog Search (n=1).

The “other search engines” category consists of Bing (n=7), Baidu (n=4), Yandex (n=4), Daum (n=2), and Yahoo (n=3), and the “other social media” category consists of YouTube (n=5), Yelp (n=5), Google+ (n=4), Foursquare (n=1), SoundCloud (n=1), Tumblr (n=1), Pinterest (n=1), and MySpace (n=1). Yahoo answers (n=2) was included in the blogs, forums, and communities category.

Although many health topics have been examined in infodemiology and infoveillance, some are significantly more popular. Figure 5 depicts the general categories in terms of number of publications, while Figure A1 of Multimedia Appendix 2 consists of the pie charts of their subcategories. All individual topics and subtopics identified in this review by number of publications can be found in Table A1 of Multimedia Appendix 2.

In the 338 publications examined in this review, the most popular subjects were conditions and diseases with 17.2% (n=58) and epidemics and outbreaks with 15.7% (n=53), followed by health care with 11.5% (n=39), drugs with 10.4% (n=35), smoking and alcohol with 8.6% (n=29), and mental health with 8.3% (n=28). Infectious diseases with 8.0% (n=27) and cancer with 6.8% (n=23) were also featured in several publications. The categories of diet and fitness with 4.1% (n=14) and mother and child with 2.7% (9) were less popular.

Figure 5. Health categories by number of publications (2009-2018).



Discussion

Principal Findings

In this scoping review, the most popular web-based data source as well as the most studied health categories in infodemiology and infoveillance were identified through a systematic search of the existing literature.

Time Line and Journals

Based on the results, the use of web-based data in health informatics is significantly increasing year by year, with half of the publications from 2009 to 2018 being in the last couple of years. The JMIR Publications database is the most significant contributor in terms of the number of publications on the subject (n=282/338, 83.4%), especially given that the most significant infodemiology-oriented journal (ie, JMIR Public Health and Surveillance) is published with JMIR Publications. The concept of infodemiology, introduced by Gunther Eysenbach in 2002, has been gaining significant recognition since its birth, and it is evident that it will play an even more significant role in health informatics in the years to come, especially as internet penetration increases along with the average age of the users.

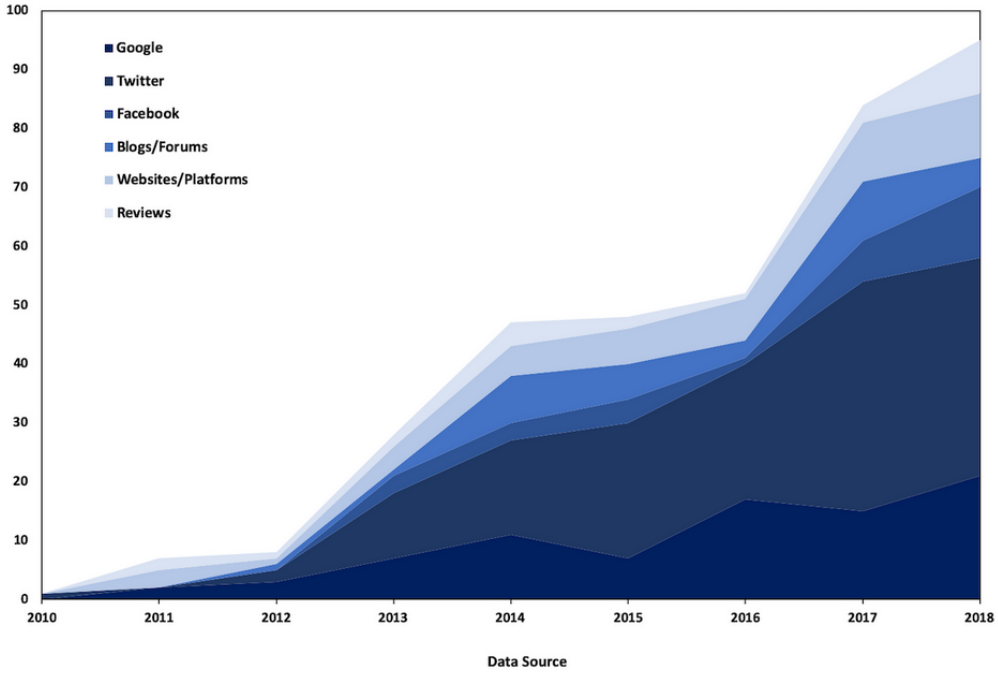
Data Sources and Tools

Figure 6 depicts the yearly changes in number of publications for the most popular data sources over the examined period. As evident, there was a significant increase in the use of web-data sources over the last couple of years, with Twitter in the lead in assessing health-related topics by health informatics researchers.

Despite the increasingly large number of users and the fact that Twitter is used significantly more than Google, Twitter has the limitation of not being universal. Its pros include that it is an outlet for official reports and news (eg, governmental, politicians), but a significant con is that it is not used by all; furthermore, not everyone interacts on the site (ie, tweets or retweets). The analysis of internet search traffic data—mainly from Google but from other search engines as well (eg, Bing, Yahoo)—is more universal in the sense that internet penetration has increased to a point where the large majority of people have access to and use the internet and searching for keywords in search engines is the most used internet feature. Apart from this, it also ensures anonymity, deeming it more reliable, as it uses the revealed and not the stated users' preferences. However, the choosing of the keywords (queries) as well as the methodology for selecting the retrieved data is much more complicated than with Twitter. In addition, more than one search engine exists, and thus, not all queries (data) on the respective selected topic can be retrieved.

On the other hand, there is a significant rise in the percentage of publications in the last couple of years using data from other social media such as Facebook and Instagram. This could be showing that younger internet users' preferences in the use of social media may be revealing a trend of said platforms over original search queries and websites. Researchers in this field should closely follow any potential shifts in internet use, along with the correspondence and age of users, to ensure—to the point that is possible with web-based data—that the sample is representative and the research aims to change along with what is trending. The latter is crucial for infodemiology research to continue to be valid, and it should be given significant focus.

Figure 6. Yearly changes in number of publications for the most popular data sources (2009-2018).



Health Topics

As is evident, there is a wide variety of topics that have been studied up to this point. The most popular topics identified in this review were illegal drugs, breast cancer, smoking, fitness, HIV/AIDS, depression, diabetes, influenza, HPV, multiple sclerosis, Zika virus, suicide, STDs, and e-cigarettes, and significant attention has been given to the evaluation of health care such as hospital ratings and patients' experiences and health topics' and medical institutions' dissemination strategies.

Approaches include nowcasting epidemics and outbreaks, surveillance of infectious diseases, assessment of chronic conditions, and basically everything traditional surveillance methods have aimed to do. Thus, the results of this review show not only the increasing popularity of web-based data but also their significant contribution to the existing literature, as well as the value of infodemiology in health informatics.

Advantages and Disadvantages

The difference in using web-based data—infodemiology's main advantage—is that it tackles the issue of traditional surveillance methods not providing real time assessments. Even in the health sector, where data are generally available compared to other topics, the gathering, assessing, and publishing of health data can sometimes take years to process. This is less of an issue for topics such as chronic diseases that are not infectious, but it makes the assessment and forecasting of epidemics and outbreaks much more complicated.

Another significant advantage of infoveillance compared to traditional surveillance methods is the anonymity that web-based data offers. Online search traffic data are completely anonymous, and in most social media and forums, an individual has the option of anonymity. In this way, data retrieved from said sources are the revealed and not the stated preferences, which can be a plus for sensitive topics such as AIDS or STDs.

Despite the many advantages that web-based data sources have to offer, several limitations have been identified in the use of infodemiology sources. The main disadvantage of using web-based sources is that the data can be affected by sudden incidents or events, which, especially in nowcasting or when the number of observations is low, could provide biased results. Similarly, the sample cannot be shown to be representative, especially in the assessment of online search traffic data; although, this is less of an issue as internet penetration increases.

With real time data that can be retrieved from web sources, disease surveillance has become much faster than with traditional methods, and web sources also have the benefit of assessing large populations, which contrasts with most traditional methods that are based on data retrieved from significantly smaller groups, such as with interviews or questionnaires. Overall, what health informatics should aim toward in the future is to combine web-data sources with traditional data assessment to provide an even more complete assessment.

Limitations

The main limitation of this scoping review was that not all infodemiology papers could be included. Though the selection of publications for this work was thorough and followed the guidelines for proper selection and included the main outlets for infodemiology papers (ie, JMIR Publications and PubMed), some publications may have been left out; a limitation that all reviews have. Specifically, articles using the two most popular infodemiology sources (ie, Twitter and Google) were only searched for in the JMIR database. Studies using, for example, Google Trends and Twitter constitute a large body of the relevant literature, and a significant number of said studies were not included as they did not use the specific searched for terms (ie, infodemiology and infoveillance); the latter being the main difference from the original Bernardo et al [6] scoping review.

However, despite the possible reduced number of included publications that use infodemiology and infoveillance sources but not the infodemiology or infoveillance terms, as JMIR is the main outlet for such themed publications, this scoping review gives a valuable qualitative and quantitative overview of how the concept has progressed over the last decade, as well as identifying the main sources and topics that have been used and assessed. Future work should focus on expanding the present results, as well as recording infodemiology papers based on tools used. As the search by source yields many results, focus should be given to future systematic reviews on the subject by source as has been done, for example, for the use of Google Trends [92].

Conclusions

Using web-based sources in epidemiology and disease surveillance has shown to be valuable and valid over the past decade, and the results of this scoping review clearly point to this direction. Data sources cover a wide variety of tools, social media, platforms, websites, blogs, and search engines, and the topics that are the most studied vary from chronic disease prevalence to nowcasting epidemics. Infodemiology and infoveillance tackle several of the issues that arise with traditional assessment methods, and, as internet penetration increases, employing web data sources for health assessment could be the future in health informatics.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Infodemiology and infoveillance publications from JMIR, Scopus, and PubMed (2009-2018).
[PDF File (Adobe PDF File), 1420 KB - jmir_v22i4e16206_app1.pdf]

Multimedia Appendix 2
Topics and subtopics (#publications) in infodemiology and infoveillance (2009- 2018).
[PDF File (Adobe PDF File), 1430 KB - jmir_v22i4e16206_app2.pdf]

References

1. 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. [doi: 10.2196/jmir.1157] [Medline: 19329408]
2. Eysenbach G. Infodemiology: the epidemiology of (mis)information. The American Journal of Medicine 2002 Dec;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0]
3. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
4. Kind T, Wheeler KL, Robinson B, Cabana MD. Do the leading children's hospitals have quality web sites? A description of children's hospital web sites. J Med Internet Res 2004 Jun 25;6(2):e20. [doi: 10.2196/jmir.6.2.e20] [Medline: 15249269]
5. 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]
6. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013 Jul 18;15(7):e147. [doi: 10.2196/jmir.2740] [Medline: 23896182]
7. Pew Research Center. Social Media Fact Sheet URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-09-09]
8. Mavragani A, Tsagarakis KP. Predicting referendum results in the Big Data Era. J Big Data 2019 Jan 14;6(1):e. [doi: 10.1186/s40537-018-0166-z]
9. Mavragani A, Tsagarakis KP. YES or NO: predicting the 2015 GReferendum results using Google Trends. Technological Forecasting and Social Change 2016 Aug;109:1-5. [doi: 10.1016/j.techfore.2016.04.028]
10. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439. [doi: 10.2196/13439] [Medline: 31144671]
11. Bian J, Zhao Y, Salloum RG, Guo Y, Wang M, Prosperi M, et al. Using social media data to understand the impact of promotional information on laypeople's discussions: a case study of lynch syndrome. J Med Internet Res 2017 Dec 13;19(12):e414. [doi: 10.2196/jmir.9266] [Medline: 29237586]
12. Chen T, Dredze M. Vaccine images on Twitter: analysis of what images are shared. J Med Internet Res 2018 Apr 03;20(4):e130. [doi: 10.2196/jmir.8221] [Medline: 29615386]
13. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring online discussions about suicide among Twitter users with schizophrenia: exploratory study. JMIR Ment Health 2018 Dec 13;5(4):e11483. [doi: 10.2196/11483] [Medline: 30545811]

14. Odlum M, Yoon S, Broadwell P, Brewer R, Kuang D. How Twitter can support the HIV/AIDS response to achieve the 2030 eradication goal: in-depth thematic analysis of World AIDS Day tweets. JMIR Public Health Surveill 2018 Nov 22;4(4):e10262. [doi: 10.2196/10262] [Medline: 30467102]
15. Chen S, Xu Q, Buchenberger J, Bagavathi A, Fair G, Shaikh S, et al. Dynamics of health agency response and public engagement in public health emergency: a case study of CDC tweeting patterns during the 2016 Zika epidemic. JMIR Public Health Surveill 2018 Nov 22;4(4):e10827. [doi: 10.2196/10827] [Medline: 30467106]
16. Wakamiya S, Kawai Y, Aramaki E. Twitter-based influenza detection after flu peak via tweets with indirect information: text mining study. JMIR Public Health Surveill 2018 Sep 25;4(3):e65. [doi: 10.2196/publichealth.8627] [Medline: 30274968]
17. Tufts C, Polsky D, Volpp KG, Groeneveld PW, Ungar L, Merchant RM, et al. Characterizing tweet volume and content about common health conditions across Pennsylvania: retrospective analysis. JMIR Public Health Surveill 2018 Dec 06;4(4):e10834. [doi: 10.2196/10834] [Medline: 30522989]
18. 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 May 08;19(5):e156. [doi: 10.2196/jmir.6815] [Medline: 28483739]
19. Smith RJ, Crutchley P, Schwartz HA, Ungar L, Shofer F, Padrez KA, et al. Variations in Facebook posting patterns across validated patient health conditions: a prospective cohort study. J Med Internet Res 2017 Jan 06;19(1):e7. [doi: 10.2196/jmir.6486] [Medline: 28062392]
20. Gittelman S, Lange V, Gotway Crawford CA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res 2015 Apr 20;17(4):e98. [doi: 10.2196/jmir.3970] [Medline: 25895907]
21. Keller MS, Park HJ, Cunningham ME, Fouladian JE, Chen M, Spiegel BMR. Public perceptions regarding use of virtual reality in health care: a social media content analysis using Facebook. J Med Internet Res 2017 Dec 19;19(12):e419. [doi: 10.2196/jmir.7467] [Medline: 29258975]
22. 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. [doi: 10.2196/publichealth.7217] [Medline: 29592849]
23. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of codeine misuse on Instagram: content analysis. JMIR Public Health Surveill 2018 Mar 20;4(1):e22. [doi: 10.2196/publichealth.8144] [Medline: 29559422]
24. Muralidhara S, Paul MJ. #Healthy Selfies: exploration of health topics on Instagram. JMIR Public Health Surveill 2018 Jun 29;4(2):e10150. [doi: 10.2196/10150] [Medline: 29959106]
25. Phillips CA, Barz Leahy A, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship between state-level Google online search volume and cancer incidence in the United States: retrospective study. J Med Internet Res 2018 Jan 08;20(1):e6. [doi: 10.2196/jmir.8870] [Medline: 29311051]
26. Radin M, Sciascia S. Infodemiology of systemic lupus erythematous using Google Trends. Lupus 2017 Jul;26(8):886-889. [doi: 10.1177/0961203317691372] [Medline: 28162030]
27. Seidl S, Schuster B, Rüth M, Biedermann T, Zink A. What do Germans want to know about skin cancer? a nationwide Google search analysis from 2013 to 2017. J Med Internet Res 2018 May 02;20(5):e10327. [doi: 10.2196/10327] [Medline: 29698213]
28. Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits with Google Trends: evidence from Taiwan. J Med Internet Res 2015 Nov 19;17(11):e264. [doi: 10.2196/jmir.4516] [Medline: 26586281]
29. Berlinberg EJ, Deiner MS, Porco TC, Acharya NR. Monitoring interest in herpes zoster vaccination: analysis of Google search data. JMIR Public Health Surveill 2018 May 02;4(2):e10180. [doi: 10.2196/10180] [Medline: 29720364]
30. Mavragani A, Ochoa G. Infoveillance of infectious diseases in USA: STDs, tuberculosis, and hepatitis. J Big Data 2018 Sep 6;5(1):30. [doi: 10.1186/s40537-018-0140-9]
31. Mavragani A, Sampri A, Sypsa K, Tsagarakis KP. Integrating smart health in the US health care system: infodemiology study of asthma monitoring in the Google Era. JMIR Public Health Surveill 2018 Mar 12;4(1):e24. [doi: 10.2196/publichealth.8726] [Medline: 29530839]
32. 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. [doi: 10.2196/jmir.5802] [Medline: 27637361]
33. Zheluk A, Gillespie JA, Quinn C. Searching for truth: internet search patterns as a method of investigating online responses to a Russian illicit drug policy debate. J Med Internet Res 2012 Dec 13;14(6):e165. [doi: 10.2196/jmir.2270] [Medline: 23238600]
34. Zheluk A, Quinn C, Meylakhs P. Internet search and krokodil in the Russian Federation: an infoveillance study. J Med Internet Res 2014 Sep 18;16(9):e212. [doi: 10.2196/jmir.3203] [Medline: 25236385]
35. Domnich A, Arbuzova EK, Signori A, Amicizia D, Panatto D, Gasparini R. Demand-based web surveillance of sexually transmitted infections in Russia. Int J Public Health 2014 Oct;59(5):841-849. [doi: 10.1007/s00038-014-0581-7] [Medline: 25012799]
36. 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. [doi: 10.2196/jmir.6240] [Medline: 27655225]

37. Yang H, Li S, Sun L, Zhang X, Hou J, Wang Y. Effects of the ambient fine particulate matter on public awareness of lung cancer risk in China: evidence from the internet-based big data platform. JMIR Public Health Surveill 2017 Oct 03;3(4):e64. [doi: 10.2196/publichealth.8078] [Medline: 28974484]
38. Cartwright AF, Karunaratne M, Barr-Walker J, Johns NE, Upadhyay UD. Identifying national availability of abortion care and distance from major US cities: systematic online search. J Med Internet Res 2018 May 14;20(5):e186. [doi: 10.2196/jmir.9717] [Medline: 29759954]
39. 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 Jun 18;15(6):e124. [doi: 10.2196/jmir.2614] [Medline: 23778053]
40. Seo D, Jo M, Sohn CH, Shin S, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. J Med Internet Res 2014 Dec 16;16(12):e289. [doi: 10.2196/jmir.3680] [Medline: 25517353]
41. Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. J Med Internet Res 2016 Jul 04;18(7):e177. [doi: 10.2196/jmir.4955] [Medline: 27377323]
42. Leung R, Guo H, Pan X. Social media users' perception of telemedicine and mHealth in China: exploratory study. JMIR Mhealth Uhealth 2018 Sep 25;6(9):e181. [doi: 10.2196/mhealth.7623] [Medline: 30274969]
43. Li Q, Wang C, Liu R, Wang L, Zeng DD, Leischow SJ. Understanding users' vaping experiences from social media: initial study using sentiment opinion summarization techniques. J Med Internet Res 2018 Aug 15;20(8):e252. [doi: 10.2196/jmir.9373] [Medline: 30111530]
44. Sadah SA, Shahbazi M, Wiley MT, 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. [doi: 10.2196/jmir.4308] [Medline: 26250986]
45. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. Demographic-based content analysis of web-based health-related social media. J Med Internet Res 2016 Jun 13;18(6):e148. [doi: 10.2196/jmir.5327] [Medline: 27296242]
46. Abbe A, Falissard B. Stopping antidepressants and anxiolytics as major concerns reported in online health communities: a text mining approach. JMIR Ment Health 2017 Oct 23;4(4):e48. [doi: 10.2196/mental.7797] [Medline: 29061554]
47. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36. [doi: 10.2196/publichealth.6577] [Medline: 28642212]
48. Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S. Detection of cases of noncompliance to drug treatment in patient forum posts: topic model approach. J Med Internet Res 2018 Mar 14;20(3):e85. [doi: 10.2196/jmir.9222] [Medline: 29540337]
49. Dyson MP, Newton AS, Shave K, Featherstone RM, Thomson D, Wingert A, et al. Social media for the dissemination of Cochrane child health evidence: evaluation study. J Med Internet Res 2017 Sep 01;19(9):e308. [doi: 10.2196/jmir.7819] [Medline: 28864427]
50. Jung Y, Hur C, Jung D, Kim M. Identifying key hospital service quality factors in online health communities. J Med Internet Res 2015 Apr 07;17(4):e90. [doi: 10.2196/jmir.3646] [Medline: 25855612]
51. Koh S, Gordon AS, Wienberg C, Sood SO, Morley S, Burke DM. Stroke experiences in weblogs: a feasibility study of sex differences. J Med Internet Res 2014 Mar 19;16(3):e84. [doi: 10.2196/jmir.2838] [Medline: 24647327]
52. Konheim-Kalkstein YL, Miron-Shatz T, Israel LJ. How women evaluate birth challenges: analysis of web-based birth stories. JMIR Pediatr Parent 2018 Dec 18;1(2):e12206. [doi: 10.2196/12206] [Medline: 31518300]
53. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113. [doi: 10.2196/mhealth.7177] [Medline: 28768606]
54. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 02;2(1):e10057. [doi: 10.2196/10057] [Medline: 31758762]
55. Hendriks H, Van den Putte B, Gebhardt WA, Moreno MA. Social drinking on social media: content analysis of the social aspects of alcohol-related posts on Facebook and Instagram. J Med Internet Res 2018 Jun 22;20(6):e226. [doi: 10.2196/jmir.9355] [Medline: 29934290]
56. Roccetti M, Marfia G, Salomoni P, Prandi C, Zagari RM, Gningaye Kengni FL, et al. Attitudes of Crohn's disease patients: infodemiology case study and sentiment analysis of Facebook and Twitter posts. JMIR Public Health Surveill 2017 Aug 09;3(3):e51. [doi: 10.2196/publichealth.7004] [Medline: 28793981]
57. Manchaiah V, Ratinaud P, Andersson G. Representation of tinnitus in the US newspaper media and in Facebook pages: cross-sectional analysis of secondary data. Interact J Med Res 2018 May 08;7(1):e9. [doi: 10.2196/ijmr.9065] [Medline: 29739734]
58. Sciascia S, Radin M. What can Google and Wikipedia can tell us about a disease? Big Data trends analysis in Systemic Lupus Erythematosus. Int J Med Inform 2017 Nov;107:65-69. [doi: 10.1016/j.ijmedinf.2017.09.002] [Medline: 29029693]
59. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using bayesian change point analysis: a comparative analysis. JMIR Public Health Surveill 2016 Oct 20;2(2):e161. [doi: 10.2196/publichealth.5901] [Medline: 27765731]

60. Timpka T, Spreco A, Dahlström Ö, Eriksson O, Gursky E, Ekberg J, et al. Performance of eHealth data sources in local influenza surveillance: a 5-year open cohort study. J Med Internet Res 2014 Apr 28;16(4):e116. [doi: 10.2196/jmir.3099] [Medline: 24776527]
61. Wagner M, Lampos V, Yom-Tov E, Pebody R, Cox IJ. Estimating the population impact of a new pediatric influenza vaccination program in England using social media content. J Med Internet Res 2017 Dec 21;19(12):e416. [doi: 10.2196/jmir.8184] [Medline: 29269339]
62. Aslam AA, Tsou M, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. J Med Internet Res 2014 Nov 14;16(11):e250. [doi: 10.2196/jmir.3532] [Medline: 25406040]
63. Baltrusaitis K, Santillana M, Crawley AW, Chunara R, Smolinski M, Brownstein JS. Determinants of participants' follow-up and characterization of representativeness in Flu Near You, a participatory disease surveillance system. JMIR Public Health Surveill 2017 Apr 07;3(2):e18. [doi: 10.2196/publichealth.7304] [Medline: 28389417]
64. Broniatowski DA, Dredze M, Paul MJ, Dugas A. Using social media to perform local influenza surveillance in an inner-city hospital: a retrospective observational study. JMIR Public Health Surveill 2015;1(1):e5. [doi: 10.2196/publichealth.4472] [Medline: 27014744]
65. Chen B, Shao J, Liu K, Cai G, Jiang Z, Huang Y, et al. Does eating chicken feet with pickled peppers cause avian influenza? observational case study on Chinese social media during the avian influenza A (H7N9) outbreak. JMIR Public Health Surveill 2018 Mar 29;4(1):e32. [doi: 10.2196/publichealth.8198] [Medline: 29599109]
66. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20. [doi: 10.2196/jmir.2911] [Medline: 24440770]
67. Hill S, Mao J, Ungar L, Hennessy S, Leonard CE, Holmes J. Natural supplements for H1N1 influenza: retrospective observational infodemiology study of information and search activity on the Internet. J Med Internet Res 2011 May 10;13(2):e36. [doi: 10.2196/jmir.1722] [Medline: 21558062]
68. Kagashe I, Yan Z, Suheryani I. Enhancing seasonal influenza surveillance: topic analysis of widely used medicinal drugs using Twitter data. J Med Internet Res 2017 Sep 12;19(9):e315. [doi: 10.2196/jmir.7393] [Medline: 28899847]
69. Kandula S, Hsu D, Shaman J. Subregional nowcasts of seasonal influenza using search trends. J Med Internet Res 2017 Nov 06;19(11):e370. [doi: 10.2196/jmir.7486] [Medline: 29109069]
70. Klembczyk JJ, Jalalpour M, Levin S, Washington RE, Pines JM, Rothman RE, et al. Google Flu Trends spatial variability validated against emergency department influenza-related visits. J Med Internet Res 2016 Jun 28;18(6):e175. [doi: 10.2196/jmir.5585] [Medline: 27354313]
71. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the Boston Metropolis. JMIR Public Health Surveill 2018 Jan 09;4(1):e4. [doi: 10.2196/publichealth.8950] [Medline: 29317382]
72. Mao C, Wu X, Fu X, Di M, Yu Y, Yuan J, et al. An internet-based epidemiological investigation of the outbreak of H7N9 Avian influenza A in China since early 2013. J Med Internet Res 2014 Sep 25;16(9):e221. [doi: 10.2196/jmir.3763] [Medline: 25257217]
73. Pervaiz F, Pervaiz M, Abdur Rehman N, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res 2012 Oct 04;14(5):e125. [doi: 10.2196/jmir.2102] [Medline: 23037553]
74. Poirier C, Lavenu A, Bertaud V, Campillo-Gimenez B, Chazard E, Cuggia M, et al. Real time influenza monitoring using hospital big data in combination with machine learning methods: comparison study. JMIR Public Health Surveill 2018 Dec 21;4(4):e11361. [doi: 10.2196/11361] [Medline: 30578212]
75. Samaras L, García-Barriocanal E, Sicilia M. Syndromic surveillance models using web data: the case of influenza in Greece and Italy using Google Trends. JMIR Public Health Surveill 2017 Nov 20;3(4):e90. [doi: 10.2196/publichealth.8015] [Medline: 29158208]
76. Santos JC, Matos S. Analysing Twitter and web queries for flu trend prediction. Theor Biol Med Model 2014 May 07;11 Suppl 1:S6 [FREE Full text] [doi: 10.1186/1742-4682-11-S1-S6] [Medline: 25077431]
77. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. JMIR Public Health Surveill 2015;1(2):e7. [doi: 10.2196/publichealth.4488] [Medline: 27227141]
78. Zheluk A, Quinn C, Hercz D, Gillespie JA. Internet search patterns of human immunodeficiency virus and the digital divide in the Russian Federation: infoveillance study. J Med Internet Res 2013 Nov 12;15(11):e256. [doi: 10.2196/jmir.2936] [Medline: 24220250]
79. Mavragani A, Ochoa G. Forecasting AIDS prevalence in the United States using online search traffic data. J Big Data 2018 May 19;5(1):5. [doi: 10.1186/s40537-018-0126-7]
80. Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. J Med Internet Res 2015 May 26;17(5):e128. [doi: 10.2196/jmir.3863] [Medline: 26013683]

81. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. JMIR Public Health Surveill 2016;2(1):e1. [doi: 10.2196/publichealth.5059] [Medline: 27227144]
82. Du J, Tang L, Xiang Y, Zhi D, Xu J, Song H, et al. Public perception analysis of Tweets during the 2015 measles outbreak: comparative study using convolutional neural network models. J Med Internet Res 2018 Jul 09;20(7):e236. [doi: 10.2196/jmir.9413] [Medline: 29986843]
83. Mavragani A, Ochoa G. The internet and the anti-vaccine movement: tracking the 2017 EU measles outbreak. BDCC 2018 Jan 16;2(1):2. [doi: 10.3390/bdcc2010002]
84. Farhadloo M, Winneg K, Chan MS, Hall Jamieson K, Albarracin D. Associations of topics of discussion on Twitter with survey measures of attitudes, knowledge, and behaviors related to Zika: probabilistic study in the United States. JMIR Public Health Surveill 2018 Feb 09;4(1):e16. [doi: 10.2196/publichealth.8186] [Medline: 29426815]
85. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015-2016 Colombian Zika virus disease outbreak. JMIR Public Health Surveill 2016 Jun 01;2(1):e30. [doi: 10.2196/publichealth.5814] [Medline: 27251981]
86. 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. [doi: 10.2196/publichealth.7157] [Medline: 28630032]
87. 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. [doi: 10.2196/publichealth.6925] [Medline: 28428164]
88. Sanz-Lorente M, Wanden-Berghe C, Castejón-Bolea R, Sanz-Valero J. Web 2.0 tools in the prevention of curable sexually transmitted diseases: scoping review. J Med Internet Res 2018 Mar 22;20(3):e113. [doi: 10.2196/jmir.8871] [Medline: 29567633]
89. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun 29;19(6):e228. [doi: 10.2196/jmir.7215] [Medline: 28663166]
90. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 02;2(1):e10057. [doi: 10.2196/10057] [Medline: 31758762]
91. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: review of the methods used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43. [doi: 10.2196/publichealth.5789] [Medline: 29685871]
92. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. J Med Internet Res 2018 Nov 06;20(11):e270. [doi: 10.2196/jmir.9366] [Medline: 30401664]
93. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013 Sep 06;15(9):e189. [doi: 10.2196/jmir.2741] [Medline: 24014109]
94. Katsuki T, Mackey TK, Cuomo R. Establishing a link between prescription drug abuse and illicit online pharmacies: analysis of Twitter data. J Med Internet Res 2015 Dec 16;17(12):e280. [doi: 10.2196/jmir.5144] [Medline: 26677966]
95. Peiper NC, Baumgartner PM, Chew RF, Hsieh YP, Bieler GS, Bobashev GV, et al. Patterns of Twitter behavior among networks of cannabis dispensaries in California. J Med Internet Res 2017 Jul 04;19(7):e236. [doi: 10.2196/jmir.7137] [Medline: 28676471]
96. Yom-Tov E, Lev-Ran S. Adverse reactions associated with cannabis consumption as evident from search engine queries. JMIR Public Health Surveill 2017 Oct 26;3(4):e77. [doi: 10.2196/publichealth.8391] [Medline: 29074469]
97. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. J Med Internet Res 2014 Jun 27;16(6):e157. [doi: 10.2196/jmir.3247] [Medline: 24974893]
98. Ricard BJ, Marsch LA, Crosier B, Hassanpour S. Exploring the utility of community-generated social media content for detecting depression: an analytical study on Instagram. J Med Internet Res 2018 Dec 06;20(12):e11817. [doi: 10.2196/11817] [Medline: 30522991]
99. Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. J Med Internet Res 2018 May 08;20(5):e168. [doi: 10.2196/jmir.9267] [Medline: 29739736]
100. Tana JC, Kettunen J, Eirola E, Paakkonen H. Diurnal variations of depression-related health information seeking: case study in Finland using Google Trends data. JMIR Ment Health 2018 May 23;5(2):e43. [doi: 10.2196/mental.9152] [Medline: 29792291]
101. DeJohn AD, Schulz EE, Pearson AL, Lachmar EM, Wittenborn AK. Identifying and understanding communities using Twitter to connect about depression: cross-sectional study. JMIR Ment Health 2018 Nov 05;5(4):e61. [doi: 10.2196/mental.9533] [Medline: 30401662]
102. Jung H, Park H, Song T. Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. J Med Internet Res 2017 Jul 24;19(7):e259. [doi: 10.2196/jmir.7452] [Medline: 28739560]
103. Lachmar EM, Wittenborn AK, Bogen KW, McCauley HL. #MyDepressionLooksLike: examining public discourse about depression on Twitter. JMIR Ment Health 2017 Oct 18;4(4):e43. [doi: 10.2196/mental.8141] [Medline: 29046270]

104. Schlichthorst M, King K, Turnure J, Sukunesan S, Phelps A, Pirkis J. Influencing the conversation about masculinity and suicide: evaluation of the Man Up multimedia campaign using Twitter data. JMIR Ment Health 2018 Feb 15;5(1):e14. [doi: 10.2196/mental.9120] [Medline: 29449203]
105. Wong PW, Fu K, Yau RS, Ma HH, Law Y, Chang S, et al. Accessing suicide-related information on the internet: a retrospective observational study of search behavior. J Med Internet Res 2013 Jan 11;15(1):e3. [doi: 10.2196/jmir.2181] [Medline: 23305632]
106. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for Twitter data against established measures of suicidality. JMIR Ment Health 2016 May 16;3(2):e21. [doi: 10.2196/mental.4822] [Medline: 27185366]
107. 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. [doi: 10.2196/jmir.7276] [Medline: 28694239]
108. Lee D, Lee H, Choi M. Examining the relationship between past orientation and US suicide rates: an analysis using big data-driven Google search queries. J Med Internet Res 2016 Feb 11;18(2):e35. [doi: 10.2196/jmir.4981] [Medline: 26868917]
109. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of menthol cigarettes among Twitter users: content and sentiment analysis. J Med Internet Res 2017 Feb 27;19(2):e56. [doi: 10.2196/jmir.5694] [Medline: 28242592]
110. Ayers JW, Westmaas JL, Leas EC, Benton A, Chen Y, Dredze M, et al. Leveraging big data to improve health awareness campaigns: a novel evaluation of the great American smokeout. JMIR Public Health Surveill 2016;2(1):e16. [doi: 10.2196/publichealth.5304] [Medline: 27227151]
111. de Viron S, Suggs LS, Brand A, Van Oyen H. Communicating genetics and smoking through social media: are we there yet? J Med Internet Res 2013 Sep 09;15(9):e198. [doi: 10.2196/jmir.2653] [Medline: 24018012]
112. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014 Jul 10;16(7):e169. [doi: 10.2196/jmir.3430] [Medline: 25014311]
113. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174. [doi: 10.2196/jmir.2534] [Medline: 23989137]
114. Rocheleau M, Sadasivam RS, Baquis K, Stahl H, Kinney RL, Pagoto SL, et al. An observational study of social and emotional support in smoking cessation Twitter accounts: content analysis of tweets. J Med Internet Res 2015 Jan 14;17(1):e18. [doi: 10.2196/jmir.3768] [Medline: 25589009]
115. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for coding tobacco-related Twitter data: a systematic review. J Med Internet Res 2017 Mar 31;19(3):e91. [doi: 10.2196/jmir.7022] [Medline: 28363883]
116. Staal YC, van de Nobelen S, Havermans A, Talhout R. New tobacco and tobacco-related products: early detection of product development, marketing strategies, and consumer interest. JMIR Public Health Surveill 2018 May 28;4(2):e55. [doi: 10.2196/publichealth.7359] [Medline: 29807884]
117. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. J Med Internet Res 2017 Jan 20;19(1):e24. [doi: 10.2196/jmir.5780] [Medline: 28108428]
118. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98. [doi: 10.2196/publichealth.8641] [Medline: 29263018]
119. Chen AT, Zhu S, Conway M. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. J Med Internet Res 2015 Sep 29;17(9):e220. [doi: 10.2196/jmir.4517] [Medline: 26420469]
120. Chu K, Sidhu AK, Valente TW. Electronic cigarette marketing online: a multi-site, multi-product comparison. JMIR Public Health Surveill 2015;1(2):e11. [doi: 10.2196/publichealth.4777] [Medline: 27227129]
121. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015 Oct 27;17(10):e243. [doi: 10.2196/jmir.4969] [Medline: 26508089]
122. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. J Med Internet Res 2015 Aug 25;17(8):e208. [doi: 10.2196/jmir.4392] [Medline: 26307512]
123. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014 Oct 16;16(10):e238. [doi: 10.2196/jmir.3622] [Medline: 25320863]
124. Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter data to gain insights into e-cigarette marketing and locations of use: an infoveillance study. J Med Internet Res 2015 Nov 06;17(11):e251. [doi: 10.2196/jmir.4466] [Medline: 26545927]
125. Kim A, Miano T, Chew R, Eggers M, Nonnemaker J. Classification of Twitter users who tweet about e-cigarettes. JMIR Public Health Surveill 2017 Sep 26;3(3):e63. [doi: 10.2196/publichealth.8060] [Medline: 28951381]
126. Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM. E-cigarette social media messages: a text mining analysis of marketing and consumer conversations on Twitter. JMIR Public Health Surveill 2016 Dec 12;2(2):e171. [doi: 10.2196/publichealth.6551] [Medline: 27956376]

127. Allem J, Ramanujam J, Lerman K, Chu K, Boley Cruz T, Unger JB. Identifying sentiment of hookah-related posts on Twitter. JMIR Public Health Surveill 2017 Oct 18;3(4):e74. [doi: 10.2196/publichealth.8133] [Medline: 29046267]
128. Cawkwell PB, Lee L, Weitzman M, Sherman SE. Tracking hookah bars in New York: utilizing Yelp as a powerful public health tool. JMIR Public Health Surveill 2015;1(2):e19. [doi: 10.2196/publichealth.4809] [Medline: 27227137]
129. Zhang Y, Allem J, Unger JB, Boley Cruz T. Automated identification of hookahs (waterpipes) on Instagram: an application in feature extraction using convolutional neural network and support vector machine classification. J Med Internet Res 2018 Nov 21;20(11):e10513. [doi: 10.2196/10513] [Medline: 30452385]
130. Allem J, Dharmapuri L, Leventhal AM, Unger JB, Boley Cruz T. Hookah-related posts to Twitter from 2017 to 2018: thematic analysis. J Med Internet Res 2018 Nov 19;20(11):e11669. [doi: 10.2196/11669] [Medline: 30455162]
131. Liu Y, Mei Q, Hanauer DA, Zheng K, Lee JM. Use of social media in the diabetes community: an exploratory analysis of diabetes-related tweets. JMIR Diabetes 2016 Nov 07;1(2):e4. [doi: 10.2196/diabetes.6256] [Medline: 30291053]
132. Martinez M, Park SB, Maison I, Mody V, Soh LS, Parihar HS. iOS appstore-based phone apps for diabetes management: potential for use in medication adherence. JMIR Diabetes 2017 Jul 11;2(2):e12. [doi: 10.2196/diabetes.6468] [Medline: 30291096]
133. Oser TK, Oser SM, McGinley EL, Stuckey HL. A novel approach to identifying barriers and facilitators in raising a child with type 1 diabetes: qualitative analysis of caregiver blogs. JMIR Diabetes 2017 Oct 26;2(2):e27. [doi: 10.2196/diabetes.8966] [Medline: 30291073]
134. Sinnenberg L, Mancheno C, Barg FK, Asch DA, Rivard CL, Horst-Martz E, et al. Content analysis of metaphors about hypertension and diabetes on Twitter: exploratory mixed-methods study. JMIR Diabetes 2018 Dec 21;3(4):e11177. [doi: 10.2196/11177] [Medline: 30578222]
135. Xu X, Litchman ML, Gee PM, Whatcott W, Chacon L, Holmes J, et al. Predicting prediabetes through Facebook postings: protocol for a mixed-methods study. JMIR Res Protoc 2018 Dec 14;7(12):e10720. [doi: 10.2196/10720] [Medline: 30552084]
136. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 09;16(4):e104. [doi: 10.2196/jmir.2968] [Medline: 24718852]
137. Koschack J, Weibezahl L, Friede T, Himmel W, Makedonski P, Grabowski J. Scientific versus experiential evidence: discourse analysis of the chronic cerebrospinal venous insufficiency debate in a multiple sclerosis forum. J Med Internet Res 2015 Jul 01;17(7):e159. [doi: 10.2196/jmir.4103] [Medline: 26133525]
138. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of treatment switching in multiple sclerosis therapies in US patients active on social media: application of social media content analysis to health outcomes research. J Med Internet Res 2016 Mar 17;18(3):e62. [doi: 10.2196/jmir.5409] [Medline: 26987964]
139. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23. [doi: 10.2196/medinform.7779] [Medline: 28760725]
140. Vasconcellos-Silva PR, Carvalho DBF, Trajano V, de La Rocque LR, Sawada ACMB, Juvanhol LL. Using Google Trends data to study public interest in breast cancer screening in Brazil: why not a pink February? JMIR Public Health Surveill 2017 Apr 06;3(2):e17. [doi: 10.2196/publichealth.7015] [Medline: 28385679]
141. Huesch M, Chetlen A, Segel J, Schetter S. Frequencies of private mentions and sharing of mammography and breast cancer terms on Facebook: a pilot study. J Med Internet Res 2017 Jun 09;19(6):e201. [doi: 10.2196/jmir.7508] [Medline: 28600279]
142. Jones J, Pradhan M, Hosseini M, Kulanthaivel A, Hosseini M. Novel approach to cluster patient-generated data into actionable topics: case study of a web-based breast cancer forum. JMIR Med Inform 2018 Nov 29;6(4):e45. [doi: 10.2196/medinform.9162] [Medline: 30497991]
143. Vickey T, Breslin JG. Online influence and sentiment of fitness tweets: analysis of two million fitness tweets. JMIR Public Health Surveill 2017 Oct 31;3(4):e82. [doi: 10.2196/publichealth.8507] [Medline: 29089294]
144. 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. [doi: 10.2196/10911] [Medline: 30530449]
145. Madden KM. The seasonal periodicity of healthy contemplations about exercise and weight loss: ecological correlational study. JMIR Public Health Surveill 2017 Dec 13;3(4):e92. [doi: 10.2196/publichealth.7794] [Medline: 29237582]
146. Wang H, Chen D. Economic recession and obesity-related internet search behavior in Taiwan: analysis of Google Trends data. JMIR Public Health Surveill 2018 Apr 06;4(2):e37. [doi: 10.2196/publichealth.7314] [Medline: 29625958]
147. Sugawara Y, Narimatsu H, Tsuya A, Tanaka A, Fukao A. Medical institutions and Twitter: a novel tool for public communication in Japan. JMIR Public Health Surveill 2016;2(1):e19. [doi: 10.2196/publichealth.4831] [Medline: 27227154]
148. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging social media to promote public health knowledge: example of cancer awareness via Twitter. JMIR Public Health Surveill 2016;2(1):e17. [doi: 10.2196/publichealth.5205] [Medline: 27227152]
149. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016 Aug 29;18(8):e232. [doi: 10.2196/jmir.6045] [Medline: 27573910]

150. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144. [doi: 10.2196/jmir.4343] [Medline: 26063290]
151. Lama Y, Chen T, Dredze M, Jamison A, Quinn SC, Broniatowski DA. Discordance between human papillomavirus Twitter images and disparities in human papillomavirus risk and disease in the United States: mixed-methods analysis. J Med Internet Res 2018 Sep 14;20(9):e10244. [doi: 10.2196/10244] [Medline: 30217792]
152. Mahoney LM, Tang T, Ji K, Ulrich-Schad J. The digital distribution of public health news surrounding the human papillomavirus vaccination: a longitudinal infodemiology study. JMIR Public Health Surveill 2015;1(1):e2. [doi: 10.2196/publichealth.3310] [Medline: 27227125]
153. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318. [doi: 10.2196/jmir.6670] [Medline: 27919863]
154. Nakada H, Yuji K, Tsubokura M, Ohsawa Y, Kami M. Development of a national agreement on human papillomavirus vaccination in Japan: an infodemiology study. J Med Internet Res 2014 May 15;16(5):e129. [doi: 10.2196/jmir.2846] [Medline: 24834471]
155. PRISMA. PRISMA statement URL: http://www.prisma-statement.org/PRISMAStatement/
156. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
157. Journal of Medical Internet Research. E-collection 'Infodemiology and Infoveillance' URL: https://www.jmir.org/themes/69
158. JMIR Public Health and Surveillance. E-collection 'Infoveillance, Infodemiology and Digital Disease Surveillance' URL: https://publichealth.jmir.org/collection/view/304

Abbreviations

e-cigarettes: electronic cigarettes
e-collections: electronic collections
HPV: human papillomavirus
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
STD: sexually transmitted disease

Edited by G Eysenbach; submitted 10.09.19; peer-reviewed by T Bernardo, C Liang, L Sun, W Marlicz; comments to author 02.02.20; revised version received 05.02.20; accepted 08.02.20; published 28.04.20.

Please cite as:
Mavragani A
Infodemiology and Infoveillance: Scoping Review
J Med Internet Res 2020;22(4):e16206
URL: http://www.jmir.org/2020/4/e16206/
doi:10.2196/16206
PMID:32310818

Viewpoint

Bringing Virtual Reality From Clinical Trials to Clinical Practice for the Treatment of Eating Disorders: An Example Using Virtual Reality Cue Exposure Therapy

Theresa Brown[1*], MS; Emily Nauman Vogel[1*], MS; Sarah Adler[2], PsyD; Cara Bohon[2], PhD; Kim Bullock[2], MD; Katherine Nameth[2], BS; Giuseppe Riva[3,4], PhD; Debra L Safer[2*], MD; Cristin D Runfola[2*], PhD

[1]PGSP-Stanford PsyD Consortium, Palo Alto, CA, United States

[2]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States

[3]Applied Technology for Neuro-Psychology Lab, Istituto Auxologico Italiano, Milan, Italy

[4]Centro Studi e Ricerche di Psicologia della Comunicazione, Università Cattolica del Sacro Cuore, Milan, Italy

[*]these authors contributed equally

Corresponding Author:
Debra L Safer, MD
Department of Psychiatry and Behavioral Sciences
Stanford University
401 Quarry Rd
Stanford, CA, 94304
United States
Phone: 1 650 498 9111
Fax: 1 650 723 9087
Email: dlsafer@stanford.edu

Abstract

Novel treatment options for eating disorders (EDs) are critically needed to enhance treatment outcomes and reduce the rates of treatment dropouts. On average, only 50% of individuals receiving evidence-based care remit, whereas 24% drop out before treatment completion. One particularly promising direction involves integrating virtual reality (VR) with existing evidence-based treatments (EBTs) such as cue exposure therapy (CET). Across psychiatric disorders, VR-based interventions are demonstrating at least preliminary efficacy and noninferiority to traditional treatments. Furthermore, VR technology has become increasingly portable, resulting in improved acceptance, increased access, and reductions in cost. However, more efficient research processes may be needed to uncover the potential benefits of these rapid technological advances. This viewpoint paper reviews existing empirical support for integrating VR with EBTs (with a focus on its use with EDs) and proposes key next steps to more rapidly bring this innovative technology-based intervention into real-world clinic settings, as warranted. VR-CET for EDs is used to illustrate a suggested process for developing such treatment enhancements. We recommend following a deployment-focused model of intervention development and testing to enable rapid implementation of robust, practice-ready treatments. In addition, our review highlights the need for a comprehensive clinical protocol that supports clinicians and researchers in the implementation and testing of VR-CET and identifies key missing protocol components with rationale for their inclusion. Ultimately, this work may lead to a more complete understanding of the full potential of the applications and integrations of VR into mental health care globally.

(J Med Internet Res 2020;22(4):e16386) doi:10.2196/16386

KEYWORDS

virtual reality; exposure therapy; eating disorders; translational research; technological innovation

Introduction

Eating disorders (EDs) pose serious risks to psychological and physical health [1,2], resulting in an elevated risk of death for those with EDs compared with the general population [3-5]. Although evidence-based treatments (EBTs) for EDs demonstrate reasonable efficacy, on average, only 50% of individuals receiving evidence-based care remit fully [6-8], and 24% are estimated to drop out before the completion of

treatment [9]. Innovation is needed to improve treatment outcomes.

Virtual reality (VR) is a technology that can augment existing EBTs with the possibility of enhancing treatment outcomes. Through the creation of immersive computer-generated experiences, users interact naturally with stimuli representing the real world while simultaneously benefiting from a clinical, supervised setting [10]. Moreover, VR shares with the brain the same basic mechanism: embodied simulations [11]. In fact, a VR system, similar to the brain, maintains a model (simulation) of the body and the space around it.

These features offer different potential advantages for using VR to augment existing EBTs for EDs [12-16]. Advantages may include automating and standardizing ED psychoeducation, opportunities to practice emotion regulation skills, reprogramming attentional biases, enhancing insight by identifying body distortion, changing implicit and explicit perceptual bodily distortions, reducing weight stigma and biases, enhancing empathy among support persons, and augmenting exposure therapy. In particular, three different randomized controlled trials [14,17,18] have shown at long-term follow-ups (6 and 12 months) that VR-enhanced cognitive behavioral therapy (CBT) for EDs had a higher efficacy on some outcomes than CBT alone (eg, greater improvement in body image disturbances; increased reduction in the frequency of binge, purge, and overeating episodes; and more weight loss for obese individuals with binge eating disorder, BED).

Nevertheless, traditional models for testing treatments, which include case series and randomized controlled trials to test efficacy under highly controlled research conditions in academic centers, followed by effectiveness and implementation studies, may render these rapidly changing technologies obsolete by the time they are clinic-ready. For example, since the early 2000s, immersive VR has become increasingly sophisticated with improved 3-dimensional constructed environments; compatibility with mobile phones; and a wide variety of head-mounted display devices equipped with increased field of view, higher-resolution images, and lightweight and comfortable designs [19]. Testing these interventions within real-world clinic settings quickly, before they are replaced by newer technology, may be crucial to benefit from the existing content and devices [20].

In this paper, we discuss the great potential of VR-based interventions in the treatment of psychiatric disorders and research efforts needed to support the development of effective clinic-ready treatments in a timely manner. Of the many applications of VR in the treatment of EDs [21,22], virtual reality cue exposure therapy (VR-CET) has the strongest empirical support at present [18,23-25]. Therefore, we focus on VR-CET to anchor the discussion. Specifically, this paper provides: (1) an overview of VR-CET and its potential advantages to augment EBTs; (2) a brief review of the empirical support for VR-based exposures in psychiatric disorders, with a focus on VR-CET for EDs; (3) an argument for small-scale effectiveness trials early in treatment development (following the deployment-focused model of intervention development and testing by Weisz et al [20]); and (4) ideas on how VR-CET can be more rapidly developed and implemented in real-world clinic settings, highlighting key missing protocol components, with rationale for their inclusion.

Advantages of Virtual Reality–Cue Exposure Therapy

VR-CET provides traditional cue exposure therapy (CET) in a virtual environment. CET is a subcategory of exposure therapy that is specific to eating and substance-related disorders. Both techniques involve repeated, controlled exposures to relevant stimuli. However, CET was initially specifically designed to target cravings (vs fear) [26]. CET is based on classical conditioning, a learning theory that explains how maladaptive behavior can develop in response to previously neutral stimuli. According to classical conditioning, repeatedly pairing a neutral stimulus (conditioned stimuli) with a stimulus (unconditioned stimulus) that naturally evokes a biologically potent response (unconditioned response) may eventually result in responding similarly to the neutral stimulus as one would for the biologically potent stimulus (conditioned response). For example, in the classical conditioning model of binge eating, Jansen [27] conceptualizes the intake of food as the unconditioned stimulus and its metabolic effects as the unconditioned response. Cues that reliably signal food intake, such as sight, smell, taste, and the context in which one eats, act as conditioned stimuli. The presence of these cues (conditioned stimuli) elicits physiological responses that are experienced as craving (ie, an almost irresistible urge to eat), which can increase the probability of binge episodes (conditioned response) [28-31]. In CET, the main objective is to weaken the bond between the cues (conditioned stimuli) and the maladaptive responses (conditioned responses), which may mask or inhibit initial learning [32]. Specifically, VR can reduce eating-related anxiety during and after exposure to virtual food, helping to disrupt the reconsolidation of adverse, food-related memories [24,26].

In VR-CET for EDs, patients are repeatedly exposed to emotionally provoking eating-related situations that typically result in maladaptive behaviors (eg, binge eating and avoidance). Exposure is planned gradually and designed to eliminate the ED behavior. New associations develop with eating-related anxiety and cravings in response to stimuli decreasing over time [24,33]. Emotional changes may occur, in part, because of the modifications to dysfunctional thinking and increased self-efficacy [34]. Other VR exposure programs created in the context of CBT extend the exposure activities to include cognitive restructuring and practice using alternative emotion regulation strategies in response to triggering stimuli. These strategies may help address core ED symptomatology [31].

VR offers several promising advantages to in vivo exposure that may result in reduced therapist burden, more rapid symptom improvement, improved acceptance of treatment, reduced treatment dropouts, and more accurate measurement-based care. VR environments may more closely approximate the settings in which problematic eating behaviors take place compared with the clinician's office or imaginal exercises. A therapist can manipulate a larger number of stimuli within the VR

environment than in the real world (eg, intensity of the feared stimulus) and personalize both contextual cues (eg, social setting and room type) and sensorial cues (eg, smell, sound, and tactile effects). VR environments can be tailored precisely to a patient's fear hierarchy (ie, the client's rank order list of cues that elicit anxiety on a scale from lowest to highest). Extensive evidence exists showing VR environments produce emotional, behavioral, and physiological responses in patients similar to those observed in real-life situations [15]. Thus, VR enables the therapist to guide the patient through an exposure or use skills within a more ecologically valid environment, efficiently challenging dysfunctional mental representations [34] and potentially facilitating improved application of the acquired skills and subsequent generalization. In addition, VR-based exposures may be a more acceptable strategy than the traditional leap from an in-office imaginal exposure to an in vivo exposure. Gradually moving from virtual to in vivo exposures offers a middle step and a more palliative intervention that may reduce treatment dropout rates [31,33]. The appeal of a new and exciting technology may also contribute to the patients' willingness to engage with exposures in VR. Indeed, one study found that 89% of patients with specific phobia chose VR over in vivo exposures when given a choice [35]. Preference for VR over in vivo exposures may very well be a manifestation of avoidance of the feared stimuli in patients with specific phobia. For example, about 25% of patients with specific phobia refuse exposure therapy [35]. However, reviews [36,37] show limited evidence that safety behaviors may not be detrimental to exposure efficacy based on the inhibitory learning model and thus conclude that judicious use of safety behavior can be clinically indicated. It is recommended to eliminate safety behavior as soon as the clients are willing. Thus, in cases in which avoidance appears to drive the patient's rationale for in virtuo exposure, VR may serve as a stepping-stone to in vivo exposure for patients who would not otherwise receive care. Furthermore, a review of VR in the treatment of EDs indicated increased motivation for change in VR treatments and lower rates of loss to follow-up compared with in vivo active comparisons across several studies [10]. Finally, data collection and measurements can occur in session more effectively when using VR. This can eliminate potential adherence and validity problems with in vivo exposure compliance tracking. VR can enable tracking of many ecologically valid variables (eg, level of anxiety and urge to binge), which can be recorded and viewed in real time [38].

Unfortunately, studies indicate that many therapists hold negative beliefs about in vivo exposure therapy for anxiety disorders (eg, exposure may harm patients, result in symptom exacerbation, or harm therapists because of vicarious trauma), which may pose a barrier to treatment dissemination [39]. In contrast, there is growing interest and increasing acceptability among therapists regarding VR-based exposure. For instance, recent studies suggest that therapists have an overall positive attitude toward VR exposure therapy (pros rated higher than cons) and view VR as applicable to conditions other than anxiety [40], including EDs [41]. Lindner et al [40] also noted that high financial costs and technical difficulties were no longer top-rated negative aspects with the release of consumer VR platforms. However, the authors note that VR has not yet been widely implemented in routine care, for unclear reasons.

Empirical Support for Virtual Reality–Based Exposure With a Focus on Eating Disorders

The intention of this paper is not to systematically review the literature on VR-based exposure therapy because that would be outside the scope of this viewpoint paper. Instead, we provide an overview of the empirical support for VR-based exposure with a focus on EDs to highlight the foundational work and provide context for our arguments in the subsequent sections. A substantial body of research supports the efficacy of VR-CET and VR-based exposures for numerous psychiatric disorders, notably those that frequently co-occur with EDs such as anxiety, substance use disorders, and posttraumatic stress disorder (PTSD) [42-44]. VR exposure therapy is consistently more effective than nonactive control groups (ie, waitlist), whereas comparisons between VR exposure therapy and manualized CBT and/or in vivo exposures tend to show equivalence in outcomes [43-49]. Considerable research has focused on anxiety disorders (eg, social phobia, panic disorder with agoraphobia, and specific phobias), with meta-analytic studies reporting large effect sizes (eg, Cohen d=0.95, ranging from 0.87 in PTSD to 1.79 in panic disorder with agoraphobia) [50]. Meta-analyses also reveal moderate effect sizes for VR exposure therapy over waitlist controls in PTSD [48,49] and show the benefit of VR exposure therapy in the assessment or treatment of acute stress disorder and paranoia [51]. These reviews note that although more research is needed to better understand the mechanisms of action, VR-specific variables such as a sense of presence within virtual environments may affect treatment outcomes [43,45]. Additional important takeaways are that modern VR systems are becoming increasingly affordable, accessible, and user-friendly [44] and that VR may be especially useful in exploring hypotheses related to the processes and mechanisms involved in exposure therapy because of the high degree of control and manipulation of specific variables that this technology allows [45]. Finally, the existing reviews illustrate the insufficient state of the current literature, given inconsistent reporting of key variables [43] and the fact that although studies have been conducted in controlled research contexts, research within real-world clinical settings are lacking [45,51].

Compared with research on the use of VR-based exposures for anxiety and related disorders, the number of studies that have specifically investigated the use of VR-CET for EDs is small. With so few individual studies conducted to date, publication of meta-analyses or reviews of VR-CET for EDs has not yet been warranted. However, the limited research conducted thus far is favorable. For example, patients with bulimia nervosa (BN) and BED who remained symptomatic after CBT showed a significantly greater reduction in binge eating and higher percentages of abstinence from binge eating and purging after randomization to six sessions of VR-CET vs six sessions of additional CBT [23]. Results from a 6-month follow-up of this same study revealed that reductions in binge, purge, and overeating episodes were greater after treatment with VR-CET [18]. Such results suggest VR-CET is not only efficacious posttreatment but appears to have lasting effects.

A total of three case reports using a similar nonimmersive VR-CET program as a complementary tool to CBT demonstrated positive effects with patients diagnosed with restrictive anorexia nervosa (AN-R), binge/purge anorexia nervosa (AN-B/P), and BN [52-54]. All patients reported lower levels of anxiety as well as reduced frequency of safety and avoidance behaviors related to food after completing a short (six or seven sessions) VR-CET module. Notably, the patient diagnosed with AN-R increased her BMI from 15 kg/m^2 to 16.8 kg/m^2 [52], the patient diagnosed with AN-B/P reduced her binge/purge behaviors from two to three times per day to once per week [53], and the patient diagnosed with BN completely eliminated binge and purge episodes [54]. Furthermore, the patients in these case reports reported that VR treatment was acceptable and helpful [52-54]. Of note, VR exposure was particularly helpful for exploring thoughts and emotions experienced in the moment while *eating* a virtual food, enabling productive therapeutic discussions to take place in an "'ecological' environment, one that was clinically significant but also safe" [53].

Given the aforementioned research, VR-CET for EDs may have value for advancing the field of EDs. The patients' responses to VR-CET outlined in the research above are rapid. Faster acting treatments may reduce treatment length and, as a result, reduce clinic wait times—thereby positively addressing access to care issues. Given these potential advantages, we discuss the importance of pilot testing in real-world clinical settings at this stage in the developmental process.

Importance of Small-Scale Effectiveness Trials

A recent systematic review of VR in the assessment, understanding, and treatment of mental health disorders broadly argues that the progress in implementing and disseminating VR applications has been slow despite growing interest [55]. The authors propose that VR in mental health care could be *revolutionary* in that the results of VR treatment may surpass those of a standard course of EBT [55]. Reilly et al [56] specifically argue for the expansion of exposure approaches and techniques, including VR-CET, for EDs. However, to date, the studies employing VR-CET for EDs have been conducted in Europe under the highly controlled settings of research trials. To translate this research into clinical practice, VR-CET needs to be tested in real-world clinic settings.

Historically, across disorders, implementation of VR in treatment was limited by the cost of technology and concerns about acceptability (eg, motion sickness experienced by a high percentage of participants) [57,58]. However, over time, VR technology has become increasingly portable (with the advent of handheld devices), resulting in improved acceptance, increased access, and reduction in cost. Market trends suggest that the demand for VR will rise steeply, with estimates of 55 million headset orders for 2022 alone [59], making VR nearly ubiquitous in American homes, similar to the personal computer. This increased access increases the likelihood of bringing VR-CET into the clinic, with supporting use at home. Investigation of how best to translate the existing research into real-world application is timely.

Given the limited number of efficacy trials of VR-CET for EDs to date, some may argue that translating this treatment into clinical practice at this stage of development is premature. In traditional models of EBT development, efficacy trials that provide substantial evidence for the treatment's success under controlled conditions are conducted before studying the treatment in real-world settings. However, as other researchers have observed [20], the gap between research and real-world conditions is sometimes so big that regardless of the treatment's robustness under controlled conditions, the treatment is not able to endure the conditions of the real world; the task of closing this gap at the end of a series of efficacy trials becomes unnecessarily complex and inefficient. As the former National Institute of Mental Health (NIMH) Director Thomas Insel [60] pointed out, interventions developed in highly controlled research settings often fail to take into account the key characteristics of the patients, providers, and settings in which the intervention will ultimately be implemented, which limits their practical value. For example, patients with comorbid issues are often excluded, and interventions are conducted by highly trained and closely supervised clinicians. However, in practice settings, comorbid issues tend to be the rule rather than the exception, and up to 40% of mental health providers in the public sector do not have a graduate or professional degree [60]. In addition, setting factors such as billing constraints and typical patterns of patient service use (eg, number of sessions typically attended in the setting) tend to be overlooked. As a result, treatments developed under highly controlled settings may be scientifically valid but not necessarily clinically meaningful or generalizable to real-world clinic settings [60].

As an alternative to traditional models of treatment development, Weisz et al [20] describe a "deployment-focused model of intervention development and testing" that integrates testing of treatments in practice settings early and throughout the treatment development process, rather than as a final phase, to ensure that they are applicable to and successful in the settings in which they will be delivered. Notably, the NIMH also shows interest in supporting clinic-based treatment development approaches in the NIMH Strategic Plan for Research [61]. We argue that deployment-focused model is not only particularly applicable to VR-CET but also an integral component of its successful development and testing—particularly given the likelihood of the technologies involved to become obsolete. Estimates of the time lag between when a treatment is initially developed and when it reaches the stage of dissemination, following the traditional research models, average about 17 years [62]. We thus identify a next direction for researchers and practitioners to translate this research into practice.

Translating Virtual Reality–Based Clinical Trials to Clinical Practice: Future Directions

Weisz et al [20] propose that the first stage of deployment-focused treatment development involves the

creation of a treatment protocol. Indeed, bringing VR-CET into clinical practice at this stage of development requires the creation of a protocol that is easily disseminated; adaptable across cultures and settings; and clearly outlines acceptability, feasibility, and implementation factors. As such, characteristics of providers and the usual settings in which care is provided must be considered. Many factors that would inform the creation of this protocol and ensure that VR-CET is practice-ready are noted in the literature as important but not consistently reported or described in sufficient detail. More comprehensive and consistent reporting of these factors has the potential to inform a VR-CET protocol that would support researchers and clinicians in becoming comfortable in implementing a technological innovation that has potential to advance the treatment for EDs. These factors include (1) patient and clinician reactions to and satisfaction with VR technology [40,63]; (2) logistical parameters such as space required, costs, and institutional buy-in [63]; (3) considerations in creating VR content, including working with software and/or third-party vendors and the optimal level of *presence* (the extent to which the individual interacts with the VR environment as if it is reality) for treatment effects to occur [64,65]; (4) clinical guidelines and outcome data, including outcomes compared with traditional EBTs, how to determine patient suitability for VR, optimal frequency and number of sessions, and the amount of time within the VR environment [66-68]; and (5) parameters regarding safety and acceptability, including cybersickness (side effects resulting from our physiological motion detection systems when in VR) [66]. Addressing these factors in a treatment protocol and providing examples for how to successfully execute each point will help move VR-CET from research settings to clinical practice.

In addition, future research should take advantage of the unique methods of data collection and assessment available within VR, including the collection of real-time self-reported data in virtuo and biological measurements that can track eye movements, facial gestures, and the movement of body parts. Given the large number of stimuli that can be manipulated and tightly controlled within VR environments (eg, intensity of stimuli, contextual, and sensorial cues), basic science research can utilize VR applications to advance the understanding of ED mechanisms of change.

Given that we are suggesting a deployment-focused model of treatment development for a novel iteration (VR-based) of an already validated therapy (CET), it is important to consider the risks compared with potential benefits of treatment implementation. One important risk associated with the utilization of any VR platform includes cybersickness, a side effect that 20% to 80% of VR users may experience [69,70]. The symptoms of cybersickness are similar to motion sickness and can include nausea, headaches, and dizziness [71]. For most people, cybersickness occurs about 15 min into the VR immersion, is worst in the first session, and becomes negligible by the third session [66]. In addition, clear procedures to address cybersickness within VR-based exposure protocols can mitigate the risks associated with these unpleasant symptoms and should always be included when implementing VR-based therapies [66]. Another risk is the potential loss of effectiveness when translating an already validated therapy into VR-based therapy. A recent review examined the negative effects of VR-based treatments for anxiety disorders using deterioration rates (rates of worsening symptomatology) as the primary outcome [72]. Deterioration rates for VR were found to coincide with other therapeutic approaches; the authors concluded that VR appears to be a nondeleterious treatment for patients with anxiety disorders [72]. In addition, meta-analytic studies show that VR exposure therapy for anxiety-related and trauma-related or stress-related disorders is not significantly less effective than in vivo treatment [43-49].

Overall, the evidence thus far suggests minimal known risk associated with VR exposure therapy, which may reduce concern for deployment-focused treatment development. In addition, the standard of care in clinical practice involves the use of an informed consent process by which patients are provided with clear information regarding their treatment options, including the relative risks and potential benefits of each approach. Given the minimal known risks associated with VR-based treatments and the ability of VR-based exposures to provide an acceptable and effective treatment option for those who refuse, drop out of, or do not progress via in vivo exposures, we believe the potential benefits outweigh the costs of offering VR-based exposures as an alternative treatment option. To further minimize any associated risks of offering a novel treatment alternative to an already validated treatment approach, we suggest that the real-world trials of this technology should assess the patients' progress and suggest in vivo exposures or a different active treatment if a patient is not making progress.

Given evidence that integrating VR technology with EBTs for EDs such as CBT leads to significantly improved outcomes [12,15], with faster effects and better maintenance than standard treatment alone [12,16], following the above recommendations to create and implement a comprehensive VR-CET protocol may help make evidence-based care more accessible and cost-effective for patients with an ED. We hypothesize that implementing VR-CET in real-world clinical settings may also foster increased patient participation in and excitement about treatment through their involvement in designing VR-based exposures. Furthermore, a patient's experience of success (eg, tolerating the urge to binge) within VR may increase their perception of self-efficacy, enhancing confidence to translate therapy skills to real-world experiences.

Overall, comprehensive study of VR-CET in real-world clinic settings is a fruitful direction that may advance intervention protocols for EDs. In particular, a clear protocol for VR-CET for EDs will help translate the exciting research supporting the applications of VR into a clinic-ready intervention, providing a model use and, ultimately, a more comprehensive understanding of the full potential of the applications of VR on mental health care globally. We suggest that continued research efforts focus on advancing VR-CET following the clinic-based treatment development approach to more rapidly move technology-based interventions from research settings into the real world, as warranted.

Acknowledgments

The authors thank Marta Ferrer-Garcia, PhD, and Realiteer Corp for their respective contributions to this research as well as Micro-Star International Co, Ltd, for their product loan. Drs Runfola, Adler, Safer, and Bullock would like to acknowledge the financial support from the National Eating Disorders Association. Giuseppe Riva would like to acknowledge the financial support by the Italian Ministry of Education, University and Research research project *Unlocking the memory of the body: Virtual Reality in Anorexia Nervosa* (201597WTTM).

Authors' Contributions

TB and EV should be considered joint first author. Drs CR and DS should be considered joint senior author.

Conflicts of Interest

None declared.

References

1. Kaye WH, Klump KL, Frank GK, Strober M. Anorexia and bulimia nervosa. Annu Rev Med 2000;51:299-313. [doi: 10.1146/annurev.med.51.1.299] [Medline: 10774466]
2. Smink FR, van Hoeken D, Hoek HW. Epidemiology, course, and outcome of eating disorders. Curr Opin Psychiatry 2013 Nov;26(6):543-548. [doi: 10.1097/YCO.0b013e328365a24f] [Medline: 24060914]
3. Chesney E, Goodwin GM, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. World Psychiatry 2014 Jun;13(2):153-160 [FREE Full text] [doi: 10.1002/wps.20128] [Medline: 24890068]
4. Papadopoulos FC, Ekbom A, Brandt L, Ekselius L. Excess mortality, causes of death and prognostic factors in anorexia nervosa. Br J Psychiatry 2009 Jan;194(1):10-17. [doi: 10.1192/bjp.bp.108.054742] [Medline: 19118319]
5. Suokas JT, Suvisaari JM, Gissler M, Löfman R, Linna MS, Raevuori A, et al. Mortality in eating disorders: a follow-up study of adult eating disorder patients treated in tertiary care, 1995-2010. Psychiatry Res 2013 Dec 30;210(3):1101-1106. [doi: 10.1016/j.psychres.2013.07.042] [Medline: 23958333]
6. Brownley KA, Berkman ND, Peat CM, Lohr KN, Cullen KE, Bann CM, et al. Binge-eating disorder in adults: A systematic review and meta-analysis. Ann Intern Med 2016 Sep 20;165(6):409-420 [FREE Full text] [doi: 10.7326/M15-2455] [Medline: 27367316]
7. Knott S, Woodward D, Hoefkens A, Limbert C. Cognitive behaviour therapy for bulimia nervosa and eating disorders not otherwise specified: Translation from randomized controlled trial to a clinical setting. Behav Cogn Psychother 2015 Nov;43(6):641-654. [doi: 10.1017/S1352465814000393] [Medline: 25331090]
8. Södersten P, Bergh C, Leon M, Brodin U, Zandian M. Cognitive behavior therapy for eating disorders versus normalization of eating behavior. Physiol Behav 2017 May 15;174:178-190 [FREE Full text] [doi: 10.1016/j.physbeh.2017.03.016] [Medline: 28322911]
9. Linardon J, Hindle A, Brennan L. Dropout from cognitive-behavioral therapy for eating disorders: A meta-analysis of randomized, controlled trials. Int J Eat Disord 2018 May;51(5):381-391. [doi: 10.1002/eat.22850] [Medline: 29493805]
10. Clus D, Larsen ME, Lemey C, Berrouiguet S. The use of virtual reality in patients with eating disorders: systematic review. J Med Internet Res 2018 Apr 27;20(4):e157 [FREE Full text] [doi: 10.2196/jmir.7898] [Medline: 29703715]
11. Riva G, Wiederhold BK, Mantovani F. Neuroscience of virtual reality: From virtual exposure to embodied medicine. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):82-96 [FREE Full text] [doi: 10.1089/cyber.2017.29099.gri] [Medline: 30183347]
12. Riva G, Bacchetta M, Cesa G, Conti S, Molinari E. Six-month follow-up of in-patient experiential cognitive therapy for binge eating disorders. Cyberpsychol Behav 2003 Jun;6(3):251-258. [doi: 10.1089/109493103322011533] [Medline: 12855080]
13. Riva G, Bacchetta M, Cesa G, Conti S, Molinari E. The use of VR in the treatment of eating disorders. Stud Health Technol Inform 2004;99:121-163. [Medline: 15295149]
14. Marco JH, Perpiñá C, Botella C. Effectiveness of cognitive behavioral therapy supported by virtual reality in the treatment of body image in eating disorders: one year follow-up. Psychiatry Res 2013 Oct 30;209(3):619-625. [doi: 10.1016/j.psychres.2013.02.023] [Medline: 23499231]
15. de Carvalho MR, Dias TR, Duchesne M, Nardi AE, Appolinario JC. Virtual reality as a promising strategy in the assessment and treatment of bulimia nervosa and binge eating disorder: A systematic review. Behav Sci (Basel) 2017 Jul 9;7(3):pii: E43 [FREE Full text] [doi: 10.3390/bs7030043] [Medline: 28698483]
16. Gutiérrez-Malsonado J, Ferrer-García M, Dakanalis A, Riva G. Virtual reality: applications to eating disorders. In: Agras WS, Robinson A, editors. The Oxford Handbook of Eating Disorders. Second Edition. New York: Oxford University Press; 2018.
17. Cesa GL, Manzoni GM, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality for enhancing the cognitive behavioral treatment of obesity with binge eating disorder: randomized controlled study with one-year follow-up. J Med Internet Res 2013 Jun 12;15(6):e113 [FREE Full text] [doi: 10.2196/jmir.2441] [Medline: 23759286]

18. Ferrer-Garcia M, Pla-Sanjuanelo J, Dakanalis A, Vilalta-Abella F, Riva G, Fernandez-Aranda F, et al. A randomized trial of virtual reality-based cue exposure second-level therapy and cognitive behavior second-level therapy for bulimia nervosa and binge-eating disorder: Outcome at six-month followup. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):60-68. [doi: 10.1089/cyber.2017.0675] [Medline: 30059240]
19. Jerdan SW, Grindle M, van Woerden HC, Boulos MN. Head-mounted virtual reality and mental health: Critical review of current research. JMIR Serious Games 2018 Jul 6;6(3):e14 [FREE Full text] [doi: 10.2196/games.9226] [Medline: 29980500]
20. Weisz JR, Chu BC, Polo AJ. Treatment dissemination and evidence-based practice: Strengthening intervention through clinician-researcher collaboration. Clin Psychol Sci Pract 2004;11(3):300-307. [doi: 10.1093/clipsy.bph085]
21. Serino S, Chirico A, Pedroli E, Polli N, Cacciatore C, Riva G. Two-phases innovative treatment for anorexia nervosa: The potential of virtual reality body-swap. Ann Rev CyberTher Telemed 2017;15:111-115 [FREE Full text]
22. Serino S, Polli N, Riva G. From avatars to body swapping: The use of virtual reality for assessing and treating body-size distortion in individuals with anorexia. J Clin Psychol 2019 Feb;75(2):313-322 [FREE Full text] [doi: 10.1002/jclp.22724] [Medline: 30552669]
23. Ferrer-García M, Gutiérrez-Maldonado J, Pla-Sanjuanelo J, Vilalta-Abella F, Riva G, Clerici M, et al. A randomised controlled comparison of second-level treatment approaches for treatment-resistant adults with bulimia nervosa and binge eating disorder: Assessing the benefits of virtual reality cue exposure therapy. Eur Eat Disord Rev 2017 Nov;25(6):479-490. [doi: 10.1002/erv.2538] [Medline: 28804985]
24. Pla-Sanjuanelo J, Ferrer-García M, Gutiérrez-Maldonado J, Riva G, Andreu-Gracia A, Dakanalis A, et al. Identifying specific cues and contexts related to bingeing behavior for the development of effective virtual environments. Appetite 2015 Apr;87:81-89. [doi: 10.1016/j.appet.2014.12.098] [Medline: 25526828]
25. Pla-Sanjuanelo J, Ferrer-García M, Vilalta-Abella F, Riva G, Dakanalis A, Ribas-Sabaté J, et al. Testing virtual reality-based cue-exposure software: which cue-elicited responses best discriminate between patients with eating disorders and healthy controls? Eat Weight Disord 2019 Aug;24(4):757-765. [doi: 10.1007/s40519-017-0419-4] [Medline: 28752497]
26. Koskina A, Campbell IC, Schmidt U. Exposure therapy in eating disorders revisited. Neurosci Biobehav Rev 2013 Feb;37(2):193-208. [doi: 10.1016/j.neubiorev.2012.11.010] [Medline: 23201859]
27. Jansen A. A learning model of binge eating: cue reactivity and cue exposure. Behav Res Ther 1998 Mar;36(3):257-272. [doi: 10.1016/s0005-7967(98)00055-2] [Medline: 9642846]
28. Boswell RG, Kober H. Food cue reactivity and craving predict eating and weight gain: a meta-analytic review. Obes Rev 2016 Feb;17(2):159-177 [FREE Full text] [doi: 10.1111/obr.12354] [Medline: 26644270]
29. García-García I, Narberhaus A, Marqués-Iturria I, Garolera M, Rădoi A, Segura B, et al. Neural responses to visual food cues: insights from functional magnetic resonance imaging. Eur Eat Disord Rev 2013 Mar;21(2):89-98. [doi: 10.1002/erv.2216] [Medline: 23348964]
30. Lyu Z, Zheng P, Jackson T. Attention disengagement difficulties among average weight women who binge eat. Eur Eat Disord Rev 2016 Jul;24(4):286-293. [doi: 10.1002/erv.2438] [Medline: 26856539]
31. Wolz I, Sauvaget A, Granero R, Mestre-Bach G, Baño M, Martín-Romera V, et al. Subjective craving and event-related brain response to olfactory and visual chocolate cues in binge-eating and healthy individuals. Sci Rep 2017 Feb 3;7:41736 [FREE Full text] [doi: 10.1038/srep41736] [Medline: 28155875]
32. Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B. Maximizing exposure therapy: an inhibitory learning approach. Behav Res Ther 2014 Jul;58:10-23 [FREE Full text] [doi: 10.1016/j.brat.2014.04.006] [Medline: 24864005]
33. Bulik CM, Sullivan PF, Carter FA, McIntosh VV, Joyce PR. The role of exposure with response prevention in the cognitive-behavioural therapy for bulimia nervosa. Psychol Med 1998 May;28(3):611-623. [doi: 10.1017/s0033291798006618] [Medline: 9626717]
34. Albano AM, Botella C, Bouchard S, Draheim A, Hoffman L, Khanna M. More than Meets the Eye: Implications of Virtual Reality for Improving and Disseminating Anxiety and Depression Treatment. 2019 Presented at: ABCT 53rd Annual Convention; November 21-24, 2019; Atlanta URL: https://www.eventscribe.com/2019/ABCT/fsPopup.asp?Mode=sessioninfo&PresentationID=603846
35. Garcia-Palacios A, Hoffman HG, See SK, Tsai A, Botella C. Redefining therapeutic success with virtual reality exposure therapy. Cyberpsychol Behav 2001 Jun;4(3):341-348. [doi: 10.1089/109493101300210231] [Medline: 11710258]
36. Blakey SM, Abramowitz JS. The effects of safety behaviors during exposure therapy for anxiety: Critical analysis from an inhibitory learning perspective. Clin Psychol Rev 2016 Nov;49:1-15. [doi: 10.1016/j.cpr.2016.07.002] [Medline: 27475477]
37. Weisman JS, Rodebaugh TL. Exposure therapy augmentation: A review and extension of techniques informed by an inhibitory learning approach. Clin Psychol Rev 2018 Feb;59:41-51. [doi: 10.1016/j.cpr.2017.10.010] [Medline: 29128146]
38. Parsons TD. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci 2015;9:660 [FREE Full text] [doi: 10.3389/fnhum.2015.00660] [Medline: 26696869]
39. Deacon BJ, Farrell NR, Kemp JJ, Dixon LJ, Sy JT, Zhang AR, et al. Assessing therapist reservations about exposure therapy for anxiety disorders: the Therapist Beliefs about Exposure Scale. J Anxiety Disord 2013 Dec;27(8):772-780. [doi: 10.1016/j.janxdis.2013.04.006] [Medline: 23816349]

40. Lindner P, Miloff A, Zetterlund E, Reuterskiöld L, Andersson G, Carlbring P. Attitudes toward and familiarity with virtual reality therapy among practicing cognitive behavior therapists: A cross-sectional survey study in the era of consumer VR platforms. Front Psychol 2019;10:176 [FREE Full text] [doi: 10.3389/fpsyg.2019.00176] [Medline: 30800086]
41. Nauman E, Brown T, Runfola C, Safer D, Bullock K, Nameth K, et al. Virtual reality cue exposure therapy for eating disorders: development of a culturally specific manual in the U.S. for use in a real world clinic setting. 2019 Presented at: International Conference on Eating Disorders; March 14-16, 2019; New York URL: https://www.aedweb.org/aed-events/previous-conferences/iced-2019
42. Hone-Blanchet A, Wensing T, Fecteau S. The use of virtual reality in craving assessment and cue-exposure therapy in substance use disorders. Front Hum Neurosci 2014;8:844 [FREE Full text] [doi: 10.3389/fnhum.2014.00844] [Medline: 25368571]
43. Turner WA, Casey LM. Outcomes associated with virtual reality in psychological interventions: where are we now? Clin Psychol Rev 2014 Dec;34(8):634-644. [doi: 10.1016/j.cpr.2014.10.003] [Medline: 25455627]
44. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
45. Botella C, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños R. Recent progress in virtual reality exposure therapy for phobias: A systematic review. Curr Psychiatry Rep 2017 Jul;19(7):42. [doi: 10.1007/s11920-017-0788-4] [Medline: 28540594]
46. Wechsler TF, Kümpers F, Mühlberger A. Inferiority or even superiority of virtual reality exposure therapy in phobias? - A systematic review and quantitative meta-analysis on randomized controlled trials specifically comparing the efficacy of virtual reality exposure to gold standard exposure in agoraphobia, specific phobia, and social phobia. Front Psychol 2019;10:1758 [FREE Full text] [doi: 10.3389/fpsyg.2019.01758] [Medline: 31551840]
47. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
48. Kothgassner OD, Goreis A, Kafka JX, van Eickels RL, Plener PL, Felnhofer A. Virtual reality exposure therapy for posttraumatic stress disorder (PTSD): a meta-analysis. Eur J Psychotraumatol 2019;10(1):1654782 [FREE Full text] [doi: 10.1080/20008198.2019.1654782] [Medline: 31489138]
49. Deng W, Hu D, Xu S, Liu X, Zhao J, Chen Q, et al. The efficacy of virtual reality exposure therapy for PTSD symptoms: A systematic review and meta-analysis. J Affect Disord 2019 Oct 1;257:698-709. [doi: 10.1016/j.jad.2019.07.086] [Medline: 31382122]
50. Parsons TD, Rizzo AA. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 2008 Sep;39(3):250-261. [doi: 10.1016/j.jbtep.2007.07.007] [Medline: 17720136]
51. Oing T, Prescott J. Implementations of virtual reality for anxiety-related disorders: systematic review. JMIR Serious Games 2018 Nov 7;6(4):e10965 [FREE Full text] [doi: 10.2196/10965] [Medline: 30404770]
52. Cardi V, Krug I, Perpiñá C, Mataix-Cols D, Roncero M, Treasure J. The use of a nonimmersive virtual reality programme in anorexia nervosa: a single case-report. Eur Eat Disord Rev 2012 May;20(3):240-245. [doi: 10.1002/erv.1155] [Medline: 21932282]
53. Perpiñá C, Roncero M. Similarities and differences between eating disorders and obese patients in a virtual environment for normalizing eating patterns. Compr Psychiatry 2016 May;67:39-45. [doi: 10.1016/j.comppsych.2016.02.012] [Medline: 27095333]
54. Roncero M, Perpiñá C. Normalizing the eating pattern with virtual reality for bulimia nervosa: a case report. Rev Mex Trastor Aliment 2015 Jul;6(2):152-159. [doi: 10.1016/j.rmta.2015.11.001]
55. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400. [doi: 10.1017/S003329171700040X] [Medline: 28325167]
56. Reilly EE, Anderson LM, Gorrell S, Schaumberg K, Anderson DA. Expanding exposure-based interventions for eating disorders. Int J Eat Disord 2017 Oct;50(10):1137-1141 [FREE Full text] [doi: 10.1002/eat.22761] [Medline: 28815659]
57. Rizzo AS, Koenig ST. Is clinical virtual reality ready for primetime? Neuropsychology 2017 Nov;31(8):877-899. [doi: 10.1037/neu0000405] [Medline: 29376669]
58. Wiederhold BK, Riva G. Virtual reality therapy: Emerging topics and future challenges. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):3-6. [doi: 10.1089/cyber.2018.29136.bkw] [Medline: 30649958]
59. Sallomi P, Lee P. Deloitte US.: Deloitte Touche Tohmatsu Kimited; 2016. Technology, Media & Telecommunications Predictions URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/technology-media-telecommunications/deloitte-au-tmt-predictions-2016-report-050218.pdf [accessed 2020-02-17]
60. Insel T. The National Institute of Mental Health.: NIMH; 2011 May 20. Making the Most of our Interventions Research URL: https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2011/making-the-most-of-our-interventions-research.shtml [accessed 2020-02-17]

61. National Institute of Mental Health.: NIMH Strategic Objective 3 URL: https://www.nimh.nih.gov/about/strategic-planning-reports/strategic-objective-3.shtml [accessed 2020-02-17]
62. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
63. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: Iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
64. Regenbrecht HT, Schubert TW, Friedmann F. Measuring the sense of presence and its relations to fear of heights in virtual environments. Int J Hum-Comput Interact 1998 Sep;10(3):233-249. [doi: 10.1207/s15327590ijhc1003_2]
65. Robillard G, Bouchard S, Fournier T, Renaud P. Anxiety and presence during VR immersion: a comparative study of the reactions of phobic and non-phobic participants in therapeutic virtual environments derived from computer games. Cyberpsychol Behav 2003 Oct;6(5):467-476. [doi: 10.1089/109493103769710497] [Medline: 14583122]
66. Gregg L, Tarrier N. Virtual reality in mental health : a review of the literature. Soc Psychiatry Psychiatr Epidemiol 2007 May;42(5):343-354. [doi: 10.1007/s00127-007-0173-4] [Medline: 17431528]
67. Ferrer-García M, Gutiérrez-Maldonado J. The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples: a review of the literature. Body Image 2012 Jan;9(1):1-11. [doi: 10.1016/j.bodyim.2011.10.001] [Medline: 22119329]
68. Pericot-Valverde I, Secades-Villa R, Gutiérrez-Maldonado J. A randomized clinical trial of cue exposure treatment through virtual reality for smoking cessation. J Subst Abuse Treat 2019 Jan;96:26-32. [doi: 10.1016/j.jsat.2018.10.003] [Medline: 30466545]
69. Cobb SV, Nichols S, Ramsey A, Wilson JR. Virtual reality-induced symptoms and effects (VRISE). Presence 1999 Apr;8(2):169-186. [doi: 10.1162/105474699566152]
70. Munafo J, Diedrick M, Stoffregen TA. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. Exp Brain Res 2017 Mar;235(3):889-901. [doi: 10.1007/s00221-016-4846-7] [Medline: 27915367]
71. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. Virtual Real 2016;20(2):101-125. [doi: 10.1007/s10055-016-0285-9]
72. Fernández-Álvarez J, Rozental A, Carlbring P, Colombo D, Riva G, Anderson PL, et al. Deterioration rates in Virtual Reality Therapy: An individual patient data level meta-analysis. J Anxiety Disord 2019 Jan;61:3-17. [doi: 10.1016/j.janxdis.2018.06.005] [Medline: 30057346]

Abbreviations

AN-B/P: anorexia nervosa, binge/purge type
AN-R: anorexia nervosa, restrictive type
BED: binge eating disorder
BN: bulimia nervosa
CBT: cognitive behavioral therapy
CET: cue exposure therapy
EBT: evidence-based treatment
ED: eating disorder
NIMH: National Institute of Mental Health
PTSD: posttraumatic stress disorder
VR: virtual reality
VR-CET: virtual reality cue exposure therapy

Edited by G Eysenbach; submitted 24.09.19; peer-reviewed by M Lippmann, S DeForte; comments to author 02.11.19; revised version received 07.12.19; accepted 15.12.19; published 23.04.20.

Please cite as:
Brown T, Nauman Vogel E, Adler S, Bohon C, Bullock K, Nameth K, Riva G, Safer DL, Runfola CD
Bringing Virtual Reality From Clinical Trials to Clinical Practice for the Treatment of Eating Disorders: An Example Using Virtual Reality Cue Exposure Therapy
J Med Internet Res 2020;22(4):e16386
URL: http://www.jmir.org/2020/4/e16386/
doi:10.2196/16386
PMID:32324145

Original Paper

Efficacy of a Theory-Based Cognitive Behavioral Technique App-Based Intervention for Patients With Insomnia: Randomized Controlled Trial

Nilofar Rajabi Majd[1], MSc; Anders Broström[2,3], PhD; Martin Ulander[3], PhD; Chung-Ying Lin[4], PhD; Mark D Griffiths[5], PhD; Vida Imani[6], MD; Daniel Kwasi Ahorsu[4], MPhil; Maurice M Ohayon[7], PhD; Amir H Pakpour[1,2], PhD

[1]Social Determinants of Health Research Center, Research Institute for Prevention of Non-Communicable Diseases, Qazvin University of Medical Sciences, Qazvin, Iran

[2]Department of Nursing, School of Health and Welfare, Jönköping University, Jönköping, Sweden

[3]Department of Clinical Neurophysiology, Linköping University Hospital, Linköping, Sweden

[4]Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hung Hom, China (Hong Kong)

[5]International Gaming Research Unit, Psychology Department, Nottingham Trent University, Nottingham, United Kingdom

[6]Pediatric Health Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

[7]Division of Public Mental Health and Population Sciences, Stanford Sleep Epidemiology Research Center, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Chung-Ying Lin, PhD
Department of Rehabilitation Sciences
Hong Kong Polytechnic University
ST534 Department of Rehabilitation Sciences
Hung Hom,
China (Hong Kong)
Phone: 852 27666755
Email: cylin36933@gmail.com

Abstract

Background: Sleep hygiene is important for maintaining good sleep and reducing insomnia.

Objective: This study examined the long-term efficacy of a theory-based app (including cognitive behavioral therapy [CBT], theory of planned behavior [TPB], health action process approach [HAPA], and control theory [CT]) on sleep hygiene among insomnia patients.

Methods: The study was a 2-arm single-blind parallel-group randomized controlled trial (RCT). Insomnia patients were randomly assigned to a treatment group that used an app for 6 weeks (ie, CBT for insomnia [CBT-I], n=156) or a control group that received only patient education (PE, n=156) through the app. Outcomes were assessed at baseline and 1 month, 3 months, and 6 months postintervention. Primary outcomes were sleep hygiene, insomnia, and sleep quality. Secondary outcomes included attitudes toward sleep hygiene behavior, perceived behavioral control, behavioral intention, action and coping planning, self-monitoring, behavioral automaticity, and anxiety and depression. Linear mixed models were used to evaluate the magnitude of changes in outcomes between the two groups and across time.

Results: Sleep hygiene was improved in the CBT-I group compared with the PE group (P=.02 at 1 month, P=.04 at 3 months, and P=.02 at 6 months) as were sleep quality and severity of insomnia. Mediation analyses suggested that perceived behavioral control on sleep hygiene as specified by TPB along with self-regulatory processes from HAPA and CT mediated the effect of the intervention on outcomes.

Conclusions: Health care providers might consider using a CBT-I app to improve sleep among insomnia patients.

Trial Registration: ClinicalTrials.gov NCT03605732; https://clinicaltrials.gov/ct2/show/NCT03605732

(J Med Internet Res 2020;22(4):e15841) doi:10.2196/15841

KEYWORDS

app-based intervention; cognitive behavioral therapy, insomnia; sleep hygiene; theory of planned behavior

Introduction

Sleep and Insomnia

Inadequate sleep and sleep disorders are among the most frequent problems worldwide [1]. Insomnia is the most common sleep disorder affecting approximately one-third of the general population [2]. It can have negative consequences on one or several spheres in daily life: psychosocial (eg, depression, daytime dysfunction, reduced quality of life), occupational functioning (eg, job absence, reduced ability to do tasks, poor job satisfaction, and inappropriate decisions and choices), or elevated burden to society (eg, increased health costs, reduced job productivity) [3-5]. Insomnia is described as a dissatisfying sleep associated with difficulty initiating or maintaining sleep or early morning awakening despite good opportunities for sleep leading to various daytime symptoms.

Introducing good sleep hygiene practices is often one of the first steps in treating insomnia. Sleep hygiene refers to a variety of behaviors that promote sleep quality [6]. Sleep promoting behaviors include avoiding going to bed hungry or thirsty, avoiding stress and anxiety, avoiding physical activity before going to bed, preparing a bedroom that provides a relaxed environment, and limiting activities in the bedroom to sleep and sexual activities [7]. However, behavioral modifications to improve sleep quality can be hard to achieve; many patients with insomnia report they have tried to modify their poor sleep habits without any effect. Dysfunctional or unrealistic sleep expectations and excessive worrying over sleep loss appear to contribute to poor sleep hygiene. For instance, a Japanese study found that sleep hygiene behaviors are confounded by sleep beliefs [8]. Additionally, due to the rise of mobile technologies, many individuals now use electronic devices in bed, and this may restrict individuals with insomnia from practicing good sleep hygiene (eg, being too excited to sleep due to the use of media) [9]. Consequently, interventions are needed to help individuals with insomnia actually practice sleep hygiene.

Simply providing intervention on sleep hygiene behaviors for individuals with insomnia is insufficient. A review paper of qualitative studies showed that individuals suffering from insomnia observe the inefficiency of sleep hygiene education delivered by health care providers. In particular, health care providers are found to have limited knowledge in sleep hygiene (ie, they only know a few basic principles), and most providers are unable to incorporate theoretical models or techniques to deliver sleep hygiene [10]. Therefore, interventions concerning sleep hygiene behaviors should incorporate robust and effective theoretical models. Indeed, in a systematic review and meta-analysis, Webb et al [11] found that using theories as a framework in designing online interventions led to a substantial effect on outcome variables. Theoretical models can help to select the components of the intervention and help in evaluation of the intervention's impact [12-22]. By designing interventions based on empirically derived theoretical principles, researchers can identify the most powerful determinants of a given construct [23]. Despite evidence-based recommendations that support the utility of theory-based approaches for designing interventions [21], very few studies have considered this aspect [22,24]. Therefore, studies using theory to support treatment efficacy are much needed.

There are four empirically validated theoretical models (cognitive behavioral therapy [CBT], theory of planned behavior [TPB], health action process approach [HAPA], and control theory [CT]) that may assist individuals with insomnia in engaging good sleep hygiene.

Cognitive Behavioral Therapy

CBT focuses on the effect of individuals' beliefs, thoughts, and attitudes on their feelings and behaviors. The purpose of CBT is to educate individuals on how to proactively deal with problems or events during their lives [25]. CBT for insomnia (CBT-I) relies on the general principles of CBT and is designed to relieve insomnia symptoms [26]. The efficacy of CBT-I has been demonstrated in studies [27,28], and a randomized controlled trial (RCT) found that CBT-I outperformed medication in the long run [28]. The principles of CBT are presented to insomnia patients as a set of scientific methods with assessed efficacy [29]. The principles of CBT-I include behavioral modifications such as stimulus control, sleep restriction, sleep hygiene, relaxation training, and cognitive therapy. The goal is to adjust sleeping habits using cognitive strategies to improve thoughts, feelings, and expectations concerning sleep [30,31]. CBT-I can be held in group, individual face-to-face, or online (eg, app-based) sessions based on the therapist's priority, patient's progress, or both [32]. Although the efficacy of online CBT-I has been well established for general populations [25], there is a lack of understanding of how the efficacy of online CBT-I is explained by psychological mechanisms [33,34]. Therefore, this study incorporated CBT-I techniques using an app and investigated the underlying psychological mechanisms in the effectiveness of CBT-I, for which evidence has been extensively presented in the literature [26,28-33].

Theory of Planned Behavior

TPB is one of the most widely applied sociocognitive theories. According to TPB, behavioral intention—an individual's expression of their decisions to perform or not perform a specific behavior—is the most proximal determinant of an individual's behavior. Behavioral intention, in turn, is determined by three constructs: attitudes toward the behavior (ie, whether an individual prefers or values the specific behavior), subjective norms (ie, the opinions of the individual's significant others on the specific behavior), and perceived behavioral control (ie, how confident the individual feels in performing the specific behavior) [35]. The TPB has been used for understanding a variety of behaviors including sleep hygiene behaviors [7,36-38]. Based on TPB, health care providers may try to improve the attitude, subjective norm, and perceived behavioral control of individuals on their sleep hygiene behaviors. Subsequently, the individual may have elevated intention to practice good sleep hygiene.

Health Action Process Approach

Despite the success of TPB, there is evidence that intentions do not necessarily translate into actual behavior (ie, there is an intention-behavior gap) [39]. For this reason, HAPA proposes

a series of volitional, or postintentional, processes that aim to bridge the intention-behavior gap. HAPA focuses more on postintentional processes than TPB. The volitional phase specifies both action planning and coping planning as mediators of the intention-behavior relationship. Action planning indicates how individuals move their intention into action (in this case, sleep hygiene behaviors). More specifically, individuals make plans to practice their sleep hygiene behavior. Coping planning indicates how individuals design any plan to overcome the possible obstacles in the intended behavior (again in this case, sleep hygiene). More specifically, individuals make plans to avoid any potential obstructions that hinder their sleep hygiene behaviors [38]. Using action and coping planning, individuals can better equip their intention to initiate sleep hygiene behaviors. Health care providers may assist individuals with insomnia to facilitate action and coping plans to ensure that individuals' sleep hygiene behaviors can be implemented and actualized.

Control Theory

CT explains self-regulation systems, which are critical elements in analyzing human behaviors. The fundamental idea of CT is the negative feedback loop, and a behavior is performed when the discrepancy between current state (eg, drink a cup of tea before sleep) and reference value (eg, should not drink caffeine prior to sleep) is observed by individuals [40]. In order to assess the discrepancy, the technique of self-monitoring can be incorporated to the theoretical models mentioned above (CBT, TPB, and HAPA) to enhance the behavioral intention in performing sleep hygiene behaviors.

Related Work on App

The insomnia intervention was delivered via a smartphone app to make it available and accessible to meet population needs; app-based interventions overcome some of the pragmatic problems of the face-to-face therapy, including time limitations and high costs [41,42]. Indeed, app-based programs have been developed with highly personalized, tailored, and fully automated features [43-45]. In relation to exposure of in-person contact, a CBT-I app can be used as supplementary support to in-person therapy or as a primary intervention for insomnia through fully automated CBT-I (ie, the therapeutic program can be offered without any human support) [45]. The app designed in this study is similar to the CBT-i Coach [46]. More specifically, the app used in this study was designed to increase patients' knowledge and skills using CBT-I (for detailed information please refer to the app overview section in the Methods). Both apps attempt to reduce insomnia symptoms. Although the CBT-i Coach includes sleep hygiene recommendations in the session, the app used in our study directly addresses practicing sleep hygiene behaviors. Specifically, some theoretical constructs (eg, behavior change techniques [BCTs]) for changing participants' sleep behaviors were applied. Also, some techniques derived from behavioral theories (eg, action planning and coping planning) were used to help participants achieve behavioral goals [47]. The app used in this study focuses on changing people's sleep behaviors whereas other available sleep apps emphasize sleep mentoring [44]. Additionally, the apps summarized by Choi et al [44] generally assess sleep quantitatively (eg, sleep efficacy, sleep time) rather than taking sleep quality into account.

Study Aims

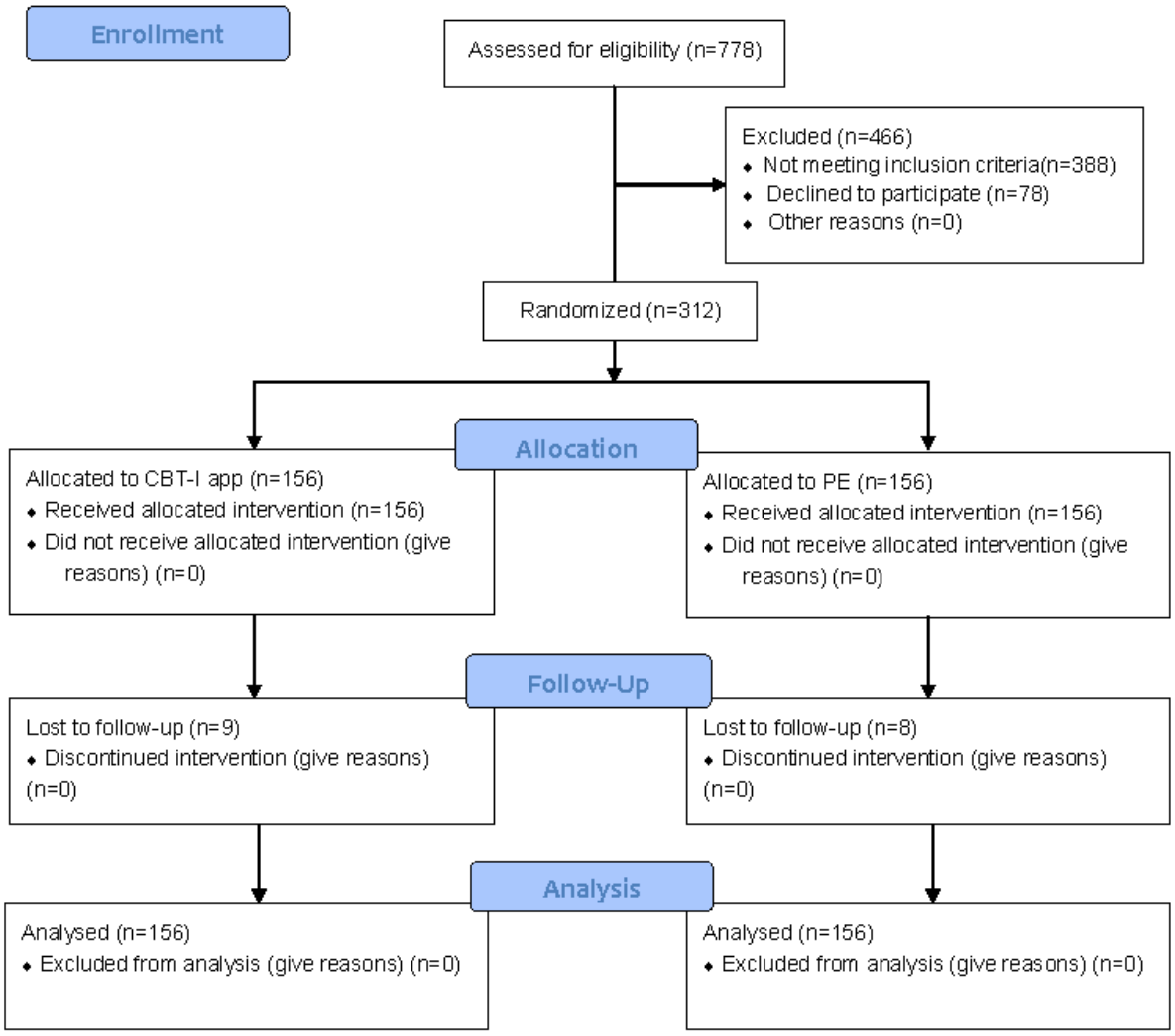
This study aimed to determine the long-term treatment efficacy of a theory-based app using CBT, TPB, HAPA, and CT on sleep hygiene among insomnia patients. The primary evaluation involved comparing sleep hygiene behaviors and sleep quality with a control group that received patient education (PE).

Methods

Design

A 2-arm single-blind parallel-group RCT was launched to compare CBT-I and PE groups via an app over a 6-month period. Potential participants were included after initial screening and signing informed consent forms. The flowchart of the study design is shown in Figure 1. Primary and secondary outcomes were assessed online at baseline, 1 month postintervention, 3 months postintervention, and 6 months postintervention.

Figure 1. Flowchart of study design. CBT-I: cognitive behavioral therapy for insomnia; PE: patient education.



Setting, Participants, and Recruitment

This Iranian-based study was advertised using brochures and posters at 3 universities, 5 colleges, and 10 general physicians' offices as well as on social media. Interested participants were asked to access an online platform specifically designed for the study to complete a screening questionnaire assessing insomnia symptoms comprising the Insomnia Severity Index (ISI), sleep characteristics, and medical history as well as their time spent on computers and online. If participants met the initial criteria and expressed interest, a phone appointment was scheduled to conduct additional eligibility assessments. All telephone interviews were performed by three research assistants who had been trained by an experienced sleep medicine specialist during two 3-hour sessions. The sleep medicine specialist observed each research assistant during three actual patient encounters. The sleep medicine specialist and research assistant scored the patients independently. The scores were then compared to ensure that data collection procedures were consistent. There was more than 98% agreement between the sleep medicine specialist and research assistant in all cases.

Selection Criteria

The inclusion criteria were (1) being age 18 years or older, (2) having insomnia disorder according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) [48], (3) having an ISI score of 10 or higher [49], (4) understanding Persian, and (5) having access to a smartphone and/or desktop computer with internet access.

Participants were excluded if they (1) had an uncontrolled medical condition that interfered with sleep or required immediate treatment (eg, obstructive sleep apnea requiring continuous positive airway pressure treatment), (2) did shift work, (3) were pregnant, (4) were participating in other research and/or clinical trials, (5) had received psychotherapy in the past 6 months, (6) had current major depressive disorder based on the Structured Clinical Interview for DSM-5 disorders, (7) had a self-reported diagnosis of schizophrenia or psychosis, (8) showed evidence of alcohol abuse (more than 3 glasses of alcohol per day at least 21 days a month), (9) misused marijuana (use more than once per week), (10) appeared suicidal, or (11) had children aged under 2 years. After baseline assessments, participants were randomly assigned to a control group receiving

PE or a treatment group receiving CBT-I. Both study groups had access to their related content on the app.

Sample Size Calculation

The sample size was estimated based on previous studies on internet-based self-help insomnia interventions with a moderate effect size [42,50]. Using a 2-tailed test with a small-to-medium effect size (Cohen d=0.40) and significance level of P=.05, a total sample size of 266 (ie, 133 per group) had 90% power. To include an estimated dropout rate of 15%, entire sample size was increased from 266 to 312 participants (ie, 156 per group).

Randomization and Allocation Procedures

Participants who met the inclusion criteria and signed the informed consent were randomly assigned to a control group (PE) or a treatment group (CBT-I) at a 1:1 ratio. Randomization was performed by an independent researcher via a random list generated using SPSS Statistics 24.0 (IBM Corp). Because blinding participants from treatment condition was impossible, the study blinded the data analyst using a code system for treatment condition. Moreover, the data analyst did not have access to the key document. Participants in both groups received assistance from a trained research assistant to help them install and unlock the app. To avoid contamination, participants in the PE group could not access the CBT-I content, which was locked using a personal code.

App Overview

The app was designed based on a combination of TPB, HAPA, and CBT-I [29] and used the self-help concept [19] in order to treat insomnia. Several BCTs are integrated in the app, including information about health consequences, information about social and environmental consequences, habit formation, habit reversal, pros and cons of performing sleep hygiene behaviors, reconstructing the physical environment, reconstructing the social environment, self-monitoring of behavior, action planning, and problem solving [51]. The intervention contents were designed across 6 weeks, with exercises and subtutorials automatically provided each week (Table 1). Detailed information with screenshots is provided in Multimedia Appendix 1.

The content of the CBT-I was designed for treating patients with insomnia individually using weekly sessions, with each session lasting approximately 1 hour. However, participants could complete their sessions more quickly (ie, less than 1 hour) if they improved faster. The content of session 1 was provided for participants using plain text. For sessions 2 and 3, we used a variety of relaxation tools via audio guided meditation exercises as well as images. Session 4 provided correct information concerning sleep in plain text and tables, addressing incorrect information about insomnia participants may have heard. In sessions 5 and 6, plain text and tables were used to help participants plan their sleep hygiene behaviors. A sleep diary was used along with the ISI completed by participants to help them understand their daily sleep improvements. Development of the current app is still at an early stage, and customizing content to suit individual specific situations is under consideration for an updated version of the app. The current version does not have individually tailored content.

Access to the app content was provided to the participants each week; participants could not view the app content for weeks 2 through 6 (ie, sessions 2-6) when they were in week 1 of the intervention, but as participants continued, previous sessions remained unlocked. Every week, there was homework or an exercise, and participants were encouraged to complete it weekly. The content in the next session was opened the following week. Most of the participants used the app as recommended, and a reminder text message was sent to those who did not open the session content on time.

Table 1. Key intervention points for cognitive behavior therapy (BCT) used in the insomnia app.

Behavior change technique[a]	Targeted outcomes
BCT 5.3: information about social and environmental consequences (week 1)	Attitude and intentions to perform sleep hygiene behaviors
BCT 9.2: pros and cons (week 1)	Attitude and intentions to perform sleep hygiene behaviors
BCT 8.3: habit formation (week 2)	Behavioral automaticity
BCT 12.1: reconstructing the physical environment (week 2)	Perceived behavioral control
BCT 8.4: habit reversal (week 3)	Behavioral automaticity
BCT 12.3: reconstructing the social environment (week 3)	Perceived behavioral control
BCT 2.3: self-monitoring of behavior (week 4)	Self-monitoring
BCT 1.4: action planning (week 5)	Action planning
BCT 1.2: problem solving (week 5)	Coping planning

[a]Behavior change techniques are sourced from the taxonomy of Michie et al [51]. For instance, BCT 5.3 (information about social and environmental consequences) represents the subgroup of the taxonomy natural consequence (group 5).

Patient Education Procedure

Participants in the PE group received written weekly information on accurate and relevant information regarding insomnia symptoms (week 1), physiological controls of sleep (week 2), sleep hygiene practices (week 3), healthy sleep behaviors (eg, reduce time in bed, get up at the same time every day, go to bed only if sleepy, and do not stay in bed unless asleep; week 4), and changing lifestyles to promote sleep health (week 5). The information was presented as separate content from the CBT-I in the app. This weekly information was unlocked for participants on a weekly basis. Participants in the PE group

were informed that they could access the CBT-I content on the app at the end of the study (6 months after completing the intervention).

Measures

All participants completed a sociodemographic profile questionnaire (age, gender, educational status, occupational status) at the baseline assessment and primary and secondary outcomes were assessed. The complete measures used to assess primary and secondary outcomes are shown in Multimedia Appendix 2.

Primary Outcomes

Sleep Hygiene Behavior

Sleep hygiene behavior was assessed using three items. The participants were asked to report how many days they had good sleep hygiene behavior over the past week (How many days did you make your bedroom restful over the past week?), whether they avoided going to bed feeling hungry or thirsty, and whether they avoided anxiety and stress-provoking activity before bed. Participants were asked to respond on an 8-point scale ranging from 0 to 7. The internal consistency of the three items was found to be acceptable in a previous study (α=.78) [38].

Insomnia Severity Index

The ISI is a 7-item self-report scale that assesses participant level of insomnia over the past 2 weeks. All items are rated on a 5-point Likert-type scale ranging from 0=no problem to 4=very severe problem. A total score is generated by summing all 7 items ranging from 0-28 with 5 subscores: 0-7 (absence of insomnia), 8-14 (subthreshold insomnia), 15-21 (moderate insomnia), and 22-28 (severe insomnia). The ISI has been translated into Persian, and its psychometric properties have been documented among Iranian adults [52].

Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI) is a subjective measure of sleep quality and disturbances over the past month that contains 19 items grouped into 7 separate component scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction. Scores are summed to provide a global sleep quality score. The Persian version of the PSQI has been validated and described in detail [53].

Secondary Outcomes

Attitude

Attitude toward good sleep hygiene was assessed using 5-point evaluative semantic differential scales (eg, “To make my bedroom/sleep environment restful would be: unpleasant-pleasant, good-bad, wise-foolish, correct-incorrect, unenjoyable-enjoyable, satisfying-unsatisfying, useful-useless”). Internal consistency of this 12-item scale has been found acceptable in previous studies [24,38].

Perceived Behavioral Control

Perceived behavioral control was assessed using three items (eg, “I am confident that every day I can prevent anxiety-provoking activity before bedtime”). All items were scored on a 5-point Likert scale, ranging from 1=totally disagree to 5=totally agree. The scale has been proved internally consistent in previous studies [24,38].

Behavioral Intention

Behavioral intention was assessed using 6 items (eg, “Over the next week, I intend to make my bedroom restful”). All items were scored on a 5-point Likert scale ranging from 1=totally disagree to 5=totally agree. Internal consistency of this scale has been found acceptable in previous studies [24,38].

Action Planning

Action planning was assessed using 4 items. The participants were asked to indicate if they have made a detailed plan regarding (1) when, (2) where, (3) how, and (4) how often they will perform sleep hygiene behaviors over the next 6 months. All items were scored on a 5-point Likert scale ranging from 1=totally disagree to 5=totally agree. The internal consistency of this scale has been found acceptable in previous studies [24,38,54].

Coping Planning

Coping planning was assessed using 5 items (eg, “I have made a detailed plan regarding what to do if something interferes with my plans”). All items were scored on a 5-point Likert scale ranging from 1=totally disagree to 5=totally agree. Internal consistency of this scale has been found acceptable in previous studies [24,38,55].

Self-Monitoring

Self-monitoring was assessed by 3 items (“I keep track of how much time I spend sleeping,” “I pay attention to how tired or rested I feel each day,” and “I take care to note the time that I go to bed and wake each day”). Responses were rated on a scale ranging from 1=never to 5=always.

Self-Report Behavioral Automaticity Index

The extent to which sleep hygiene behaviors are habitual for an individual was assessed using the Self-Report Behavioral Automaticity Index (SRBAI). The SRBAI contains 4 items that start with the stem “Sleep hygiene behavior is something...” followed by “I do automatically,” “I do without having to consciously remember,” “I do without thinking,” and “I start doing before I realize I’m doing it.” The reliability of the Persian SRBAI has been reported [56].

Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale is a 14-item scale that assesses anxiety (7 items) and depression (7 items) in patients with both somatic and mental problems. All items were scored on a 0 to 3 scale with higher score representing higher levels of anxiety and depression. The Iranian version of the scale has been validated in different clinical patients [57].

Data Management

All data were collected and stored using the FileMaker Pro 15 (Claris International Inc) database. Participants’ names, phone numbers, and addresses were recorded. Each participant received a unique code. All data were password protected and encrypted to ensure that confidentiality of the data was maintained throughout the study. The study was monitored (via quality

control and audits) by the Social Determinants of Health Research Center at Qazvin University of Medical Sciences to ensure compliance with the protocol and applicable regulations was maintained and data were collected in a timely, accurate, and complete manner.

Statistical Methods

All statistical analyses were performed in accordance with Consolidated Standards of Reporting Trials (CONSORT) guidelines. Intention-to-treat analyses were performed to take care of attrition. Descriptive statistics were used to summarize the characteristics of the participants. To evaluate the magnitude of changes in primary and secondary outcomes over time across the two groups, linear mixed models (PROC MIXED) were performed controlling for baseline variables and other covariates that may relate to the outcome. The mixed modeling approach is a powerful statistical tool to evaluate group differences over time with unequal numbers of participants at baseline and follow-up. It is also a helpful way to handle missing data using full information maximum likelihood estimation. The analysis incorporated two between-participant effects (between groups and between participants within groups) and three within-participant effects (between times, group by time interactions, and random variation). All *P* values were 2-sided and were evaluated as statistically significant at the .05 level. All statistical analyses were performed using SAS version 9.3 (SAS Institute Inc).

Ethics and Dissemination

This study protocol was approved by the ethics committee of Qazvin University of Medical Sciences (IR.QUMS.REC. 1396.455) and is registered with ClinicalTrials.gov [NCT03605732], last updated July 2018. This study was performed based on the Helsinki Declaration principles. Required permissions were obtained from authorities of Qazvin University of Medical Sciences. After research assistants expressed objectives and assured participants about the confidentiality of their data and possibility of withdrawing from the study, informed consent forms were signed by those participants willing to participate in this research.

Results

General Use of the App

App engagement was assessed using an online database that recorded the number of log-ins and the duration of each log-in for all participants. The results showed that over the 6-month study period, the mean number of log-ins in both groups was 41.21 (SD 39.50). Participants in the CBT group had a higher number of log-ins (mean 42.65 [SD 42.58]) as compared with the PE group (mean 40.93 [SD 45.01]); however, this difference was not statistically significant (*P*=.38). Mean length of time using the app in all participants was 1580 minutes (SD 1460.37). The participants in the CBT group had significantly longer duration of app use (mean 1780.34 [SD 1770.07] minutes) compared with the PE group (mean 1370 [SD 450.01] minutes; *P*=.03).

Descriptive Statistics in Demographics, Primary Outcomes, and Secondary Outcomes

Table 2 shows that both CBT-I and PE groups shared similar demographic characteristics. Specifically, the mean age of the insomnia patients in the CBT-I group was 36.21 (SD 5.81) years, and the mean age of those in the PE was 35.29 (SD 5.76) years. Slightly less than half of the participants were male in each group (46.1% [72/156] in the CBT-I group and 42.3% [66/156] in the PE group).

Table 3 describes the sleep quality (using PSQI), severity of insomnia (using ISI), sleep hygiene behaviors, and health status of both groups at baseline and 1-month, 3-month, and 6-month follow-ups.

Table 2. Demographic characteristics by condition (n=312).

Characteristic	PE[a]	CBT-I[b]	*P* value
Age in years, mean (SD)	35.29 (5.76)	36.21 (5.81)	.16
Male, n (%)	66 (42.3)	72 (46.1)	.49
Education in years, mean (SD)	14.22 (4.28)	14.70 (5.13)	.37
Insomnia duration in months, mean (SD)	16.71 (5.33)	15.89 (7.02)	.25

[a]PE: patient education.

[b]CBT-I: cognitive behavioral therapy for insomnia.

Table 3. Descriptive statistics for all outcome measures by condition and time.

Outcome	PE[a] mean (SD)				CBT-I[b] mean (SD)			
	Baseline	Month 1	Month 3	Month 6	Baseline	Month 1	Month 3	Month 6
PSQI[c]	13.34 (2.72)	12.23 (3.79)	12.42 (3.88)	12.34 (3.92)	13.25 (3.13)	9.35 (4.82)	9.33 (3.82)	9.09 (4.02)
ISI[d]	19.21 (4.57)	16.69 (4.98)	16.65 (4.77)	16.70 (5.63)	19.26 (4.57)	12.67 (5.60)	12.50 (5.57)	12.38 (5.55)
Sleep hygiene behavior	9.27 (2.46)	12.84 (5.74)	12.90 (5.27)	13.00 (4.25)	9.86 (3.14)	15.61 (5.95)	15.57 (5.12)	15.82 (5.10)
Attitude	2.69 (0.77)	3.22 (0.97)	3.20 (1.02)	3.23 (1.19)	2.74 (0.59)	3.61 (0.76)	3.66 (0.63)	3.68 (0.72)
PBC[e]	2.81 (0.94)	2.97 (0.74)	2.94 (1.01)	2.88 (0.68)	2.77 (0.86)	4.20 (0.84)	4.17 (0.59)	4.24 (0.89)
Intention	2.82 (1.15)	3.08 (1.08)	3.12 (1.14)	3.14 (1.11)	2.87 (1.02)	4.33 (0.71)	4.37 (1.06)	4.40 (1.09)
Action planning	2.12 (0.63)	2.36 (1.14)	2.38 (1.18)	2.34 (1.22)	2.18 (0.87)	3.56 (1.11)	3.67 (0.91)	3.64 (1.02)
Coping planning	2.48 (0.68)	2.83 (1.16)	2.86 (1.14)	2.89 (1.21)	2.43 (0.98)	3.75 (0.81)	3.92 (0.71)	3.96 (0.77)
Self-monitoring	2.29 (0.90)	2.84 (0.97)	2.85 (0.99)	2.88 (0.91)	2.32 (0.94)	3.32 (1.04)	3.41 (1.11)	3.43 (1.16)
SRBAI[f]	3.37 (1.20)	3.05 (1.16)	3.24 (1.06)	3.20 (1.24)	3.23 (1.12)	4.50 (0.89)	4.48 (0.18)	4.52 (0.79)
Anxiety[g]	8.35 (3.62)	7.82 (4.13)	7.93 (4.91)	7.96 (4.88)	8.22 (3.89)	5.58 (3.84)	5.27 (3.63)	5.50 (3.69)
Depression[g]	6.03 (3.30)	5.98 (3.13)	5.94 (3.28)	5.97 (3.25)	6.30 (3.82)	3.81 (0.89)	3.36 (0.94)	3.22 (0.88)

[a]PE: patient education.

[b]CBT-I: cognitive behavioral therapy for insomnia.

[c]PSQI: Pittsburgh Sleep Quality Index.

[d]ISI: Insomnia Severity Index.

[e]PBC: perceived behavioral control.

[f]SRBAI: Self-Report Behavioral Automaticity Index.

[g]Anxiety and depression were measured using the Hospital Anxiety and Depression Scale.

Effects of the Intervention on Primary Outcomes

The intervention had promising effects on the three primary outcomes: sleep quality, severity of insomnia, and sleep hygiene behaviors. Specifically, sleep quality was improved among insomnia patients in the CBT-I group compared with those in the PE at all follow-ups, as suggested by scores on the PSQI (P<.001 for 1, 3, and 6 months). Similar improvements were shown in severity of insomnia (ie, significant decrease in the severity of insomnia among the CBT-I group compared with those in the PE group at 1, 3, and 6 months following the intervention, all P values <.001) and sleep hygiene behaviors (insomnia patients in the CBT-I group significantly engaged in more good sleep hygiene behaviors than those in the PE group at 1 [P=.02], 3 [P=.04], and 6 [P=.02] months following the intervention). Table 4 shows the findings of the linear mixed models predicting primary outcomes between two groups after controlling for their age, gender, and education.

Table 4. Results of the linear mixed models for primary outcomes.

Variables	PSQI[a]		ISI[b]		Sleep hygiene behaviors	
	B[c] (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)
CBT-I[d] (vs PE[e] at baseline)	–0.02 (0.55)	.97 (–1.10 to 1.06)	0.05 (0.51)	.92 (–0.95 to 1.05)	0.40 (0.66)	.55 (–0.89 to 1.69)
Month 1 (vs baseline)	–1.11 (0.46)	.02 (–2.01 to –0.21)	–2.52 (0.37)	<.001 (–3.25 to –1.80)	3.56 (0.62)	<.001 (2.35 to 4.78)
Month 3 (vs baseline)	–0.92 (0.42)	.03 (–1.74 to –0.10)	–2.55 (0.36)	<.001 (–3.26 to –1.84)	3.64 (0.63)	<.001 (2.41 to 4.88)
Month 6 (vs baseline)	–1.0 (0.39)	.01 (–1.76 to –0.24)	–2.51 (0.44)	<.001 (–3.37 to –1.65)	3.73 (0.69)	<.001 (2.38 to 5.08)
CBT-I (vs PE at 1 month)	–2.76 (0.64)	<.001 (–4.01 to –1.51)	–4.01 (0.51)	<.001 (–5.01 to –3.01)	2.19 (0.93)	.02 (0.37 to 4.01)
CBT-I (vs PE at 3 months)	–3.0 (0.66)	<.001 (–4.29 to –1.71)	–4.21 (0.56)	<.001 (–5.31 to –3.11)	2.01 (0.96)	.04 (0.13 to 3.89)
CBT-I (vs PE at 6 months)	–3.16 (0.61)	<.001 (–4.36 to –1.96)	–4.36 (0.50)	<.001 (–5.34 to –3.38)	2.23 (0.91)	.02 (0.45 to 4.01)
Age	0.03 (0.03)	.32 (–0.03 to 0.09)	0.02 (0.04)	.62 (–0.06 to 0.10)	0.09 (0.04)	.03 (0.01 to 0.17)
Female	0.31 (0.40)	.44 (–0.47 to 1.09)	0.04 (0.42)	.92 (–0.78 to 0.86)	0.13 (0.35)	.71 (–0.56 to 0.82)
Education	0.44 (0.72)	.54 (–0.97 to 1.85)	0.22 (0.38)	.56 (–0.53 to 0.97)	0.70 (0.54)	.20 (–0.36 to 1.76)
Intercept	14.14 (1.23)	<.001 (11.73 to 16.55)	18.67 (1.27)	<.001 (16.18 to 21.16)	5.99 (1.13)	<.001 (3.78 to 8.21)

[a]PSQI: Pittsburgh Sleep Quality Index.

[b]ISI: Insomnia Severity Index.

[c]B: unstandardized coefficient.

[d]CBT-I: cognitive behavioral therapy for insomnia.

[e]PE: patient education.

Effects of the Intervention on Secondary Outcomes

In terms of secondary outcomes, analyses indicated that insomnia patients in the CBT-I group had better attitude (*P*=.03 at 1 month; *P*=.04 at 3 months; *P*=.03 at 6 months), stronger perceived behavioral control (*P* values <.001 at 1, 3, and 6 months), higher intention (*P* values <.001 at 1, 3, and 6 months) to develop good sleep hygiene behavior, and were more likely to have formed relevant action and coping plans (*P* values <.001 at 1, 3, and 6 months; Table 5). Additionally, insomnia patients in the CBT-I group had higher behavioral automaticity (*P* values <.001 at 1, 3, and 6 months), better self-monitoring (*P* values ≤.001 at 1, 3, and 6 months), and less anxiety (*P*=.003 at 1 month and *P*<.001 at both 3 and 6 months), and depression (*P* values <.001 at 1, 3, and 6 months) relative to those in the PE group (Table 6).

Table 5. Results of the linear mixed models for measures on theory of planned behavior and health action process approach concepts.

Variables	Attitude		PBC[a]		Intention		Action planning		Coping planning	
	B[b] (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)
CBT-I[c] (vs PE[d])	0.02 (0.13)	.88 (–0.24 to 0.28)	0.05 (0.14)	.72 (–0.22 to 0.32)	0.03 (0.16)	.85 (–0.28 to 0.34)	0.04 (0.16)	.80 (–0.27 to 0.35)	0.04 (0.15)	.79 (–0.25 to 0.33)
Month 1 (vs baseline)	0.53 (0.11)	<.001 (0.31 to 0.75)	0.16 (0.12)	.18 (–0.08 to 0.40)	0.25 (0.14)	.08 (–0.02 to 0.52)	0.24 (0.12)	.046 (0.005 to 0.48)	0.36 (0.13)	.01 (0.11 to 0.62)
Month 3 (vs baseline)	0.51 (0.16)	.002 (0.20 to 0.82)	0.15 (0.16)	.35 (–0.16 to 0.46)	0.29 (0.15)	.05 (–0.004 to 0.58)	0.26 (0.14)	.06 (–0.01 to 0.53)	0.39 (0.12)	.001 (0.16 to 0.63)
Month 6 (vs baseline)	0.53 (0.18)	.003 (0.18 to 0.88)	0.08 (0.11)	.47 (–0.14 to 0.30)	0.31 (0.18)	.09 (–0.04 to 0.66)	0.22 (0.11)	.046 (0.004 to 0.44)	0.42 (0.14)	.003 (0.15 to 0.69)
CBT-I (vs PE at 1 month)	0.34 (0.16)	.03 (0.03 to 0.65)	1.27 (0.17)	<.001 (0.94 to 1.60)	1.21 (0.20)	<.001 (0.82 to 1.60)	1.13 (0.17)	<.001 (0.80 to 1.46)	0.91 (0.17)	<.001 (0.58 to 1.24)
CBT-I (vs PE at 3 months)	0.40 (0.19)	.04 (0.03 to 0.77)	1.25 (0.15)	<.001 (0.96 to 1.54)	1.11 (0.22)	<.001 (0.68 to 1.54)	1.22 (0.13)	<.001 (0.97 to 1.48)	1.06 (0.16)	<.001 (0.75 to 1.37)
CBT-I (vs PE at 6 months)	0.39 (0.18)	.03 (0.04 to 0.74)	1.40 (0.14)	<.001 (1.13 to 1.67)	1.02 (0.21)	<.001 (0.61 to 1.43)	1.26 (0.19)	<.001 (0.89 to 1.63)	1.10 (0.15)	<.001 (0.81 to 1.39)
Age	–0.01 (0.01)	.32 (–0.03 to 0.01)	0.03 (0.07)	.67 (–0.11 to 0.17)	–0.05 (0.09)	.58 (–0.23 to 0.13)	0.01 (0.05)	.84 (–0.09 to 0.11)	0.04 (0.09)	.66 (–0.14 to 0.22)
Female	–0.08 (0.09)	.38 (–0.26 to 0.10)	0.16 (0.09)	.08 (–0.02 to 0.34)	0.14 (0.11)	.20 (–0.08 to 0.36)	0.11 (0.12)	.36 (–0.13 to 0.35)	0.09 (0.11)	.41 (–0.13 to 0.31)
Education	0.27 (0.14)	.06 (–0.004 to 0.54)	0.06 (0.08)	.45 (–0.10 to 0.22)	0.28 (0.18)	.12 (–0.07 to 0.63)	0.27 (0.21)	.20 (–0.14 to 0.68)	0.05 (0.18)	.78 (–0.30 to 0.40)
Intercept	2.53 (0.26)	<.001 (2.02 to 3.04)	2.66 (0.28)	<.001 (2.11 to 3.21)	2.82 (0.35)	<.001 (2.13 to 3.51)	1.76 (0.40)	<.001 (0.98 to 2.54)	2.46 (0.34)	<.001 (1.79 to 3.13)

[a]PBC: perceived behavioral control.

[b]B: unstandardized coefficient.

[c]CBT-I: cognitive behavioral therapy for insomnia.

[d]PE: patient education.

Table 6. Results of the linear mixed models for measures on control theory concept and psychological distress (anxiety and depression).

Variables	SRBAI[a]		Self-monitoring		Anxiety		Depression	
	B[b] (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)	B (SE)	*P* value (95% CI)
CBT-I[c] (vs PE[d])	0.12 (0.19)	.53 (–0.25 to 0.49)	0.03 (0.14)	.83 (–0.24 to 0.30)	–0.02 (0.56)	.97 (–1.12 to 1.08)	–0.27 (0.47)	.57 (–1.19 to 0.65)
Month 1 (vs baseline)	0.32 (0.16)	.046 (0.01 to 0.63)	0.56 (0.09)	<.001 (0.38 to 0.74)	–0.53 (0.51)	.30 (–1.53 to 0.47)	–0.04 (0.44)	.93 (–0.90 to 0.82)
Month 3 (vs baseline)	0.43 (0.18)	.02 (0.08 to 0.78)	0.57 (0.11)	<.001 (0.35 to 0.79)	–0.41 (0.50)	.41 (–1.39 to 0.57)	–0.08 (0.37)	.83 (–0.81 to 0.65)
Month 6 (vs baseline)	0.47 (0.21)	.03 (0.06 to 0.88)	0.58 (0.10)	<.001 (0.38 to 0.78)	–0.38 (0.46)	.41 (–1.28 to 0.52)	–0.05 (0.35)	.89 (–0.74 to 0.64)
CBT-I (vs PE at 1 month)	1.59 (0.23)	<.001 (1.14 to 2.04)	0.43 (0.13)	.001 (0.18 to 0.69)	–2.12 (0.72)	.003 (–3.53 to –0.71)	–2.44 (0.62)	<.001 (–3.66 to –1.23)
CBT-I (vs PE at 3 months)	1.68 (0.22)	<.001 (1.25 to 2.11)	0.52 (0.12)	<.001 (0.29 to 0.76)	–2.53 (0.70)	<.001 (–3.90 to –1.16)	–2.86 (0.61)	<.001 (–4.06 to –1.66)
CBT-I (vs PE at 6 months)	1.76 (0.26)	<.001 (1.25 to 2.27)	0.54 (0.16)	.001 (0.23 to 0.85)	–2.64 (0.75)	<.001 (–4.11 to –1.17)	–3.03 (0.65)	<.001 (–4.30 to –1.76)
Age	–0.01 (0.01)	.32 (–0.03 to 0.01)	–0.06 (0.10)	.55 (–0.26 to 0.14)	–0.05 (0.03)	.10 (–0.11 to 0.01)	–0.01 (0.02)	.62 (–0.05 to 0.03)
Female	0.14 (0.13)	.28 (–0.12 to 0.40)	0.18 (0.12)	.14 (–0.06 to 0.42)	0.03 (0.27)	.91 (–0.50 to 0.56)	0.26 (0.30)	.39 (–0.33 to 0.85)
Education	0.06 (0.10)	.55 (–0.14 to 0.26)	0.53 (0.26)	.042 (0.02 to 1.04)	0.53 (0.61)	.39 (–0.67 to 1.73)	0.54 (0.45)	.23 (–0.34 to 1.42)
Intercept	3.81 (0.41)	<.001 (3.01 to 4.61)	2.23 (0.37)	<.001 (1.51 to 2.96)	9.86 (1.18)	<.001 (7.55 to 12.17)	5.92 (0.94)	<.001 (4.08 to 7.76)

[a]SRBAI: Self-Reported Behavioral Automaticity Index.

[b]B: unstandardized coefficient.

[c]CBT-I: cognitive behavioral therapy for insomnia.

[d]PE: patient education.

Mediation analyses were conducted to investigate whether the effects of CBT-I on the primary outcome of sleep hygiene behaviors were mediated by changes in relevant beliefs about sleep hygiene behaviors as specified by the three proposed theories (TPB, HAPA, and CT). Regarding sleep hygiene behaviors at 3 months, only action planning was a significant mediator (coefficient 1.11; 95% CI 0.54-1.69). Regarding sleep hygiene behaviors at 6 months, perceived behavioral control (coefficient 0.44; 95% CI 0.10-0.78), coping planning (coefficient 0.24; 95% CI 0.12-0.36), self-monitoring (coefficient 0.58; 95% CI 0.13-1.03), and behavioral automaticity (coefficient 1.00; 95% CI 0.17-1.83) were significant mediators (Table 7).

Table 7. Mediated effects of variables in theory of planned behavior, health action process approach, and control theory in the impacts of the cognitive behavioral therapy for insomnia app intervention on self-reported sleep hygiene behaviors at 3 months postintervention.

Outcome and mediator	Intervention effect on outcome (C) (SE/95% CI)	Intervention effect on mediator (A) (SE/95% CI)	Mediator effect on outcome (B) (SE/95% CI)	Mediated effect (A*B) (SE/95% CI)
At 3 months				
Sleep hygiene behaviors	2.01 (0.96/0.12 to 3.09)			
Attitude		0.34 (0.16/0.02 to 0.66)	1.71 (0.65/0.43 to 2.99)	0.58 (0.35/–0.11 to 1.27)
PBC[a]		1.27 (0.17/1.07 to 1.47)	1.67 (1.11/–0.52 to 3.86)	2.12 (1.44/–0.72 to 4.96)
Intention		1.21 (0.20/0.82 to 1.60)	0.24 (0.67/–1.08 to 1.56)	0.29 (0.81/–1.31 to 1.89)
Action planning		1.13 (0.17/0.79 to 1.47)	0.98 (0.21/0.57 to 1.39)	1.11 (0.29/0.54 to 1.69)
Coping planning		0.91 (0.17/0.57 to 1.25)	0.35 (0.62/–0.87 to 1.57)	0.32 (0.57/–0.80 to 1.45)
Self-monitoring		0.43 (0.13/0.17 to 0.69)	1.79 (0.48/0.84 to 2.74)	0.77 (0.42/–0.06 to 1.60)
SRBAI[b]		1.59 (0.23/1.14 to 2.04)	0.10 (0.25/–0.39 to 0.59)	0.16 (0.40/–0.63 to 0.95)
At 6 months				
Sleep hygiene behaviors	2.23 (0.91/0.43 to 4.02)			
Attitude		0.34 (0.16/0.02 to 0.66)	0.93 (0.37/0.21 to 1.66)	0.32 (0.19/–0.05 to 0.69)
PBC		1.27 (0.17/1.07 to 1.47)	0.34 (0.13/0.08 to 0.60)	0.44 (0.17/0.10 to 0.78)
Intention		1.21 (0.20/0.82 to 1.60)	0.07 (0.33/0.58 to 0.71)	0.08 (0.40/–0.71 to 0.87)
Action planning		1.13 (0.17/0.79 to 1.47)	0.25 (0.24/0.22 to 0.73)	0.28 (0.27/–0.25 to 0.81)
Coping planning		0.91 (0.17/0.57 to 1.25)	0.26 (0.05/0.16 to 0.36)	0.24 (0.06/0.12 to 0.36)
Self-monitoring		0.43 (0.13/0.17 to 0.69)	1.36 (0.33/0.71 to 2.01)	0.58 (0.23/0.13 to 1.03)
SRBAI		1.59 (0.23/1.14 to 2.04)	0.63 (0.25/0.14 to 1.12)	1.00 (0.42/0.17 to 1.83)

[a]PBC: perceived behavioral control.

[b]SRBAI: Self-Reported Behavioral Automaticity Index.

Discussion

Principal Findings

This research found that a theory-based intervention improved sleep outcomes among insomnia patients in Iran, as evidenced by improved sleep hygiene behaviors, increased sleep quality, and decreased insomnia severity. These beneficial effects were mediated by several changes in the putative determinants of behavior, including perceived behavioral control in TPB, action and coping planning in HAPA, and self-monitoring and behavioral automaticity in CT. Moreover, the entire treatment effects were guided by CBT (ie, the use of BCTs). Furthermore, as a consequence of improvements in sleep, positive effects of CBT-I were demonstrated by reduced anxiety and depression among insomnia patients. Researchers and practitioners interested in improving sleep quality, particularly among insomnia patients, might draw on insights provided by several components in TPB [7,35-38], HAPA [38,39], and CT [40] in order to design effective interventions for insomnia patients. Moreover, use of the CBT-I app may enhance the feasibility of providing CBT-I treatments incorporating the components in TPB, HAPA, and CT mentioned above.

The finding that the PE group did not have substantial improvement could be due to the following barriers. The written information (1) did not have interactive components, commonly used in apps, and may not have triggered the motivation of participants to obtain the information; (2) did not have any psychological mechanism to support its use whereas the CBT-I app incorporated and improved psychological components (eg, attitudes) for participants to actively engage in sleep, and (3) may have been hard to translate into daily practice for the PE group. Consequently, use of the PE group as a control is appropriate because PE only provides passive engagement and does not have any psychological mechanism to support its use. On the other hand, the CBT-I app involved active engagement and included psychological mechanism to support its use.

Apart from illustrating the effectiveness of the CBT-I app, our findings support the notion that the improvements in sleep accrued (in part) from changes in several components specified by TPB (perceived behavioral control), HAPA (action and coping planning), and CT (self-monitoring and behavioral automaticity). Therefore, our findings partially support the notions from previous studies: health care providers may want to use TPB to effectively understand the performance of sleep hygiene behaviors [37,38]. Additionally, these findings go beyond those correlational studies [37,38] and agree with another RCT study [24] that used theory-based (TPB, HAPA, and CBT) interventions to improve sleep hygiene behaviors among adolescents. More specifically, changes in the putative determinants of action lead to changes in the respective behaviors (eg, sleep hygiene behaviors, sleep quality, and severity of insomnia). Consequently, this research provides

experimental support for the combination of TPB, HAPA, CT, and CBT as a framework for improving sleep among insomnia patients.

Our results also support the idea that it is important for CBT-I interventions to incorporate self-regulatory processes such as those specified in HAPA and CT. More specifically, action planning, coping planning, self-monitoring, and behavioral automaticity were significant mediators in this study. Interestingly, mediated effects of action planning on sleep hygiene behaviors were found at 3 months after completing the CBT-I program but not at 6 months after completing the program. In contrast, mediated effects of coping planning, self-monitoring, and behavioral automaticity on sleep hygiene behaviors were observed at 6 months after completing the CBT-I program but not at 3 months after completing the program.

A possible explanation is the cognitive progress during the CBT-I program. Insomnia patients first need to generate effective strategies (ie, using action planning) to assist them in developing good sleep hygiene behaviors because they do not have sufficient ability to devise good strategies [24]. Therefore, action planning was an important mediator for sleep hygiene behaviors at 3 months. Meanwhile, coping planning, self-monitoring, and behavioral automaticity are not mediators at this stage because insomnia patients put all their efforts in action planning and do not have additional efforts to work on coping planning, self-monitoring, and behavioral automaticity. In terms of sleep hygiene behaviors at 6 months, insomnia patients are very likely to have satisfactory capability in action planning. Therefore, action planning loses its importance as mediator because insomnia patients can use the good strategies they have generated. In contrast, insomnia patients need to deal with all the unforeseen barriers and keep their practice. Therefore, coping planning (ie, thinking ahead to tackle possible barriers) [39], self-monitoring (ie, reflection on the difference between preferred and actual performance) [40], and behavioral automaticity (ie, developing the behaviors into habitual pattern) [56] become important mediators at this stage.

This study's findings on self-regulatory processes also support those of correlational studies that point to the importance of self-regulatory processes in predicting sleep hygiene behaviors and related outcomes, although in a different population [38]. In their study with obstructive sleep apnea patients, Deng et al [58] found that an intervention based on HAPA improved participant sleep quality. Our study suggests that targeting different cognitive processes (ie, motivational and self-regulatory processes) specified in TPB, HAPA, and HAPA can lead to changes in the respective behaviors (eg, sleep hygiene in this study).

Strengths and Limitations

Our study has several strengths. First, few trials using psychological interventions built on empirically based theories have been designed for sleep problems. Therefore, results derived from this study provided information about relevant mechanisms of change. Second, development of an app may improve the effects of a sleep intervention in insomnia. More specifically, providing a technology-based intervention could be an effective and accessible way to overcome numerous barriers to CBT-I, such as low access, high costs, time constraints, social stigma, and the lack of therapists and trained mental health experts, as well as providing an intervention overcoming economic and geographical barriers. Third, this study controlled for many potential biases. For example, the blindness of the analyst and the avoidance of contamination were addressed. Fourth, although all the measures were self-reported, instruments used in this study had satisfactory psychometric properties. Consequently, the reliability and validity of outcome measures were ensured.

There are some limitations in the study. First, participants were not representative of the entire population of Iranian patients with insomnia. For example, participants in this study needed to use the app to receive the intervention. Therefore, insomnia patients who have little literacy in using smartphones or desktop computers may not gain benefit from CBT-I. Given that the mean age of the studied sample was mid-thirties; the possibility to generalize our findings to all inhabitants and age groups is low. Second, the use of self-report measures for sleep and mental health outcomes could be biased by social desirability or memory recall. Although mental health outcomes are hard to measure using non-self-reports, future studies may want to use objectively measured instruments on sleep quality (eg, actigraphy). Third, participants could not be blinded because the intervention was obvious to them. Therefore, placebo effects were hard to exclude, especially because most of the outcome measures were self-reported. However, given that the promising effects were found across all outcome measures and placebo effects would be unlikely to last 6 months after treatment ends, we tentatively concluded that CBT-I is an effective program to treat insomnia. Last, as the study only followed participants 6 months after completing the intervention, it is unclear whether CBT-I can provide effects lasting longer than 6 months. Future studies are thus warranted to examine the long-term effect of CBT-I.

Conclusion

The theory-based CBT-I intervention shows promising effects in treating sleep problems for insomnia patients. After receiving the feasible and short (ie, 6 weeks) CBT-I, insomnia patients showed improved sleep hygiene behaviors, enhanced sleep quality, and less insomnia severity. Moreover, other psychological distress outcomes (ie, anxiety and depression) of the insomnia patients who received CBT-I showed improvement lasting 6 months after the intervention ended.

Acknowledgments

Partial financial support for the project by the Research Vice Chancellor of Qazvin University of Medical Sciences is appreciated.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Behavior change techniques, content, and screenshots of app.
[DOCX File , 438 KB - jmir_v22i4e15841_app1.docx]

Multimedia Appendix 2
Questionnaires used to assess primary and secondary outcomes.
[DOCX File , 26 KB - jmir_v22i4e15841_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2697 KB - jmir_v22i4e15841_app3.pdf]

References

1. Barati Dowom P, Roshanaei K, Darvishi M. Neurophysiological mechanism of sleep and wakefulness regulation. Neurosci J Shefaye Khatam 2015 Sep 01;3(3):121-135. [doi: 10.18869/acadpub.shefa.3.3.121]
2. Septianto R. Relationship Between Depression Level With Score of Clinical Reasoning Module I in the Medical Students UIN Syarif Hidayatullah Jakarta Class 2013 [Thesis]. Jakarta, Indonesia: State Islamic University Syarif Hidayatullah Jakarta; 2014.
3. Ohayon MM, Roth T. Place of chronic insomnia in the course of depressive and anxiety disorders. J Psychiatr Res 2003;37(1):9-15. [doi: 10.1016/s0022-3956(02)00052-3] [Medline: 12482465]
4. Simon GE, VonKorff M. Prevalence, burden, and treatment of insomnia in primary care. Am J Psychiatry 1997 Oct;154(10):1417-1423. [doi: 10.1176/ajp.154.10.1417] [Medline: 9326825]
5. Babamiri M, Moeini B, Tahmasian H, Barati M, Roshanai G. The study of sleep health education effect on sleep quality among Lorestan nursing personnel. J Ergon 2017 Mar 01;4(4):8-13. [doi: 10.21859/joe-04042]
6. National Sleep Foundation. Sleep hygiene URL: https://www.sleepfoundation.org/articles/sleep-hygiene [accessed 2020-02-09]
7. Kor K, Mullan BA. Sleep hygiene behaviours: an application of the theory of planned behaviour and the investigation of perceived autonomy support, past behaviour and response inhibition. Psychol Health 2011 Sep;26(9):1208-1224. [doi: 10.1080/08870446.2010.551210] [Medline: 21678170]
8. Williams R. The Prevalence of Insomnia on School Principals and Superintendents in Missouri [Thesis]. Jonesboro: Arkansas State University; 2017.
9. Bjorvatn B, Mrdalj J, Saxvig IW, Aasnæs T, Pallesen S, Waage S. Age and sex differences in bedroom habits and bedroom preferences. Sleep Med 2017 Apr;32:157-161. [doi: 10.1016/j.sleep.2017.01.003] [Medline: 28366328]
10. Araújo T, Jarrin DC, Leanza Y, Vallières A, Morin CM. Qualitative studies of insomnia: current state of knowledge in the field. Sleep Med Rev 2017 Feb;31:58-69 [FREE Full text] [doi: 10.1016/j.smrv.2016.01.003] [Medline: 27090821]
11. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
12. Morin CM, Beaulieu-Bonneau S, LeBlanc M, Savard J. Self-help treatment for insomnia: a randomized controlled trial. Sleep 2005 Oct;28(10):1319-1327. [Medline: 16295218]
13. Bonnar D, Gradisar M, Moseley L, Coughlin A, Cain N, Short MA. Evaluation of novel school-based interventions for adolescent sleep problems: does parental involvement and bright light improve outcomes? Sleep Health 2015 Mar;1(1):66-74. [doi: 10.1016/j.sleh.2014.11.002] [Medline: 29073417]
14. Cain N, Gradisar M, Moseley L. A motivational school-based intervention for adolescent sleep problems. Sleep Med 2011 Mar;12(3):246-251. [doi: 10.1016/j.sleep.2010.06.008] [Medline: 21292553]
15. Gooneratne N. Complementary and Alternative Medicine for Sleep Disturbances [Thesis]. Tehran: University of Tarbiat Modares; 2007.
16. Roehrs T, Roth T. Sleep, sleepiness, and alcohol use. Alcohol Res Health 2001;25(2):101-109 [FREE Full text] [Medline: 11584549]
17. van Straten A, Cuijpers P. Self-help therapy for insomnia: a meta-analysis. Sleep Med Rev 2009 Feb;13(1):61-71. [doi: 10.1016/j.smrv.2008.04.006] [Medline: 18952469]
18. Cushing CC, Jensen CD, Miller MB, Leffingwell TR. Meta-analysis of motivational interviewing for adolescent health behavior: efficacy beyond substance use. J Consult Clin Psychol 2014 Dec;82(6):1212-1218. [doi: 10.1037/a0036912] [Medline: 24841861]

19. Gruber R, Somerville G, Bergmame L, Fontil L, Paquin S. School-based sleep education program improves sleep and academic performance of school-age children. Sleep Med 2016 May;21:93-100. [doi: 10.1016/j.sleep.2016.01.012] [Medline: 27448478]
20. Rigney G, Blunden S, Maher C, Dollman J, Parvazian S, Matricciani L, et al. Can a school-based sleep education programme improve sleep knowledge, hygiene and behaviours using a randomised controlled trial. Sleep Med 2015 Jun;16(6):736-745. [doi: 10.1016/j.sleep.2015.02.534] [Medline: 25979180]
21. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
22. Prestwich A, Webb TL, Conner M. Using theory to develop and test interventions to promote changes in health behaviour: evidence, issues, and recommendations. Curr Opin Psychol 2015 Oct;5:1-5. [doi: 10.1016/j.copsyc.2015.02.011]
23. Rothman AJ. "Is there nothing more practical than a good theory?" Why innovations and advances in health behavior change will arise if interventions are used to test and refine theory. Int J Behav Nutr Phys Act 2004 Jul 27;1(1):11 [FREE Full text] [doi: 10.1186/1479-5868-1-11] [Medline: 15279674]
24. Lin C, Strong C, Scott A, Broström A, Pakpour A, Webb T. A cluster randomized controlled trial of a theory-based sleep hygiene intervention for adolescents. Sleep 2018 Nov:1. [doi: 10.1093/sleep/zsy170]
25. Drake CL. The promise of digital CBT-I. Sleep 2016 Jan 01;39(1):13-14 [FREE Full text] [doi: 10.5665/sleep.5304] [Medline: 26564140]
26. Manber R, Carney C. Treatment Plans and Interventions for Insomnia: A Case Formulation Approach. New York: Guilford Press; 2015.
27. Merrigan JM, Buysse DJ, Bird JC, Livingston EH. Insomnia. JAMA 2013 Feb 20;309(7):733. [doi: 10.1001/jama.2013.524] [Medline: 23423421]
28. Morin CM, Vallières A, Guay B, Ivers H, Savard J, Mérette C, et al. Cognitive behavioral therapy, singly and combined with medication, for persistent insomnia: a randomized controlled trial. JAMA 2009 May 20;301(19):2005-2015 [FREE Full text] [doi: 10.1001/jama.2009.682] [Medline: 19454639]
29. Richardson R, Richards DA. Self help: towards the next generation. Behav Cognit Psychother 2005 Nov 22;34(01):13-23. [doi: 10.1017/S1352465805002481]
30. Jernelöv S, Lekander M, Blom K, Rydh S, Ljótsson B, Axelsson J, et al. Efficacy of a behavioral self-help treatment with or without therapist guidance for co-morbid and primary insomnia: a randomized controlled trial. BMC Psychiatry 2012 Jan 22;12:5 [FREE Full text] [doi: 10.1186/1471-244X-12-5] [Medline: 22264332]
31. Insel TR, Cuthbert BN. Endophenotypes: bridging genomic complexity and disorder heterogeneity. Biol Psychiatry 2009 Dec 01;66(11):988-989. [doi: 10.1016/j.biopsych.2009.10.008] [Medline: 19900610]
32. Leichsenring F, Hiller W, Weissberg M, Leibing E. Cognitive-behavioral therapy and psychodynamic psychotherapy: techniques, efficacy, and indications. Am J Psychother 2006;60(3):233-259. [doi: 10.1176/appi.psychotherapy.2006.60.3.233] [Medline: 17066756]
33. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-delivered cognitive behavioral therapy to treat insomnia: a systematic review and meta-analysis. PLoS One 2016;11(2):e0149139 [FREE Full text] [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
34. Ye Y, Zhang Y, Chen J, Liu J, Li X, Liu Y, et al. Internet-based cognitive behavioral therapy for insomnia (ICBT-i) improves comorbid anxiety and depression-a meta-analysis of randomized controlled trials. PLoS One 2015;10(11):e0142258 [FREE Full text] [doi: 10.1371/journal.pone.0142258] [Medline: 26581107]
35. Damghanian M, Alijanzadeh M. Theory of planned behavior, self-stigma, and perceived barriers explains the behavior of seeking mental health services for people at risk of affective disorders. Soc Health Behav 2018;1(2):54. [doi: 10.4103/shb.shb_27_18]
36. Knowlden AP, Sharma M, Bernard AL. A theory of planned behavior research model for predicting the sleep intentions and behaviors of undergraduate college students. J Prim Prev 2012 Feb;33(1):19-31. [doi: 10.1007/s10935-012-0263-2] [Medline: 22293980]
37. Lao HC, Tao VY, Wu AM. Theory of planned behaviour and healthy sleep of college students. Aust J Psychol 2015 May 07;68(1):20-28. [doi: 10.1111/ajpy.12094]
38. Strong C, Lin C, Jalilolghadr S, Updegraff JA, Broström A, Pakpour AH. Sleep hygiene behaviours in Iranian adolescents: an application of the Theory of Planned Behavior. J Sleep Res 2017 Jun 08;27(1):23-31. [doi: 10.1111/jsr.12566]
39. Walsh JJ, da Fonseca RS, Banta A. Watching and participating in exercise videos: a test of the theory of planned behaviour, conscientiousness, and the role of implementation intentions. Psychol Health 2005 Dec;20(6):729-741. [doi: 10.1080/14768320500187866]
40. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]
41. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]

42. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia: a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
43. Andersson G, Carlbring P. Internet-assisted cognitive behavioral therapy. Psychiatr Clin North Am 2017 Dec;40(4):689-700. [doi: 10.1016/j.psc.2017.08.004] [Medline: 29080594]
44. Choi YK, Demiris G, Lin S, Iribarren SJ, Landis CA, Thompson HJ, et al. Smartphone applications to support sleep self-management: review and evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790 [FREE Full text] [doi: 10.5664/jcsm.7396] [Medline: 30353814]
45. Luik AI, van der Zweerde T, van Straten A, Lancee J. Digital delivery of cognitive behavioral therapy for insomnia. Curr Psychiatry Rep 2019 Jun 04;21(7):50 [FREE Full text] [doi: 10.1007/s11920-019-1041-0] [Medline: 31161406]
46. Kuhn E, Weiss BJ, Taylor KL, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I Coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. J Clin Sleep Med 2016;12(4):597-606. [doi: 10.5664/jcsm.5700] [Medline: 26888586]
47. Mead MP, Irish LA. Application of health behaviour theory to sleep health improvement. J Sleep Res 2019 Nov 22:e12950. [doi: 10.1111/jsr.12950]
48. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. Arlington: American Psychiatric Association; 2017.
49. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
50. Lancee J, van den Bout J, Sorbi MJ, van Straten A. Motivational support provided via email improves the effectiveness of internet-delivered self-help treatment for insomnia: a randomized trial. Behav Res Ther 2013 Dec;51(12):797-805. [doi: 10.1016/j.brat.2013.09.004] [Medline: 24121097]
51. 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]
52. Yazdi Z, Sadeghniiat-Haghighi K, Zohal MA, Elmizadeh K. Validity and reliability of the Iranian version of the insomnia severity index. Malays J Med Sci 2012 Oct;19(4):31-36 [FREE Full text] [Medline: 23613647]
53. Farrahi Moghaddam J, Nakhaee N, Sheibani V, Garrusi B, Amirkafi A. Reliability and validity of the Persian version of the Pittsburgh Sleep Quality Index (PSQI-P). Sleep Breath 2012 Mar;16(1):79-82. [doi: 10.1007/s11325-010-0478-5] [Medline: 21614577]
54. Pakpour AH, Hidarnia A, Hajizadeh E, Plotnikoff RC. Action and coping planning with regard to dental brushing among Iranian adolescents. Psychol Health Med 2012;17(2):176-187. [doi: 10.1080/13548506.2011.596551] [Medline: 21777092]
55. Pakpour AH, Zeidi IM, Chatzisarantis N, Molsted S, Harrison AP, Plotnikoff RC. Effects of action planning and coping planning within the theory of planned behaviour: a physical activity study of patients undergoing haemodialysis. Psychol Sport Exerc 2011 Nov;12(6):609-614. [doi: 10.1016/j.psychsport.2011.06.008]
56. Lin C, Yaseri M, Pakpour AH, Malm D, Broström A, Fridlund B, et al. Can a multifaceted intervention including motivational interviewing improve medication adherence, quality of life, and mortality rates in older patients undergoing coronary artery bypass surgery? A multicenter, randomized controlled trial with 18-month follow-up. Drugs Aging 2017 Feb;34(2):143-156. [doi: 10.1007/s40266-016-0429-3] [Medline: 28004259]
57. Lin C, Pakpour AH. Using Hospital Anxiety and Depression Scale (HADS) on patients with epilepsy: confirmatory factor analysis and Rasch models. Seizure 2017 Feb;45:42-46 [FREE Full text] [doi: 10.1016/j.seizure.2016.11.019] [Medline: 27915110]
58. Deng T, Wang Y, Sun M, Chen B. Stage-matched intervention for adherence to CPAP in patients with obstructive sleep apnea: a randomized controlled trial. Sleep Breath 2013 May;17(2):791-801. [doi: 10.1007/s11325-012-0766-3] [Medline: 22945541]

Abbreviations

BCT: behavior change technique
CBT: cognitive behavioral therapy
CBT-I: cognitive behavioral therapy for insomnia
CONSORT: Consolidated Standards of Reporting Trials
CT: control theory
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
HAPA: health action process approach
ISI: Insomnia Severity Index
PE: patient education
PSQI: Pittsburgh Sleep Quality Index

RCT: randomized controlled trial
SRBAI: Self-Report Behavioral Automaticity Index
TPB: theory of planned behavior

Edited by G Eysenbach; submitted 11.08.19; peer-reviewed by D Ben Neriah, J Martínez-Miranda; comments to author 23.11.19; revised version received 31.12.19; accepted 23.01.20; published 01.04.20.

Please cite as:
Rajabi Majd N, Broström A, Ulander M, Lin CY, Griffiths MD, Imani V, Ahorsu DK, Ohayon MM, Pakpour AH
Efficacy of a Theory-Based Cognitive Behavioral Technique App-Based Intervention for Patients With Insomnia: Randomized Controlled Trial
J Med Internet Res 2020;22(4):e15841
URL: http://www.jmir.org/2020/4/e15841/
doi:10.2196/15841
PMID:32234700

Original Paper

Optimizing Text Messages to Promote Engagement With Internet Smoking Cessation Treatment: Results From a Factorial Screening Experiment

Amanda L Graham[1,2], PhD; George D Papandonatos[3], PhD; Megan A Jacobs[1], MPH; Michael S Amato[1,2], PhD; Sarah Cha[1], MSPH; Amy M Cohn[4], PhD; Lorien C Abroms[5], PhD; Robyn Whittaker[6], PhD

[1]Innovations Center, Truth Initiative, Washington, DC, United States

[2]Mayo Clinic College of Medicine and Science, Rochester, MN, United States

[3]Center for Statistical Sciences, Brown University, Providence, RI, United States

[4]Oklahoma Tobacco Research Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[5]Department of Prevention and Community Health, Milken Institute School of Public Health, The George Washington University, Washington, DC, United States

[6]National Institute for Health Innovation, University of Auckland, Auckland, New Zealand

Corresponding Author:
Amanda L Graham, PhD
Innovations Center
Truth Initiative
900 G Street, NW
4th Floor
Washington, DC, 20001
United States
Phone: 1 12024545938
Email: agraham@truthinitiative.org

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/7/e21027/

Abstract

Background: Smoking remains a leading cause of preventable death and illness. Internet interventions for smoking cessation have the potential to significantly impact public health, given their broad reach and proven effectiveness. Given the dose-response association between engagement and behavior change, identifying strategies to promote engagement is a priority across digital health interventions. Text messaging is a proven smoking cessation treatment modality and a powerful strategy to increase intervention engagement in other areas of health, but it has not been tested as an engagement strategy for a digital cessation intervention.

Objective: This study examined the impact of 4 experimental text message design factors on adult smokers' engagement with an internet smoking cessation program.

Methods: We conducted a 2×2×2×2 full factorial screening experiment wherein 864 participants were randomized to 1 of 16 experimental conditions after registering with a free internet smoking cessation program and enrolling in its automated text message program. Experimental factors were *personalization* (on/off), *integration* between the web and text message platforms (on/off), *dynamic tailoring* of intervention content based on user engagement (on/off), and *message intensity* (tapered vs abrupt drop-off). Primary outcomes were 3-month measures of engagement (ie, page views, time on site, and return visits to the website) as well as use of 6 interactive features of the internet program. All metrics were automatically tracked; there were no missing data.

Results: Main effects were detected for *integration* and *dynamic tailoring*. *Integration* significantly increased interactive feature use by participants, whereas *dynamic tailoring* increased the number of features used and page views. No main effects were found for *message intensity* or *personalization* alone, although several synergistic interactions with other experimental features were observed. Synergistic effects, when all experimental factors were active, resulted in the highest rates of interactive feature use

and the greatest proportion of participants at high levels of engagement. Measured in terms of standardized mean differences (SMDs), effects on interactive feature use were highest for Build Support System (SMD 0.56; 95% CI 0.27 to 0.81), Choose Quit Smoking Aid (SMD 0.38; 95% CI 0.10 to 0.66), and Track Smoking Triggers (SMD 0.33; 95% CI 0.05 to 0.61). Among the engagement metrics, the largest effects were on overall feature utilization (SMD 0.33; 95% CI 0.06 to 0.59) and time on site (SMD 0.29; 95% CI 0.01 to 0.57). As no SMD >0.30 was observed for main effects on any outcome, results suggest that for some outcomes, the combined intervention was stronger than individual factors alone.

Conclusions: This factorial experiment demonstrates the effectiveness of text messaging as a strategy to increase engagement with an internet smoking cessation intervention, resulting in greater overall intervention dose and greater exposure to the core components of tobacco dependence treatment that can promote abstinence.

Trial Registration: ClinicalTrials.gov NCT02585206; https://clinicaltrials.gov/ct2/show/NCT02585206.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2015-010687

(J Med Internet Res 2020;22(4):e17734) doi:10.2196/17734

KEYWORDS

smoking cessation; tobacco dependence; internet; text messaging

Introduction

Background

Internet-based interventions for smoking cessation have the potential to significantly impact public health, given their broad reach and proven effectiveness. Nine of 10 adults in the United States have internet access [1], more than one-third of all smokers—12.4 million individuals—look online each year to quit smoking [2], and hundreds of thousands enroll in freely available programs [3,4]. Systematic reviews and meta-analyses have demonstrated the effectiveness of tailored and interactive internet interventions for smoking cessation [5]. However, a sizable proportion of smokers disengage early from internet programs without being exposed to the content or features that can promote abstinence. Indeed, low levels of engagement with internet interventions have been documented across a range of health behaviors [6]. Given the evidence of a dose-response association between engagement and behavior change outcomes [7-9], identifying strategies to promote engagement has been noted as a priority across digital health interventions [10-12].

The relationship of engagement to outcomes is complex and includes both behavioral and cognitive dimensions [11,13]. Engagement with an internet intervention can be usefully conceptualized into 3 phases [14], although users' progress through the phases is often nonlinear. In the first phase, an individual decides to visit a website to determine its relevance and potential utility. In the second phase, the individual uses a part of the intervention. In the third phase, the individual returns to engage more fully with the intervention. This study aimed to influence engagement at this third phase by using prompts and reminders delivered via text messages. Text messaging is a proven intervention modality to promote smoking cessation [15] and a powerful strategy to increase intervention engagement [16,17]. Across most demographic groups, a majority of individuals own a mobile phone and use text messaging [18], including economically disadvantaged groups among whom tobacco use is more prevalent. However, to date, text messaging has not been tested specifically as an engagement strategy for internet cessation interventions [10], and little is known about how best to design such text messages [19].

The Multiphase Optimization Strategy (MOST) is a method for systematically building and evaluating interventions to ensure they comprise active components delivered in optimal doses [20]. The screening phase of MOST is designed to determine which intervention components are active (ie, make a difference in the target outcome) and should be retained, and which are inactive and should be discarded. This is accomplished efficiently through a randomized experiment involving a factorial design, which allows for the examination of several design factors simultaneously. As they are scalable and can automate the delivery of many experimental conditions, internet interventions are well suited to conduct such experiments [21]. Guided by the principles of MOST, we conducted a screening experiment to evaluate 4 experimental factors hypothesized to promote engagement with an internet smoking cessation program.

Personalization incorporates user-specific elements (eg, name) to enhance the personal relevance of messaging. People are more likely to actively process information if they perceive it to be personally relevant [22]. Personalization can increase smokers' attention to written information and the perceived quality of that information [23,24] and is a desired and expected feature of text messaging [25]. Supported by prior literature [26], we hypothesized that text messages that incorporate personalized content would be more efficacious than generic ones.

Integration refers to the ability to interact with an intervention platform through the web and text messages, accomplished by sharing data between systems. This approach leverages the unique and combined advantages of these 2 different modalities to create a seamless user experience. Internet interventions can be used to deliver multimedia content but require users to initiate contact, whereas text messages are a powerful form of push notification that have a 98% open rate, with 90% of messages being read within 3 min [27]. A 2010 meta-analysis found that the effectiveness of internet interventions for a variety of health behaviors was enhanced by adding text messaging [28], but these early studies and others [29] most often delivered text messaging in parallel to a web-based intervention with little integration between the 2 modalities. We hypothesized that an

intervention that allows smokers to interact with the tools and content of an internet program via text messages would be more effective in promoting treatment engagement than delivering text messages in parallel to an internet intervention.

Dynamic Tailoring delivers individually tailored feedback that adapts over time to a smoker's needs. Research over several decades supports the superiority of individually tailored messaging over generic, one-size-fits-all messaging in improving behavior change outcomes [26] and in promoting intervention engagement [30]. However, it typically involves messaging around static, theory-driven psychosocial constructs (eg, readiness to quit and self-efficacy) gathered at the outset of an intervention. Few studies have dynamically tailored communications to deliver feedback based on a user's pattern of intervention engagement [31]. We hypothesized that messages tailored to a user's pattern of engagement to encourage the exploration of components they have not yet used and prompt continued engagement with the ones they have would yield higher engagement than messages without this kind of feedback.

Message intensity refers to the delivery schedule of text messages. One of the largest randomized smoking cessation trials demonstrating the effectiveness of a text message intervention [32] delivered 5 messages a day for the first 5 weeks, followed by an abrupt drop to just 3 messages per week for the next 26 weeks. However, a meta-analysis of health behavior change interventions found that the largest effect size was observed for text message interventions with tapered intensity (ie, gradually decreasing content delivery [26]). We set tapered intensity as the active form of this factor and hypothesized that it would make message delivery more salient and impactful than a fixed schedule of messages followed by an abrupt drop-off.

Objective

To summarize, this factorial screening experiment evaluated the impact of *personalization*, *integration*, *dynamic tailoring*, and *message intensity* on engagement with an internet smoking cessation program. We hypothesized that the active form of each experimental factor would yield higher rates of engagement with one or more features of the program and overall metrics of engagement.

Methods

Experimental Design

This full factorial experiment had 4 factors, each of which was implemented at 2 levels: *personalization* (on/off), *integration* (on/off), *dynamic tailoring* (on/off), and *message intensity* (tapered/abrupt). The factors were designed to be compatible with each other but also to deliver a coherent intervention when implemented singularly. All participants had full access to the website to use as they desired. In addition, they were randomized by a computer algorithm to one of the 2^4=16 experimental conditions, stratified by whether they enrolled on a desktop/mobile because mobile access to the website may influence engagement. The use of the website and text messages was automatically tracked for 3 months to allow sufficient time to examine the impact of the text message intervention on intervention engagement because most nonusage attrition happens within this period [33]. There was no involvement by research staff, and there were no missing data.

Procedure

This fully automated experiment was conducted within BecomeAnEX (*EX*), a free, evidence-based smoking cessation program developed in 2008 by Truth Initiative in collaboration with Mayo Clinic. Since it launched, more than 800,000 tobacco users have registered on the site. Approximately 80% of newly registered users sign up for text messaging. As described previously [34], newly registered users who met study eligibility criteria were randomized to 1 of the 16 experimental arms. Eligibility criteria were current smoking (every day/some days), US residence, aged 18 years or older, and enrollment in the EX text message program during website registration. To register on EX, users must agree to the terms of use and privacy policy, which state that (1) Truth Initiative automatically collects information about use of the site, (2) information is used for research and quality improvement purposes, and (3) personal information is kept confidential. To enroll in the EX text message program, users enter their mobile number and explicitly consent to receive text messages during website registration. The study was conducted as a quality improvement project, meaning that eligible individuals were automatically randomized; no recruitment information was presented, and no study informed consent was solicited. The Chesapeake institutional review board approved the trial protocol (CR00086431).

Web-Based Cessation Program

EX, which is accessible on any web-enabled device, was designed around tobacco dependence treatment guidelines [35], Social Cognitive Theory [36], and the Mayo Clinic model for engaging tobacco users in cessation treatment [37]. At the time of this study, users could engage with 6 interactive features: (1) Set Quit Date assists users in selecting a quit date, (2) Track Smoking Triggers allows users to track cigarettes and identify personal smoking triggers associated with smoking, (3) Beat Smoking Triggers encourages identification of strategies to dissociate cigarettes from triggers, (4) Choose Quit Smoking Aid educates users about medication and helps them create a medication plan, (5) Build Support System discusses the importance of social support and encourages users to identify supportive friends/family, and (6) EX Community introduces users to a large online social network of current and former smokers. Higher levels of engagement with the site and its features are associated with greater odds of quitting [8,38].

Text Messaging Intervention

The standard EX text message program is a 90-day program that is fully automated. Also designed around tobacco cessation treatment guidelines [35] and Social Cognitive Theory [36], like the EX website, the program is designed to enhance motivation for quitting, educate users about addiction, encourage the use of evidence-based cessation methods (eg, nicotine replacement therapy and peer support), prepare them for quitting by building confidence and coping skills, praise success, reinforce the benefits of quitting, and support users in

maintaining abstinence. The program is interactive, with bidirectional messages (eg, true/false, multiple-choice, and yes/no questions) and keywords that allow users to request on-demand support (eg, CRAVE, SLIP, and MOOD), additional detail about specific topics (MORE), or to unenroll (STOP). Approximately 10% of messages contain URLs to the EX website for additional information. The program is tailored to a participant's quit date, should they choose to set one. Users can set and reset their quit date as often as they like either via the website or text messages. This basic level of integration is expected by users and was common across all experimental conditions.

Text messages focused on promoting engagement were added to this standard program according to the experimental factors described in the following section. These messages were designed to encourage or facilitate the use of the interactive features of EX and promote sustained engagement.

Experimental Factors

Personalization

Half of the participants received text messages personalized with their EX username, chosen during registration on the website. The remaining participants received generic messages with no personalization.

Integration

Participants with this experimental factor turned *on* were able to engage with interactive features of EX through special keywords and interactive messages. For example, the use of Track Smoking Triggers and Beat Smoking Triggers was facilitated as follows:

> *System: Quitting is a journey—not an event. If you've had a slip, learn about your triggers by texting back SMOKED or going to http://bit.ly/trackcig and noting your high-risk situations.*
>
> *User: SMOKED.*
>
> *System: Ok, you smoked—we saved it in your EX Plan. Reply with the trigger that made you want to smoke.*
>
> *User: Sitting in traffic.*
>
> *System: Got it, we've noted it in your EX Plan. To make a plan to separate from this trigger, text back COPE.*
>
> *User: COPE.*
>
> *System: You said that Sitting in traffic was a trigger for you. What else can you do instead of smoke the next time that trigger comes up? Reply with your plan, the more detail the better!*
>
> *User: Listen to music.*
>
> *System: Great idea! It's now in your EX Plan at http://bit.ly/beattrigger so you can always reference it if that trigger pops up.*

As seen in this exchange, interaction with the platform via text messages was reflected on the website at the user's next visit. This approach was used for the first 5 interactive features described earlier. To facilitate engagement with the EX Community via text message, a slightly different approach was required for technical reasons. Once a participant had accessed the community via the website, a special keyword (TIPS) allowed the user to request advice and guidance from community members that had been manually curated from actual community posts. For participants with the integration feature turned off, there was no integration of the website and text message programs beyond the quit date feature, and special keywords were not available.

Dynamic Tailoring

Half of the participants received messages tailored to real-time engagement data from EX. Messages reinforced actions that they had already taken or prompted the exploration of features they had not yet used. The remaining participants received standard messages that were agnostic to prior use of EX.

Message Intensity

The intervention duration was 12 weeks for both groups. Both groups received 2 messages per day for the first 3 days of the program to ensure a standardized onboarding experience, and in both program versions, approximately half of all engagement messages solicited a response from users. Participants randomized to *tapered* intensity received a total of 69 engagement messages delivered as follows: weeks 1 to 2, 14 messages per week; weeks 3 to 4, 7 messages per week; weeks 5 to 8, 4 messages per week; weeks 9 to 11, 3 messages per week; and week 12, 2 messages. Participants randomized to an *abrupt* intensity drop-off received a total of 28 engagement messages, which were delivered as follows: week 1, 8 messages; week 2, 4 messages; weeks 3 to 8, 2 messages per week; weeks 9 to 12, 1 message per week.

Sources of Data

Gender, age, and smoking status (*every day* or *some days*) entered during website registration were extracted from the EX database. Website utilization metrics were extracted at 3-month postrandomization and included measures of website engagement (number of website visits, time on site in minutes, and page views) and the use of the 6 interactive features described earlier. Text message data were also extracted at 3-month postrandomization and included the number of messages received and sent by participants, the use of the 6 interactive features described earlier via text messages, the use of keywords, and the date of unenrollment.

Analytic Plan

A full factorial design was used in the study design phase [34]. The primary outcome used for sample size calculations was a composite engagement score, with weights given by the regression coefficients of a logistic regression model developed to measure the effects of website engagement on 3-month abstinence rates in the control arm of a previous randomized trial by our group [7]. This composite engagement score had the advantage of being continuously distributed, even if some of the individual engagement metrics were binary or count data. A priori sample size calculations based on a normal approximation to the distribution of the composite score determined that a sample of 864 (n=432 per factor level)

participants would allow us to detect small main effects (d=0.25) or moderate second-order interactions (d=0.50) on normalizing transformations of this composite outcome. Power was set at 80% at a 2-sided significance level of alpha of .05 out of 10 (multiplicity adjustment based on 4 main effects and 6 two-way interactions in a factorial model, with third- and fourth-order interactions excluded a priori).

In carrying out our original analytic plan [34], we made 3 post hoc modifications. First, ongoing enhancements to the EX website led us to question the applicability of the weights of the original composite engagement metric. We decided to analyze our engagement metrics separately and to identify common patterns in standardized factorial effects across the full set of engagement metrics. Our primary outcomes were (1) number of interactive features used through the web or text messages, (2) total time spent on the website, (3) number of page views, and (4) return visits to the website (ie, postregistration). This approach is consistent with the multidimensional nature of engagement and numerous systematic reviews and meta-analyses that have called for more standardized engagement metrics to advance the field [10-12,30,39,40]. Our intent was to ensure that study effect sizes could be included in pooled analyses.

Second, evidence of synergistic interactions led us to supplement analyses focused on individual experimental factors with between-arm comparisons that capture the joint effect of multiple terms in the full factorial model, with the hope of identifying an arm with superior performance across all engagement metrics. If such an arm could be identified, the need to reestimate weights for a composite metric would become moot: any arm that dominated each available metric would also dominate their weighted average. To facilitate such between-arm comparisons, we presented CIs for each engagement metric for all 16 arm-specific means, with the confidence level adjusted so that an overlap in the respective CIs can be interpreted as lack of significant pairwise differences between the arms being compared at the alpha value of .05 significance level. Unlike cases where an arm-specific mean is being compared with constant, pairwise comparisons of means based on the overlap method involve uncertainty in the centers of both CIs under inspection [41]. In such cases, 2 arms may still be significantly different from one another at an alpha value of .05, even if their 95% CIs overlap. To correctly assess the presence of significant pairwise differences based on the overlap rule, we have employed narrower intervals whose individual confidence levels were set to about 83.5% under normality [42]. CIs were first calculated in the scale of the continuously and normally distributed linear predictor and then back-transformed to the original outcome scale.

Third, large skewness observed in time on site, page views, and website visits led us to transform the data to reduce the impact of outliers in the final model. Although a logarithmic transformation would have served this purpose, it would have changed the interpretation of the main effect of each factor in the original scale to an average of arm-specific medians. We dichotomized these variables instead at cut points that bifurcated the sample at an approximately 1:2 ratio. The cut points were one or more unique website visits (333/864, 38.5% of the sample), 15 or more minutes of website use (257/864, 29.8% of the sample), and 25 or more page views (278/864, 32.2% of the sample). Additional sensitivity analyses examined the robustness of the findings to variation in these cutoffs.

Use (yes/no) of specific interactive features and overall engagement metrics were analyzed via logistic regression except for interactive feature utilization, a count variable analyzed via quasi-Poisson regression with a scale parameter, ϕ, to account for under- or overdispersion. CIs for marginal factor effects were estimated via a parametric percentile bootstrap procedure [43] with 1 million replicates. Effect size calculations were based on standardized mean differences (SMDs) between high and low levels of experimental factors when calculating main effects and between arms 1 and 16 when calculating the full impact of the intervention, including main effects and higher order interactions. Arm 1 was defined with all experimental factors *off*, whereas Arm 16 was defined by all experimental factors *on*. SMDs for frequency counts were calculated as $(\mu_1-\mu_2)/[\phi(\mu_1+\mu_2)]^{1/2}$, where μ_1 and μ_2 were the sample means of each comparison group. SMDs for binary outcomes were calculated as $(p_1-p_2)/[p_1\times q_1+p_2\times q_2]^{1/2}$, where $p_1=1-q_1$ and $p_2=1-q_2$ were the sample outcome prevalence of each comparison group. All analyses were conducted using the *glm* function in R version 3.6.1 (R Foundation).

Results

Participants

Between March 29 and June 5, 2018, 864 newly registered users on EX who met the study eligibility criteria were randomized. Of those, 83.4% (721/864) enrolled on a mobile device, and 16.6% (143/864) enrolled on a desktop. Most (844/864, 97.7%) participants were every day smokers, and 2.3% (20/864) of the participants were some day smokers. The sample was predominantly female (637/864, 73.7%). Age distribution was as follows: 18 to 30 years (182/864, 21.1%), 31 to 44 years (322/864, 37.3%), 45 to 65 years (313/864, 36.2%), and 65 years and older (47/864, 5.4%). No between-arm differences were observed for any of the abovementioned variables (all values for P>.15).

Intervention Engagement

Of the 864 participants randomized, 461 (53.5%) completed the full 90-day text message program. Among participants who unenrolled, the median day of unenrollment was 8 days postrandomization (IQR 3-22). On average, study participants used, on average, 2.40 (SD 1.41) of the 6 targeted interactive features. Use by feature was as follows: Set Quit Date, 85.8% (741/864); Track Smoking Triggers, 55.4% (479/864); Choose Quit Smoking Aid, 37.0% (320/864); Visit Community, 33.4% (289/864); Beat Smoking Triggers, 15.3% (132/864); and Build Support System, 12.7% (110/864). Study participants received a median of 87 text messages (IQR 27-160) during the 3-month intervention period and sent a median of 4 text messages (IQR 3-12). One-fourth (213/864, 24.7%) of the sample used one or more keywords: MOOD (76/864, 8.8%), HELP (74/864, 8.6%), CRAVE (70/864, 8.1%), SLIP (60/864, 6.9%), SOS (34/864, 3.9%). Among participants randomized to the active integration

arms, 11.6% used at least one special keyword: SMOKED (61/864, 7.1%), COPE (46/864, 5.3%), TIPS (27/864, 3.1%), TRIGGER (25/864, 2.9%), MEDS (14/864, 1.6%). Participants in the active integration arms used standard keywords at similar rates to other participants (all differences <1 percentage point).

Engagement Outcomes

Table 1 (interactive features) and Table 2 (key engagement metrics) reflect the study findings under the original analytic plan. They show average response levels at *on* and *off* levels of each experimental factor and raw mean differences that correspond to marginal factor effects. SMDs are also included, as they allow us to calibrate the clinical significance of our nominal *P* values. In this study, *P*<.001 corresponds to SMDs ranging from 0.17 to 0.28 (small effects in Cohen nomenclature [44]), whereas *P* values greater than .001 and less than .05 correspond to SMDs ranging from 0.09 to 0.13.

As seen in Table 1, *integration* is the strongest experimental factor affecting interactive feature utilization, raising usage rates of Choose Quit Smoking Aid by 18.7 percentage points (95% CI 12.5 to 24.8) and Build Support System by 11.8 percentage points (95% CI 7.2 to 16.4). Altogether, *integration* raised the average number of interactive features used by participants by 0.36 (95% CI 0.16 to 0.57) when the study-wide mean did not exceed 2.4 features (median 2, IQR 1-3).

Table 2 suggests that *dynamic tailoring* was the experimental factor with broadest impact, in that it raised both the average number of interactive features used by 0.29 (95% CI 0.09 to 0.50) and the probability of higher engagement levels by 7.3 percentage points for page views (95% CI 1.1 to 13.5). Its beneficial effect on interactive feature use appears driven by similar increases in the rates of Build Support System (6.3 points; 95% CI 1.7 to 10.9), Track Smoking Triggers (6.6 points; 95% CI 0.1 to 13.1), and Beat Smoking Triggers (5.4 points, 95% CI 0.5 to 10.3).

Table 1. Marginal effects of experimental design factors on interactive feature utilization rates (95% CI).

Factor	Interactive feature					
	Set Quit Date	Choose Quit Smoking Aid	Build Support System	Track Smoking Triggers	Beat Smoking Triggers	Visit Community
Personalization						
On	84.5 (81.1 to 87.9)	35.2 (30.9 to 39.6)	14.6 (11.2 to 17.9)	55.0 (50.4 to 59.5)	16.8 (13.3 to 20.4)	32.5 (28.2 to 36.8)
Off	85.7 (82.3 to 89.0)	39.3 (34.9 to 43.7)	12.3 (9.1 to 15.5)	55.7 (51.1 to 60.2)	15.0 (11.6 to 18.4)	35.0 (30.6 to 39.4)
Raw difference	−1.1 (−5.9 to 3.6)	−4.1 (−10.3 to 2.1)	2.3 (−2.3 to 6.9)	−0.7 (−7.1 to 5.8)	1.8 (−3.1 to 6.7)	−2.5 (−8.7 to 3.7)
SMD[a]	−0.02 (−0.12 to 0.07)	−0.06 (−0.15 to 0.03)	0.05 (−0.05 to 0.14)	−0.01 (−0.10 to 0.08)	0.04 (−0.06 to 0.13)	−0.04 (−0.13 to 0.06)
Integration						
On	84.5 (81.1 to 88.0)	46.6 (42.0 to 52.2)	19.3 (15.6 to 23.0)	58.2 (53.6 to 62.7)	17.7 (14.1 to 21.3)	32.3 (28.0 to 36.6)
Off	85.7 (82.3 to 89.0)	27.9 (23.8 to 32.1)	7.5 (4.8 to 10.3)	52.5 (47.9 to 57.1)	14.1 (10.8 to 17.4)	35.2 (30.8 to 39.6)
Raw difference	−1.1 (−5.9 to 3.7)	18.7 (12.5 to 24.8)[b]	11.8 (7.2 to 16.4)[b]	5.7 (−0.8 to 12.1)	3.6 (−1.3 to 8.5)	−2.9 (−9.1 to 3.2)
SMD	−0.02 (−0.12 to 0.07)	0.28 (0.18 to 0.38)[b]	0.25 (0.15 to 0.34)[b]	0.08 (−0.01 to 0.17)	0.07 (−0.03 to 0.17)	−0.04 (−0.14 to 0.05)
Dynamic Tailoring						
On	87.0 (83.8 to 90.3)	37.7 (33.3 to 42.1)	16.6 (13.1 to 20.1)	58.7 (54.1 to 63.1)	18.6 (15.0 to 22.3)	36.6 (32.1 to 41.0)
Off	83.1 (79.7 to 86.7)	36.8 (32.5 to 41.2)	10.3 (7.3 to 13.3)	52.1 (47.4 to 56.7)	13.2 (9.9 to 16.4)	30.9 (26.6 to 35.2)
Raw difference	3.9 (−0.9 to 8.6)	0.9 (−53 to 7.1)	6.3 (1.7 to 10.9)[c]	6.6 (0.1 to 13.1)[d]	5.4 (0.5 to 10.3)[d]	5.7 (−0.5 to 11.9)
SMD	0.08 (−.02 to 0.17)	0.01 (−0.08 to 0.10)	0.13 (0.04 to 0.23)[c]	.09 (0.00 to 0.19)[d]	.11 (0.01 to 0.20)[d]	.09 (−.01 to 0.18)
Intensity						
Tapered	83.4 (79.9 to 86.9)	38.2 (33.7 to 42.6)	14.6 (11.2 to 17.9)	56.4 (51.8 to 60.9)	18.2 (14.6 to 21.8)	31.4 (27.1 to 35.7)
Abrupt	86.8 (83.6 to 90.1)	36.4 (32.0 to 40.7)	12.3 (9.1 to 15.5)	54.3 (49.7 to 58.9)	13.6 (10.3 to 16.9)	36.1 (31.7 to 40.6)
Raw difference	−3.4 (−8.2 to 1.4)	1.8 (−4.4 to 8.0)	2.3 (−2.3 to 6.8)	2.1 (−4.4 to 8.5)	4.5 (−0.4 to 9.5)	−4.8 (−10.9 to 1.4)
SMD	−0.07 (−0.16 to 0.03)	0.03 (−0.06 to 0.12)	0.05 (−0.05 to 0.14)	0.03 (−0.06 to 0.12)	0.09 (−0.01 to 0.18)	−0.07 (−0.16 to 0.02)

[a]SMD: standardized mean difference.

[b]$P<.001$.

[c]$P<.01$.

[d]$P<.05$.

Table 2. Marginal effects of experimental design factors on key engagement metrics (95% CI).

Factor	Key Engagement Metric			
	Feature utilization, mean	Page views ≥25, %	Time on site ≥15 min, %	Returned to website, %
Personalization				
On	2.38 (2.24 to 2.53)	31.6 (27.2 to 36.0)	31.3 (27.0 to 35.6)	37.9 (33.3 to 42.4)
Off	2.43 (2.28 to 2.58)	32.9 (28.5 to 37.3)	28.4 (24.1 to 32.6)	39.1 (34.6 to 43.7)
Raw difference	−0.05 (−0.25 to 0.16)	−1.3 (−7.5 to 4.9)	2.9 (−3.2 to 9.0)	−1.2 (−7.7 to 5.2)
Standard mean difference	−0.021 (−0.115 to 0.074)	−0.017 (−0.111 to 0.077)	0.045 (−0.049 to 0.139)	−0.019 (−0.113 to 0.075)
Integration				
On	2.59 (2.44 to 2.74)	30.8 (26.5 to 35.2)	29.7 (25.4 to 34.0)	37.0 (32.4 to 41.5)
Off	2.22 (2.08 to 2.37)	33.7 (29.3 to 38.1)	29.9 (25.6 to 34.0)	40.0 (35.5 to 44.6)
Raw difference	0.36 (0.16 to 0.57)[a]	−2.9 (−9.1 to 3.3)	−0.2 (−6.3 to 5.9)	−3.0 (−9.5 to 3.4)
Standard mean difference	0.166 (0.071 to 0.260)[a]	−0.043 (−0.137 to 0.051)	−0.003 (−0.097 to 0.091)	−0.044 (−0.138 to 0.048)
Dynamic Tailoring				
On	2.55 (2.40 to 2.70)	35.9 (31.4 to 40.4)	32.3 (28.0 to 36.7)	41.6 (36.9 to 46.2)
Off	2.26 (2.12 to 2.40)	28.6 (24.3 to 32.9)	27.3 (23.1 to 31.0)	35.4 (30.9 to 39.9)
Raw difference	0.29 (0.09 to 0.50)[b]	7.3 (1.1 to 13.5)[c]	5.0 (−1.0 to 11.1)	6.2 (−0.3 to 12.6)
Standard mean difference	0.134 (0.040 to 0.229)[b]	0.110 (0.016 to 0.205)[c]	0.078 (−0.016 to 0.173)	0.089 (−0.004 to 0.184)
Intensity				
Tapered	2.42 (2.27 to 2.57)	33.1 (28.72 to 37.6)	31.3 (27.1 to 35.7)	40.0 (35.5 to 44.6)
Abrupt	2.39 (2.25 to 2.54)	31.4 (27.0 to 35.8)	28.3 (24.1 to 32.6)	36.9 (32.5 to 41.5)
Raw difference	0.03 (−0.18 to 0.23)	1.7 (−4.5 to 8.0)	3.0 (−3.0 to 9.1)	3.1 (−3.3 to 9.5)
Standard mean difference	0.012 (−0.083 to 0.106)	0.028 (−0.066 to 0.122)	0.047 (−0.047 to 0.141)	0.045 (−0.049 to 0.139)

[a]SMD: standardized mean difference.

[b]$P<.001$.

[c]$P<.01$.

The main effects of *intensity* and *personalization* failed to attain even nominal levels of statistical significance on any engagement metrics. However, the impact of these 2 experimental factors was still beneficial as a whole via their synergistic interactions with *dynamic tailoring* and *integration*. Examination of the factorial models for each interactive feature in isolation revealed a synergistic interaction between *integration* and *personalization* on Set Quit Date ($P<.001$) and a synergistic interaction between *dynamic tailoring*, *integration*, and *personalization* on Track Smoking Triggers ($P=.02$). A synergistic interaction of *dynamic tailoring* × *integration* × *intensity* × *personalization* was also detected for time on site ($P=.01$), while a synergistic interaction of *integration* and *personalization* was detected for time on return visits to the website ($P=.04$).

To better understand the joint effect of all 4 experimental factors, Table 3 (interactive features) and Table 4 (key engagement metrics) present point estimates and 95% CIs for arm-specific means based on simulation findings also depicted in Multimedia Appendices 1 and 2. As seen in Table 3, Arm 16 yielded the highest engagement rates for Set Quit Date, Build Support System, Track Smoking Triggers, and Beat Smoking Triggers. It lagged behind other arms in terms of the proportions of users at the high engagement level for Choose Quit Smoking Aid and Visit Community. Overall, Arm 16 had the highest rate of interactive feature use (2.95 out of 6).

As seen in Table 4, Arm 16 also had the greatest proportion of participants at the high engagement level for page views (45%) and return visits (51%), although it ranked second with regard to the proportion at the high engagement level for time on site (41%). Sensitivity analyses that varied the cutoffs for page views and time spent on site from the 55th to the 75th percentile (ie, from 15 to 32 pages and 8 to 20 min, respectively) confirmed the superiority of Arm 16, suggesting that these findings are robust to the choice of cutoff.

Table 3. Arm-specific interactive feature utilization rates (95% CI).

Arm	Factor[a]				Interactive Feature					
	P[b]	IG[c]	T[d]	IS[e]	Set quit date	Choose quit smoking aid	Build support system	Track smoking triggers	Beat smoking triggers	Visit community
1	−	−	−	−	95 (86-99)	21 (12-33)	3 (0-12)	48 (35-61)	14 (6-25)	41 (29-54)
2	−	−	−	+	88 (77-95)	43 (30-56)	6 (2-16)	48 (35-61)	6 (2-16)	32 (21-45)
3	−	−	+	−	86 (75-94)	28 (17-41)	8 (3-18)	55 (42-68)	12 (5-23)	39 (27-52)
4	−	−	+	+	90 (80-96)	30 (19-43)	12 (5-23)	61 (48-73)	21 (12-33)	45 (32-58)
5	−	+	−	−	85 (73-92)	54 (40-66)	10 (4-20)	57 (44-70)	8 (3-18)	34 (22-47)
6	−	+	−	+	75 (63-85)	45 (32-58)	7 (9-29)	65 (51-76)	21 (12-33)	23 (13-35)
7	−	+	+	−	81 (69-90)	43 (30-56)	23 (13-35)	52 (39-65)	17 (9-29)	37 (25-51)
8	−	+	+	+	85 (73-92)	52 (39-65)	19 (10-31)	59 (46, 71)	21 (12-33)	30 (19-43)
9	+	−	−	−	83 (71-91)	25 (15-37)	5 (1-14)	52 (39-65)	12 (5-23)	26 (16-39)
10	+	−	−	+	75 (63-85)	25 (15-37)	5 (1-14)	48 (35-61)	14 (6-25)	30 (19-43)
11	+	−	+	−	86 (75-94)	23 (13-35)	10 (4-20)	54 (40-66)	14 (6-25)	35 (24-49)
12	+	−	+	+	81 (69-90)	30 (19-43)	12 (5-23)	54 (40-66)	21 (12-33)	34 (22-47)
13	+	+	−	−	86 (75-94)	46 (34-60)	21 (12-33)	52 (39-65)	14 (6-25)	34 (22-47)
14	+	+	−	+	77 (65-87)	37 (25-51)	15 (8-27)	46 (34-60)	17 (9-29)	28 (17-41)
15	+	+	+	−	92 (82-97)	52 (39-65)	19 (10-31)	65 (51-76)	19 (10-31)	43 (30-56)
16	+	+	+	+	95 (86-99)	45 (32-58)	30 (19-43)	70 (57-81)	25 (15-37)	30 (19-43)

[a]For P, IG, and T, + implies On and − implies Off. For IS, + implies Tapered and − implies Abrupt.

[b]P: Personalization.

[c]IG: Integration.

[d]T: Dynamic Tailoring.

[e]IS: Intensity.

Table 4. Arm-specific summaries of key engagement metrics (95% CI).

Arm	Factor[a]				Key Engagement Metric			
	P[b]	IG[c]	T[d]	IS[e]	Feature utilization (mean)	Page views ≥25 (percent)	Time on site ≥15 min (percent)	Returned to website (percent)
1	−	−	−	−	2.21 (1.84-2.64)	34 (23-48)	23 (13-36)	47 (34-60)
2	−	−	−	+	2.23 (1.86-2.66)	28 (18-42)	19 (10-32)	36 (24-49)
3	−	−	+	−	2.29 (1.91-2.72)	35 (23-49)	24 (14-37)	49 (36-62)
4	−	−	+	+	2.58 (2.18-3.04)	44 (31-58)	44 (31-58)	44 (31-58)
5	−	+	−	−	2.47 (2.08-2.92)	28 (18-42	30 (19-44)	26 (16-40)
6	−	+	−	+	2.45 (2.06-2.9)	28 (18-4)	26 (16-39)	37 (25-51)
7	−	+	+	−	2.53 (2.13-2.98)	33 (22-47)	29 (19-43)	37 (25-51)
8	−	+	+	+	2.66 (2.25-3.12)	29 (18-42)	29 (18-42)	35 (23-48)
9	+	−	−	−	2.01 (1.66-2.41)	26 (16-39)	26 (16-39)	30 (19-43)
10	+	−	−	+	1.95 (1.61-2.35)	31 (20-44)	33 (21-46)	37 (25-50)
11	+	−	+	−	2.21 (1.84-2.64)	32 (21-46)	34 (23-48)	32 (21-46)
12	+	−	+	+	2.31 (1.93-2.74)	37 (25-50)	35 (23-48)	44 (31-58)
13	+	+	−	−	2.53 (2.13-2.98)	29 (19-43)	37 (25-51)	33 (22-47)
14	+	+	−	+	2.21 (1.84-2.64)	22 (12-35)	22 (12-35)	35 (23-49)
15	+	+	+	−	2.90 (2.47-3.38)	30 (19-43)	21 (12-33)	39 (27-52)
16	+	+	+	+	2.95 (2.52-3.44)	45 (32-59)	41 (29-55)	51 (38-64)

[a]For P, IG, and T, + implies On and − implies Off. For IS, + implies Tapered and − implies Abrupt.

[b]P: Personalization.

[c]IG: Integration.

[d]T: Dynamic Tailoring.

[e]IS: Intensity.

Significance of pairwise differences at an alpha of .05 can be evaluated using the overlap method applied to the 83.5% CIs shown in Figures 1 and 2. The results confirmed the conclusion that Arm 16 was either the top-ranked arm or did not differ from the top-ranked arm for any of the key engagement metrics and individual interactive features across all 16 experimental conditions. However, this begs the question of whether the intervention as a whole led to a significant improvement in utilization outcomes over no intervention at all. To answer this question, we calculated Arm 1 versus Arm 16 SMDs for all outcomes of interest. We found that the combined effect of our 4 experimental factors on interactive utilization features was highest for Build Support System (SMD 0.56; 95% CI 0.27 to 0.81), followed by Choose Quit Smoking Aid (SMD 0.38; 95% CI 0.10 to 0.66) and Track Smoking Triggers (SMD 0.33; 95% CI 0.05 to 0.61). No significant effect was found for Beat Smoking Triggers (SMD 0.20; 95% CI −0.07 to 0.47), Set Quit Date (SMD 0.00; 95% CI −0.28 to 0.28), or Visit Community (SMD −0.17; 95% CI −0.44 to 0.10). As for our key engagement metrics, the largest effect was on overall feature utilization (SMD 0.33; 95% CI 0.06 to 0.59), followed by time on site (SMD 0.29; 95% CI 0.01 to 0.57), page views (SMD 0.16; 95% CI −0.11 to 0.44), and return visits to the website (SMD 0.05; 95% CI −0.22 to 0.33). As no SMD >.30 was observed for main factor effects on any of the outcomes of interest, these results also suggest that, for at least some outcomes, the combined intervention was stronger than individual factors alone.

Figure 1. Arm-specific interactive feature utilization rates.

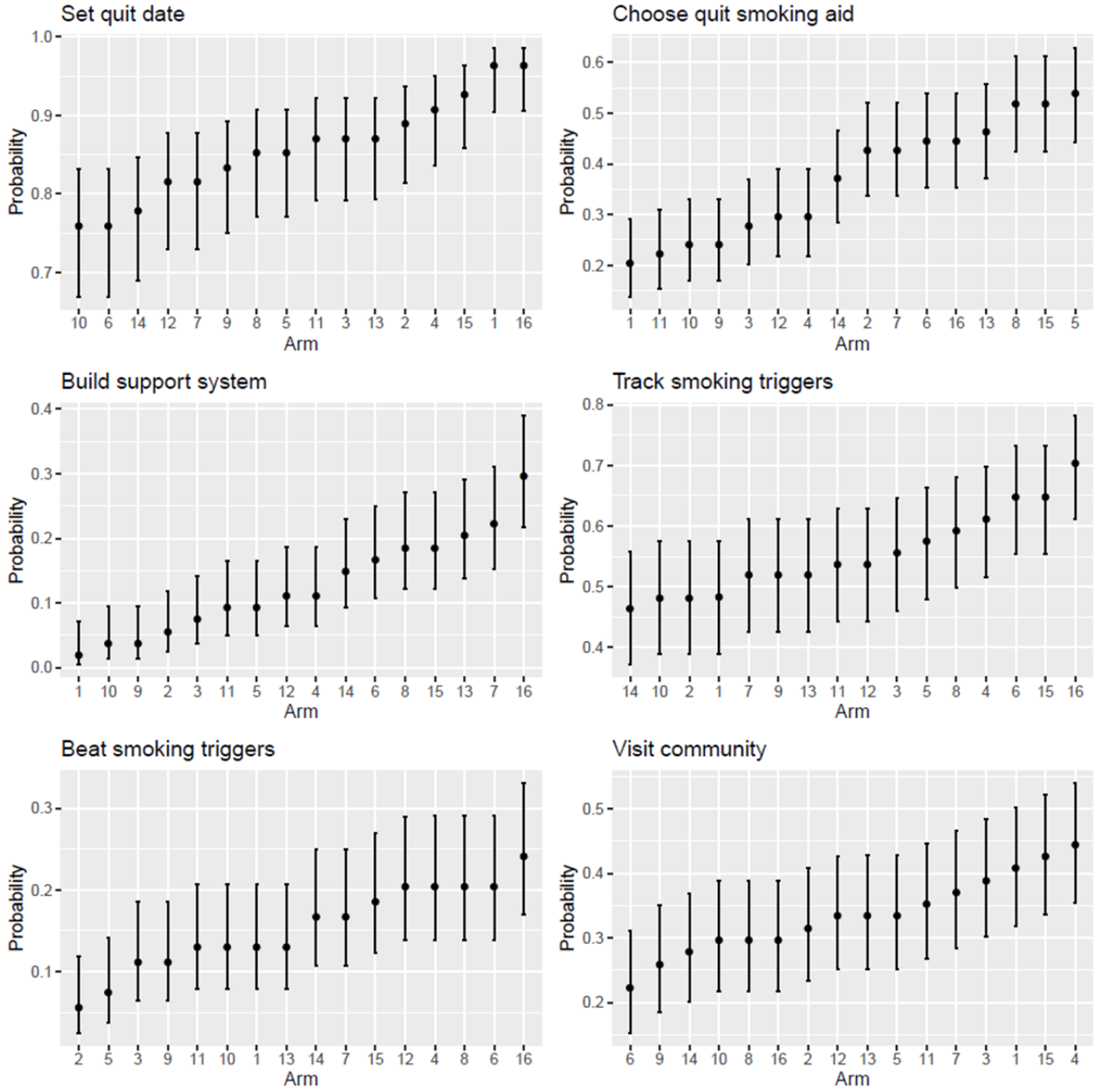
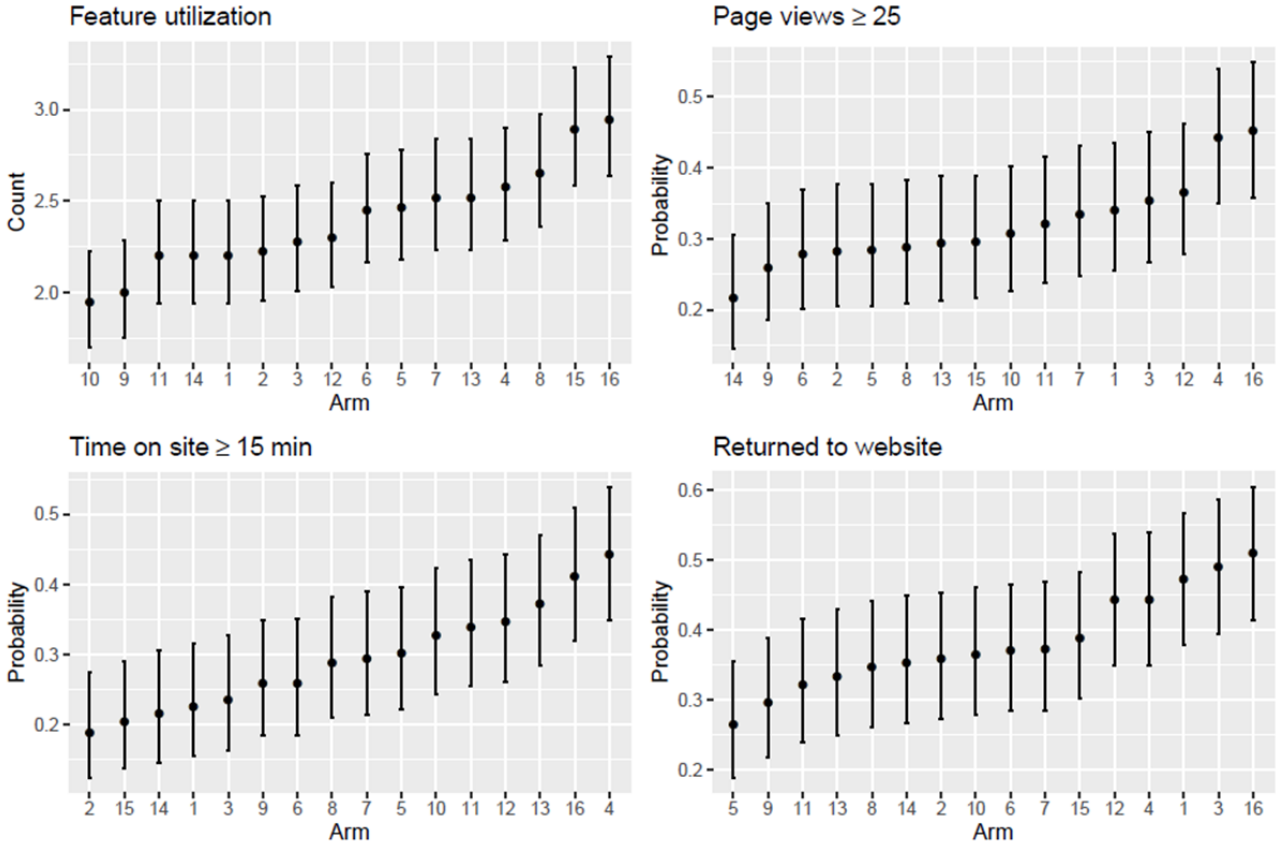


Figure 2. Arm-specific summaries of key engagement metrics (95% CIs).



Discussion

Principal Findings

The aim of this factorial screening experiment was to test the effectiveness of text message design factors in increasing treatment engagement among adult smokers who enrolled in an internet smoking cessation intervention. We examined general metrics of engagement (ie, page views, time on site, and return visits) and specific engagement metrics for core intervention components. As hypothesized, synergistic factor effects in Arm 16, in which all 4 experimental factors were active simultaneously, resulted in the highest rates of interactive feature use. Nearly all participants (95%) set a quit date, 70% tracked their triggers, and approximately half (45%) designated a medication plan. Arm 16 also yielded the greatest proportion of participants at high levels of engagement, with 40% to 50% of the sample engaged at the highest levels of page views, time on site, and return visits.

To our knowledge, this is the first study that has examined the impact of tailoring content based on treatment engagement. *Dynamic tailoring*, aimed at showcasing intervention features that participants had not yet used and encouraging ongoing utilization of those they had, was the most powerful of the 4 experimental factors tested in increasing engagement. It resulted in more participants engaging with the core components of tobacco dependence treatment, namely identifying and rallying the support of key people in their social network, identifying triggers for smoking, and developing coping strategies for those triggers. These results are consistent with previous research showing the effectiveness of individually tailored content and demonstrate that text messages can be an effective strategy to help shepherd and guide users through an intervention, much like would happen in a face-to-face encounter.

Although *dynamic tailoring* encouraged participants to explore the most program features, enabling users' access to these features via text messages had the most dramatic impact on individual feature use. With a relatively simple mechanism to engage users in an interactive fashion via text messaging, *integration* was effective at increasing the utilization of interactive features above the study-wide mean. It increased the use of tools related to medication selection and planning for social support by 19 and 12 percentage points, respectively. The fact that we did not observe an impact of integration on setting a quit date likely results from a ceiling effect, given that 85% of all participants set a quit date, often immediately after website registration.

The lack of significant main effects for *personalization* and *intensity* is worth noting. Messages designed to feel individually tailored by using a person's name, but where the content is clearly generic, did not appear to enhance program utilization. These findings are consistent with previous research that has shown that the use of a person's name alongside generic information that is not perceived as personally relevant may even have counterproductive effects [45]. The fact that we only detected the synergistic effects of *personalization* when implemented alongside *dynamic tailoring* and *integration* is consistent with this notion. *Intensity* in this study was operationalized as the schedule of message delivery, and it was hypothesized that messages sent at less frequent and changing intervals over the 12-week intervention period would be more

impactful than messages sent at a fixed interval with an abrupt drop-off. This hypothesis was not supported. It is possible that differences between the tapered and abrupt arms yielded variations in the dose of text messages received (eg, number of days enrolled, number of messages received), which we intend to explore in secondary analyses.

Finally, it is also worth noting that none of the experimental factors we tested increased engagement with the online community. For technical reasons, this was the only interactive feature that required the user to first visit the website to subsequently engage with community content via an SMS text message (ie, TIPS keyword). It may be that a text message approach that did not require a website action may have yielded different findings. Alternatively, it may be that interest in and use of social support resources—whether online or offline—may be a more trait-like characteristic that is not subject to external manipulation [46]. Other research has also failed to increase the use of an online social network in an experimental design [47,48].

Limitations

Several limitations should be considered. This study does not allow us to draw conclusions about the impact of an engagement strategy versus none on smoking outcomes [10]. This factorial screening experiment was conducted as the first phase of 2-phase trial. Whereas all 16 arms were compared in terms of their ability to increase engagement with a smoking cessation intervention, the next phase of this study involves a comparative effectiveness trial (currently underway), which will allow us to evaluate the impact of the presence versus absence of a comprehensive engagement strategy (ie, Arm 16) in increasing abstinence rates. In addition, we cannot disentangle the effect of engagement messages alone because they were delivered as part of a broader text message intervention. This was a deliberate design decision because text messages solely focused on promoting engagement without reference to a user's progress in quitting would likely have been perceived as irrelevant. Finally, we are cognizant of the fact that more does not always equal better when it comes to digital engagement [49,50]. Our classification of high levels of engagement was based on empirical distributions, which may not necessarily correlate with clinically meaningful engagement (ie, capable of promoting abstinence). In previous research, McClure et al [51] found positive effects of prescriptive message tone, dictated content viewing order, and reminder emails in a factorial screening experiment focused on internet engagement, but none of these features enhanced cessation outcomes [52]. Phase 2 of this study will enable us to evaluate the impact of the level of engagement on smoking outcomes, to investigate the role of complex issues such as reverse causality and confounding factors on the causal pathway from engagement to outcomes, and to determine what constitutes clinically meaningful engagement.

Given the proliferation of mobile apps for smoking cessation, one may question our use of text messaging as an engagement strategy over push notifications via a mobile app. Several factors support our decision. Text messaging is a recommended cessation modality [53], whereas the evidence for smartphone apps is lacking [15]. A majority of apps that are downloaded are either never opened or used only once [54]. In addition, text messaging may feel less intrusive to users and be more widely accepted: 57% of app users uninstall/decline to install apps because of privacy concerns [55]. Finally, smartphone penetration lags behind high rates of cellphone ownership [18].

Comparison With Prior Work

Program completion results compare favorably with a large study from SmokefreeTXT [56], a US-based text message program from the National Cancer Institute. Among 25,283 individuals who subscribed to SmokefreeTXT, 38.3% (n=9686) completed the entire 42-day program. In our trial, 53.5% (461/864) of all study participants completed the full 90-day text message program. In both programs, a sizable number of participants disengaged early in treatment. We included all participants randomized to treatment in our analyses, whereas Augustson et al [56] restricted analyses to those that fully initiated treatment (ie, set a quit date and received first full day of treatment). Understanding patterns of early opt out from text message interventions and identifying opportunities for improving program delivery remain important areas of inquiry.

This study addresses several gaps in the literature on digital interventions. Previous studies on improving engagement have largely focused on the use of email and telephone calls [10]. The use of text messaging as an engagement strategy is novel in this regard. In addition, previous studies have suffered from small sample sizes and lack of statistical significance [10] and provided few insights into the characteristics of digital intervention approaches that make them effective for promoting engagement [19]. This study was conducted as a full factorial, which provided a reasonably powered and efficient opportunity to test for the presence of both main factor effects and pairwise interactions. Other smoking cessation trials involving a full factorial design [57] have also found that higher order interactions can account for more variance than the marginal effects. When interactions are significant, factors should not be examined in isolation, but one should consider their joint effects (ie, the sum of their main effects and multiway interactions). Although the MOST framework has been primarily described as an efficient approach for evaluating the main effects of intervention components, it can be easily adapted to accommodate higher order terms as well. Our results and those of Cook et al [57] suggest that synergistic interactions may often be present in smoking cessation trials and should be taken into account at the study design phase.

Conclusions

In conclusion, this factorial screening experiment demonstrates the effectiveness of a theory-driven text message intervention in boosting overall engagement and use of the core features of an internet smoking cessation program among adult smokers. The results suggest that enabling users to engage with the tools and content of an internet intervention via text messages and tailoring the experience based on a user's pattern of program use can boost the overall levels of engagement. These findings have relevance to improving engagement in internet health behavior change interventions more broadly and for future research into the complex relationship between engagement and outcomes. This study can serve as a model for conducting

rigorous, fully powered research on engagement as a first step in understanding how to optimize behavior change outcomes in digital interventions.

Acknowledgments

This work was supported by the National Institute on Drug Abuse of the National Institutes of Health (R01DA038139). The sponsor was not involved in the conduct of this study or the preparation of this paper. The authors would like to acknowledge the contributions of Fiona Mortell and Monica-Ann Ortiguera Mendoza.

Conflicts of Interest

AG, MJ, MA, and SC are employees of Truth Initiative, a nonprofit public health foundation, which sells enterprise digital tobacco cessation programs to support its mission-driven work. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Simulation-Based Histograms of Arms-Specific Utilization Rates: Interactive Features.
[PDF File (Adobe PDF File), 44 KB - jmir_v22i4e17734_app1.pdf]

Multimedia Appendix 2
Simulation-Based Histograms of Arms-Specific Utilization Rates: Key Engagement Metrics.
[PDF File (Adobe PDF File), 35 KB - jmir_v22i4e17734_app2.pdf]

References

1. Pew Research Center. 2019 Jun 12. Internet/Broadband Fact Sheet, 2019 URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2020-03-09]
2. Graham AL, Amato MS. Twelve million smokers look online for smoking cessation help annually: health information national trends survey data, 2005-2017. Nicotine Tob Res 2019 Jan 4;21(2):249-252 [FREE Full text] [doi: 10.1093/ntr/nty043] [Medline: 29660037]
3. McCausland K, Curry L, Mushro A, Carothers S, Xiao H, Vallone D. Promoting a Web-Based Smoking Cessation Intervention: Implications for Practice. Cases in Public Health Communication & Marketing. Proc 2011;5:3-26.
4. Rudie M, Bailey L. North American Quitline Consortium. 2018. North American Quitline Consortium FY2018 Annual Survey: Progress Update on State Quitlines URL: https://www.naquitline.org/resource/resmgr/2018_survey/NAQC_FY2018_Annual_Survey_Da.pdf [accessed 2020-03-09]
5. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 4;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
6. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
7. Graham A, Papandonatos G, Cha S, Erar B, Amato M. Improving adherence to smoking cessation treatment: smoking outcomes in a web-based randomized trial. Ann Behav Med 2018 Mar 15;52(4):331-341 [FREE Full text] [doi: 10.1093/abm/kax023] [Medline: 29878062]
8. Graham AL, Zhao K, Papandonatos GD, Erar B, Wang X, Amato MS, et al. A prospective examination of online social network dynamics and smoking cessation. PLoS One 2017;12(8):e0183655 [FREE Full text] [doi: 10.1371/journal.pone.0183655] [Medline: 28832621]
9. Danaher B, Smolkowski K, Seeley J, Severson H. Mediators of a successful web-based smokeless tobacco cessation program. Addiction 2008 Oct;103(10):1706-1712 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02295.x] [Medline: 18715238]
10. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
11. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
12. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 6;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
13. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]

14. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008 Apr 16;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]
15. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611. [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
16. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
17. van Dulmen S, Sluijs E, van Dijk L, de Ridder D, Heerdink R, Bensing J. Patient adherence to medical treatment: a review of reviews. BMC Health Serv Res 2007 Apr 17;7:55 [FREE Full text] [doi: 10.1186/1472-6963-7-55] [Medline: 17439645]
18. Pew Research Center. 2019 Jun 12. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2020-03-09]
19. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
20. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
21. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
22. Petty RE, Cacioppo JT, Schumann D. Central and peripheral routes to advertising effectiveness: the moderating role of involvement. J Comsum Res 1983;10(2):135-146. [doi: 10.1086/208954]
23. Dijkstra A. Working mechanisms of computer-tailored health education: evidence from smoking cessation. Health Educ Res 2005 Oct;20(5):527-539. [doi: 10.1093/her/cyh014] [Medline: 15701665]
24. Webb MS, Simmons VN, Brandon TH. Tailored interventions for motivating smoking cessation: using placebo tailoring to examine the influence of expectancies and personalization. Health Psychol 2005 Mar;24(2):179-188. [doi: 10.1037/0278-6133.24.2.179] [Medline: 15755232]
25. Jamison J, Naughton F, Gilbert H, Sutton S. Delivering smoking cessation support by mobile phone text message: what information do smokers want? A focus group study. J Appl Behav Res 2013;18(1):1-23. [doi: 10.1111/jabr.12004]
26. Head KJ, Noar SM, Iannarino NT, Harrington NG. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
27. Garcia Z. Chromis Technology. 2018 Feb 8. Text Messaging Stats That Matter URL: https://www.chromis.com/text-stats-that-matter/ [accessed 2020-03-09]
28. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
29. Heminger CL, Boal AL, Zumer M, Abroms LC. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Jul;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]
30. Milward J, Drummond C, Fincham-Campbell S, Deluca P. What makes online substance-use interventions engaging? A systematic review and narrative synthesis. Digit Health 2018;4:2055207617743354 [FREE Full text] [doi: 10.1177/2055207617743354] [Medline: 29942622]
31. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
32. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
33. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
34. Graham AL, Jacobs MA, Cohn AM, Cha S, Abroms LC, Papandonatos GD, et al. Optimising text messaging to improve adherence to web-based smoking cessation treatment: a randomised control trial protocol. BMJ Open 2016 Mar 30;6(3):e010687 [FREE Full text] [doi: 10.1136/bmjopen-2015-010687] [Medline: 27029775]
35. Fiore M, Jaén C, Baker T, Tobacco Use and Dependence Guideline Panel. Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: US Department of Health and Human Services; 2008.
36. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
37. Mayo Clinic. Mayo clinic-mayo foundation courses and meetings to be held in Rochester, MN, 1984. Mayo Clinic Proceedings 1984 Jul;59(7):516 [FREE Full text] [doi: 10.1016/s0025-6196(12)60447-9]

38. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013 Jan 28;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
39. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2 Suppl):TC9-T20. [doi: 10.1097/NCN.0b013e3182155274] [Medline: 21372645]
40. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of technology-based strategies to promote engagement with digital interventions: a systematic review protocol. JMIR Res Protoc 2015 Apr 28;4(2):e47 [FREE Full text] [doi: 10.2196/resprot.3990] [Medline: 25921274]
41. Cumming G, Finch S. Inference by eye: confidence intervals and how to read pictures of data. Am Psychol 2005;60(2):170-180. [doi: 10.1037/0003-066X.60.2.170] [Medline: 15740449]
42. Goldstein H, Healy MJ. The graphical presentation of a collection of means. J R Stat Soc Ser A Stat Soc 1995;158(1):175-177. [doi: 10.2307/2983411]
43. Efron B, Tibshirani R. An Introduction to the Bootstrap. London, UK: Chapman & Hall; 1993.
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. Hillsdale, NJ: Erlbaum; 1988.
45. Dijkstra A, Ballast K. Personalization and perceived personal relevance in computer-tailored persuasion in smoking cessation. Br J Health Psychol 2012 Feb;17(1):60-73. [doi: 10.1111/j.2044-8287.2011.02029.x] [Medline: 22106987]
46. Graham AL, Papandonatos GD, Zhao K. The failure to increase social support: it just might be time to stop intervening (and start rigorously observing). Transl Behav Med 2017 Dec;7(4):816-820 [FREE Full text] [doi: 10.1007/s13142-016-0458-9] [Medline: 28070778]
47. 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 1;19(3):324-332 [FREE Full text] [doi: 10.1093/ntr/ntw282] [Medline: 27836982]
48. Cutrona SL, Sadasivam RS, DeLaughter K, Kamberi A, Volkman JE, Cobb N, National Dental PBRN Collaborative Group comprises practitioners, faculty and staff who contributed to this activity. A list of these persons is at http://nationaldentalpbrn.org/publication.php. Online tobacco websites and online communities-who uses them and do users quit smoking? The quit-primo and national dental practice-based research network Hi-Quit studies. Transl Behav Med 2016 Dec;6(4):546-557 [FREE Full text] [doi: 10.1007/s13142-015-0373-5] [Medline: 27379777]
49. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an internet-based cognitive behaviour therapy program. Aust N Z J Psychiatry 2006 Jan;40(1):59-62. [doi: 10.1080/j.1440-1614.2006.01743.x] [Medline: 16403040]
50. Saul JE, Amato MS, Cha S, Graham AL. Engagement and attrition in internet smoking cessation interventions: Insights from a cross-sectional survey of 'one-hit-wonders'. Internet Interv 2016 Sep;5:23-29 [FREE Full text] [doi: 10.1016/j.invent.2016.07.001] [Medline: 30135803]
51. McClure JB, Shortreed SM, Bogart A, Derry H, Riggs K, St John J, et al. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. J Med Internet Res 2013 Mar 25;15(3):e69 [FREE Full text] [doi: 10.2196/jmir.2508] [Medline: 23529377]
52. McClure J, Peterson D, Derry H, Riggs K, Saint-Johnson J, Nair V, et al. Exploring the 'active ingredients' of an online smoking intervention: a randomized factorial trial. Nicotine Tob Res 2014 Aug;16(8):1129-1139 [FREE Full text] [doi: 10.1093/ntr/ntu057] [Medline: 24727369]
53. The Community Guide. 2011. Tobacco Use and Secondhand Smoke Exposure: Mobile Phone-Based Cessation Interventions URL: https://www.thecommunityguide.org/findings/tobacco-use-and-secondhand-smoke-exposure-mobile-phone-based-cessation-interventions [accessed 2020-03-09]
54. Perez S. Tech Crunch. 2016 May 31. Nearly 1 in 4 People Abandon Mobile Apps After Only One Use URL: https://techcrunch.com/2016/05/31/nearly-1-in-4-people-abandon-mobile-apps-after-only-one-use/ [accessed 2020-03-09]
55. Boyles J, Smith A, Madden M. Pew Research Center. 2012 Sep 5. Privacy and Data Management on Mobile Devices URL: https://www.pewresearch.org/internet/2012/09/05/privacy-and-data-management-on-mobile-devices/ [accessed 2020-03-09]
56. Augustson E, Cole-Lewis H, Sanders A, Schwarz M, Geng Y, Coa K, et al. Analysing user-reported data for enhancement of SmokefreeTXT: a national text message smoking cessation intervention. Tob Control 2017 Nov;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]
57. Cook JW, Collins LM, Fiore MC, Smith SS, Fraser D, Bolt DM, et al. Comparative effectiveness of motivation phase intervention components for use with smokers unwilling to quit: a factorial screening experiment. Addiction 2016 Jan;111(1):117-128 [FREE Full text] [doi: 10.1111/add.13161] [Medline: 26582140]

Abbreviations

EX: BecomeAnEX
MOST: Multiphase Optimization Strategy
SMD: standardized mean difference

Edited by G Eysenbach; submitted 12.01.20; peer-reviewed by V Do, O Perski; comments to author 31.01.20; revised version received 09.02.20; accepted 22.02.20; published 02.04.20.

Please cite as:
Graham AL, Papandonatos GD, Jacobs MA, Amato MS, Cha S, Cohn AM, Abroms LC, Whittaker R
Optimizing Text Messages to Promote Engagement With Internet Smoking Cessation Treatment: Results From a Factorial Screening Experiment
J Med Internet Res 2020;22(4):e17734
URL: https://www.jmir.org/2020/4/e17734
doi:10.2196/17734
PMID:32238338

Original Paper

Effectiveness of Message Frame-Tailoring in a Web-Based Smoking Cessation Program: Randomized Controlled Trial

Maria Altendorf[1], MSc; Ciska Hoving[2], PhD; Julia CM Van Weert[1], PhD; Eline Suzanne Smit[1], PhD

[1]Department of Communication Science, Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands
[2]Department of Health Promotion, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Maria Altendorf, MSc
Department of Communication Science
Amsterdam School of Communication Research
University of Amsterdam
P O Box 15791
Amsterdam, 1001 NG
Netherlands
Phone: 31 627967317
Email: m.b.altendorf@uva.nl

Abstract

Background: The content of online computer-tailored interventions is often determined to match an individual's characteristics, beliefs, and behavioral factors. These content-tailored interventions lead to better message processing and a higher likelihood of behavior change such as smoking cessation. However, a meta-analysis of online computer-tailored interventions showed that effect sizes, albeit positive, remain small, suggesting room for improvement. A promising strategy to enhance the effectiveness of online computer-tailored interventions is to tailor the message frame (ie, how a message is communicated) based on the preferred communication style of the user in addition to content-tailoring. One factor that determines an individual's communication style preference is the need for autonomy; some individuals prefer an autonomy-supportive communication style (offering choice and use of suggestive language), whereas others might prefer a directive communication style, which is replete with imperatives and does not provide choice. Tailoring how messages are presented (eg, based on the need for autonomy) is called message frame-tailoring.

Objective: The aim of the present study was to test the effectiveness of message frame-tailoring based on the need for autonomy, in isolation and in combination with content-tailoring, within the context of an online computer-tailored smoking cessation intervention. The primary outcome measure was the 7-day point-prevalence of smoking abstinence. Secondary outcomes were perceived message relevance, self-determined motivation to quit smoking, and sociocognitive beliefs.

Methods: A randomized controlled trial with a 2 (message frame-tailoring vs no message frame-tailoring) by 2 (content-tailoring vs no content-tailoring) design was conducted among adult smokers intending to quit smoking (N=273).

Results: Structural equation modeling revealed that the content-tailored condition increased smoking abstinence rates 1 month after the start of the intervention (beta=.57, *P*=.02). However, neither message frame-tailoring nor its interaction with content-tailoring significantly predicted smoking abstinence. In our model, message frame-tailoring, content-tailoring, as well as their interaction significantly predicted perceived relevance of the smoking cessation messages, which consequently predicted self-determined motivation. In turn, self-determined motivation positively affected attitudes and self-efficacy for smoking cessation, but only self-efficacy consequently predicted smoking abstinence. Participants in the control condition perceived the highest level of message relevance (mean 4.78, SD 1.27). However, messages that were frame-tailored for individuals with a high need for autonomy in combination with content-tailored messages led to significantly higher levels of perceived message relevance (mean 4.83, SD 1.03) compared to those receiving content-tailored messages only (mean 4.24, SD 1.05, *P*=.003).

Conclusions: Message frame-tailoring based on the need for autonomy seems to be an effective addition to conventional content-tailoring techniques in online smoking cessation interventions for people with a high need for autonomy; however, this is not effective in its current form for people with a low need for autonomy.

Trial Registration: Dutch Trial Register (NL6512/NRT-6700); https://www.trialregister.nl/trial/6512

(J Med Internet Res 2020;22(4):e17251) doi:10.2196/17251

KEYWORDS

online computer tailoring; smoking cessation; message frame tailoring; content tailoring; need for autonomy; randomized controlled trial

Introduction

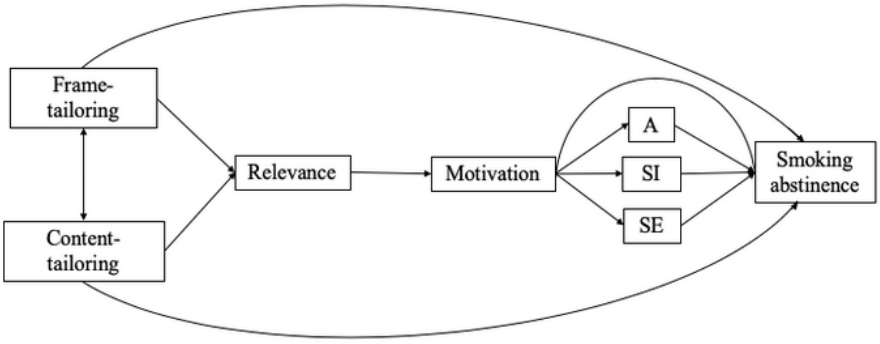
Smoking tobacco is the single most preventable cause of noncommunicable diseases such as cancer [1]. Behavioral support through online computer content-tailored (CCT) smoking cessation interventions can be effective in improving quit rates among smokers, substantially exceeding the success rates of more static interventions such as generic online smoking cessation information [2]. Online CCT smoking cessation interventions aim to provide smokers with individualized cessation information, which is assessment-based (eg, a computerized survey assesses participants' current behavioral beliefs, characteristics, and other attributes) and automatically created by computer software [3-6]. In content-tailored messages, an individual's responses are automatically matched with the relevant message content only. Previous studies have shown that content-tailored messages increase the perceived message relevance and enhance desired behavior [4,6-10]. Although online CCT smoking cessation interventions lead to better message processing and a higher likelihood of performance of advocated behaviors [5,8,11], effect sizes tend to remain small [2]. To enhance the effectiveness of online CCT health interventions, it is suggested to also use message frame-tailoring in which the message frame is matched (ie, tailor *how* a message is presented or formulated) based on a person's preferred communication style in addition to message content-tailoring [10-12]. However, no smoking cessation interventions that incorporate both content-tailoring and frame-tailoring have been rigorously tested to date.

A promising factor for message frame-tailoring is people's need for autonomy (NFA), which determines one's preference for an autonomy-supportive or more directive communication style, as shown in several studies conducted in face-to-face and other offline health settings in the fields of cancer prevention and healthy nutrition [11,13,14]. In self-determination theory [15], it is theorized that the satisfaction of a person's NFA is essential for the development of self-determined motivation, well-being, and behavioral engagement [16-18]. In turn, motivations to change are more likely to be translated into actions via sociocognitive beliefs (ie, attitudes, subjective norms, and self-efficacy perceptions) when this motivation is self-determined [19]. People with a higher NFA prefer to choose their own way of how to obtain a goal such as to quit smoking, whereas those with a lower NFA instead prefer to be told through clearcut expert advice how best to reach their goal [11,13,14,19,20]. To illustrate this difference, two studies on the effects of printed health communication showed that people's preference for a certain communication style moderated the intervention impact [11,13]. That is, people who received messages that were frame-tailored according to their communication style preference (eg, with a high NFA) and were presented with messages in an autonomy-supportive message style using suggestive language (eg, words such as "may" or "could") more often performed the desired behavior than those who received no frame-tailored messages or messages in a controlling message style (eg, messages in directive wording such as "must" or "should").

However, to the best of our knowledge, there has been no study investigating whether message frame-tailoring based on the NFA enhances the effectiveness of a content-tailored smoking cessation intervention in an online context. Therefore, the aim of the present study was to test the effectiveness of message frame-tailoring based on the NFA, in isolation and in combination with content-tailoring, within the context of an online CCT smoking cessation intervention. The online environment is specifically promising to enhance intervention effectiveness, as it has a great reach and is thus an "easy to access" medium compared to tailored print health information [21].

Specifically, we set out to test the following three main hypotheses: (H1a) frame-tailoring based on people's NFA will lead to higher smoking abstinence rates than no frame-tailoring, (H1b) content-tailoring will lead to higher abstinence rates than no content-tailoring, and (H1c) the combination of message frame-tailoring and content-tailoring will lead to the overall highest abstinence rates. In addition, we hypothesized that the above-described effects of message frame-tailoring, content-tailoring, and their combination are mediated by perceived relevance of the message (H2a), self-determined motivation (H2b), and sociocognitive beliefs (H2c). Figure 1 depicts the full conceptual model.

Figure 1. Conceptual model. Smoking abstinence was measured as the 7-day point prevalence of absence of smoking. A, attitudes; SE, self-efficacy; SI, social influence.



Methods

Study Design and Procedure

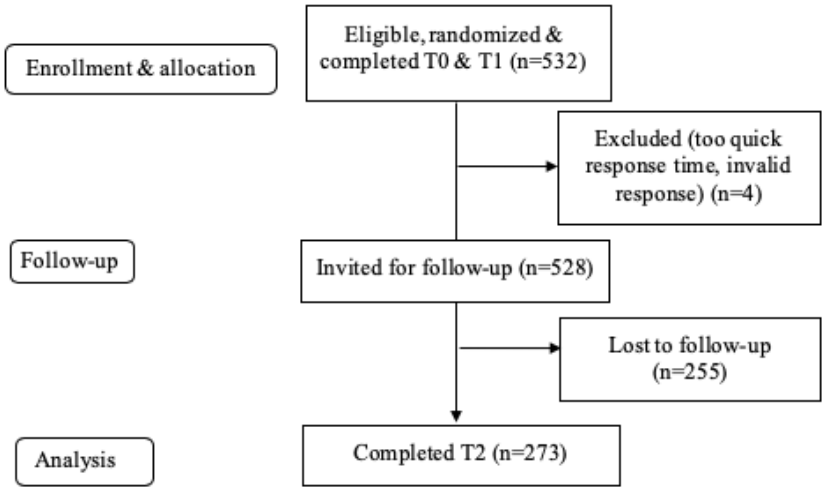
To test the hypotheses, we relied on data collected within a 6-month randomized controlled trial (RCT) using a 2 (frame-tailoring vs no frame-tailoring) by 2 (content-tailoring vs no content-tailoring) between-subjects design. We here present the data of T0 (baseline measurement), Tl (immediate postintervention follow-up), and T2 (1-month postintervention follow-up) measurements in the context of the Web-based CCT smoking cessation program Personal Advice in Stopping smoking (PAS). PAS was exclusively accessible via the project website [22] and was suitable for computers, laptops, as well as for mobile phones and tablets. Prior to study enrolment, smokers in the Dutch general public were targeted through social media (eg, Facebook, Twitter, LinkedIn), Google advertisements, and Dutch (online) newspapers and radio. Once smokers were willing to participate, they were provided with study information and could give their online informed consent, after which they could create their own username and password. Subsequently (T0), participants were automatically assigned to one of the four conditions through computer randomization and asked to complete the baseline questionnaire (T0), invited to use the intervention, and asked to complete the immediate postintervention evaluation (T1). One month later, they were prompted via email to fill out a brief follow-up questionnaire (T2). Participants received a 10 Euro voucher for their total 45-minute participation when completing the last and third follow-up questionnaire after 6 months.

The study was approved by the Institutional Review Board of the Amsterdam School for Communication Research, University of Amsterdam (2017-PC-7599), and is registered with the Dutch Trial Register (NL6512/NRT-6700).

Participants

At baseline, 534 participants were recruited from mid-December 2018 to March 2019, 273 (51.1%) of whom could be followed-up after 1 month, including 85 (31.1%) in the frame-tailored and content-tailored group, 58 (21.2%) in the frame-tailored and no-content-tailored group, 55 (20.1%) in the control group, and 75 (27.5%) in the no frame-tailored and content-tailored group. The participant flow throughout the study is shown in Figure 2. Inclusion criteria for participants were: 18 years or older, intending to quit smoking within the upcoming 6 months, providing online informed consent, and not having smoked during the last 7 days. An a priori power analysis using G*Power software [23] estimated that a sample size of a minimum of 198 participants should be sufficient to detect small effects and interaction effects (power=.80, odds ratio=1.68, R^2 content-tailoring=.03) based on an earlier study [24].

Figure 2. Flow chart of participants. T0, enrollment, allocation, and baseline measurements; T1, measurement of perceived relevance; T2, measurement of motivation, attitude toward smoking cessation, social influence, self-efficacy, smoking abstinence.



Experimental Conditions

Content-Tailoring

In the content-tailored condition, participants received smoking cessation advice adapted according to their answers in the questionnaire, which was grounded in the I-Change Model [25]. Questions concerned participants' smoking behavior, attitude, self-efficacy, social influence, action and coping planning, as well as their intention to quit smoking or to remain a nonsmoker [26,27].

Frame-Tailoring

Message frame-tailoring was based on an assessment of the participants' individual NFA. Participants with a high NFA received autonomy-supportive message frames that encouraged people to accept responsibility for their own behavior by taking the message recipient's perspective into account through reflective feedback, using language that minimized pressure on the reader, and providing choice (eg, by choosing whether or not to receive additional information on smoking cessation or by choosing whether or not to decide on a quit smoking date) [11,13,28,29]. Participants with a low NFA received controlling message frames that consisted of directive and forceful sentences with imperatives and commands. In addition, authoritative statements such as "experts say" were included and positive filling terms (eg, "luckily", "good") were avoided. In this case, the participants were not provided with choice, but rather received all smoking cessation information and a quit date within the next 2 weeks.

As in the frame-tailoring–only condition, in which we tailored both the content and message frames, the style or tone was adjusted based on the NFA throughout all intervention messages.

Control

In the control condition, participants received generic smoking cessation advice, which was neither tailored to their preassessed answers nor to their NFA. A smoking cessation advice example used for each of the conditions is shown in Multimedia Appendix 1.

Pilot Testing

Previous to this study, we conducted an extensive usability test of PAS among smoking cessation experts (N=5) and smokers from different sociodemographic backgrounds (N=7) (personal communication with van Strien-Knippenberg, Faculty of Social and Behavioral Sciences, University of Amsterdam). The questionnaire and stimulus materials were pilot-tested and used in previous online experiments (details can be obtained from the corresponding author MA on request and in our previous study [24]).

Measures

Overall Measures and Evaluation Timeline

At baseline (T0), we measured demographic variables along with the frame-tailoring and content-tailoring variables (ie, NFA and I-Change Model variables). Immediately postintervention (T1), the manipulation assessment and participants' perceived relevance were measured. At 1-month follow-up (T2), self-determined motivation, sociocognitive beliefs, and smoking abstinence were assessed. All items were measured on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree), unless indicated otherwise. Full descriptions of the scales, including item wording, are listed in Multimedia Appendix 2.

Demographics

Age, gender, living arrangement, educational level, presence of respiratory or cardiovascular diseases and (in the case of female gender) pregnancy, and smoking-related behaviors (eg, cigarettes smoked per day) were assessed.

Dependent Variable

We measured the 7-day point prevalence abstinence from smoking (smoking abstinence) by asking participants whether they had smoked in the last 7 days (yes=0, no=1).

Mediators

The perceived relevance of the smoking cessation message was assessed using 3 items described by Zhao and Peterson [9]. This scale was proven to be reliable, in which higher scores signified higher perceived relevance (Cronbach alpha=.87, mean 4.44,

SD 0.08). Self-determined motivation to quit smoking was measured using the 6-item Treatment Self-Regulation Questionnaire (TSRQ) [30], which also showed good reliability (Cronbach alpha=.92, mean 5.34, SD 0.21). Higher scores on the response scale denoted higher levels of self-determined motivation to quit.

Attitudes toward smoking cessation, social influence beliefs, self-efficacy, and intention to quit smoking were assessed based on the I-Change Model [25]. Twelve items were used to measure attitude toward smoking cessation, which were answered on a 5-point Likert scale (1=completely disagree, 5=completely agree). Higher scores indicated higher perceptions of the pros or cons of smoking cessation, respectively. Two subscales were formed with each of the 6 items assessing the perceived pros and cons of smoking cessation, respectively. Both subscales appeared to have good reliability (Cronbach $alpha_{pros}$=.79, mean 3.55, SD 0.62; Cronbach $alpha_{cons}$=.76, mean 2.36, SD 0.51).

Social influence was measured using the concepts of social support (3 items) and social norms (3 items). Answers were given within 6 response categories. The subscales for social support and social norms had poor reliability (Cronbach $alpha_{social\ support}$=.58; Cronbach $alpha_{social\ norms}$=.48) and therefore neither of these scales could be used.

Self-efficacy was measured by 9 items, which were answered on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). The scale was found to be reliable (Cronbach alpha=.91, mean 3.51, SD 0.29) and higher scores indicated higher perceived self-efficacy for smoking cessation.

Tailoring Variables

NFA was assessed with the Health Causality Orientations Scale (HCOS) [14,31]. In the HCOS, participants receive 4 scenarios for changing their health behavior with each of 3 different statements of how they would act in the scenario (eg, methods of quitting smoking). The participants then have to indicate how they would quit smoking by choosing one of the 3 statements. Each statement comprises a motivation orientation (ie, self-determined, controlled orientation toward friends and family, controlled orientation toward experts). Responses were given on a 5-point Likert scale (1=very unlikely, 5=very likely). Four items from the HCOS reflect people's autonomous orientation, which were used to determine the participant's NFA; higher mean scores indicated a higher NFA. For tailoring, the cutoff point to determine a high or low NFA was 3.8 on the HCOS, which was based on results from an earlier online experiment (more details can be obtained from the corresponding author MA on request).

Manipulation Check

To assess the validity of our frame-tailoring approach, we used 4 items that assessed the degree to which participants perceived the tone of the advice as controlling or autonomy-supportive (eg, "The advice was formulated in a pressuring tone"). The validity of our content-tailored manipulation was measured with 3 items asking whether participants felt that the smoking cessation advice was specifically written for them (eg, "In this program, I received advice based on the responses that I gave to the questions"). Responses were given on a 5-point Likert scale (1=strongly disagree, 5=strongly agree).

Statistical Analysis

Descriptive analyses with SPSS version 25 (SPSS Inc., Chicago, IL, USA) were conducted to determine sample characteristics and to check for differences in background variables and smoking-related behaviors (eg, number of cigarettes smoked on an average day) between conditions. We used two-sided *t* tests, Chi square tests, and analysis of variance (ANOVA) as appropriate. In addition, a nonresponse analysis with two-sided *t* tests and Chi square tests was conducted to determine whether selective dropout had occurred. We compared complete with lost-to-follow-up cases at T2 with regard to the same set of variables. Structural equation modeling (SEM) was conducted in R (R Foundation for Statistical Computing, Vienna, Austria) with the lavaan package version 0.6-3 [32]. Manifest variables were used for data analysis owing to the rather small sample size (N=273) for SEM analysis. Covariances were added among the two subscales measuring attitude toward smoking cessation, as these subscales measured different parts of the same concept. Next, we built a path model with smoking abstinence (measured at T2) as the main outcome. Based on our hypotheses, we added direct paths from the exogenous variables (ie, frame-tailoring, content-tailoring, and their combination) to smoking abstinence. We then added directs paths from the exogenous variables to perceived relevance and to self-determined motivation, along with a direct path from perceived relevance to self-determined motivation. Direct paths were added from self-determined motivation to attitudes and self-efficacy perceptions and to smoking abstinence. In addition, direct paths were added from attitudes and self-efficacy perceptions to smoking abstinence. The significance level was set at 5% and only the direct unstandardized effects are reported.

The data that support the findings of this study are available via Open Science Framework [33].

Results

Randomization and Manipulation Check

There were no significant differences between participants in the experimental conditions and control condition with regard to their demographics such as age and educational level, chronic diseases, and smoking behaviors. In terms of the manipulation, as expected, the frame-tailored and content-tailored conditions were significantly more often perceived as such by participants compared to the nonframe-tailored and noncontent-tailored conditions, respectively. Thus, the manipulation succeeded. An overview of all items assessing our manipulations, together with their mean values in each of the experimental conditions, is provided in Multimedia Appendix 3.

Sample Characteristics and Attrition

Comparisons of the 273 participants who completed the study and the 255 participants who were lost to follow-up after 1 month showed no significant differences in gender, educational level, smoking behaviors, and chronic diseases, but the Chi square test for condition and intervention drop-out was significant (Chi-$square_3$=11.15, N=528, *P*=.11): less participants

in the message frame-tailoring and content-tailoring condition were lost to follow-up (n=50, 9.5%) compared to participants who received message frame-tailoring without content-tailoring (n=65, 12.3%), generic advice (ie, the control condition; n=71, 13.4%), or no message frame-tailoring but content-tailoring only (n= 69, 13.1%). Participants who dropped out (mean 40.11 years, SD 14.28) were also significantly (F_1=5.89, P=.02) younger than those who completed the follow-up measurement (mean 43.05 years, SD 13.52).

Participant age was added as a covariate to our structural model because it was significantly correlated with smoking abstinence and with intervention drop-out. As our model with the covariate was very complex, for clarity purposes, we here only report the results of variables that were of substantial interest based on the theory. Table 1 provides an overview of the sample characteristics of the participants who completed the study and those lost to follow-up. The assumptions of multivariate normality and linearity were met and no multicollinearity existed. We conducted our SEM analysis with the diagonal weighted least squares (DWLS) estimator, which provides robust values from the full weight matrix to compute standard errors. No missing data among endogenous variables were observed.

Table 1. Comparison of participants who completed the study with those who dropped out.

Participant characteristics	T0 (N=528)	Completed T2 (N=273)	Dropout at T2 (N=255)
Demographics			
Female, n (%)	187 (35.4)	106 (38.8)	81 (31.8)
Age (years), mean (SD)	41.63 (13.95)	43.05 (13.526)	40.11 (14.28)[a]
Educational level, n (%)			
High	233 (44.1)	133 (48.7)	100 (39.2)
Middle	228 (43.2)	111 (40.7)	117 (45.9)
Low	67 (12.7)	29 (10.6)	38 (14.9)
Other/missing	0 (0)	0 (0)	0 (0)
Living arrangement, n (%)			
With partner	110 (20.8)	61 (22.3)	49 (19.2)
With partner and child(ren)	119 (22.5)	61 (22.3)	58 (22.7)
With child(ren)	55 (10.4)	25 (9.2)	30 (11.8)
Alone	208 (39.4)	110 (40.3)	98 (38.4)
Other/missing	36 (6.8)	16 (5.9)	20 (7.8)
Number of daily smoked, mean (SD)			
Cigarettes	11.13 (8.34)	10.6 (7.92)	11.63 (8.75)
Shags	3.69 (8.26)	3.8 (8.40)	3.75 (8.12)
Cigars	0.40 (2.68)	0.24 (1.85)	0.57 (3.34)
Cigarillos	0.18 (1.38)	0.19 (1.34)	0.18 (1.42)
Pipes	0.13 (1.17)	0.07 (0.78)	0.18 (1.49)
Earlier quit attempts, mean (SD)	5.74(0.58)	5.29 (11.41)	6.23 (15.27)
Existence of (chronic) disease, n (%)			
Heart disease	39 (7.4)	21 (7.7)	18 (7.1)
COPD[b]	106 (20.1)	58 (21.2)	48 (18.8)
Diabetes	25 (4.7)	13 (4.8)	12 (4.7)
Cancer	29 (5.5)	12 (4.4)	17 (6.7)

[a]Significant mean difference from T2: $F_{1, 526}$=5.893, P=.02.

[b]COPD: chronic obstructive pulmonary disorder.

We identified outliers among endogenous variables (ie, perceived relevance, self-determined motivation, perceived pros of smoking cessation, self-efficacy), which were checked and considered random, and therefore not removed.

Model Testing

Our hypothesized path model appeared to have a poor model fit according to conventional goodness-of-fit indices [34]. Based on the modification indices and the residual covariance matrix, we assumed it to be necessary to trim our path model by

discarding a variable from the model (ie, the cons of smoking cessation), which subsequently led to good model fit. Table 2 provides an overview of model fit indices for the hypothesized and fitted model.

The results from SEM analysis are depicted in the structural model in Figure 3. For clarity, we present the results only for the significant regression coefficients.

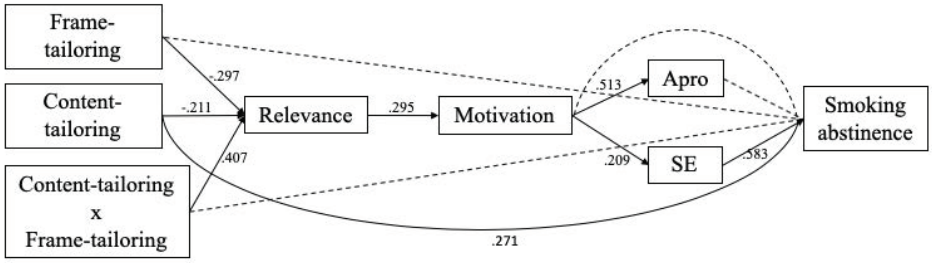
Table 2. Fit indices of the path model with smoking abstinence as outcome.

Fit indices	Hypothesized model	*P* value	Final (trimmed) model	*P* value
Chi-square	75.43_{18}	<.001	20.249_{13}	.09
Comparative fit	.814	N/A[a]	.967	N/A
RMSEA[b] (90% CI)	0.10 (0.08, 0.13)	N/A	0.04 (0.00- 0.08)	N/A

[a]Not applicable.

[b]RMSEA: root mean square error of approximation.

Figure 3. Final model with significant paths only. Results are presented as standardized direct effects. Dotted lines represent nonsignificant paths. Straight lines represent significant paths (*P*<.05). Apro, pros of smoking cessation; SE, self-efficacy.



Hypothesis Testing

Effects of Message Frame-Tailoring and Content-Tailoring on Smoking Abstinence

In contrast to our expectations, neither message frame-tailoring based on a smoker's NFA nor the combination of message frame-tailoring and content-tailoring significantly affected smoking abstinence. However, as expected, we identified a significant positive effect from content-tailoring on smoking abstinence (beta=.57, *P*=.02). In the frame-tailored and content-tailored condition, 23 (30.3%) smokers refrained from smoking, while 12 (15.8%) smokers in the frame-tailored–only, 11 (14.5%) smokers in the control condition, and 30 (39.5%) smokers in the content-tailored–only condition refrained from smoking. Thus, we could only partly confirm the first hypothesis (H1b).

Mediation of Perceived Relevance, Self-Determined Motivation, and Sociocognitive Beliefs About Smoking Cessation

As shown in the model (Figure 3), we identified a significant main effect of content-tailoring and message frame-tailoring based on the users' NFA as well as their combination on smokers' perceived relevance of the smoking cessation message, the first mediator in our model. ANOVA showed significant differences between the conditions ($F_{3,269}$=4.82, *P*=.003) and Tukey's post hoc test was used to identify the conditions with significant differences. The control condition (ie, no content- or message frame-tailoring) was perceived as significantly more relevant than message frame-tailoring or content-tailoring alone (mean difference 0.65, SE 0.20 and mean difference 0.54, SE 0.19, respectively) and was similarly relevant as the condition with both content- and message frame-tailoring. Thus, surprisingly, generic smoking cessation advice led to similarly high levels of perceived message relevance as a message that was tailored both in terms of content and message framing, and led to higher perceived relevance than messages tailored in terms of only one of these aspects.

As these findings were against the expected direction, we decided to inspect the data even more closely by comparing participants with a high and low NFA within the message frame-tailored conditions. This comparison showed that smokers with a high NFA generally perceived their message as more relevant compared to participants who had a low NFA when they received a frame-tailored smoking cessation message, both with and without content-tailoring. Moreover, the combination of message frame-tailoring and content-tailoring led to significantly higher perceived relevance than content-tailored messages only, but only for smokers with a high NFA (mean difference 0.59, SE 0.19; *P*=.04). In addition, those with a high NFA in the frame-tailored and content-tailored condition perceived the messages as significantly more relevant compared to those with a low NFA who received frame-tailored but not content-tailored messages. To illustrate these findings, the means per condition for all continuous variables in the SEM are provided in Table 3.

Table 3. Mean (SD) values per condition for all endogenous variables (N=273).

Dependent variable	Frame-tailoring and content-tailoring			Frame-tailoring and no content-tailoring			No frame-tailoring and no content-tailoring	No frame-tailoring and content-tailoring	Overall mean	F_{df}; P value[a]
	All participants	High NFA[b]	Low NFA	All participants	High NFA	Low NFA				
Relevance	4.59 (1.05)	4.83 (1.03)	4.26 (1.01)	4.13 (0.96)	4.29 (1.07)	3.97 (.83)	4.78 (1.27)	4.24 (1.05)	4.43 (1.10)	4.36_5; .008[c], .03[d], .01[e]
Motivation	5.33 (1.32)	5.57 (1.42)	5.00 (1.10)	5.16 (1.11)	5.36 (1.00)	4.98 (1.19)	5.42 (1.14)	5.42 (1.26)	5.34 (1.22)	1.56_5; .17
SE[f]	3.52 (0.84)	3.70 (0.82)	3.27 (0.82)	3.47 (0.69)	3.50 (0.56)	3.44 (0.81)	3.48 (0.83)	3.54 (0.96)	3.51 (0.84)	1.12_5; .35
Apro[g]	3.64 (76)	3.87 (0.74)	3.33 (0.69)	3.43 (0.96)	3.55 (0.84)	3.32 (1.06)	3.55 (0.91)	3.52 (0.86)	3.55 (0.86)	2.26_5; .05
Acon[h]	2.36 (0.83)	2.21 (0.84)	2.56 (0.80)	2.22 (0.77)	2.13 (0.79)	2.31 (0.76)	2.43 (0.90)	2.39 (0.94)	2.35 (0.86)	1.16_5; .33

[a]Analysis of variance based on the six groups of subtailoring.

[b]NFA: need for autonomy.

[c]High NFA vs Low NFA frame-tailoring and content-tailoring.

[d]High NFA frame-tailoring and content-tailoring vs no frame-tailoring and content-tailoring.

[e]Low NFA and content-tailoring vs no frame-tailoring and no-content tailoring.

[f]SE: self-efficacy.

[g]Apro: attitudes about pros of smoking cessation.

[h]Acon: attitudes about cons of smoking cessation.

Subsequently, perceived relevance had a positive effect on the self-determined motivation to quit smoking (beta=.32, P<.001). Furthermore, although self-determined motivation did not have a direct effect on smoking abstinence, there was a positive effect of self-determined motivation on the perceived pros of smoking cessation (ie, positive attitudes) (beta=.37, P<.001) and a positive effect on self-efficacy perceptions (beta=.15, P<.001). Moreover, we confirmed a positive effect from self-efficacy perceptions on smoking abstinence (beta=.72, P<.001).

In summary, the effect of message frame-tailoring, content-tailoring, as well as their combined effect on smoking abstinence was mediated by perceived relevance, self-determined motivation to quit, and self-efficacy on smoking abstinence. As such, we could confirm the second hypothesis. An overview of the hypothesized direct and indirect effects from our final model is provided in Table 4, and the correlation matrix of standardized effects is provided in Multimedia Appendix 4.

Table 4. Standardized indirect and direct effects of the trimmed model[a].

Independent variable	Perceived relevance	Self-determined motivation	Attitudes	Self-efficacy	Smoking abstinence
Content-tailoring					
Indirect effect	_[b]	–0.062	0.089	–0.001	0.230
Direct effect	–0.474	–	–	–	0.271
Frame-tailoring					
Indirect effect	–	-0.088	0.064	–0.026	–0.043
Direct effect	–0.297	–	–	–	0.023
Perceived relevance					
Indirect effect	–	–	0.151	0.062	0.021
Direct effect	–	0.295	–	–	–
Self-determined motivation					
Indirect effect	–	–	–	–	0.052
Direct effect	–	–	0.513	0.209	0.072
Attitudes					
Indirect effect	–	–	–	–	–
Direct effect	–	–	–	–	–0.173
Self-efficacy					
Indirect effect	–	–	–	–	–
Direct effect	–	–	–	–	0.583
Smoking abstinence					

[a]The model controlled for age, and only the paths to self-determined motivation, social norms, and smoking abstinence were significant.

[b]Not applicable.

Discussion

Effect of Tailoring on Smoking Abstinence and Perceived Relevance

The aim of this study was to test the effectiveness of message frame-tailoring based on smokers' NFA in isolation and in combination with content-tailoring in the context of an online CCT smoking cessation intervention. Our results confirm findings from earlier research on content-tailoring [2,24], as we could identify a positive effect of content-tailoring on 7-day point prevalence abstinence rates 1 month after the start of the intervention. However, in contrast to our expectations, message that were frame-tailored based on the NFA did not lead to higher smoking abstinence rates in isolation and in combination with content-tailoring as compared to the no frame-tailoring condition.

Mediating Roles of Perceived Relevance, Self-Determined Motivation, and Sociocognitive Beliefs

Overall, our findings were in line with our hypothesis that perceived relevance, self-determined motivation to quit, and sociocognitive beliefs mediate the effects of content-tailoring, message frame-tailoring, and their combination on smoking abstinence. That is, we demonstrated a positive effect of perceived message relevance on the self-determined motivation to quit smoking. Therefore, we could confirm earlier findings concerning elaboration likelihood model research [35] demonstrating that people who perceive their messages as relevant are also more motivated to devote more cognitive effort on processing the messages. Moreover, self-determined motivation positively predicted positive attitudes and self-efficacy beliefs of smoking cessation. Finally, we found a positive effect of self-efficacy on smoking abstinence, which was also observed in a meta-analysis on the integration of self-determination theory and the theory of planned behavior [19].

However, in contrast to our expectations, attitudes toward smoking cessation did not significantly predict smoking abstinence, which is not supported by earlier research [19]. According to the theory of planned behavior [36], positive attitudes lead to an enhanced intention for behavior change, which in turn predicts behavior change. However, previous smoking cessation research showed that self-efficacy perceptions were the main predictors of smoking cessation among smokers intending to quit [37]. A potential explanation for the lack of a significant effect of attitude on smoking abstinence could be that we only used one subscale of attitudes (ie, the pros of smoking cessation). The subscale related to the cons of smoking cessation had to be discarded during SEM analysis owing to the noise it caused in the data, which prevented reaching model fit. This may have led to the attitude variable, as included in our model, not being fully representative of the theoretical construct in its entirety. In addition, we did not assess intention to quit smoking in our study, and therefore cannot state whether

attitudes might have indirectly—instead of directly—predicted smoking abstinence, which could have been assumed based on theory and evidence [37,38].

Moreover, we had to discard the social influence scale [39,40] as it had poor reliability (see Multimedia Appendix 2). Thus, we could not test for a possible mediation effect of social influence, which might be an important factor in explaining variance in smoking abstinence rates [41]. Based on comments made by smokers and experts in the pilot testing phase of our intervention, we recommend that efforts be made in future research to improve the comprehensibility, and subsequently the reliability, of the social influence scale by adapting the response categories of the subscales (eg, specification of terms such as "majority of your children" that seemed to be difficult to answer when having two children) and to include the resulting reliable scale in further analyses similar to those presented here.

Exploring Message Frame-Tailoring on the Need for Autonomy

We found that content-tailoring, message frame-tailoring, and their combination had a significant effect on participants' perceived message relevance. However, these effects were against expectation as both message frame-tailoring and content-tailoring led to significantly lower perceived message relevance compared to the control condition (ie, generic smoking cessation messages). This finding also conflicts with earlier tailoring-based research [4,8,11], which demonstrated that content-tailored messages led to better message processing, better message recall, and more positive behavioral outcomes via more perceived message relevance. To gain better understanding of this finding, we checked the time the participants took to finish the intervention. Although the control condition messages were similar in length and contained generic smoking cessation information, we wanted to explore whether smokers in the control condition had processed the messages longer and perhaps more thoroughly, resulting in their perceptions of relevance being higher than in the other conditions. However, as participants between conditions did not differ significantly in the time they used to finish the intervention (data not shown), this possibility is unlikely.

An alternative potential explanation comes from exploratory data analyses conducted with the participants' showing a high and low NFA separated within the message frame-tailored conditions. These analyses showed that participants with a high NFA who received message frame-tailoring (ie, messages in an autonomy-supportive frame) as an addition to content-tailoring did perceive their messages as significantly more relevant compared to participants with a low NFA who received message frame-tailoring and to participants that received messages that were content-tailored only. This finding supports the results of Resnicow et al [11] who showed that participants with a preference for an autonomous form of communication perceived autonomy-supportive messages as more relevant compared to those with a preference for controlling forms of communication. In addition, although not significant, when comparing the means among all other mediators, participants with a high NFA generally had a higher level of self-determined motivation, more positive attitudes about smoking cessation, and higher self-efficacy perceptions compared to those with a lower NFA, regardless of whether or not they had received content-tailored messages.

This pattern raises the question as to whether smokers with a lower NFA might prefer different message frames than those provided in our study (ie, message frames using controlling language and without the provision of choice). It could be that the controlling language that was used might have been too controlling, resulting in message resistance and an insufficient ability to motivate participants with a low NFA to refrain from smoking [42]. Negative message evaluations such as resistance have been shown to lead to less deep message processing [8], resulting in lower message effectiveness. Thus, participants with a low NFA might need differently tailored message frames (eg, clearcut expert advice about smoking cessation using less controlling language such as without imperatives or terms like "must" and "should"). Furthermore, future research would benefit from validating the HCOS scale, which was used to assess the NFA in this study, and further investigate whether the dichotomization of people into high and low NFA based on the 4 items that assess autonomous orientation proofs is a valid categorization approach. Perhaps there is another group (eg, with a moderate NFA) of people or it is better to take into account the HCOS items that assess people's controlled or impersonal orientations [14]. Such research efforts would help to optimize message frame-tailoring based on the NFA and enable further research into its effectiveness.

Another possibility is that participants with a lower NFA might be less susceptible to autonomy-supportive messages framed in a controlling message style than those with a high NFA owing to, for instance, different message processing needs that potentially correlate with the NFA such as the need for cognition. For example, message frames for low NFA participants might not have sufficiently met their relatively lower autonomy needs or their preferences for (low levels of) information processing. In a similar vein, it could be expected that those with a higher NFA might have better health outcomes than their low NFA counterparts, as individuals with a high NFA might also report higher needs for cognition and thus be more motivated to process health messages and eventually change their health behavior than low NFA individuals [8]. In support of this reasoning, in an exploratory data analysis (data not shown), we found that participants with a high NFA also had higher, but nonsignificant, rates of smoking abstinence than those with a low NFA. Accordingly, it seems to be necessary to study whether the NFA indeed correlates with the need for cognition and how their potential interaction might influence message effectiveness. Moreover, we recommend future research to (qualitatively) investigate how messages should be formulated among those with a lower NFA (and potentially a low need for cognition) to meet their communication preferences and message processing needs so that these messages can lead to optimal health outcomes.

Study Strengths and Limitations

Findings from this study contribute to a growing understanding of the effects of message frame-tailoring; however, approximately 50% of the participants could not be followed

up 1 month after the baseline measurement. Although this is common for RCTs, and in particular for internet-based longitudinal studies [43,44], the high rates of attrition (while also differing between conditions) may have reduced the internal validity of the results presented, consequently introducing potential bias to the estimates of effectiveness. This could result in an overestimation of the effectiveness for several reasons. First, participants who were lost to follow-up were significantly younger than those that completed the intervention. However, this age difference was only 3 years (ie, drop-out age 40 years vs completer age 43 years) and also within a rather homogenous age group. Therefore, realistically, we do not assume major differences among completers and those lost to follow-up. Second, a significantly lower rate of study dropout was observed in the message frame-tailoring and content-tailoring condition, which was expected to impact smoking cessation and underlying mechanisms (eg, attitude change and self-efficacy perceptions) most positively (as compared to other conditions) (see Sample Characteristics and Attrition in the Results section). However, overall, the message frame-tailoring and content-tailoring condition had no significant effect on smoking cessation and underlying mechanisms. Thus, we assume the risk of bias due to study attrition to be small.

To prevent high attrition rates in the first place, we used strategies such as sending several email reminders to participants, and only including participants who were sufficiently motivated to participate, as well as offering shopping vouchers after completion of the intervention and follow-up questionnaires. Moreover, we used a forced data entry option, so that we only had missing values on outcome variables when the entire case was missing and there were no missing values, such as for any of the mediating variables while data on smoking abstinence was present or vice versa. Although we acknowledge that the complete case analysis we conducted might have led to a bias in the results presented, imputing the missing values for the nearly 50% of cases that dropped out of the study would have introduced a high degree of uncertainty that would further reduce the reliability of the presented results. However, a major strength of this study was that we were able to recruit a large sample of eligible smokers, resulting in a sample size that was still sufficient for analysis according to our a priori power analysis.

Conclusion

This study extends the tailoring literature by providing first evidence for the effects of message frame-tailoring based on the NFA in isolation and in combination with content-tailoring. Based on our findings, we can conclude that message frame-tailoring based on the NFA seems to be an effective addition to conventional content-tailoring techniques in online health interventions for people with a high NFA, but is not effective in its current form for people with a low NFA. To enhance the effectiveness of message frame-tailoring, future research efforts might therefore want to focus on (qualitatively) investigating which type of message frame might be most beneficial for smokers with a low NFA.

Acknowledgments

We would like to thank Inge Van Strien-Knippenberg for her great support in the preparations of the data collection for this RCT. This work was supported by the Dutch Cancer Society (KWF Kankerbestrijding) under grant number KWF 2015-7913 and by the Innovational Research Incentives Scheme Veni from Netherlands Organization for Scientific Research-Division for Social Sciences (NWO-MaGW) accredited to ES (project number 451-15-028).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example advice.
[DOCX File , 26 KB - jmir_v22i4e17251_app1.docx]

Multimedia Appendix 2
Item wording and reliability.
[DOCX File , 21 KB - jmir_v22i4e17251_app2.docx]

Multimedia Appendix 3
Manipulation check.
[DOCX File , 19 KB - jmir_v22i4e17251_app3.docx]

Multimedia Appendix 4
Correlation matrix.
[DOCX File , 22 KB - jmir_v22i4e17251_app4.docx]

Multimedia Appendix 5

CONSORT-eHealth checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1308 KB - jmir_v22i4e17251_app5.pdf]

References

1. World Health Organization (WHO). Europe Office: World Health Organization; 2014. European Tobacco Control Status Report 2014 URL: http://www.euro.who.int/__data/assets/pdf_file/0009/248418/European-Tobacco-Control-Status-Report-2014-Eng.pdf [accessed 2019-10-31]
2. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
3. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [doi: 10.1016/s0738-3991(98)00127-x] [Medline: 10223015]
4. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
5. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Nov 01;27(Suppl 3):S227-S232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]
6. Noar SM, Harrington NG, Aldrich RS. The Role of Message Tailoring in the Development of Persuasive Health Communication Messages. Annal Int Commun Assoc 2016 May 18;33(1):73-133. [doi: 10.1080/23808985.2009.11679085]
7. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
8. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. In: Communication And Persuasion: Central And Peripheral Routes To Attitude Change (Springer Series In Social Psychology). New York: Springer; 2020:1-24.
9. Zhao X, Peterson E. Effects of Temporal Framing on Response to Antismoking Messages: The Mediating Role of Perceived Relevance. J Health Commun 2017 Jan;22(1):37-44. [doi: 10.1080/10810730.2016.1250844] [Medline: 28001492]
10. Rimer BK, Kreuter MW. Advancing tailored health communication: A persuasion and message effects perspective. J Commun 2006;56:S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]
11. Resnicow K, Davis RE, Zhang G, Konkel J, Strecher VJ, Shaikh AR, et al. Tailoring a fruit and vegetable intervention on novel motivational constructs: results of a randomized study. Ann Behav Med 2008 Apr 10;35(2):159-169. [doi: 10.1007/s12160-008-9028-9] [Medline: 18401673]
12. Smit E, Linn A, vanWeert J. Taking online computer-tailoring forward. Eur Health Psychol 2015;17(1):25-31 [FREE Full text]
13. Resnicow K, Zhou Y, Hawley S, Jimbo M, Ruffin MT, Davis RE, et al. Communication preference moderates the effect of a tailored intervention to increase colorectal cancer screening among African Americans. Patient Educ Couns 2014 Dec;97(3):370-375 [FREE Full text] [doi: 10.1016/j.pec.2014.08.013] [Medline: 25224317]
14. Smit ES, Bol N. From self-reliers to expert-dependents: identifying classes based on health-related need for autonomy and need for external control among mobile users. Media Psychol 2019 Apr 30:1-24. [doi: 10.1080/15213269.2019.1604235]
15. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
16. Ng JYY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-Determination Theory Applied to Health Contexts: A Meta-Analysis. Perspect Psychol Sci 2012 Jul 29;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
17. Williams G, McGregor H, Sharp D, Levesque C, Kouides R, Ryan R, et al. Testing a self-determination theory intervention for motivating tobacco cessation: supporting autonomy and competence in a clinical trial. Health Psychol 2006 Jan;25(1):91-101. [doi: 10.1037/0278-6133.25.1.91] [Medline: 16448302]
18. Williams G, Gagné M, Ryan R, Deci E. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002;21(1):40-50. [doi: 10.1037/0278-6133.21.1.40]
19. Hagger M, Chatzisarantis N. Integrating the theory of planned behaviour and self-determination theory in health behaviour: a meta-analysis. Br J Health Psychol 2009 May;14(Pt 2):275-302. [doi: 10.1348/135910708X373959] [Medline: 18926008]
20. Gagné M, Deci E. Self-determination theory and work motivation. J Organiz Behav 2005 Apr 14;26(4):331-362 [FREE Full text] [doi: 10.1002/job.322]
21. Taylor G, Dalili M, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
22. PAS. URL: https://persoonlijkstopadvies.nl/ [accessed 2020-02-21]
23. 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. [doi: 10.3758/bf03193146] [Medline: 17695343]
24. Smit ES, de Vries H, Hoving C. Effectiveness of a Web-based multiple tailored smoking cessation program: a randomized controlled trial among Dutch adult smokers. J Med Internet Res 2012 Jun 11;14(3):e82 [FREE Full text] [doi: 10.2196/jmir.1812] [Medline: 22687887]

25. de Vries H. An Integrated Approach for Understanding Health Behavior; The I-Change Model as an Example. Psychol Behav Sci Int J 2017 Mar 9;2(2):1-6. [doi: 10.19080/pbsij.2017.02.555585]
26. Smit E, Candel M, Hoving C, de Vries H. Results of the PAS Study: A Randomized Controlled Trial Evaluating the Effectiveness of a Web-Based Multiple Tailored Smoking Cessation Program Combined With Tailored Counseling by Practice Nurses. Health Commun 2016 Sep;31(9):1165-1173 [FREE Full text] [doi: 10.1080/10410236.2015.1049727] [Medline: 26934538]
27. Smit E, de Vries H, Hoving C. The PAS study: a randomized controlled trial evaluating the effectiveness of a web-based multiple tailored smoking cessation programme and tailored counselling by practice nurses. Contemp Clin Trials 2010 May;31(3):251-258. [doi: 10.1016/j.cct.2010.03.001] [Medline: 20215047]
28. Deci EL, Eghrari H, Patrick BC, Leone DR. Facilitating internalization: the self-determination theory perspective. J Pers 1994 Mar;62(1):119-142. [doi: 10.1111/j.1467-6494.1994.tb00797.x] [Medline: 8169757]
29. Markland D, Ryan R, Tobin V, Rollnick S. Motivational Interviewing and Self–Determination Theory. J Soc Clin Psychol 2005 Sep;24(6):811-831 [FREE Full text] [doi: 10.1521/jscp.2005.24.6.811]
30. Levesque CS, Williams GC, Elliot D, Pickering MA, Bodenhamer B, Finley PJ. Validating the theoretical structure of the Treatment Self-Regulation Questionnaire (TSRQ) across three different health behaviors. Health Educ Res 2007 Oct 23;22(5):691-702. [doi: 10.1093/her/cyl148] [Medline: 17138613]
31. Smit E, Zeidler C, Resnicow K, de Vries H. Identifying the Most Autonomy-Supportive Message Frame in Digital Health Communication: A 2x2 Between-Subjects Experiment. J Med Internet Res 2019 Oct 30;21(10):e14074 [FREE Full text] [doi: 10.2196/14074] [Medline: 31670693]
32. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. J Stat Soft 2012;48(2):1-5 [FREE Full text] [doi: 10.18637/jss.v048.i02]
33. Altendorf M. Open Science Framework. Effectiveness of message frame-tailoring URL: https://osf.io/k5n4t/ [accessed 2020-02-21]
34. Bryne B. Structural Equation Modeling With AMOS. New York: The Guilford Press; 2016:1-460.
35. Petty RE, Heesacker M, Hughes JN. The elaboration likelihood model: Implications for the practice of school psychology. J School Psychol 1997 Jun;35(2):107-136. [doi: 10.1016/s0022-4405(97)00003-4]
36. Ajzen I. The theory of planned behavior. Organiz Behav Hum Decision Proc 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
37. Smit ES, Hoving C, Schelleman-Offermans K, West R, de Vries H. Predictors of successful and unsuccessful quit attempts among smokers motivated to quit. Addict Behav 2014 Sep;39(9):1318-1324. [doi: 10.1016/j.addbeh.2014.04.017] [Medline: 24837754]
38. Vangeli E, Stapleton J, Smit E, Borland R, West R. Predictors of attempts to stop smoking and their success in adult general population samples: a systematic review. Addiction 2011 Dec;106(12):2110-2121. [doi: 10.1111/j.1360-0443.2011.03565.x] [Medline: 21752135]
39. de Vries H, Mudde AN, Dijkstra A, Willemsen MC. Differential beliefs, perceived social influences, and self-efficacy expectations among smokers in various motivational phases. Prev Med 1998 Sep;27(5 Pt 1):681-689. [doi: 10.1006/pmed.1998.0344] [Medline: 9808799]
40. de Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, et al. The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. Health Educ Res 2003 Oct 01;18(5):611-626. [doi: 10.1093/her/cyg031] [Medline: 14572020]
41. van den Putte B, Yzer MC, Brunsting S. Social influences on smoking cessation: a comparison of the effect of six social influence variables. Prev Med 2005 Jul;41(1):186-193. [doi: 10.1016/j.ypmed.2004.09.040] [Medline: 15917010]
42. Rains S. The Nature of Psychological Reactance Revisited: A Meta-Analytic Review. Hum Commun Res 2013;39(1):47-73. [doi: 10.1111/j.1468-2958.2012.01443.x]
43. DeLeeuw E. Dropout in longitudinal studies: Strategies to limit the problem. Encycl Stat Behav Sci 2005:515-518.
44. Van Horn PS, Green KE, Martinussen M. Survey Response Rates and Survey Administration in Counseling and Clinical Psychology. Educ Psychol Meas 2008 Oct 15;69(3):389-403. [doi: 10.1177/0013164408324462]

Abbreviations

ANOVA: analysis of variance
CCT: computer content tailoring
DWLS: diagonal weighted least squares
HCOS: Health Causality Orientations Scale
NFA: need for autonomy
PAS: Personal Advice in Stopping smoking
RCT: randomized controlled trial
RMSEA: root mean square error of approximation
SEM: structural equation modeling

TSRQ: Treatment Self-Regulation Questionnaire

Edited by G Eysenbach; submitted 28.11.19; peer-reviewed by K Resnicow, B Nievas Soriano; comments to author 04.01.20; revised version received 27.01.20; accepted 27.01.20; published 03.04.20.

Please cite as:
Altendorf M, Hoving C, Van Weert JCM, Smit ES
Effectiveness of Message Frame-Tailoring in a Web-Based Smoking Cessation Program: Randomized Controlled Trial
J Med Internet Res 2020;22(4):e17251
URL: https://www.jmir.org/2020/4/e17251
doi:10.2196/17251
PMID:32242826

Original Paper

Fibromyalgia Impact Reduction Using Online Personal Health Informatics: Longitudinal Observational Study

William Collinge[1*], MSW, MPH, PhD; Robert Soltysik[2*], MS; Paul Yarnold[3*], PhD

[1]Collinge and Associates, Inc, Eugene, OR, United States

[2]Optimal Data Analysis LLC, Glen Burnie, MD, United States

[3]Optimal Data Analysis LLC, Chicago, IL, United States

[*]all authors contributed equally

Corresponding Author:
William Collinge, MSW, MPH, PhD
Collinge and Associates, Inc
3480 Kincaid Street
Eugene, OR, 97405-0000
United States
Phone: 1 5416323502
Fax: 1 2075108060
Email: william@collinge.org

Abstract

Background: Personal health informatics have the potential to help patients discover personalized health management strategies that influence outcomes. Fibromyalgia (FM) is a complex chronic illness requiring individualized strategies that may be informed by analysis of personal health informatics data. An online health diary program with dynamic feedback was developed to assist patients with FM in identifying symptom management strategies that predict their personal outcomes, and found reduced symptom levels associated with program use.

Objective: The aim of this study was to determine longitudinal associations between program use and functional impact of FM as measured by scores on a standardized assessment instrument, the Fibromyalgia Impact Questionnaire (FIQ).

Methods: Participants were self-identified as diagnosed with FM and recruited via online FM advocacy websites. Participants used an online health diary program ("SMARTLog") to report symptom ratings, behaviors, and management strategies used. Based on single-subject analysis of the accumulated data over time, individualized recommendations ("SMARTProfile") were then provided by the automated feedback program. Indices of program use comprised of cumulative numbers of SMARTLogs completed and SMARTProfiles received. Participants included in this analysis met a priori criteria of sufficient program use to generate SMARTProfiles (ie, ≥22 SMARTLogs completed). Users completed the FIQ at baseline and again each subsequent month of program use as follow-up data for analysis. Kendall tau-b, a nonparametric statistic that measures both the strength and direction of an ordinal association between two repeated measured variables, was computed between all included FIQ scores and both indices of program use for each subject at the time of each completed FIQ.

Results: A total of 76 users met the a priori use criteria. The mean baseline FIQ score was 61.6 (SD 14.7). There were 342 FIQ scores generated for longitudinal analysis via Kendall tau-b. Statistically significant inverse associations were found over time between FIQ scores and (1) the cumulative number of SMARTLogs completed (tau-b=–0.135, P<.001); and (2) the cumulative number of SMARTProfiles received (tau-b=–0.133, P<.001). Users who completed 61 or more SMARTLogs had mean follow-up scores of 49.9 (n=25, 33% of the sample), an 18.9% drop in FM impact. Users who generated 11 or more new SMARTProfiles had mean follow-up scores of 51.8 (n=23, 30% of the sample), a 15.9% drop.

Conclusions: Significant inverse associations were found between FIQ scores and both indices of program use, with FIQ scores declining as use increased. Based on established criteria for rating FM severity, the top one-third of users in terms of use had clinically significant reductions from "severe" to "moderate" FM impact. These findings underscore the value of self-management interventions with low burden, high usability, and high perceived relevance to the user.

Trial Registration: ClinicalTrials.gov NCT02515552; https://clinicaltrials.gov/ct2/show/NCT02515552

(J Med Internet Res 2020;22(4):e15819) doi:10.2196/15819

KEYWORDS

fibromyalgia; fibromyalgia impact; health informatics; predictive analytics; symptom reduction; functional status; personalized medicine; health diary

Introduction

Health informatics is a promising field that includes the use of internet technology to create new models of patient care and disease management by analyses of large bodies of health-related data. Health informatics are traditionally associated with large population databases (eg, the insurance industry or "big data"); however, patient-level interventions that employ health informatics also have the potential to empower individuals to take greater control over their symptoms and outcomes [1]. Combining personal health informatics with emerging techniques of predictive analytics represents an intriguing and potentially powerful approach to personalized medicine for some of the most confounding health conditions.

Such innovative strategies may be particularly helpful in complex, multisystemic chronic illnesses that have inconsistent or unreliable treatment responses across patients, such as in conditions where no effective treatments are available or where complementary therapies, drugs, dosages, or self-care strategies have widely varying effects across different individuals with the same diagnosis. Fibromyalgia (FM) is one such condition in which optimal symptom management calls for highly individualized integration of pharmacological and nonpharmacological approaches [2].

We have developed an electronic health (eHealth) intervention consisting of a daily health diary designed to enhance an individual's awareness of health-related actions and their impacts in daily living. In addition to the heightened self-awareness that may be anticipated from regular use of any such diary, the program incorporates a predictive analytics algorithm that processes the user's accumulated data to produce personalized intervention guidance driven by those data. We previously reported data on effects of the program (now revised and called AwareHealth) on *symptom ratings* in users with FM. Use was found associated with significant reductions in pain, stiffness, fatigue, concentration problems, memory problems, feeling anxious, feeling depressed, gastrointestinal problems, and sleep difficulties [3].

This paper reports the effects of program use on the *functional impact* of FM as defined by the Fibromyalgia Impact Questionnaire (FIQ), a standardized instrument used widely in FM clinical outcome research [4]. The FIQ was developed to measure the status, progress, and outcomes of patients with FM by measuring the components of health status believed to be most affected by FM using a 1-week recall period. The FIQ contains 11 questions related to physical functioning, each rated on a 4-point Likert type scale; two items asking the patient to mark the number of days they felt well and the number of days they were unable to work (including housework) because of FM symptoms; and seven horizontal linear scales, marked in 10 increments, on which the patient rates work difficulty, pain, fatigue, morning tiredness, stiffness, anxiety, and depression. Scoring provides a global measure of FM impact from 0 (lowest impact) to 100 (highest impact).

Methods

Subjects and Recruitment

In an online study sponsored by the US National Institute of Arthritis, Musculoskeletal and Skin Diseases [5], participants were recruited via public announcements on several FM advocacy websites. Enrollment and consent were processed online with electronic attestation of being 18 years or older and having been diagnosed with FM by a health professional. No financial or other incentive was offered. Human participant oversight was provided by the New England Institutional Review Board.

Intervention

The intervention program consisted of a brief health diary ("SMARTLog") completed by the user in about 5 minutes, three or more times per week over several weeks [6]. For a 24-hour recall period, users entered symptom ratings. They then used drop-down menus and radio buttons to report on four domains of behavioral and lifestyle variables informed by research, theory, and clinical observation as potentially influencing symptoms in FM, and a fifth domain for tracking user-defined inputs such as pharmacological or nonpharmacological therapies, or other variables of interest to the individual such as foods or activities. Textbox 1 shows the SMARTLog contents of the original version, which was used for the data reported in this paper. The revised version, which is in current use, provides an expanded array of 75 user-selectable symptoms and wellness outcomes, and adds a new domain, "Technology use", to report duration of screen time (eg, TV, movie, computer, mobile device) and wireless device exposure (see Multimedia Appendix 1).

Textbox 1. SMARTLog Contents (Original Version).

Outcomes

Pain, stiffness, fatigue, concentration problems, memory problems, feeling anxious, feeling depressed, gastrointestinal problems, and sleep difficulties, all rated on a 0-10 scale for "how bothersome" for the 24-hour recall period

Sleep and rest

Time to bed the previous night, wake-up time, and times and durations of daytime naps

Meals and snacks

Time and size of meals and snacks

Self-care practices

Time and duration of bathing; mind, body, or spirit practices; and exercise including the type, duration, and exertion level

Daily activities

Duration, exertion level, satisfaction level, and stressfulness of the following types of activities: vocational, domestic activity, social activity, recreation, travel or commuting time, and the day's overall activity level

Personal inputs (optional)

Users may self-define up to ten additional inputs to track (eg, specific medications, complementary therapies, dosages, foods, nutritional supplements, or activities not listed earlier), with user-selected metrics for each, such as quantities or durations.

The data from SMARTLog accumulates in the user's personal database and are processed daily in the background by a proprietary automated algorithm using single-subject (N-of-1) analytic methods to find statistically significant (alpha=.05 level) associations between symptom levels and specific health-related actions. This algorithm applies optimal discriminant analysis, a nonparametric method for maximizing classification accuracy in models with two classes and an ordered attribute (independent variable) [7]. The likelihood of finding significant associations increases as the volume of data accumulates. Users are encouraged to experiment by employing various strategies (eg, different bedtimes, dosages of a drug, complementary therapies, durations of exercise), as variability of data helps the program discover models that predict symptom reduction. When associations are found, the algorithm uses a natural language generator to report them to the user as "SMARTProfile" statements (eg, "My pain will improve if my bed time is no later than 9:40 pm", "My fatigue will improve if my dosage of Lyrica is no more than 50 mg", or "My sleep will improve if my yoga is at least 20 minutes"). See Multimedia Appendix 1.

Data Collection

All participation and data collection was online via the project website. Participants completed an FIQ at baseline, and then every subsequent 30 days of program use as "follow-ups" during an 11-month open-use period. Users were free to use the SMARTLog program as often and for as many or as few months as they wished during the study period. In addition to the monthly FIQ, cumulative counts of the number of SMARTLogs and SMARTProfiles accruing for each subject were used for analysis.

Data Analysis

To determine longitudinal associations between FIQ follow-up scores and program use we used Kendall tau-b, a nonparametric statistic that measures both the strength and direction of an ordinal association between two repeated measured variables [8]. We computed this between all FIQ scores and the *indices of program use* for each subject in the sample (ie, the cumulative numbers of SMARTLogs and SMARTProfiles produced for each subject at the time of each completed FIQ).

Since the use data were skewed, with relatively few users showing high rates of use compared to others, Kendall tau-b was more appropriate for these data than the more commonly used Pearson correlation *r*, as the latter inflates Type I error rates and reduces power when the data are not normally distributed [9]. Kendall tau-b was also preferred to the Spearman rank correlation coefficient r_S, as tau-b exhibits improved performance in the presence of contaminated normal distributions with outliers [10] and has better statistical properties; as explained by Howell, "Kendall's τ has generally been given preference of Spearman's r_S because it is a better estimate of the corresponding population parameter, and its standard error is known," [11]. It was also felt that this statistic better characterized the individualized analysis in this case.

All associations were 2-tailed and evaluated at the alpha=.05 level of statistical significance. Prior studies found no demographic predictors of study participation vs nonparticipation. Software used was JMP statistical software (version 14.0).

Results

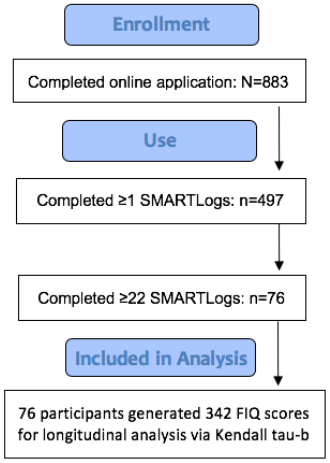
Sample

As is common in remote online studies, applicants varied widely in their levels of commitment to participate, with many applying out of curiosity. This analysis focuses on users who had sufficient levels of program use over the 11-month open-access study period to activate the analytic functionality provided by the SMARTProfile feature of the program. In pilot testing we determined that at least 22 completed SMARTLogs were needed to generate at least 1 SMARTProfile statement [12]. Thus, we set 22 or more SMARTLogs as the a priori criterion for selecting the sample for this analysis.

Of 497 study applicants completing at least 1 SMARTLog, a sample of 76 participants met the use criteria for this analysis Figure 1. Of these, we classified those who completed 22 to 47 SMARTLogs (n=46) as moderate users, and ≥48 SMARTLogs (n=30) as heavy users. We found no demographic predictors of use. Moderate users completed a mean of 32.9 (SD 6.6) SMARTLogs, and heavy users completed a mean of 95.3 (SD 52.5) SMARTLogs. The mean number of follow-up FIQs provided by moderate users (n=46) was 2.54 (SD 1.46), and for heavy users (n=30), 4.97 (SD 2.20).

Of the 76 participants, 99% (75) were female and 96% (73) were white, with a mean age of 47 (SD 12) years. About 70% (53) were college graduates, 32% (24) described themselves as disabled, 43% (33) were working part or full time, 5% (4) were seeking employment, and 20% (15) were retired. Mean reported duration of being diagnosed with FM was 13.6 (SD 10.8) years. The most commonly reported concurrent medical conditions were osteoarthritis (26, 34%), hypertension (22, 29%), chronic fatigue syndrome (21, 28%), and gastroesophageal problems (17, 22%). The number of monthly follow-up FIQs available for analysis per subject ranged from 1-9.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. FIQ: Fibromyalgia Impact Questionnaire.



Change in FIQ scores

The sample had a mean baseline FIQ score of 61.6 (SD 14.7, n=76) and provided a total of 342 FIQ scores for longitudinal analysis via Kendall tau-b. As seen in Table 1, statistically significant inverse correlations were found over time between FIQ scores and both indices of program use (ie, as SMARTLogs or SMARTProfiles accumulated, FM impact was reduced).

Users who completed 61 or more SMARTLogs had mean follow-up scores of 49.9 (25/76, 33% of the sample), an 18.9% drop in FM impact. Users who generated 11 or more new SMARTProfiles had mean follow-up scores of 51.8 (23/76, 30% of the sample), a 15.9% drop.

Table 1. Longitudinal associations of program use with FIQ scores (n=76).

Index of use	Tau-b	*P* value
Cumulative number of SMARTLogs completed	–0.135	<.001
Cumulative number of SMARTProfiles received	–0.133	<.001

Discussion

Principal Results

Statistically significant longitudinal associations were found between both indices of use and FIQ scores; specifically, as cumulative numbers of each index for a user increased, FIQ scores declined, indicating reduced functional impact of FM. In terms of duration of program use, we noted that these indices function as proxies of time, as users could accumulate a given number of SMARTLogs or SMARTProfiles over variable lengths of time. For example, the fastest a user could accumulate 50 SMARTLogs is 50 days, but the accumulated SMARTLogs could also be spread over 6 months or longer; the statistical model does not treat time duration per se as a variable.

Clinical Relevance

As a frame of reference for clinical relevance of these associations, we refer to a pooled analysis by Bennett et al [13] of 2228 patients with FM participating in three clinical trials. In that analysis, participants had a mean baseline FIQ score of 62; across the three studies, the estimated minimal clinically important difference using a 95% confidence interval was found to cause a 14% drop in FIQ scores, which the investigators defined as a “clinically relevant outcome”. In their severity analysis, an FIQ total score ≤38 was found to represent “mild” FM impact, 39-58 a “moderate” FM impact, and 59-100 a “severe” FM impact.

Thus, applying the Bennett et al [13] criteria to the present sample, roughly the top one-third of users in terms of frequency of use experienced mean reductions in FM impact levels,

moving their scores from the “severe” category to the “moderate” category.

Limitations

In any self-directed behavioral health intervention, compliance and use are important contributors to positive outcomes. For such interventions to be successful, they must have high usability and high relevance to the user [14]. In this study, the ratio of users who met the a priori use criteria vs total number of project enrollees (76/497, 15.3%) indicates that a relatively high level of use may be required to achieve the greatest clinical benefit from the program.

We do not know what individual participant variables predicted use levels that were sufficient to satisfy our a priori criteria and be included in this analysis. It is possible that limitations imposed by the illness itself are an obstacle to use for many users. Further exploration is needed to determine how a higher level of compliance and use could be encouraged in this patient population.

Conclusions

The functional impacts of FM in daily living are a significant source of distress. Interventions that aid discovery of personalized strategies to reduce these impacts can make an important contribution to quality of life and well-being in this population, as well as for others with complex chronic illnesses. The findings reported here suggest that clinically significant functional improvement—along with reduced symptoms as previously reported—is possible through an eHealth intervention using personal health informatics. Future research will seek to identify additional predictors of outcomes of the AwareHealth program and explore its use with other chronic conditions.

Conflicts of Interest

The authors are co-owners of the AwareHealth program.

Multimedia Appendix 1
Using AwareHealth: Quick Tour (video). This video with narration shows the main screens of the AwareHealth program and navigates through the various program functions.
[MOV File , 24797 KB - jmir_v22i4e15819_app1.mov]

References

1. Demiris G. Consumer health informatics: past, present, and future of a rapidly evolving domain. Yearb Med Inform 2016 May 20;Suppl 1:S42-S47 [FREE Full text] [doi: 10.15265/IYS-2016-s005] [Medline: 27199196]
2. Friedberg F, Williams DA, Collinge W. Lifestyle-oriented non-pharmacological treatments for fibromyalgia: a clinical overview and applications with home-based technologies. J Pain Res 2012;5:425-435 [FREE Full text] [doi: 10.2147/JPR.S35199] [Medline: 23166446]
3. Collinge W, Yarnold P, Soltysik R. Fibromyalgia symptom reduction by online behavioral self-monitoring, longitudinal single subject analysis and automated delivery of individualized guidance. N Am J Med Sci 2013 Sep;5(9):546-553 [FREE Full text] [doi: 10.4103/1947-2714.118920] [Medline: 24251273]
4. Bennett R. The Fibromyalgia Impact Questionnaire (FIQ): a review of its development, current version, operating characteristics and uses. Clin Exp Rheumatol 2005;23(5 Suppl 39):S154-S162. [Medline: 16273800]
5. Collinge W. ClinicalTrials.gov. 2015 Aug 04. Web-based program for symptom management in fibromyalgia URL: https://clinicaltrials.gov/ct2/show/NCT02515552 [accessed 2019-08-01]
6. Collinge W. AwareHealth.org. Eugene, OR: Collinge and Associates, Inc URL: http://www.awarehealth.org [accessed 2019-08-01]
7. Yarnold P, Soltysik R. Optimal Data Analysis: A Guidebook With Software For Windows. Washington, DC: American Psychological Association; 2005.
8. Kendall MG. Rank Correlation Methods, 2nd Ed. New York, New York: Hafner Publishing Company; 1955.
9. Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. Psychol Methods 2012;17(3):399-417. [doi: 10.1037/a0028087] [Medline: 22563845]
10. Xu W, Hou Y, Hung Y, Zou Y. A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. Signal Processing 2013 Jan;93(1):261-276 [FREE Full text] [doi: 10.1016/j.sigpro.2012.08.005]
11. Howell DC. Statistical Methods for Psychology (7th ed). Belmont, CA: Thomson Wadsworth; 2010:306.
12. Collinge W, Soltysik R, Yarnold P. An internet-based intervention for fibromyalgia self-management: initial design and alpha test. Optimal Data Analysis 2010 Jan 01:163-175 [FREE Full text]
13. Bennett RM, Bushmakin AG, Cappelleri JC, Zlateva G, Sadosky AB. Minimal clinically important difference in the fibromyalgia impact questionnaire. J Rheumatol 2009 Jun;36(6):1304-1311. [doi: 10.3899/jrheum.081090] [Medline: 19369473]
14. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]

Abbreviations

eHealth: electronic health
FIQ: Fibromyalgia Impact Questionnaire
FM: fibromyalgia

Edited by G Eysenbach; submitted 08.08.19; peer-reviewed by S Silverman, A Sule, UK Jamaludin, C Gao, R Mpofu, H Oh; comments to author 29.08.19; revised version received 18.01.20; accepted 27.02.20; published 07.04.20.

Please cite as:
Collinge W, Soltysik R, Yarnold P
Fibromyalgia Impact Reduction Using Online Personal Health Informatics: Longitudinal Observational Study
J Med Internet Res 2020;22(4):e15819
URL: https://www.jmir.org/2020/4/e15819
doi:10.2196/15819
PMID:32131045

Original Paper

Impact of a Web-Based Exercise and Nutritional Education Intervention in Patients Who Are Obese With Hypertension: Randomized Wait-List Controlled Trial

Juan Francisco Lisón[1,2], MD, PhD; Gonzalo Palomar[3], MD, PhD; Marinna S Mensorio[4,5], PhD; Rosa M Baños[2,5], PhD; Ausiàs Cebolla-Martí[2,5], PhD; Cristina Botella[2,6], PhD; Vicent Benavent-Caballer[7], PT, PhD; Enrique Rodilla[1,8], MD, PhD

[1]Department of Medicine, Universidad Cardenal Herrera CEU, CEU Universities, Valencia, Spain

[2]Centro de Investigación Biomédica en Red Fisiopatología Obesidad y Nutrición, Instituto Carlos III, Madrid, Spain

[3]Primary Care Health Center, Quartell, Spain

[4]Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Foundation, Brasilia, Brazil

[5]Department of Personality, Assessment and Psychological Treatment, Facultad de Psicología, Universitat de València, Valencia, Spain

[6]Universitat Jaume I, Castellón, Spain

[7]Department of Physiotherapy, Universidad Cardenal Herrera CEU, CEU Universities, Valencia, Spain

[8]Hypertension and Vascular Risk Unit, Hospital de Sagunto, Sagunto, Spain

Corresponding Author:
Juan Francisco Lisón, MD, PhD
Department of Medicine
Universidad Cardenal Herrera CEU, CEU Universities
Carrer Luis Vives 1, Alfara del Patriarca
Valencia, 46115
Spain
Phone: 34 961369000 ext 64349
Email: juanfran@uchceu.es

Abstract

Background: Internet-based interventions are a promising strategy for promoting healthy lifestyle behaviors. These have a tremendous potential for delivering electronic health interventions in scalable and cost-effective ways. There is strong evidence that the use of these programs can lead to weight loss and can lower patients' average blood pressure (BP) levels. So far, few studies have investigated the effects of internet-based programs on patients who are obese with hypertension (HTN).

Objective: The aim of this study is to investigate the short- and long-term efficacy, in terms of body composition and BP parameters, of a self-administered internet-based intervention involving different modules and learning techniques aimed at promoting lifestyle changes (both physical activity and healthy eating) in patients who are obese with HTN.

Methods: A randomized wait-list controlled trial design was used. We recruited 105 adults with HTN who were overweight or obese and randomly assigned them to either a 3-month internet-based intervention group (n=55) or the wait-list control group (n=50). We assessed BMI (primary outcome), body fat mass (BFM), systolic (S)BP and diastolic (D)BP, blood glucose and insulin levels, physical activity levels, and functional capacity for aerobic exercise at Time 0 (preintervention) and Time 1 (postintervention). All the patients in the wait-list control group subsequently received the intervention, and a secondary within-group analysis, which also included these participants, was conducted at Time 2 (12-month follow-up).

Results: A 2-way mixed analysis of covariance showed a significant decrease in BMI, BFM, and blood glucose at 3 months in the internet-based intervention group; the effect size for the BMI and BFM parameters was moderate to large, and there was also a borderline significant trend for DBP and insulin. These results were either maintained or improved upon at Time 2 and showed significant changes for BMI (mean difference −0.4, 95% CI −0.1 to −0.6; *P*=.005), BFM (mean difference −2.4, 95% CI −1.1 to −3.6; *P*<.001), DBP (mean difference −1.8, 95% CI −0.2 to −3.3; *P*=.03), and blood glucose (mean difference −2, 95% CI 0 to −4; *P*=.04).

Conclusions: Implementation of our self-administered internet-based intervention, which involved different learning techniques aimed to promote lifestyle changes, resulted in positive short- and long-term health benefits in patients who are obese with HTN.

Trial Registration: ClinicalTrials.gov NCT03396302; https://clinicaltrials.gov/ct2/show/NCT03396302

(J Med Internet Res 2020;22(4):e14196) doi:10.2196/14196

KEYWORDS

web; internet; overweight; obesity; hypertension

Introduction

Background

Cardiovascular disease is the leading cause of morbidity and mortality in developed countries and continues to be a major public health issue that accounts for over 4 million deaths per year in Europe [1]. Obesity, which has been described as a global pandemic [2], is one of the most significant medical threats. It is associated with early death [3,4] and is universally recognized as a risk factor for many health complications such as hypertension (HTN). Because a strong association between blood pressure (BP) and body weight has previously been documented [5,6], the increasing prevalence of HTN is thought to be linked to the dramatic increase of individuals who are overweight or obese [7]. As the prevalence, health consequences, and costs of obesity and HTN rise, clinicians and researchers are continuing to investigate a variety of treatment options for these patients.

Promoting healthy behaviors such as physical activity and healthy eating through lifestyle counselling is recommended as the first-line therapy for the treatment of these patients and may be an effective tool for treating obesity and preventing obesity-related health burdens [6,8,9]. However, frequent visits to outpatient clinics are costly and time-consuming for patients as well as for physicians and nurse practitioners. These patients need to develop specific self-management skills so that they can make long-lasting lifestyle changes and adhere to their treatments; however, clinics may not be an ideal environment for them because these surroundings can be perceived as intrusive and threatening [10,11]. Furthermore, in many cases physicians and nurses do not have the proper background training to provide patients with the best possible counselling about nutrition or physical activity. Therefore, an approach is needed that can be offered without imposing additional burdens on our health care workers or budget.

Internet-based interventions (IBIs) are a promising strategy for promoting healthy lifestyle behaviors and have not been entirely explored; these have a tremendous potential for delivering electronic health (eHealth) interventions in scalable and cost-effective ways. IBIs can provide immediate, easy to access, relatively inexpensive, and individually tailored support for self-management and the promotion of behavioral change to large segments of the population. Moreover, there is strong evidence that the use of these programs can lead to weight loss [12-21] and can lower average patient BP levels [21-26]. So far, few studies have investigated the effects of internet-based programs on patients who are obese with HTN [27-29]. Furthermore, to the best of our knowledge, no study has assessed the efficacy of a self-administered IBI, which entails the completion of different modules and incorporates several learning techniques aimed at promoting lifestyle changes in patients who are obese with HTN.

Aim and Hypotheses

This study aimed to examine the short- and long-term efficacy in terms of body composition and BP parameters of a self-administered IBI involving different modules and learning techniques, aimed at promoting lifestyle changes (both physical activity and healthy eating) in patients who are obese with HTN.

Compared with the wait-list control (WLC) group receiving standard medical care, we hypothesized that the IBI group will have significantly better improvements for body composition, BP, blood glucose and insulin levels, physical activity levels, and functional capacity for aerobic exercise from baseline to postintervention.

Methods

Study Design

This prospective, single-center, wait-list controlled trial (trial registration: ClinicalTrials.gov NCT03396302) with balanced randomization (1:1) was approved by the Hospital of Sagunto Human Ethics Committee and followed the ethical guidelines set out in the Declaration of Helsinki.

Eligibility Criteria

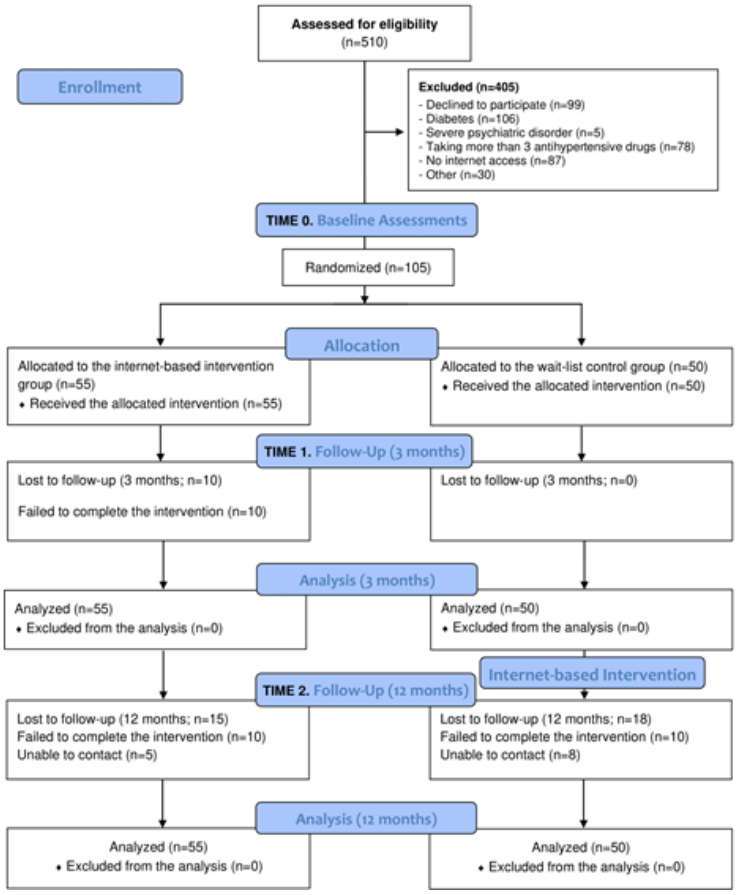
Eligible participants were all adults between 18 and 65 years of age with HTN and who were overweight (BMI>24.9 kg/m^2 and <30 kg/m^2) or had type 1 obesity (BMI>29.9 kg/m^2 and <35 kg/m^2). HTN was defined as systolic (S)BP≥140 mmHg or diastolic (D)BP≥90 mmHg, or current use of antihypertensive medication. All patients in our study were on antihypertensive treatment. The exclusion criteria included a diagnosis of diabetes, previous ischemic heart disease, cerebrovascular disease, or a severe psychiatric disorder; taking more than three antihypertensive drugs; physical impairments precluding participation in physical activity; receiving any treatment for weight loss elsewhere; or having no internet access.

Procedure

This study took place at the HTN and Vascular-Risk Unit at the hospital of Sagunto (Spain), from January 2018 to March 2019. The study included 105 participants. Before the start of the trial, Researcher 1, who was not involved in the recruitment or inclusion of the participants, generated a random sequence (based on simple randomization) using a computerized random number generator; this was concealed from all other study investigators throughout the entire study period. Upon enrollment in the study and after completing the primary and secondary outcome measures (baseline), the participants were randomly assigned either to the 3-month IBI group (n=55) or

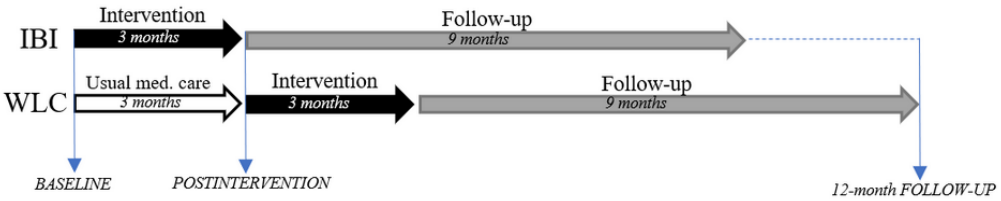
the WLC group (n=50). As shown in the participant flowchart in Figure 1, all the outcome measures were assessed at baseline (Time 0), 3 months postbaseline (Time 1), and at a 12-month follow-up (Time 2) for the IBI group. The WLC group took part in an initial baseline assessment (Time 0) followed by a second assessment 3 months postbaseline (Time 1).

Figure 1. Progression of the participants through the trial.



After the Time 1 assessment, WLC participants received the intervention, which was offered to comply with the instructions of the hospital's human ethics committee, and the patients were subsequently assessed at a 12-month follow-up (Time 2) (Figure 2). Thus, all the participants underwent a preintervention, postintervention (3 months), and a 12-month follow-up assessment, and the WLC participants were assessed twice before the intervention. All the outcome measurements were recorded in both groups by two trained researchers who were blinded to the group allocation.

Figure 2. Measurements at trial profile. IBI: internet-based intervention group; WLC: wait-list control group.



Intervention

The intervention program we used is called "Vivir mejor" (translated from Spanish as "Live better") [27,30]. For this intervention, in addition to usual medical care, participants received a 3-month multimedia, interactive, and self-administered online intervention program comprised of nine modules (Multimedia Appendix 1). These modules specifically focus on obesity and HTN, and are presented via a webpage that aims to progressively establish healthy eating habits and increase the patient's physical activity levels, as recommended by the World Health Organization's guidelines [31,32]. The first 5 modules were activated at a rate of one per week, and modules 6 to 9 were activated every 2 weeks; thus, the intervention lasted for 3 months. The program was delivered in the participants' native language (Spanish) and included psychoeducation about what a healthy lifestyle involves and taught techniques for how this can be achieved on a day-to-day basis. Some of the techniques used were self-monitoring, self-instruction, behavioral recording, stimulus control, self-reinforcement, problem-solving techniques, and homework. In addition, the webpage offered useful tools such as downloadable documents and videos. This program is described in more detail in Baños et al [30].

During the first 3 months of the study and prior to starting the intervention, the WLC group received standard medical care that focused on reducing their cardiovascular risk factors [33]. Standard medical care included antihypertensive prescription, written lifestyle advice, moderate salt restriction with a list of recommended foods, a low-calorie diet, and advice for physical activity, which aimed for 5 sessions per week and a recommended maximum heart rate during exercise adjusted to age. The usual care program did not provide any classes for the subjects. Standard medical care was carried out exclusively by the public National Health System, which the Hypertension Unit belongs to.

This IBI was delivered using a web-based platform called Wix. Wix is a self-hosted website builder and content management system with more than 90 million users [34]. This cloud-based website development platform is customizable with drag-and-drop features and includes apps, graphics, image galleries, fonts, and a responsive design that adjusts the site for viewing with mobile devices. We purchased a unique website domain URL to avoid pop-up advertising from Wix. To prevent the general public from accessing the site, a password was added to secure the proprietary information and participant responses [35].

Outcome Measures

BMI and secondary outcomes (body fat mass [BFM], SBP and DBP, plasma glucose, insulin, habitual level of physical activity, and functional capacity for aerobic exercise) were measured using validated and reliable tests. BMI (kg/m^2) was calculated using an electronic balance scale (SECA 780 with a mechanical telescopic stadiometer).

BFM was determined using a body-fat analyzer (Tanita SC-331S). SBP and DBP were measured with a validated, semi-automatic, digital tensiometer (Colin Press-Mate BP-8800) according to the European Society of Hypertension and European Society of Cardiology guidelines, as described elsewhere [36,37]. All the assessments (baseline, 3 month postintervention, and 12-month follow-up) were scheduled between 8 am and noon to minimize diurnal BP variability. Blood samples were obtained in the morning after a minimum 8 hours of fasting. Serum biochemical profiles were measured using a multiple-channel autoanalyzer. Plasma glucose was assayed using the glucose oxidase method (Beckman Glucose Analyzer).

At Time 0, the patients' habitual levels of physical activity were objectively monitored for 7 consecutive days using an accelerometer (Actigraph GT3X) [36]. The participants were carefully instructed on how to attach the activity monitor (using an elastic belt), which was to be worn each day for the whole measurement period, and were asked to wear the monitor during the daytime; exceptions were made during the performance of water activities such as swimming or bathing. The total amount of physical activity recorded by the accelerometer was expressed as the average of the total counts per minute of the registered time.

A submaximal exercise test, the 6-minute walk test (6MWT), was used to assess the participants' functional capacity for aerobic exercise. This test evaluates the maximum distance that they could cover along a 30-m long corridor during a 6-minute period. Participants were instructed to walk along the walkway as fast as possible and to stop when needed. The assessor walked alongside the participants to ensure their safety and provided them with standardized verbal encouragement at 1, 3, and 5 minutes, and the total distance covered (in meters) was recorded. Participants were instructed to avoid smoking for 48 hours, caffeine for 12 hours, and strenuous exercise for 24 hours prior to their assessment.

We directly recorded participants' participation and activity in the program (to assess their engagement) and determined adherence to the intervention by tracking the number of program activities they completed. We also registered the percentage of participants who completed all the activities in the program (ie, completed all nine modules).

Data Analysis

To detect a reduction in BMI of 1 (SD 1.7), which agrees with the data of a previous study [27], with a two-sided 5% significance level and a power of 80%, and also accounting for an anticipated dropout rate of 10%, a sample size of 52 participants per group was required.

This study used both a between-subjects controlled (analysis at 3 months) and a within-subjects uncontrolled (analysis at 12 months) design; therefore, it involved two different statistical approaches. The between-group comparison (IBI vs WLC) assessed the effect of the IBI on outcomes at 3 months; the within-subject uncontrolled analyses assessed the effect of the IBI over time at the individual level (using pooled data from both groups). This approach was selected because all the WLC participants also subsequently received the IBI to comply with the instructions of the hospital's human ethics committee. The statistical analysis was performed according to intention-to-treat. Statistical analyses were performed using SPSS version 19.0 for Windows, and the statistical significance was set at $P<.05$ for all our analyses. The data in this study are presented as mean (SD).

Analysis at 3 Months

Two-way mixed analysis of covariance (ANCOVA) tests were used to compare the study effects on the BMI, BFM, SBP, DBP, blood glucose and insulin levels, physical activity levels, and functional capacity for aerobic exercise, using time (baseline vs 3 months postintervention) as the within-group factor and group (IBI vs WLC) as the between-group factor. The analysis was adjusted for sex, age, and use of antihypertensive drugs. Effect sizes were estimated using partial eta^2 and interpreted following the Cohen guidelines [38] for small effect sizes (partial eta^2=0.01), moderate effect sizes (partial eta^2=0.06), and large effect sizes (partial eta^2=0.14).

Analysis at 12 Months

After testing the normality (using the Shapiro-Wilk test) of the pooled data from both groups (N=105) at the 12-month follow-up, the following statistical tests were carried out as follows: (1) *t* tests on related samples to compare the baseline vs 12-month follow-up values for the anthropometric variables

(ie, BMI and BFM), SBP, DBP, and blood glucose; and (2) Wilcoxon tests to compare the preintervention vs 12-month follow-up values for insulin.

We did not record the patients' levels of physical activity or their functional capacities for aerobic exercise at 12 months because of the implied complexity of asking participants to record these measurements for 1 year.

Program engagement was analyzed by calculating the percentage of participants who completed the entire program.

Results

We screened 510 participants in this randomized controlled trial. A total of 405 consecutive subjects were not allocated for randomization, because they declined to participate (n=99) or did not meet the inclusion criteria (n=306): diabetes (106), severe psychiatric disorder (5), taking more than 3 antihypertensive drugs (78), no internet access (87), or other (30). The general characteristics of the study population are shown in Table 1.

Table 1. Baseline characteristics of the participants.

Variables	WLC[a] group (n=50), mean (SD)	IBI[b] group (n=55), mean (SD)
Age (years)	51.4 (9.3)	54.9 (8.3)
Weight (kg)	81.1 (11.8)	85.4 (12)
Antihypertensive drugs (n)	1.6 (1.1)	2.1 (1.2)
BMI (kg/m^2)	29.9 (2.6)	30.1 (2.7)
Body fat mass (kg)	27.0 (6.9)	28.2 (6.3)
Systolic blood pressure (mmHg)	128.5 (13.5)	132.2 (14.2)
Diastolic blood pressure (mmHg)	75.8 (9.1)	78.7 (8.1)
Blood glucose (mg/dL)	98 (10)	99 (15)
Insulin (mg/dL)	13.9 (8.7)	15.0 (9.7)
Physical activity counts min^{-1}	227 (87)	245 (103)
6-minute walk test (meters)	554 (69)	559 (72)

[a]WLC: wait-list control.

[b]IBI: internet-based intervention.

Analysis at 3 Months

The results of the 2-way mixed ANCOVA showed a significant decrease in BMI, BFM, and blood glucose after 3 months in the IBI group, with a moderate to large effect size for BMI and BFM; the analysis also highlighted a borderline significant trend (P=.05) for DBP and insulin (Table 2). In contrast, we observed a significant increase in BMI and insulin among the WLC group. Additionally, intragroup analysis revealed a statistically significant increase in the functional capacity for aerobic exercise both in the IBI and the WLC groups (Table 2); however, no between-group differences were found. No changes were observed in either group for the level of physical activity measured with accelerometers.

Table 2. Intragroup comparisons: baseline vs postintervention (at 3 months)a.

Variables	Wait-list control group (n=50)			Internet-based intervention group (n=55)		
	Difference (95% CI)	Partial eta^2	*P* value	Difference (95% CI)	Partial eta^2	*P* value
BMI (kg/m^2)	0.3 (0.1-0.5)	.062	.01	–0.4 (–0.7 to –0.2)	.132	<.001
Body fat mass (kg)	0.3 (–0.4 to 1.1)	.018	.34	–1.1 (–1.9 to –0.3)	.125	.009
SBP[b] (mmHg)	0.0 (–3.7 to 3.7)	.000	.996	–2.6 (–6.1 to 0.9)	.021	.15
DBP[c] (mmHg)	0.6 (–1.8 to 3.0)	.002	.63	–2.2 (–4.5 to 0.0)	.037	.05
Blood glucose (mg/dL)	2.0 (–1.2 to 5.3)	.015	.22	–3.5 (–6.6 to –0.4)	.048	.03
Insulin (mg/dL)	2.3 (0.4-4.3)	.058	.02	–1.7 (–3.7 to 0.2)	.032	.07
Physical activity counts min^{-1}	–8 (–16 to 32)	.005	.52	–6 (–30 to 18)	.002	.64
6MWT[d] (meters)	21 (8-33)	.108	.001	32 (19-46)	.213	<.001

[a]Difference was calculated as 3 months minus baseline.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]6MWT: 6-minute walk test.

Analysis at 12 Months

The results at the 12-month follow-up (Table 3) showed significant improvements in BMI (mean difference [MD] −0.4, 95% CI −0.1 to −0.6; *P*=.005), BFM (MD −2.4, 95% CI −1.1 to −3.6; *P*<.001), DBP (MD −1.8, 95% CI −0.2 to −3.3. *P*=.03), and blood glucose levels (MD −2, 95% CI 0 to −4; *P*=.04) The results of the Wilcoxon tests did not show any changes in insulin values. Regarding engagement in the program, 73.3% (77 out of 105 participants) completed the entire program, comprising of nine modules.

Table 3. Comparison of the baseline vs 12-month follow-up values for all participants (N=105).

Variables	Baseline	12-month follow-up	Difference[a] (95% CI)	*P* value
BMI (kg/m^2), mean (SD)	30.2 (2.7)	29.8 (2.7)	–0.4 (–0.1 to –0.6)	.005
Body fat mass (kg), mean (SD)	27.6 (6.4)	25.2 (7.9)	–2.4 (–1.1 to –3.6)	<.001
SBP[b] (mmHg), mean (SD)	130.5 (13.7)	129.3 (14.1)	–1.2 (0.9 to –3.4)	.24
DBP[c] (mmHg), mean (SD)	77.4 (9.1)	75.6 (9.1)	–1.8 (–0.2 to –3.3)	.03
Blood glucose (mg/dL), mean (SD)	100 (14)	98 (11)	–2 (0 to –4)	.04
Insulin (mg/dL), median (IQR)	16 (12)	15 (9)	–1 (1 to –2)	.64

[a]Difference calculated as 12 months minus baseline.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

Discussion

Principal Results

The present study evaluated the efficacy of a 3-month completely self-administered IBI called *Live better* that aims to promote lifestyle changes in patients who are obese with HTN. Our results showed a significant decrease in the BMI, BFM, and blood glucose levels at 3 months in the IBI group, with a moderate to large effect size for BMI and BFM, and also highlighted a trend toward significance for DBP and insulin levels. Moreover, these improvements in BMI, BFM, and BP were sustained and reached statistical significance for DBP at the 12-month follow-up. When evaluating specific IBIs, several attributes may be associated with increased e-counselling efficacy. These include the duration [22,39], the range of behavior change techniques offered [22,40], the target behaviors to be modified [22], whether specific disease entities are targeted or not [39], and program engagement [35,41].

Duration

For any intervention to have a significant and sustained effect, a minimum follow-up time of 6 months is required [39]. Indeed, in a recent meta-analysis, Sam Liu et al [22] showed that IBIs lasting at least 6 months were associated with greater BP reductions. These authors argued that the influence of lifestyle interventions on BP may require a critical period for therapeutic changes to appear, and longer interventions might be required to facilitate comprehensive physical changes such as weight reduction. After beginning our study, we found a significant BMI reduction (0.4 kg/m^2) after 3 months that was sustained after 12 months. Moreover, the mean decrease in BFM (by 1.1

kg) achieved after 3 months had doubled 12 months later (to 2.4 kg). Although this parameter is less frequently used because of its lower feasibility, BFM may be a better marker of cardiovascular risk compared to BMI, which has been criticized because the latter does not always reflect true body fat content [42,43]. Indeed, BMI has some limitations in assessing the risk of obesity-related diseases in subjects with low muscle and high body fat content [44] and in individuals with increased body fat and a normal BMI.

Consistent with our results, a Cochrane meta-analysis concluded that interactive computer-based interventions, compared to no intervention or minimal interventions (eg, pamphlets or usual care), are an effective tool for enhancing weight loss and weight maintenance [12]. In addition, providing support oriented towards self-management while patients change their lifestyle leads to improved health outcomes [24,45] and better long-term effects [46].

Concerning BP values, we found a nonsignificant decrease in SBP and DBP after 3 months, which was sustained and even reached statistical significance after 12 months in the case of DBP (−1.8 mmHg). These results are slightly lower than those reported in the Internet Lifestyle Counselling meta-analysis (−3.8 mmHg for SBP and −2.1 mmHg for DBP) by Liu et al [22] and are also lower than the BP reductions reported by other meta-analyses that considered face-to-face lifestyle counselling [9,47]. The lack of a stronger significant effect in our study may be because the BP at baseline was already well controlled by antihypertensive medications [48-50] by the specialized HTN unit. Effectively, starting from normal mean values in treated hypertensives, even greater weight losses would probably not have induced further reductions of BP. This phenomenon is known as the “floor effect” and must be accounted for [51].

Range of Behavior Change Techniques

Two recent meta-analyses that evaluated the BP lowering [22] and weight loss [40] achieved by e-counselling lifestyle interventions reported that their efficacy depended on the number of intervention components they included. The meta-analysis by Liu et al [22] found that BP was preferentially, and significantly, reduced by interventions providing a wider range of behavioral change techniques and suggested that “a critical number of techniques (at least 5) may be required to build a flexible repertoire of skills that are necessary to overcome situational stressors that might otherwise impede therapeutic lifestyle change”. These authors reported that the behavioral change techniques present in more than 50% of the successful IBIs shared the following features: provisioning general information about the consequences of the patient’s behavior (86% of studies), incorporating feedback on performance (86%), prompting behavioral self-monitoring (71%), and giving instructions on how to perform targeted behavioral changes (71%). Likewise, Khaylis et al [52] also conducted a systematic review of efficacious technology-based weight loss interventions and identified several similar components (self-monitoring, counsellor feedback and communication, social support, structured programs, and individually tailored programs) that were key to their success. Other effective components identified in the technology-based weight loss literature as potential factors that increase intervention effectiveness are goal setting, motivational interviewing, and incentives [30,53,54].

All the behavioral change techniques mentioned by Liu et al [22] and Khaylis et al [52] were present in our intervention, except for feedback incorporation, motivational interviewing, and social support, as these were not possible given the self-administered nature of our intervention, which maintained its low monetary and time costs. Moreover, this trial was distinct insofar as it also incorporated problem-solving techniques focused on the regulation of emotional eating (module 5) and the difficulties associated with body image and assertiveness (module 7). As Katan [55] suggested, cognitions and feelings have a big impact on behavior during dieting and, thus, may strengthen or disrupt treatment engagement and compliance with clinical prescriptions. Indeed, psychological factors and processes mediate every behavior change and differently affect both the initiation and maintenance phases of change [56].

Target Behaviors to Be Modified

Exercise and diet are two major target behaviors for modification by internet-based e-counselling interventions designed to prevent or treat cardiovascular disease risk factors such as obesity and HTN. Numerous studies suggest that automated, self-guided, internet-based lifestyle counselling (e-counselling) programs can evoke meaningful improvements in daily physical activity [57,58] and dietary behaviors [58,59]. Moreover, meta-analysis reviews indicate that exercise and diet, provided by conventional programs or by IBIs, significantly decrease cardiovascular risk factors [9,16,47,60,61]. Our *Live better* intervention focused on modifying both of these target behaviors, because it has been shown that tackling them simultaneously is more effective at promoting weight loss than targeting either alone [62].

After 3 months, our results showed improvements in the patients’ functional capacities for aerobic exercise with a large effect size in both sexes (partial eta^2=0.213) in the IBI group. Nevertheless, caution must be taken in interpreting these results, as this variable also increased with a moderate effect size in the WLC group (partial eta^2=0.108). It is worth considering that a learning effect over a 2-month period has been documented with repeated administration of the 6MWT [63]. In contrast to subjective physical activity estimations such as questionnaires, in our study we used an accelerometer to obtain precise and objective measurements of activity levels in individuals who are overweight or obese [64]. We did not find significant intergroup or intragroup differences after 3 months. However, these results should be interpreted with caution since the accelerometers do not accurately quantify activities such as resistance training or swimming (included in the recommendations of the IBI). The fact that the reduction in BFM doubled in the time between the end of the intervention (3 months) and the end of the study (12 months), while simultaneously maintaining a constant BMI, may indicate that the patients had increased muscle mass secondary to higher levels of physical activity during this period. Thus, future studies will be required to measure muscle mass to test this hypothesis.

Targeting Specific Disease Entities

Live better was specifically designed to treat HTN in patients who are overweight or those with type I obesity, and its implementation resulted in modest but positive measurable results in body composition and BP. In general, internet-based programs are more successful if they are targeted at specific disease entities such as HTN [39]. Although our study listed diabetes as an exclusion criterion, and our intervention did not specifically focus on this disease, improved glucose metabolism was also relevant in our study because we detected a significant and sustained decrease in glucose levels among our patients. The close relationship between glucose metabolism variables and healthy living through physical activity and healthy eating is well known. In line with this, a recent high-quality randomized lifestyle-intervention trial conducted over 12 and 24 months in individuals with type 2 diabetes mellitus or impaired fasting glucose levels showed the IBI to be useful in reducing fasting plasma glucose [21].

Program Engagement

In terms of program engagement, the percentage of participants who completed our entire program was high (77/105, 73.3%) and was similar to the levels reported in related eHealth interventions [35,65]. Increasing knowledge about healthy lifestyles and making tailor-made prevention programs possible can empower individuals and improve their adherence to interventions [66]. In this sense, the wide and still-growing access and use of the internet has become a major resource in the assessment of health information [67]. This is especially true for adults with chronic conditions who are more likely to seek health information on the internet than their counterparts with acute ailments [68-70]. Using modern information and communication technologies to deliver physical activity and diet interventions is particularly promising considering the increased proliferation of such technologies in many developing countries.

The internet is an efficient way to prevent and treat chronic conditions by promoting healthy lifestyles because it can reach more individuals (including those with limited access to health services or with low levels of social support) and can provide more intensive contact at potentially lower costs than conventional face-to-face programs [71-74]; it can provide immediate, easy access, individually tailored (one-to-one), and "permanent" (accessible at any time) behavioral-change support by delivering care to patients in the comfort of their own homes with self-paced delivery [35,40,73]; and it is less intrusive than traditional methods and can more easily be implemented in an environment that is less threatening than a hospital [10].

However, despite the encouraging results of this study and other studies on internet-based programs, the goal of web-based interventions is not to replace in-person care but rather to maximize care. Of note, internet interventions also have potential disadvantages such as their inability to recognize comorbidities that would not have become apparent without the patient implementing lifestyle changes. Furthermore, internet use depends on age, income, education level, and digital skills, and there may be participation bias and lower response rates because of technical problems or different levels of computer experience among participants [5]. In addition, ironically, internet use is a sedentary behavior, which is known to be an independent risk factor for cardiovascular disease, so caution should be exercised when designing these programs so that they prevent, rather than encourage, further sedentary attitudes and behaviors.

Limitations

The main limitation of this study was its lack of a control group for the analysis carried out at 12 months. Although the positive results at 3 months remained at the 12-month follow-up visit, the follow-up analyses were based on uncontrolled data and, thus, should be interpreted with caution. Controls eliminate possible alternate explanations of experimental results, especially those of confounding variables and experimental bias, which allows investigators to control for threats to validity.

Furthermore, the participants we enrolled had demonstrated an initial level of motivation to engage in an eHealth program. Therefore, our findings may only be generalizable to individuals with internet access who are similarly interested in such eHealth interventions [67]. In addition, our participants were recruited from a public hospital (rather than a private one), which may have influenced our results because sociodemographic status (which could correlate with the use of public hospitals) has been related to treatment adherence for chronic conditions [75,76]. Another possible limitation of the study was the inclusion of BMI as the primary outcome. Although the waist circumference or waist-hip ratio are considered better anthropometric parameters to reflect the risk of cardiovascular disease associated with obesity, the BMI clinical tool has been shown to have the least bias during assessment [73]. Furthermore, BMI measurements in our study were complemented with segmental body-fat distribution analysis measured with bioelectrical impedance. Even though intervention acceptability is related to its eventual effectiveness [77], we did not assess this data in our study, and thus, we do not know how well the participants accepted this IBI. Finally, we did not measure physical activity levels or functional capacity for aerobic exercise at the 12-month follow-up.

Conclusions

In summary, we explored the effects of implementing a 3 month, self-administered IBI involving different learning techniques and aimed at promoting lifestyle changes (physical activity and healthy eating) in patients who are obese with HTN. Overall, the participant engagement was high, and we observed moderate to large effect sizes in relation to BMI and BFM reductions. In addition, there was a favorable trend towards a relation to BP, which reached statistical significance at the 12-month follow-up. Taken together, these findings suggest that our tailored approach for delivering a lifestyle change intervention to patients who are obese with HTN provides positive health benefits. Simple strategies that can easily be incorporated into daily living in a scalable and cost-effective way can empower patients by educating them about health, thus, increasing their confidence and promoting self-management. Future research should investigate the acceptability and cost-effectiveness of self-administered internet-based interventions in populations with different sociodemographic status.

Acknowledgments

This work was supported by grants from the Generalitat Valenciana (Subvenciones para grupos de investigación consolidables-AICO/2019/331) and from the University CEU Cardenal Herrera (convocatoria de Provectos en Consolidación CEU Banco Santander 2017-2018). Centro de Investigación Biomédica en Red de Fisiopatología de la Obesidad y Nutrición (CIBERobn) is an initiative of Instituto de Salud Carlos III. Author MM is grateful to CAPES/Brazil (Coordination for the Improvement of Higher Level Personnel) for the doctoral scholarship. Author ER thanks our unit's nurse Mari Cruz Marco for her excellent and priceless collaboration in coordinating the patients' visits.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topics and objectives covered in the modules.
[DOCX File , 5848 KB - jmir_v22i4e14196_app1.docx]

Multimedia Appendix 2
CONSORT e-HEALTH checklist (V. 1.6.1).
[PDF File (Adobe PDF File), 2499 KB - jmir_v22i4e14196_app2.pdf]

References

1. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe 2014: epidemiological update. Eur Heart J 2014 Nov 07;35(42):2929. [doi: 10.1093/eurheartj/ehu378] [Medline: 25381246]
2. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 2011 Aug 27;378(9793):804-814. [doi: 10.1016/S0140-6736(11)60813-1] [Medline: 21872749]
3. Prospective Studies Collaboration, Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. Lancet 2009 Mar 28;373(9669):1083-1096 [FREE Full text] [doi: 10.1016/S0140-6736(09)60318-4] [Medline: 19299006]
4. Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. JAMA 2005 Apr 20;293(15):1861-1867. [doi: 10.1001/jama.293.15.1861] [Medline: 15840860]
5. Doll S, Paccaud F, Bovet P, Burnier M, Wietlisbach V. Body mass index, abdominal adiposity and blood pressure: consistency of their association across developing and developed countries. Int J Obes Relat Metab Disord 2002 Jan;26(1):48-57. [doi: 10.1038/sj.ijo.0801854] [Medline: 11791146]
6. Seo D, Sa J. A meta-analysis of psycho-behavioral obesity interventions among US multiethnic and minority adults. Prev Med 2008 Dec;47(6):573-582. [doi: 10.1016/j.ypmed.2007.12.010] [Medline: 18201758]
7. Francischetti EA, Genelhu VA. Obesity-hypertension: an ongoing pandemic. Int J Clin Pract 2007 Feb;61(2):269-280. [doi: 10.1111/j.1742-1241.2006.01262.x] [Medline: 17263714]
8. Avenell A, Broom J, Brown TJ, Poobalan A, Aucott L, Stearns SC, et al. Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement. Health Technol Assess 2004 May;8(21):iii-iv, 1 [FREE Full text] [doi: 10.3310/hta8210] [Medline: 15147610]
9. Dickinson HO, Mason JM, Nicolson DJ, Campbell F, Beyer FR, Cook JV, et al. Lifestyle interventions to reduce raised blood pressure: a systematic review of randomized controlled trials. J Hypertens 2006 Feb;24(2):215-233. [doi: 10.1097/01.hjh.0000199800.72563.26] [Medline: 16508562]
10. Artinian NT. Telehealth as a tool for enhancing care for patients with cardiovascular disease. J Cardiovasc Nurs 2007;22(1):25-31. [doi: 10.1097/00005082-200701000-00004] [Medline: 17224694]
11. Goessens BMB, Visseren FLJ, de Nooijer J, van den Borne HW, Algra A, Wierdsma J, SMART Study Group. A pilot-study to identify the feasibility of an Internet-based coaching programme for changing the vascular risk profile of high-risk patients. Patient Educ Couns 2008 Oct;73(1):67-72. [doi: 10.1016/j.pec.2008.06.004] [Medline: 18639410]
12. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012 Aug 15(8):CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
13. Weinstein PK. A review of weight loss programs delivered via the Internet. J Cardiovasc Nurs 2006;21(4):251-8; quiz 259. [doi: 10.1097/00005082-200607000-00003] [Medline: 16823276]

14. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
15. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 07;285(9):1172-1177. [doi: 10.1001/jama.285.9.1172] [Medline: 11231746]
16. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of Web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012 May;36(5):675-685. [doi: 10.1038/ijo.2011.121] [Medline: 21694698]
17. Lemacks J, Wells BA, Ilich JZ, Ralston PA. Interventions for improving nutrition and physical activity behaviors in adult African American populations: a systematic review, January 2000 through December 2011. Prev Chronic Dis 2013 Jun 20;10:E99 [FREE Full text] [doi: 10.5888/pcd10.120256] [Medline: 23786910]
18. Petersen R, Sill S, Lu C, Young J, Edington DW. Effectiveness of employee internet-based weight management program. J Occup Environ Med 2008 Feb;50(2):163-171. [doi: 10.1097/JOM.0b013e31815c6cf6] [Medline: 18301173]
19. Bennett JB, Broome KM, Schwab-Pilley A, Gilmore P. A web-based approach to address cardiovascular risks in managers: results of a randomized trial. J Occup Environ Med 2011 Aug;53(8):911-918 [FREE Full text] [doi: 10.1097/JOM.0b013e3182258bd8] [Medline: 21785368]
20. Spittaels H, De Bourdeaudhuij I, Brug J, Vandelanotte C. Effectiveness of an online computer-tailored physical activity intervention in a real-life setting. Health Educ Res 2007 Jun;22(3):385-396. [doi: 10.1093/her/cyl096] [Medline: 16971674]
21. Kang JY, Cho SW, Sung SH, Park YK, Paek YM, Choi TI. Effect of a continuous diabetes lifestyle intervention program on male workers in Korea. Diabetes Res Clin Pract 2010 Oct;90(1):26-33. [doi: 10.1016/j.diabres.2010.06.006] [Medline: 20621377]
22. Liu S, Dunford SD, Leung YW, Brooks D, Thomas SG, Eysenbach G, et al. Reducing blood pressure with Internet-based interventions: a meta-analysis. Can J Cardiol 2013 May;29(5):613-621. [doi: 10.1016/j.cjca.2013.02.007] [Medline: 23618507]
23. Nunes APDOB, Rios ACDS, Cunha GAD, Barretto ACP, Negrão CE. [The effects of nonsupervised exercise program, via internet, on blood pressure and body composition in normotensive and prehypertensive individuals]. Arq Bras Cardiol 2006 Apr;86(4):289-296 [FREE Full text] [doi: 10.1590/s0066-782x2006000400009] [Medline: 16680294]
24. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
25. Liu S, Brooks D, Thomas SG, Eysenbach G, Nolan RP. Effectiveness of user- and expert-driven web-based hypertension programs: an RCT. Am J Prev Med 2018 Apr;54(4):576-583. [doi: 10.1016/j.amepre.2018.01.009] [Medline: 29456025]
26. Watson AJ, Singh K, Myint- U, Grant RW, Jethwani K, Murachver E, et al. Evaluating a web-based self-management program for employees with hypertension and prehypertension: a randomized clinical trial. Am Heart J 2012 Oct;164(4):625-631. [doi: 10.1016/j.ahj.2012.06.013] [Medline: 23067923]
27. Mensorio MS, Cebolla-Martí A, Rodilla E, Palomar G, Lisón JF, Botella C, et al. Analysis of the efficacy of an internet-based self-administered intervention ("Living Better") to promote healthy habits in a population with obesity and hypertension: An exploratory randomized controlled trial. Int J Med Inform 2019 Apr;124:13-23. [doi: 10.1016/j.ijmedinf.2018.12.007] [Medline: 30784422]
28. Park M, Kim H, Kim K. Cellular phone and Internet-based individual intervention on blood pressure and obesity in obese patients with hypertension. Int J Med Inform 2009 Oct;78(10):704-710. [doi: 10.1016/j.ijmedinf.2009.06.004] [Medline: 19643661]
29. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. Obesity (Silver Spring) 2010 Feb;18(2):308-313 [FREE Full text] [doi: 10.1038/oby.2009.242] [Medline: 19696764]
30. Banos RM, Mensorio MS, Cebolla A, Rodilla E, Palomar G, Lisón JF, et al. An internet-based self-administered intervention for promoting healthy habits and weight loss in hypertensive people who are overweight or obese: a randomized controlled trial. BMC Cardiovasc Disord 2015 Aug 04;15:83 [FREE Full text] [doi: 10.1186/s12872-015-0078-1] [Medline: 26239241]
31. World Health Organization. 2018 Oct 23. Healthy diet URL: https://www.who.int/news-room/fact-sheets/detail/healthy-diet [accessed 2020-03-13]
32. World Health Organization. 2018 Feb 16. Obesity and overweight URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2020-03-13]
33. Pascual JM, Rodilla E, Costa JA, Perez-Lahiguera F, Gonzalez C, Lurbe E, et al. Body weight variation and control of cardiovascular risk factors in essential hypertension. Blood Press 2009;18(5):247-254. [doi: 10.3109/08037050903244791] [Medline: 19919395]
34. Wix. URL: http://www.wix.com
35. Staffileno BA, Tangney CC, Fogg L. Favorable outcomes using an eHealth approach to promote physical activity and nutrition among young African American women. J Cardiovasc Nurs 2018;33(1):62-71. [doi: 10.1097/JCN.0000000000000409] [Medline: 28481823]

36. Mancia G, Fagard R, Narkiewicz K, Redon J, Zanchetti A, Böhm M, et al. 2013 ESH/ESC guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). Eur Heart J 2013 Jul;34(28):2159-2219. [doi: 10.1093/eurheartj/eht151] [Medline: 23771844]
37. Rodilla E, Costa JA, Martín J, González C, Pascual JM, Redon J. Impact of abdominal obesity and ambulatory blood pressure in the diagnosis of left ventricular hypertrophy in never treated hypertensives. Med Clin (Barc) 2014 Mar 20;142(6):235-242. [doi: 10.1016/j.medcli.2013.04.046] [Medline: 24139053]
38. Cohen J. Statistical Power Analysis For The Behavioral Sciences. New Jersey: Lawrence Erlbaum Associates; 1988.
39. Aneni EC, Roberson LL, Maziak W, Agatston AS, Feldman T, Rouseff M, et al. A systematic review of internet-based worksite wellness approaches for cardiovascular disease risk management: outcomes, challenges & opportunities. PLoS One 2014;9(1):e83594 [FREE Full text] [doi: 10.1371/journal.pone.0083594] [Medline: 24421894]
40. Seo D, Niu J. Evaluation of internet-based interventions on waist circumference reduction: a meta-analysis. J Med Internet Res 2015 Jul 21;17(7):e181 [FREE Full text] [doi: 10.2196/jmir.3921] [Medline: 26199208]
41. Müller AM, Alley S, Schoeppe S, Vandelanotte C. The effectiveness of e-& mHealth interventions to promote physical activity and healthy diets in developing countries: A systematic review. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]
42. Lavie CJ, Milani RV, Ventura HO. Obesity and cardiovascular disease: risk factor, paradox, and impact of weight loss. J Am Coll Cardiol 2009 May 26;53(21):1925-1932 [FREE Full text] [doi: 10.1016/j.jacc.2008.12.068] [Medline: 19460605]
43. Romero-Corral A, Montori VM, Somers VK, Korinek J, Thomas RJ, Allison TG, et al. Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. Lancet 2006 Aug 19;368(9536):666-678. [doi: 10.1016/S0140-6736(06)69251-9] [Medline: 16920472]
44. Thibault R, Pichard C. The evaluation of body composition: a useful tool for clinical practice. Ann Nutr Metab 2012;60(1):6-16 [FREE Full text] [doi: 10.1159/000334879] [Medline: 22179189]
45. Beishuizen CRL, Stephan BCM, van Gool WA, Brayne C, Peters RJG, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
46. Gagliardi AR, Abdallah F, Faulkner G, Ciliska D, Hicks A. Factors contributing to the effectiveness of physical activity counselling in primary care: a realist systematic review. Patient Educ Couns 2015 Apr;98(4):412-419. [doi: 10.1016/j.pec.2014.11.020] [Medline: 25499578]
47. Whelton SP, Chin A, Xin X, He J. Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. Ann Intern Med 2002 Apr 02;136(7):493-503. [doi: 10.7326/0003-4819-136-7-200204020-00006] [Medline: 11926784]
48. Southard BH, Southard DR, Nuckolls J. Clinical trial of an Internet-based case management system for secondary prevention of heart disease. J Cardiopulm Rehabil 2003;23(5):341-348. [doi: 10.1097/00008483-200309000-00003] [Medline: 14512778]
49. Zutz A, Ignaszewski A, Bates J, Lear SA. Utilization of the internet to deliver cardiac rehabilitation at a distance: a pilot study. Telemed J E Health 2007 Jun;13(3):323-330. [doi: 10.1089/tmj.2006.0051] [Medline: 17603835]
50. Vernooij JWP, Kaasjager HAH, van der Graaf Y, Wierdsma J, Grandjean HMH, Hovens MMC, SMART Study Group. Internet based vascular risk factor management for patients with clinically manifest vascular disease: randomised controlled trial. BMJ 2012 Jun 12;344:e3750 [FREE Full text] [doi: 10.1136/bmj.e3750] [Medline: 22692651]
51. Cohen N, Flamenbaum W. Obesity and hypertension. Demonstration of a "floor effect". Am J Med 1986 Feb;80(2):177-181. [doi: 10.1016/0002-9343(86)90006-9] [Medline: 3946433]
52. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
53. Morgan PJ, Lubans DR, Collins CE, Warren JM, Callister R. The SHED-IT randomized controlled trial: evaluation of an Internet-based weight-loss program for men. Obesity (Silver Spring) 2009 Nov;17(11):2025-2032. [doi: 10.1038/oby.2009.85] [Medline: 19343018]
54. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, de Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 2013 Nov 06;15(11):e233 [FREE Full text] [doi: 10.2196/jmir.2843] [Medline: 24195965]
55. Katan MB. Weight-loss diets for the prevention and treatment of obesity. N Engl J Med 2009 Feb 26;360(9):923-925. [doi: 10.1056/NEJMe0810291] [Medline: 19246365]
56. Rothman AJ. Toward a theory-based analysis of behavioral maintenance. Health Psychol 2000 Jan;19(1S):64-69. [doi: 10.1037/0278-6133.19.suppl1.64] [Medline: 10709949]
57. King AC, Bickmore TW, Campero MI, Pruitt LA, Yin JL. Employing virtual advisors in preventive care for underserved communities: results from the COMPASS study. J Health Commun 2013;18(12):1449-1464. [doi: 10.1080/10810730.2013.798374] [Medline: 23941610]
58. Nolan RP, Liu S, Shoemaker JK, Hachinski V, Lynn H, Mikulis DJ, et al. Therapeutic benefit of internet-based lifestyle counselling for hypertension. Can J Cardiol 2012 May;28(3):390-396. [doi: 10.1016/j.cjca.2012.02.012] [Medline: 22498181]

59. Wangberg SC. An Internet-based diabetes self-care intervention tailored to self-efficacy. Health Educ Res 2008 Feb;23(1):170-179. [doi: 10.1093/her/cym014] [Medline: 17412717]
60. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019 Apr;18(4):260-271. [doi: 10.1177/1474515119826510] [Medline: 30667278]
61. van den Berg MH, Schoones JW, Vliet Vlieland TPM. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007 Sep 30;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]
62. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005 Mar 31;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
63. Wu G, Sanderson B, Bittner V. The 6-minute walk test: how important is the learning effect? Am Heart J 2003 Jul;146(1):129-133. [doi: 10.1016/S0002-8703(03)00119-4] [Medline: 12851620]
64. Innerd P, Harrison R, Coulson M. Using open source accelerometer analysis to assess physical activity and sedentary behaviour in overweight and obese adults. BMC Public Health 2018 Apr 23;18(1):543 [FREE Full text] [doi: 10.1186/s12889-018-5215-1] [Medline: 29685121]
65. Joseph RP, Pekmezi D, Dutton GR, Cherrington AL, Kim Y, Allison JJ, et al. Results of a culturally adapted internet-enhanced physical activity pilot intervention for overweight and obese young Adult African American women. J Transcult Nurs 2016 Mar;27(2):136-146 [FREE Full text] [doi: 10.1177/1043659614539176] [Medline: 24934566]
66. Bhatt DL, Steg PG, Ohman EM, Hirsch AT, Ikeda Y, Mas J, REACH Registry Investigators. International prevalence, recognition, and treatment of cardiovascular risk factors in outpatients with atherothrombosis. JAMA 2006 Jan 11;295(2):180-189. [doi: 10.1001/jama.295.2.180] [Medline: 16403930]
67. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
68. Millard RW, Fintak PA. Use of the internet by patients with chronic illness. Disease Management and Health Outcomes 2002;10(3):187-194. [doi: 10.2165/00115677-200210030-00006]
69. United States Census Bureau. Census.gov. 2018 Aug. Computer and internet use in the United States: 2016 URL: https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf [accessed 2020-03-13]
70. International Telecommunication Union. Geneva: International Telecommunication Union; 2017. ICT facts and figures 2017 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [accessed 2020-03-13]
71. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
72. Pietrzak E, Cotea C, Pullman S. Primary and secondary prevention of cardiovascular disease: is there a place for Internet-based interventions? J Cardiopulm Rehabil Prev 2014;34(5):303-317. [doi: 10.1097/HCR.0000000000000063] [Medline: 25079147]
73. Richard E, Jongstra S, Soininen H, Brayne C, Moll van Charante EP, Meiller Y, et al. Healthy ageing through internet counselling in the elderly: the HATICE randomised controlled trial for the prevention of cardiovascular disease and cognitive impairment. BMJ Open 2016 Jun 10;6(6):e010806 [FREE Full text] [doi: 10.1136/bmjopen-2015-010806] [Medline: 27288376]
74. Castelnuovo G, Manzoni GM, Cuzziol P, Cesa GL, Tuzzi C, Villa V, et al. TECNOB: study design of a randomized controlled trial of a multidisciplinary telecare intervention for obese patients with type-2 diabetes. BMC Public Health 2010 Apr 23;10:204 [FREE Full text] [doi: 10.1186/1471-2458-10-204] [Medline: 20416042]
75. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [doi: 10.1097/01.mlr.0000114908.90348.f9] [Medline: 15076819]
76. Falagas ME, Zarkadoulia EA, Pliatsika PA, Panos G. Socioeconomic status (SES) as a determinant of adherence to treatment in HIV infected patients: a systematic review of the literature. Retrovirology 2008 Feb 01;5:13 [FREE Full text] [doi: 10.1186/1742-4690-5-13] [Medline: 18241330]
77. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

Abbreviations

ANCOVA: analysis of covariance
BFM: body fat mass
BP: blood pressure
D: diastolic
eHealth: electronic health
HTN: hypertension
IBI: internet-based intervention

MD: mean difference
S: systolic
WLC: wait-list control
6MWT: 6-minute walk test.

Edited by G Eysenbach; submitted 30.03.19; peer-reviewed by S Liu, SG Thomas, F Xhafa; comments to author 03.10.19; revised version received 31.10.19; accepted 09.02.20; published 14.04.20.

Please cite as:
Lisón JF, Palomar G, Mensorio MS, Baños RM, Cebolla-Martí A, Botella C, Benavent-Caballer V, Rodilla E
Impact of a Web-Based Exercise and Nutritional Education Intervention in Patients Who Are Obese With Hypertension: Randomized Wait-List Controlled Trial
J Med Internet Res 2020;22(4):e14196
URL: http://www.jmir.org/2020/4/e14196/
doi:10.2196/14196
PMID:32286232

Original Paper

Usage Metrics of Web-Based Interventions Evaluated in Randomized Controlled Trials: Systematic Review

Elena Koneska[1], BSc; Duncan Appelbe[1,2], BSc, MSc, PhD; Paula R Williamson[1], BSc, PhD; Susanna Dodd[1], BSc, MSc, PhD

[1]Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom

[2]Oxford Trauma, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Susanna Dodd, BSc, MSc, PhD
Department of Biostatistics
University of Liverpool
Block F Waterhouse Building
1-5 Brownlow Street
Liverpool, L69 3GL
United Kingdom
Phone: 44 1517949752
Email: s.r.dodd@liverpool.ac.uk

Abstract

Background: The evaluation of web-based interventions (defined as an intervention that can be downloaded or accessed on the internet through a web browser) in randomized controlled trials (RCTs) has increased over the past two decades. Little is known about how participants' use of the intervention is measured, reported, and analyzed in these studies.

Objective: This study aimed to review the evaluation of web-based interventions in RCTs, assessing study characteristics and the methods used to record, and adjust for, intervention usage.

Methods: A systematic review of the literature was undertaken to identify all published reports of RCTs that involved a web-based intervention. A random sample of 100 published trials was selected for detailed data extraction. Information on trial characteristics was extracted, including whether web usage data were recorded, and if so, the methods used to gather these data and whether these data were used to inform efficacy analyses.

Results: A PubMed search identified 812 trials of web-based interventions published up to the end of 2017 and demonstrated a growing trend over time. Of the 100 studies reviewed, 90 studies collected web usage data, but more than half (49/90, 54%) of these studies did not state the method used for recording web usage. Only four studies attempted to check on the reliability of their web usage data collection methods. A total of 39% (35/90) studies reported patterns or levels of web intervention use, of which 21% (19/90) studies adjusted for intervention use in their outcome analysis, but only two of these used appropriate statistical methods.

Conclusions: Trialists frequently report a measure of web-based intervention usage but do not always report the collection method or provide enough detail on their analysis of web usage. Appropriate statistical methods to account for intervention use are rarely used and are not well reported even in the very few trials in which they are used. The number of trialists who attempt to check on the reliability of their web usage collection methods is extremely low.

(J Med Internet Res 2020;22(4):e15474) doi:10.2196/15474

KEYWORDS

internet; web-based interventions; randomized controlled trial; web usage data; systematic review

Introduction

Randomized Controlled Trials

A randomized controlled trial (RCT) is used to assess the efficacy or effectiveness of an intervention by randomly dividing trial participants into experimental or control treatment arms, thereby providing a fair comparison for the unbiased assessment of treatment effects [1-4]. Traditionally, trials have predominantly been conducted in a clinic setting; however, with the increase of the internet as a mainstream communication

channel, there has been an increase in the use of email, SMS, and social media for the communication and delivery of interventions [5,6].

Web-Based Interventions

We defined a web-based or a web intervention as "downloadable or accessible via the internet through a web browser," which can take the form of (but not limited to) a website, an email, or a web message board. There are various definitions of web-based interventions, some of which include social media and mobile phone apps; however, for the purposes of our review (in particular, our interest in assessing web usage data), we were interested in confining our search to studies that would have been able to assess usage, which until recently was not easy with social media or phone apps. As such, we restricted our definition of web-based interventions accordingly; however, our chosen definition is very similar to that provided by Barak [7].

With an estimated 4.4 billion people being active internet users as of April 2019 [8], an increasing proportion of the global population are potential users of web-based interventions, particularly given the convenience and flexibility of such interventions. As such, these interventions have enormous potential to improve health and health care delivery and can be easily accessible to patients [1,9-11].

Monitoring Web Usage

In the same way that drug treatments may be prescribed at a certain dose, trial participants receiving a web-based intervention may be advised to use the intervention to a specified degree (eg, in terms of duration or frequency of intervention use). If it is of interest to determine whether trial participants adhered to the recommended intervention dose, it is important to be able to track participants' intervention use. There are multiple published reviews relating to web-based intervention usage. For example, Kelders et al [12] reviewed the literature to investigate whether study design predicts adherence to a web-based intervention, whereas Perski et al [13] reviewed the literature on digital behavior change interventions to identify or develop a framework linking direct and indirect influences on engagement and the relationship between engagement and intervention effectiveness.

There are numerous automated tools that can be used to track and record a participant's web intervention use [14]. These tools can be split into two categories, either client (browser) based or server based. Client-based tools, such as Google Analytics (GA) [15], rely on the web browser supporting them (eg, JavaScript being enabled) [16], whereas server-based tools, such as web server log data [17], will always be populated, as they record what data are sent to the client. These tools provide information about participants' web intervention use, such as which web pages a participant has visited and when a web page has been accessed. However, the reliability of these tools is not guaranteed. Some tools that have been adopted by researchers to measure web usage, such as GA, were not originally designed for accurate reporting of web usage but were instead developed as a marketing aid. As such, while being easily accessible and commonly used, GA may not be the most appropriate tool to use in scientific research [18]. For example, prior research by OBrien et al [19] has demonstrated that 58% of activity on a website is unreported by GA.

To link intervention usage to a particular participant, rather than just obtaining general information about overall intervention use by all participants, each participant requires a unique identifier (UID), such as the study randomization number or a username [20]. The use of a UID facilitates statistical analyses by linking intervention use with outcome data on an individual participant basis. Such data can then be used to inform statistical analysis to estimate the efficacy of the intervention received, rather than simply estimate the effectiveness of the intervention as randomized (as estimated by an intention-to-treat analysis). Commonly used methods to estimate efficacy, using participants' usage of the assigned intervention, include an as-treated, per-protocol analysis and completer analyses [21]. However, the use of these methods when a trial is subject to deviations from randomized treatment may introduce bias, and more appropriate causal methods should be used, such as complier average causal effect (CACE) analysis [22,23].

Consolidated Standards of Reporting Trials and Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth Guidelines

The Consolidated Standards of Reporting Trials (CONSORT) [24] guidelines were introduced in 1996 to improve the consistency and quality of reporting in RCTs. To address the specific challenges of web-based and mobile app–based intervention studies, the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online TeleHealth (CONSORT-EHEALTH) extension was published in 2011 [25]. This extension encourages trialists to report on participants' intervention use; subitem 6a-ii of the CONSORT-EHEALTH extension states that researchers should "explain how use and engagement was measured and defined" and subitem 17-I states that "use and usage outcomes should be reported". The intended benefit of these guidelines will, however, only be realized if they are adhered to; as such, it is important to assess their uptake in trials that have been published since their release.

Aims and Objectives

This systematic review was conducted to ascertain the extent and nature of web-based intervention use in trials and the current practice among trialists in terms of collecting, reporting, and analyzing web usage data. We were also interested in determining the characteristics of such trials, including the types of design, intervention formats, and clinical areas.

Methods

Literature Search

An initial systematic search of PubMed was conducted to ascertain whether there had already been any comprehensive systematic reviews of web-based intervention trials published to date (see Multimedia Appendix 1 for search terms) [12,25] The electronic database, PubMed [26], was then searched to

identify all web-based intervention trials published by the end of 2017 (see Textbox 1 for search terms). The protocol for this review has been published in the International Prospective Register of Systematic Reviews [27].

Textbox 1. Search terms for published Web-based intervention trials.

(online[tiab] OR digital[tiab] OR web-based OR web) AND internet[majr] AND ("Randomized Controlled Trial" [Publication Type] OR randomized control trial OR randomised control trial OR controlled trial OR controlled clinical trial OR RCT) (PLUS manual entry of upper limit of 31/12/2017 for date published)

Eligibility Screening

Following the removal of duplicate records, all remaining abstracts identified through the PubMed search were screened by an author (EK) to assess eligibility. Only RCTs involving a web-based intervention and published by the end of 2017 were eligible. Studies were excluded if they did not involve a web-based health intervention (eg, educational studies) or were nonrandomized (eg, feasibility studies that did not involve randomization, observational studies, quasi-randomized studies, and surveys), secondary analyses, trial protocols, or systematic reviews. Where there was any uncertainty regarding eligibility, authors DA and SD were consulted, and any disagreements were resolved by consensus. Five percent (77/1540) of the abstracts were randomly selected and assessed for eligibility by authors DA and SD to validate this process, on which there was 100% agreement.

Data Extraction

A total of 100 studies were randomly selected from the cohort of eligible trials identified in this search, with sampling proportional to the annual distribution of publication years across the entire set of eligible studies. The initial data extraction form was piloted on five studies and refined accordingly. The final dataset included the study characteristics, whether a CONSORT flow diagram and CONSORT-EHEALTH checklist were reported, whether treatment protocol deviations (ie, changes to randomized web-based interventions) were reported, the methods used to collect web usage data, and which statistical analysis methods were used to adjust for intervention use.

Results

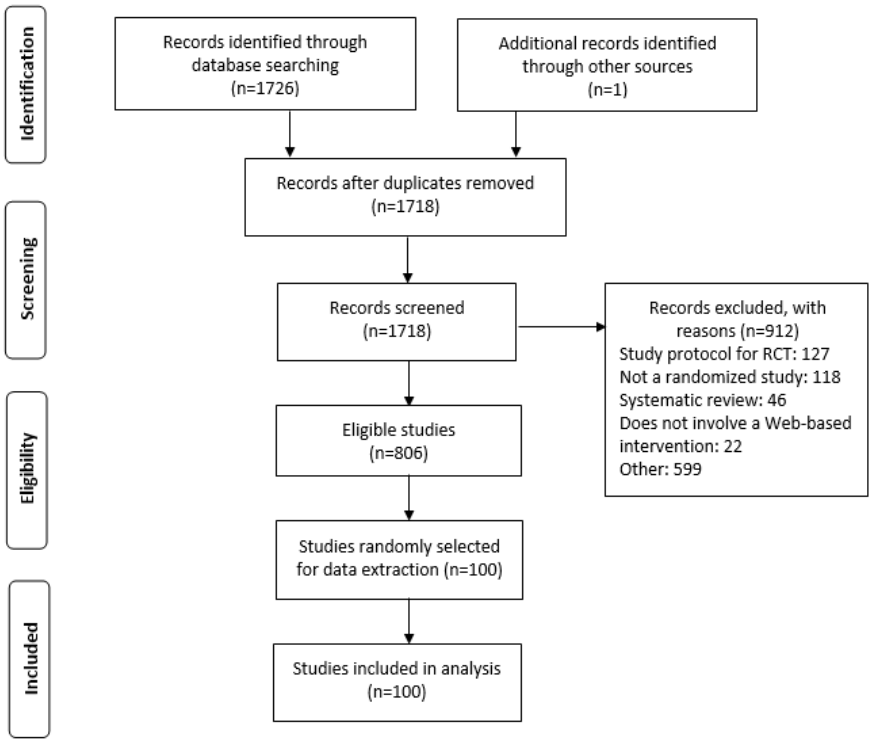
Review of Systematic Reviews of Web-Based Intervention Trials

The PubMed search for systematic reviews of web-based intervention trials identified 271 citations, 123 of which were found to be eligible following a review of titles and abstracts. These systematic reviews covered a wide range of clinical or methodological areas, most commonly health promotion (47/123, 38.2%) and mental health (40/123, 32.5%; see Multimedia Appendix 2). None of these systematic reviews included a comprehensive search of all published web-based health intervention trials.

Review of Web-Based Intervention Trials

The electronic database search for trials of web-based interventions yielded 1726 publications (Figure 1). After removing nine duplicates, there were 812 eligible and 906 ineligible studies based on the review of abstracts, including one publication identified manually as the original trial report relating to another publication identified in the search. Of the 100 eligible studies selected for data extraction, six were subsequently excluded after reading the full publication. These ineligible studies were replaced with an additional six eligible studies for data extraction.

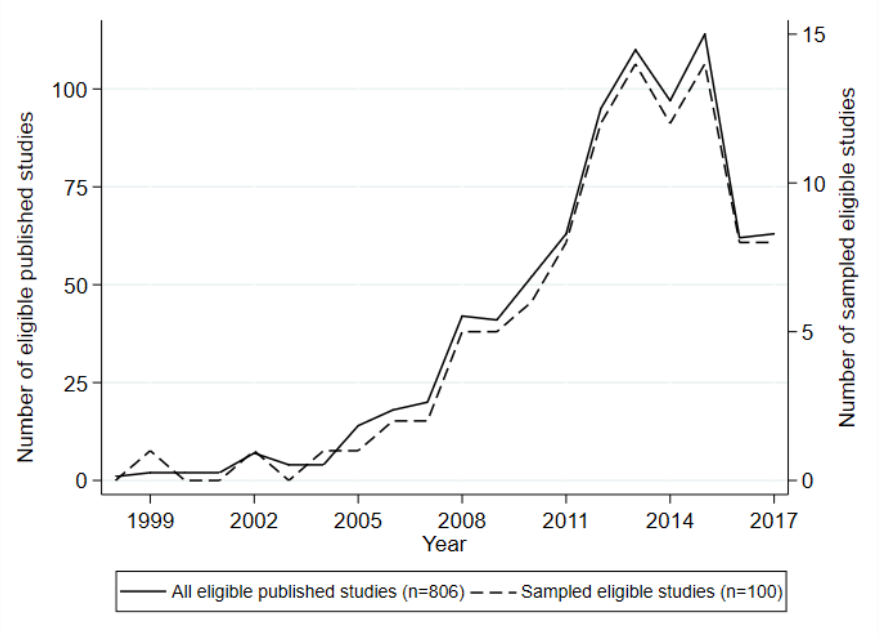
Figure 1. Search and screening process. RCT: randomized controlled trial.



Published Web-Based Intervention Trials

The number of published trials involving web-based interventions is displayed in Figure 2, demonstrating an increasing trend over time. However, despite this increase, the number of trials using web-based interventions remains proportionally low when compared with the total number of trials during this period (estimated as 496,238 from a PubMed search filtered to only include trials published up to the end of 2017). The reduction seen after 2015 is likely to be due to publications not being fully indexed or registered within the PubMed database when the search was run (Feb 12, 2018). A PubMed librarian confirmed that new publications may be posted on PubMed significantly later than their publication date.

Figure 2. Number of published and sampled trials of online intervention trials each year.



A Description of Studies and Study Characteristics

The characteristics of the 100 publications randomly selected for data extraction are given in Table 1. Most of these studies covered health promotion (42/100, 42.0%; most commonly smoking cessation, physical activity, and weight) and mental health (32/100, 32.0%).

Table 1. Characteristics of sampled trials (N=100).

Clinical area	Values
Health promotion, n	42
Smoking cessation	11
Physical activity	8
Weight	7
Alcohol	3
Eating disorder	3
Lifestyle behaviors	2
Physical activity and diet	2
Diet	2
Sexual health	1
Tanning	1
Adolescent health	1
General health management	1
Mental health, n	32
Cancer, n	4
Respiratory illnesses, n	3
Neurology, n	3
Diabetes, n	3
Dentistry, n	2
Otolaryngology, n	2
Cardiovascular, n	2
Pain, n	1
Autonomic arousal, n	1
Discharge from emergency department, n	1
Parathyroid disorder, n	1
HIV, n	1
Cancer screening, n	1
Women’s health, n	1
Design, n	
Superiority	94
Equivalence	4
Noninferiority	2
Blinding, n	
Double	1
Single	13
None	46
Not stated	40
Web-based intervention, n	
Website	77
Website plus additional element	10
Internet (other)	13
Control arm, n	

Clinical area	Values
Website	14
Internet (other)	14
Waiting list group	32
Noninternet intervention	28
No intervention	9
Not stated	3
CONSORT[a] flow diagram presented[b], n (%)	79 (79.0)
CONSORT-EHEALTH[c] checklist presented[d], n (%)	26 (38.2)

[a]CONSORT: Consolidated Standards of Reporting Trials.

[b]The denominator for percentage values is equal to 100, as all trials were published after the original CONSORT flow diagram (1996).

[c]CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online TeleHealth.

[d]The denominator for percentage values is equal to the number of trials published since the formulation of the CONSORT-EHEALTH (2011; N=68).

The vast majority of trials had a superiority design and did not use blinding or did not state whether there was any blinding. A total of 13 studies reported being single blinded (six reported blinding of the assessors, six reported blinding of the patients, and one reported blinding of the clinician), and only one study reported being double blinded (patients and assessors). In the 86 trials that stated that there was no blinding or did not mention blinding, the web-based and control interventions took different formats—most commonly a website intervention vs a waitlist (n=25) or noninternet (n=18) intervention—which would have made it difficult to blind participants.

The majority of studies involved a website as the intervention; other interventions included a podcast, emails, web applications, a web-based video camera, computer simulation, a computer-generated photoaging intervention, web message boards, an internet partner, a YouTube video, a web-based video, and an internet video conference. In total, 10 studies reported a website plus an additional element, which took the form of a mobile app, a web-based video, social media, an interactive voice response, a personal activity monitor, a personal digital assistant, or an online forum. The most common type of a control arm intervention was waiting list (delayed treatment) followed by noninternet interventions (a face-to-face intervention, written materials, and treatment as usual). Table 2 displays the cross-tabulation of web-based and control interventions in the 100 sampled trials.

Table 2. Web-based and control interventions.

Control intervention	Web-based intervention			Total, n
	Website, n	Website plus additional element, n	Internet (other), n	
Website	12	2	0	14
Internet (other)	10	0	4	14
Waiting list group	26	2	4	32
Noninternet intervention	20	4	4	28
No intervention	6	2	1	9
Not stated	3	0	0	3
Total	77	10	13	100

Of the 100 studies, 79 included a CONSORT flow diagram, whereas 38% (26/68 studies published after the CONSORT-EHEALTH guideline, 2011) of studies included a CONSORT-EHEALTH checklist (Table 1).

The publication of CONSORT-EHEALTH does not appear to have positively influenced the rate of reporting web usage (Table 3).

Of 26 trial publications that included a CONSORT-EHEALTH checklist, four did not report whether web usage data were collected. There were different reasons for not reporting usage in these four publications: one trial acknowledged collecting usage data with the intention to publish usage in a separate publication, one trial did not collect usage because of privacy protection (with no further explanation), one trial gave no explanation on why usage was not collected, and it was not possible to access the CONSORT-EHEALTH checklist in the fourth trial (because of an expired or invalid checklist hyperlink).

Table 3. Rates of reporting web usage data according to the publication year and Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online TeleHealth checklist reporting.

Publication year	Reported web usage data		Total
	Yes	No	
≤2011, n (%)	30 (94)	2 (6)	32 (100)
>2011, n (%)			
Included Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online TeleHealth checklist			
Yes	22 (85)	4 (15)	26 (100)
No	38 (91)	4 (10)	42 (100)
Total, n	90	10	100

Collection and Reporting of Web Usage Data

Commonly used formats for the web-based intervention included sessions (n=17 trials), modules (n=13), content (n=13), and assignments (n=5). Other formats included cartoons, messages, videos, photographs, and various tasks or exercises. Examples of these interventions included a brief personalized normative feedback system provided by various modes of delivery [28], identical content delivered as a podcast or via a website [29], and website information to encourage and support a personalized physical activity plan [30]. One trial [31] used a computer-generated photoaging intervention, with which participants were digitally photoaged and received a photograph of themselves as a lifelong smoker and as a nonsmoker. Exercises took the form of mindfulness exercises as a part of module completion [32] and a series of abdominal plank exercises while exercising with an internet partner [33].

Web usage data were collected in 90 of the studies, but more than half (49/90, 54%) of these studies did not state the method used for recording web usage. The most commonly reported tool used for tracking web usage was a server or electronic log files (see Table 4). Other methods included software tools, website tracking data, GA, and self-reported data. Only 4% (4/90) trial reports mentioned checking the reliability of their web usage measurement methods, two of which used more than two tools to capture and compare web usage data.

Table 4. Web usage data collection methods among 90 trials which collected web usage data.

Method and second method (if applicable)	Frequency, n (%[a])
Logs	10 (11)
Software tools	6 (7)
Website tracking	4 (4)
Google Analytics	5 (6)
Alone	3 (3)
With Logs	2 (2)
Self-reporting	5 (6)
Alone	3 (3)
With Logs	1 (1)
With tracking data	1 (1)
Others	11 (12)
Not stated	49 (54)

[a]% of 90 trials which reported web-based intervention use.

Among the 87 trials involving a website, 78 (90%) recorded web usage data, most commonly in terms of the number of log-ins (37/87, 43%), the number of individual intervention components completed (21/87, 24%; eg, assignments, exercises, lessons, and modules), measures of activity on the site (eg, answers entered, activated hyperlinks, blogs, or forum posts; 18/87, 21%), and time spent on the site (18/87, 21%; see Table 5). A total of 36% (31/87) of these trials recorded a combination of two or more usage measures, most commonly the number of log-ins and time spent on the site (15 trials). Among the 23 trials involving a web-based intervention other than a website, 20 (87%) recorded web usage data, most commonly in terms of the number of log-ins (6/23, 26%), video views (6/23, 26%), and measures of activity (5/23, 21%). A total of 26% (6/23) of these trials recorded more than one usage measure (see Table 6).

Table 5. Features of web usage recorded among trials that involved a website (N=87).

Web usage recorded among trials that involved a website	Trials[a], n (%)
No web usage data collected	9 (10)
Activity on site (eg, answers, activated hyperlinks, and blog or forum posts)	18 (21)
Communication (eg, emails, Skype calls, call logs, and messages sent)	3 (4)
Completed intervention (eg, all assignments, exercises, lessons, or modules)	3 (4)
Number of individual intervention components (eg, modules, sessions) started/accessed	3 (4)
Number of individual intervention components (eg, modules, sessions) completed	21 (24)
Number of log-ins	37 (43)
Number of page hits (individual actions, eg, audio clips, scrolling, and printing)	1 (1)
Number of page views	14 (16)
Time spent on site (including time spent listening to podcast)	18 (21)
Video views (including YouTube views)	1 (1)

[a]Note that 24 trials included two measures of web usage, four trials included three measures of web usage, and three trials included four measures of web usage.

Table 6. Features of web usage recorded among trials that involved a web-based intervention other than a website (N=23).

Web usage recorded among trials that involved a website	Trials[a], n (%)
No web usage data collected	9 (10)
Activity on site (eg, answers, activated hyperlinks, and blog or forum posts)	18 (21)
Communication (eg, emails, Skype calls, call logs, and messages sent)	3 (4)
Completed intervention (eg, all assignments, exercises, lessons, or modules)	3 (4)
Number of individual intervention components (eg, modules, sessions) started/accessed	3 (4)
Number of individual intervention components (eg, modules, sessions) completed	21 (24)
Number of log-ins	37 (43)
Number of page hits (individual actions, eg, audio clips, scrolling, and printing)	1 (1)
Number of page views	14 (16)
Time spent on site (including time spent listening to podcast)	18 (21)
Video views (including YouTube views)	1 (1)

[a]Note that five trials included two measures of web usage, and one trial included three measures of web usage.

A total of 44% (40/90) of trials that collected web usage reported using UIDs, most commonly log-in credentials or internet protocol addresses (see Table 7). An additional 12% (11/90) of publications reported the use of a server or electronic logs to record web usage, both of which have the potential to include UIDs. A total of 8% (7/90) of trials implied having UIDs but did not state what type of UID was used.

Table 7. Unique identifiers (N=100).

Unique identifiers	Values, n
Total web usage collected	90
Unique identifier	40
Potential unique identifier (server/electronic logs)	11
Implied unique identifier but not specified	7
No unique identifier	3
Not stated	29

Trialists reported changes to randomized web-based interventions (treatment protocol deviations) in 33 of the studies. Departures from randomized treatment included failing to initiate treatment (in 15 trials, eg, when participants did not

activate the account, access the site, or log in); premature discontinuation of the intervention (in 18 trials, eg, when participants withdrew from the trial or experienced difficulties using the site); switching to an alternative arm, which was reported in two trials; and switching to non-web-based treatment, reported in two trials.

Intervention Dose

A total of 69 trials from our sample specified a recommended dose of the web-based intervention, 62 (90%) of which measured web usage. The dose was specified in terms of sessions, modules, or assignments in 49% (34/69) of these studies (mean 2.8, SD 2.3; range 1-14). Of the 23 studies that reported a time frame for the use of the web-based intervention, the duration ranged from 1 to 12 weeks (mean 2.2, SD 1.3), with the exception of one study, which reported a duration of 150 days (5 months). The average dose frequency was one task per week in 36% (25/69) of studies that recommended a dose. A total of 9% (6/69) of studies reported that participants had more than one task to complete per week, and 10% (7/69) studies reported that participants were due to complete tasks less frequently than 1 per week.

Analyses Involving web Usage Data

Only 39% (35/90) of trials that collected web usage data investigated the levels of intervention use (Table 8). A total of 21% (19/90) of studies used statistical methods to adjust for intervention usage, such as a completer analysis (11 trials), regression analyses with intervention use as a covariate (six trials), and a CACE analysis (two trials). One of the two trials that used a CACE analysis did not present results or explain their method further, whereas the other trial presented CACE results and explained that the analysis estimates the potential efficacy among participants who would comply with their randomized intervention.

Table 8. Analyses involving web intervention use (N=100).

Analyses involving web intervention use	Values, n
Any analysis involving web intervention use	35
Comparison of web intervention use between randomized arms	3
Assessed patterns of web intervention use	4
Correlation between web intervention use and outcome	9
Completer analysis	11
Regression analyses with web intervention use as a covariate	6
Causal analysis (complier average causal effect)	2

Discussion

Characteristics of Web-Based Intervention Trials and Systematic Reviews

Although the use of web-based interventions in RCTs has been on the rise over the last 15 years, unsurprisingly, the number is still low in comparison with the overall number of published trials. A random sample of 100 trials suggests that web-based interventions are most commonly used for health promotion (42/100, 42.0%) or mental health issues (32/100, 32.0%), with the remaining 26.0% (26/100) of trials covering 14 clinical areas, including cancer (4/100, 4.0%), diabetes (3/100, 3.0%), and neurology (3/100, 3.0%). The review of systematic reviews of web-based intervention studies demonstrated a similar pattern, with 38.2% (47/123) of reviews relating to health promotion interventions and 32.5% (40/123) relating to mental health. All systematic reviews identified were restricted to trials within a certain clinical condition, other than the review by Mathieu et al [5], which only included trials that were fully or primarily conducted online (eg, involving web-based recruitment, consent, randomization, and follow-up), whereas Lustria et al [34] reviewed trials that defined *electronic health*. As such, this study of systematic reviews demonstrated that there were no previously published reviews of all web-based intervention studies, providing evidence of the novelty and usefulness of this study.

Adherence to CONSORT and CONSORT-EHEALTH Guidelines

Good quality reporting allows clinicians and researchers to replicate trial methods [35-37] and supports the understanding of trial methods, interventions, and outcomes. This study suggests that there is a need for greater adherence to reporting guidelines in publications of web-based intervention trials. Less than 80% of the trials in our sample presented CONSORT flow diagrams, which is considerably less than the 96% reported to have presented CONSORT flow diagrams in a sample of 100 trials published in 2008 [21]. This may be because of the fact that CONSORT is less commonly endorsed by health informatics journals than clinical journals or is less familiar to trialists assessing web-based interventions than clinical trialists, generally.

Furthermore, although the CONSORT-EHEALTH guideline is listed on the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) website [38] and has been adopted by the Journal of Medical Internet Research, less than 40% of the studies published since CONSORT-EHEALTH was published including a CONSORT-EHEALTH checklist; the authors may, therefore, want to consider some of the strategies suggested by the EQUATOR network to increase the use of guidelines [39], such as further dissemination via journal editorials or conference presentations, the provision of web-based training, or publicity via social media or blog posts. Improving awareness and uptake of the CONSORT-EHEALTH guidelines is important to ensure

that the methodological quality of web-based intervention trials is clearly communicated, thereby allowing readers to make informed judgments on the validity of inferences and conclusions drawn in such trials.

Reporting and Analysis of Web Usage Data

The CONSORT-EHEALTH guideline recommends reporting data collection methods and results relating to intervention use, but not all studies that included a CONSORT-EHEALTH checklist reported information on the collection of web usage data. Indeed, the publication of CONSORT-EHEALTH does not seem to have influenced the quality of reporting regarding web usage, as the rate of reporting web usage data was higher before the publication of CONSORT-EHEALTH.

Unlike drug interventions, the adherence to which can be summarized using uncomplicated measures of treatment intake (eg, initiation, completion, and persistence [21]), web-based interventions often involve multiple features [40,41], engagement with which may be more complex to record. For example, it may be of interest to determine typical navigation patterns through a website, which precise areas of a web page are read or whether videos are watched in their entirety, none of which would be trivial to capture. Our review demonstrated that trialists collect data on a wide variety of web usage features, most commonly the number of log-ins, the number of intervention components completed, activity, and time spent on the site. One-third of the trials that recorded web usage information collected web usage data on more than one feature, the most common combination being the number of log-ins and time spent on the site. The likelihood of measuring web usage data did not vary according to whether or not participants were recommended to follow a specific dose (eg, when participants were asked to use the web-based intervention for a specific period or to complete a certain number of modules): the proportion of trials that measured web usage was equal to 90% in those trials that did (62/69), and in those that did not (28/31), specify a recommended dose. This suggests that the high rate of measuring web usage in web-based intervention trials is not necessarily because of the trialists' interest in assessing participants' adherence to a recommended intervention dose; instead, web usage data are commonly recorded regardless of whether there is a recommended dose, demonstrating that such data appear to be of interest to trialists in their own right.

Trialists rarely provided a rationale for their choice of web usage metrics or analysis methods to adjust for web usage. Only two of the 15 trials that adjusted their outcomes for intervention use applied an appropriate method of causal analysis (CACE) to estimate efficacy, suggesting a lack of awareness regarding appropriate methods to account for the impact of participants' intervention use on their outcomes.

Assessing the Reliability of Web Usage Data

Although automated capture of participants' use of web-based interventions may be assumed to be more straightforward and reliable than the usual measures used to capture drug treatment intake (which typically involve participant self-reporting, such as pill counts and treatment diaries, and, therefore, are potentially subject to recall bias or distortion), this is not necessarily the case. Assessing the reliability of web usage data collection methods is, therefore, vital, but very few trialists in our sample mentioned checking the reliability of their web usage measurement methods. When trialists do not check the reliability of their web usage data collection methods, there is a potential for their web usage data (and any subsequent inferences based on these data) to be biased, particularly when inherent features of web usage differ between the randomized interventions. van Rosmalen-Nooijens et al [42] compared the results from GA, content management system logs, and data files with self-reported data from participants and concluded that the usage information from the different sources corresponded well. Nguyen et al [43] and Mermelstein et al [44] also aimed to assess the reliability of their methods, but both studies reported a lack of reliability of their data because of technical or logistical issues. Similar to drug trials, participants' self-reported web usage may also misrepresent the true use of the intervention [45]. For example, Fleisher et al [46] found discrepancies between self-reported data and usage data obtained from the NetTracker software tool. Fleisher [46] reported that nearly 40% of the participants who reported using the website actually did not log in, whereas 20% of those who reported they did not use the website did, in fact, log in. We are currently undertaking work to determine the reliability of different web usage collection methods, given the uncertainty regarding the accuracy of certain methods.

Strengths and Limitations

This review was not designed to identify trials that used mobile phone apps or social media interventions. This was a conscious decision because our primary aim was to determine the frequency with which trialists monitored web usage.

A large number of eligible studies prohibited data extraction on all eligible trials; as such, it was decided that a random sample of these trials would be selected (using stratified sampling according to the year of publication to ensure that the publication year profile mirrored that of the complete cohort of eligible studies). Although only 100 of the eligible trial publications were, therefore, included in the data extraction exercise, we believe that this is a sufficient number to give reliable estimates (eg, ensuring the estimation of proportions up to a maximum standard error of 0.05) and an accurate indication of trends in reporting and analysis.

The process of determining the eligibility of web-based intervention trials was based on the review of abstracts only, as such some of the studies deemed as eligible may not have been, as evidenced by the exclusion of six studies from the sample of 100 studies. In addition, only 1 reviewer carried out data extraction; however, this reviewer was able to consult the opinion of a second reviewer if in any doubt so as to appropriate classifications.

This review is limited by the search of only one web-based publication database, PubMed. The number of web-based interventions in 2016/2017 will be underestimated from this search because of delays in registration and indexing of studies within PubMed. PubMed indexes the majority of, but not all, health informatics journals; there are currently 286 health informatics journals, of which 196 are indexed in PubMed.

Therefore, a total of 806 trials cannot be taken as the absolute number of web-based intervention trials published up to the end of 2017.

Conclusions

There is an increasing trend in the use of web-based interventions in RCTs. Tracking web usage data in such trials is necessary to establish the efficacy of web-based interventions. When an intervention is found to be less effective than desired, without usage data, it is hard to determine if the problem is because of the intervention content or the lack of use of the intervention [46]. Information on participants' intervention use should, therefore, be reported within trial publications with particular focus on relevant features of participation, which are likely to have an impact on outcomes. Although the majority of studies reviewed here reported a measure of web-based intervention usage, trialists often did not report sufficient detail about how the data were collected and rarely considered the accuracy of their web usage data collection methods. There was a modest degree of interest in investigating patterns of web usage, but very few trialists used an appropriate method of analysis to account for the impact of intervention use on participant health outcomes.

Authors' Contributions

EK developed the protocol, carried out the search and data extraction, and drafted the manuscript. SD conceived the initial idea, helped to develop the protocol, acted as a second opinion on data extracted, and commented on drafts of the manuscript. DA helped to develop the protocol, acted as a second opinion on data extracted, and commented on drafts of the manuscript. PW helped to develop the protocol and commented on drafts of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for published systematic reviews of web-based intervention trials.
[DOCX File , 14 KB - jmir_v22i4e15474_app1.docx]

Multimedia Appendix 2
Clinical or methodological areas covered by systematic reviews identified in search.
[DOCX File , 14 KB - jmir_v22i4e15474_app2.docx]

References

1. Paul J, Seib R, Prescott T. The internet and clinical trials: background, online resources, examples and issues. J Med Internet Res 2005 Mar 16;7(1):e5 [FREE Full text] [doi: 10.2196/jmir.7.1.e5] [Medline: 15829477]
2. Stanley K. Design of randomized controlled trials. Circulation 2007;115(9):1164-1169. [doi: 10.1161/circulationaha.105.594945]
3. Kendall JM. Designing a research project: randomised controlled trials and their principles. Emerg Med J 2003 Mar;20(2):164-168 [FREE Full text] [doi: 10.1136/emj.20.2.164] [Medline: 12642531]
4. Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M. Randomized controlled trials: part 17 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2011 Sep;108(39):663-668 [FREE Full text] [doi: 10.3238/arztebl.2011.0663] [Medline: 22013494]
5. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc 2013 May 1;20(3):568-576 [FREE Full text] [doi: 10.1136/amiajnl-2012-001175] [Medline: 23065196]
6. Pugatch J, Grenen E, Surla S, Schwarz M, Cole-Lewis H. Information architecture of web-based interventions to improve health outcomes: systematic review. J Med Internet Res 2018 Mar 21;20(3):e97 [FREE Full text] [doi: 10.2196/jmir.7867] [Medline: 29563076]
7. 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]
8. Statista. Global Digital Population as of January 2020 (in Billions) URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2019-12-27]
9. Wasilewski MB, Stinson JN, Cameron JI. Web-based health interventions for family caregivers of elderly individuals: a scoping review. Int J Med Inform 2017 Jul;103:109-138. [doi: 10.1016/j.ijmedinf.2017.04.009] [Medline: 28550996]
10. 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]
11. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]

12. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
13. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
14. Payne P, Embi P, Johnson S, Mendonca E, Starren J. Improving clinical trial participant tracking tools using knowledge-anchored design methodologies. Appl Clin Inform 2010;1(2):177-196 [FREE Full text] [doi: 10.4338/ACI-2010-02-RA-0012] [Medline: 22132037]
15. Google Analytics. URL: https://analytics.google.com/analytics/web/ [accessed 2019-12-27]
16. Thushara Y, Ramesh V. A study of web mining application on e-commerce using Google Analytics tool. Int J Comput Appl 2016 Sep;149(11):21-26. [doi: 10.5120/ijca2016911610]
17. The Apache HTTP Server Project. Log Files URL: https://httpd.apache.org/docs/1.3/logs.html [accessed 2019-12-27]
18. Clark D, Nicholas D, Jamali HR. Evaluating information seeking and use in the changing virtual world: the emerging role of Google Analytics. Learn Publ 2014;27(3):185-194. [doi: 10.1087/20140304]
19. OBrien P, Arlitsch K, Mixter J, Wheeler J, Sterman LB. RAMP – the Repository Analytics and Metrics Portal: A prototype web service that accurately counts item downloads from institutional repositories. Library Hi Tech 2017;35(1):144-158. [doi: 10.1108/lht-11-2016-0122]
20. Olden M, Holle R, Heid IM, Stark K. IDGenerator: unique identifier generator for epidemiologic or clinical studies. BMC Med Res Methodol 2016 Sep 15;16:120 [FREE Full text] [doi: 10.1186/s12874-016-0222-3] [Medline: 27628043]
21. Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. Trials 2012 Jun 18;13:84 [FREE Full text] [doi: 10.1186/1745-6215-13-84] [Medline: 22709676]
22. Dodd S, White IR, Williamson P. A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. Trials 2017 Oct 25;18(1):498 [FREE Full text] [doi: 10.1186/s13063-017-2240-9] [Medline: 29070048]
23. Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. Clin Trials 2007;4(1):58-73. [doi: 10.1177/1740774506075549] [Medline: 17327246]
24. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. J Am Med Assoc 1996 Aug 28;276(8):637-639. [doi: 10.1001/jama.276.8.637] [Medline: 8773637]
25. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
26. NCBI - NIH. PubMed URL: https://www.ncbi.nlm.nih.gov/pubmed [accessed 2019-12-27]
27. Koneska E, Appelbe D, Dodd S. Prospero. 2018. Determining the Extent of Web-Based Intervention Use in Health Research: A Systematic Review URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018095116 [accessed 2020-02-10]
28. Andersson C. Comparison of WEB and interactive voice response (IVR) methods for delivering brief alcohol interventions to hazardous-drinking university students: a randomized controlled trial. Eur Addict Res 2015;21(5):240-252. [doi: 10.1159/000381017] [Medline: 25967070]
29. Turner-McGrievy G, Kalyanaraman S, Campbell MK. Delivering health information via podcast or web: media effects on psychosocial and physiological responses. Health Commun 2013;28(2):101-109 [FREE Full text] [doi: 10.1080/10410236.2011.651709] [Medline: 22420785]
30. Irvine AB, Philips L, Seeley J, Wyant S, Duncan S, Moore RW. Get moving: a web site that increases physical activity of sedentary employees. Am J Health Promot 2011;25(3):199-206 [FREE Full text] [doi: 10.4278/ajhp.04121736] [Medline: 21192750]
31. Burford O, Jiwa M, Carter O, Parsons R, Hendrie D. Internet-based photoaging within Australian pharmacies to promote smoking cessation: randomized controlled trial. J Med Internet Res 2013 Mar 26;15(3):e64 [FREE Full text] [doi: 10.2196/jmir.2337] [Medline: 23531984]
32. Boettcher J, Aström V, Påhlsson D, Schenström O, Andersson G, Carlbring P. Internet-based mindfulness treatment for anxiety disorders: a randomized controlled trial. Behav Ther 2014 Mar;45(2):241-253 [FREE Full text] [doi: 10.1016/j.beth.2013.11.003] [Medline: 24491199]
33. Irwin BC, Feltz DL, Kerr NL. Silence is golden: effect of encouragement in motivating the weak link in an online exercise video game. J Med Internet Res 2013 Jun 4;15(6):e104 [FREE Full text] [doi: 10.2196/jmir.2551] [Medline: 23732514]
34. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
35. Schroter S, Glasziou P, Heneghan C. Quality of descriptions of treatments: a review of published randomised controlled trials. BMJ Open 2012;2(6):pii: e001978 [FREE Full text] [doi: 10.1136/bmjopen-2012-001978] [Medline: 23180392]

36. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
37. Butcher NJ, Monsour A, Mew EJ, Szatmari P, Pierro A, Kelly LE, et al. Improving outcome reporting in clinical trial reports and protocols: study protocol for the Instrument for reporting Planned Endpoints in Clinical Trials (InsPECT). Trials 2019 Mar 6;20(1):161 [FREE Full text] [doi: 10.1186/s13063-019-3248-0] [Medline: 30841935]
38. Centre for Statistics in Medicine (CSM). The EQUATOR Network. URL: http://www.equator-network.org/ [accessed 2019-12-27]
39. Centre for Statistics in Medicine. The EQUATOR Network. Disseminating Your Reporting Guideline URL: http://www.equator-network.org/toolkits/developing-a-reporting-guideline/disseminating-your-reporting-guideline/ [accessed 2019-12-27]
40. Kelders SM, Pots WT, Oskam MJ, Bohlmeijer ET, van Gemert-Pijnen JE. Development of a web-based intervention for the indicated prevention of depression. BMC Med Inform Decis Mak 2013 Feb 20;13:26 [FREE Full text] [doi: 10.1186/1472-6947-13-26] [Medline: 23425322]
41. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
42. van Rosmalen-Nooijens K, Lo Fo Wong S, Prins J, Lagro-Janssen T. Young people, adult worries: randomized controlled trial and feasibility study of the internet-based self-support method 'Feel the ViBe' for adolescents and young adults exposed to family violence. J Med Internet Res 2017 Jun 12;19(6):e204 [FREE Full text] [doi: 10.2196/jmir.6004] [Medline: 28606893]
43. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008 Apr 16;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
44. Mermelstein R, Turner L. Web-based support as an adjunct to group-based smoking cessation for adolescents. Nicotine Tob Res 2006 Dec;8(Suppl 1):S69-S76 [FREE Full text] [doi: 10.1080/14622200601039949] [Medline: 17491173]
45. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005 Mar 26;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
46. Fleisher L, Kandadai V, Keenan E, Miller SM, Devarajan K, Ruth KJ, et al. Build it, and will they come? Unexpected findings from a study on a web-based intervention to improve colorectal cancer screening. J Health Commun 2012;17(1):41-53 [FREE Full text] [doi: 10.1080/10810730.2011.571338] [Medline: 22217118]

Abbreviations

CACE: complier average causal effect
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online TeleHealth
EQUATOR: Enhancing the QUAlity and Transparency Of health Research
GA: Google Analytics
RCT: randomized controlled trial
UID: unique identifier

Edited by G Eysenbach; submitted 16.07.19; peer-reviewed by L Le, K Villaume, D Guertler; comments to author 28.09.19; revised version received 24.10.19; accepted 26.01.20; published 16.04.20.

Please cite as:
Koneska E, Appelbe D, Williamson PR, Dodd S
Usage Metrics of Web-Based Interventions Evaluated in Randomized Controlled Trials: Systematic Review
J Med Internet Res 2020;22(4):e15474
URL: http://www.jmir.org/2020/4/e15474/
doi:10.2196/15474
PMID:32297870

Original Paper

Sustainability of mHealth Effects on Cardiometabolic Risk Factors: Five-Year Results of a Randomized Clinical Trial

Antonio Bernabe-Ortiz[1], MD, MPH, PhD; Julia Pauschardt[1], MSc; Francisco Diez-Canseco[1], BA, MPH; J Jaime Miranda[1,2], MD, MSc, PhD, FFPH

[1]CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru

[2]School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

Corresponding Author:
Antonio Bernabe-Ortiz, MD, MPH, PhD
CRONICAS Center of Excellence in Chronic Diseases
Universidad Peruana Cayetano Heredia
Avenida Armendáriz 445
Miraflores
Lima, 15074
Peru
Phone: 51 12416978
Email: antonio.bernabe@upch.pe

Abstract

Background: The long-term effects of mobile health (mHealth) interventions have not been documented, especially in resource-constrained settings.

Objective: This study aimed to assess the effects of a 1-year mHealth intervention on blood pressure levels and body weight in low-resource urban settings in Peru, 4 years after the completion of the original study.

Methods: Four years after the original Grupo de Investigación en Salud Móvil en America Latina (GISMAL) study, we attempted to contact the 212 individuals originally enrolled in the study in Peru. The primary outcomes were systolic and diastolic blood pressure levels and hypertension incidence. Secondary outcome measures were body weight, BMI, and self-reported target behaviors. The study personnel collecting the data were masked to the group assignment. Linear mixed models were used to evaluate the effects of the intervention on primary and secondary outcomes in an intention-to-treat analysis.

Results: Data from 164 (77.4%) of the 212 originally enrolled participants were available and analyzed (80 in the intervention group and 84 in the control group). The intervention did not result in changes in systolic (–2.54 mm Hg, 95% CI –8.23 to 3.15) or diastolic (3.41 mm Hg, 95% CI –0.75 to 7.57) blood pressure compared with the control group. The intervention reduced the risk of developing hypertension, but the result was not significant (risk ratio (RR) 0.76, 95% CI 0.45-1.28). However, those who received the intervention had lower body weight (–5.42 kg, 95% CI –10.4 to –0.48) and BMI (–2.56 kg/m2, 95% CI –4.46 to –0.66). In addition, compared to the control participants, those who received ≥50% of the scheduled calls during the intervention had greater reductions in body weight (–6.23 kg, 95% CI –11.47 to –0.99) and BMI (–2.81 kg/m2, 95% CI –4.77 to –0.85).

Conclusions: An mHealth intervention comprising motivational interview calls and SMS text messaging appears to have effects on health 4 years after intervention completion. Although there were no effects on blood pressure levels, important reductions in body weight and BMI were seen 5 years after randomization. Thus, mHealth appears to be a promising preventive strategy for noncommunicable diseases in resource-constrained settings.

Trial Registration: Clinicaltrials.gov NCT01295216; https://clinicaltrials.gov/ct2/show/NCT01295216

(J Med Internet Res 2020;22(4):e14595) doi:10.2196/14595

KEYWORDS

mHealth; low- and middle-income countries; blood pressure; body weight

Introduction

Hypertension is the most important risk factor for stroke and premature cardiovascular disease (CVD) [1,2]. Prehypertension is defined as a systolic blood pressure (SBP) of 120-139 mm Hg or a diastolic blood pressure (DBP) of 80-89 mm Hg [3]; however, the risks of coronary artery disease and stroke rise progressively as blood pressure increases above 115/75 mm Hg [4]. As a result, interventions focused on individuals with prehypertension may be of interest to address the burden of hypertension.

Worldwide, the number of individuals who own a mobile phone is increasing. Mobile health (mHealth), an ever-expanding concept, uses this growing technology in a wide range of health care applications [5]. Interventions using mHealth have the potential to shorten gaps to reach underserved populations [6], providing a more flexible platform for improving patient self-care. mHealth technology has been applied successfully to meet the treatment of infectious chronic diseases such as tuberculosis (treatment adherence, prevention, and education) [7] and HIV/AIDS (uptake of sexual health services and information) [8], and some mHealth interventions have been used to promote changes toward healthier lifestyles, thereby improving health outcomes [9-11]. However, the number of studies assessing the impact of mHealth on specific cardiovascular outcomes is more limited in low-income and middle-income settings [6,12], and although the success of these interventions is evident, the long-term impact of interventions involving mHealth technology is not.

Some interventions have been proved to be effective to change lifestyle behaviors. A systematic review found that interventions based on the Transtheoretical Model can reduce fat consumption, increase the consumption of fruit and vegetables, and increase physical activity depending upon the progression through the stages of change [13]. Similarly, a recent systematic review reported that telephone-based interventions that incorporate motivational interviewing are promising for weight loss [14]. The GISMAL (Grupo de Investigación en Salud Móvil en America Latina, in Spanish) study was a 1-year randomized controlled trial conducted in three Latin American countries (Argentina, Guatemala, and Peru) using the Transtheoretical Model and motivational interviewing.

The aim of the GISMAL study was to assess whether an mHealth intervention would improve the cardiometabolic profile (ie, reduce blood pressure levels and body weight) among individuals with prehypertension [15,16]. Although the intervention did not reduce blood pressure levels, it was associated with a reduction of body weight and improvement in some dietary habits, especially in Peru. One year may not be a sufficient period to observe changes in blood pressure levels due to a behavioral intervention or to assess whether the effects of the intervention can continue to provide benefit after it is stopped; therefore, we aimed to evaluate the long-term effects (ie, 4 years after completion of the original study) of the GISMAL mHealth intervention on blood pressure and body weight in participants recruited in Peru.

Methods

Original Intervention and Settings

The GISMAL study (NCT01295216) was performed in 2012. It was a multicenter, parallel-group, randomized controlled trial that was stratified by sex and age (30-44 years and 45-60 years). Details about the intervention have been published elsewhere [15]. In brief, randomization was stratified by country using minimization by sex and age group. The intervention lasted 12 months, followed a standardized protocol, which was implemented by trained nutritionists and comprised monthly phone calls in which the nutritionists used motivational interview techniques. Participants were contacted through their personal mobile phones, and the conversations focused on reduction of dietary sodium intake, reduction of high-fat and high-sugar food intake, increase in fruit and vegetable consumption, and promotion of physical activity. In addition, SMS text messages were sent weekly to participants to reinforce the calls [17]. The same nutritionist entered the information obtained during each call into a web-based platform to customize the weekly text messages delivered to participants in the following month. Several text messages were developed and culturally adapted to each country to guarantee understanding, adequacy of the message wording, and tone, as previously described [18].

Eligibility criteria included men and women aged 30-60 years who owned mobile phones and with SBP and DBP in the prehypertension range (between 120 and 139 mm Hg and between 80 and 89 mm Hg, respectively). People with an earlier diagnosis of diabetes or hypertension, illiterate individuals, and pregnant women were excluded.

The total sample of the original intervention was 636 people from Argentina, Guatemala, and Peru; however, the sample size of the trial was calculated separately for each country. Thus, researchers from Peru enrolled 212 individuals to ensure a change in systolic blood pressure levels of 5 mm Hg. Subjects from two different sites (Pampas de San Juan de Miraflores and Hospital Nacional Cayetano Heredia) were recruited; 107 were allocated to the intervention group and 105, to the control group. The outcomes of interest were blood pressure, body weight, diet quality, and physical activity; these were evaluated at baseline (randomization) and at 6 and 12 months later.

Follow-up Assessment

Between August and December 2017, on average, 5 years after randomization (ie, 4 years after completion of the intervention), participants enrolled in Peruvian sites were contacted to determine the long-term effects of the intervention (ie, whether the effects of the mHealth intervention were maintained over time and affected blood pressure levels as originally planned). In the time between the 1-year and 5-year assessments, the research team had no contact with the participants. As per the original study [15], where participants had moved or had changed their telephone number, family members or friends reported as next of kin in the original study were contacted to find the participants.

Primary and Secondary Outcomes

As in the original study, in this new assessment, the primary outcomes were SBP and DBP, both measured in millimeters of mercury; in addition, hypertension incidence was included as a primary outcome. The secondary outcomes were the same as in the original study, including body weight (in kilograms), BMI (kg/square meter), physical activity (in metabolic equivalents [METs] per minute of a task per week), and diet patterns (daily intake of fruits and vegetables, high-sodium food, and of high-fat and high-sugar foods).

SBP and DBP were measured using an automatic monitor (HEM-742 INT, Omron) as in the original study. Measurements were taken in triplicate; after a 5-minute resting period, the first blood pressure measurement was taken, and the time between subsequent measurements was at least 1 minute. The average of the second and third blood pressure measurements was used for the analyses. All measurements were taken with the participant in a seated position using the same arm where the original measurement was taken.

Body weight was measured three times, following standardized techniques. We used the same digital scales (Seca 803/Omron SC-100) used in the original study and calibrated them to an accuracy of 100 g. BMI was calculated by dividing weight (kg) by height (m) squared. Primary and secondary outcomes were evaluated as numerical variables; however, hypertension incidence was evaluated by taking into account the new cases of hypertension detected since randomization.

Changes in physical activity were evaluated using the METs obtained from the short version of the International Physical Activity Questionnaire [19]. Moreover, diet patterns were evaluated using the same food frequency questionnaire used in the original study; the questions focused on the consumption of foods with high contents of sodium, simple sugars, trans fats, and saturated fats, as well as on the consumption of fruits and vegetables [20].

Other important variables were considered, including sociodemographic data (age, sex, marital status, household income, years of education, employment status, and health insurance coverage) as well as self-reported lifestyle behaviors (smoking status, alcohol intake, physical activity, and daily dietary intake). In addition, stages of change (precontemplation, contemplation, preparation, action, and maintenance) according to the Transtheoretical Model [21] were described (physical activity, intake of servings of 5 fruits and vegetables, food with harmful fats, high-sugar food and beverages, high-sodium processed foods, salt added at the table, and salt added for cooking).

Statistical Analysis

Data were analyzed using STATA 13 software (Stata Corp). The intention-to-treat principle was used to compare the intervention and control arms. Means and standard deviations for numerical continuous variables and proportions and frequencies for categorical variables were used to describe the study population enrolled in Peru at baseline.

For the incidence analysis, we excluded patients whose hypertension status could not be confirmed, including those who were lost to follow-up and who had died. Person-years of follow-up were calculated by summing the follow-up times for the remaining participants. For participants with a new diagnosis of hypertension, only half of the time between the last assessment and the previous assessment was used in this sum since the actual date of diagnosis was unknown. Incidence rates and their 95% confidence intervals were then estimated. Poisson regression models with log link functions and robust standard errors to account for cluster effects were calculated by reporting the risk ratio (RR) and 95% confidence intervals.

Differences in primary and secondary outcomes were assessed using linear mixed models, including two levels (assessments as level 1 units and subjects as level 2 clusters). As in the original study, the a priori defined model included the interaction between the intervention and the time of follow-up as categorical variables (baseline, 6 months, 12 months, and 5 years), adjusting by sex and age to reverse the stratification during randomization. The regression model results focus on the 5-year assessment and are presented as coefficients with their respective 95% CI. A dose-response analysis was also conducted using appropriate statistical techniques but was categorized in two groups instead of the three groups in the original study: participants who received <6 (50%) and ≥6 (50%) of the scheduled motivational interviewing calls.

Ethical Aspects

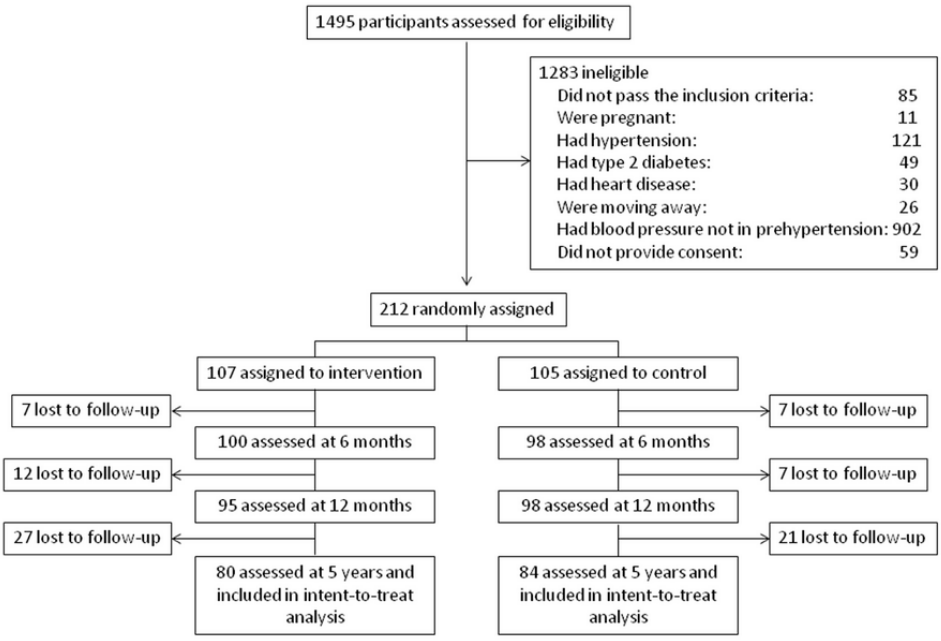
As we only contacted participants from the Peruvian sites, this new assessment was approved by the Ethics Committee of the Universidad Peruana Cayetano Heredia, Lima, Peru. To ensure the autonomy of the participants, written informed consent was obtained before we assessed the participants. The data collected were kept confidential and accessed only by the researchers who performed the study.

Results

Principal Findings

In Peru, a total of 1495 participants were assessed for eligibility; finally, 212 were randomly assigned to the intervention or control arm. After the first year of the study, 193 participants were retained in the study (n=95, 49.2% in the intervention group and n=98, 50.8% in the control group); 5 years after randomization, 1 (0.5%) participant had died, and 47 (22.2%) were lost to follow-up. Therefore, data from 164 individuals were analyzed (n=80, 48.8% in the intervention group and n=84, 51.2% in the control group; Figure 1).

Figure 1. Trial profile in Peru.



According to the baseline comparison, the randomization works relatively well, as the distributions of the characteristics of the study population were similar in the intervention and control groups (Multimedia Appendix 1).

Primary and Secondary Outcomes

In intention-to-treat analysis, the intervention did not result in changes in SBP (–2.54 mm Hg, 95% CI –8.23 to 3.15) or DBP (3.41 mm Hg, 95% CI –0.75 to 7.57) compared to controls. Among secondary outcomes, those who received the intervention had lower body weight (–5.42 kg, 95% CI –10.4 to –0.48) and BMI (–2.56 kg/m^2, 95% CI: –4.46 to –0.66) compared to the control group. However, there was no evidence of effects of the intervention on physical activity levels, intake of fruits and vegetables, or intake of high-sodium food, but there were small increases in the intake of high-fat and high-sugar foods (Table 1).

The intervention reduced the risk of developing hypertension after 5 years of follow-up (RR 0.76, 95% CI 0.45-1.28); however, the estimates were not significant (Table 2).

Table 1. Primary and secondary outcomes at 5 years of follow-up: results from mixed linear models.

Outcome	Intervention group (n=80)	Control group (n=84)	Mean difference[b] β (95% CI)	*P* value
Primary				
Systolic blood pressure	124.0 (12.1)	126.6 (12.5)	–2.54 (–8.23 to 3.15)	.38
Diastolic blood pressure	74.4 (8.8)	74.2 (9.0)	3.41 (–0.75 to 7.57)	.11
Secondary				
Body weight (kg)	78.0 (15.1)	79.8 (15.7)	–5.42 (–10.4 to –0.48)	.03
BMI (kg/m^2)	31.4 (5.2)	32.8 (5.9)	–2.56 (–4.46 to –0.66)	.008
Physical activity (METs[a]/min per week)	1077 (1641.6)	911 (1274.6)	–61.12 (–745.94 to 623.70)	.86
Daily intake of fruits and vegetables[b]	1.62 (0.95)	1.65 (0.94)	0.05 (–0.47 to 0.56)	.86
Daily intake of high-sodium food[b]	0.56 (0.60)	0.82 (0.69)	–0.22 (–0.55 to 0.11)	.20
Daily intake of high-fat and high-sugar foods[b]	3.72 (1.69)	3.93 (1.73)	0.98 (0.14 to 1.82)	.02

[a]METs: metabolic equivalents.

[b]Mean differences were estimated using linear mixed models using information from the baseline and 6-month, 12-month, and 5-year follow-ups and were controlled by sex and age.

Table 2. Incidence of hypertension by intervention group and dose-response to intervention 5 years after randomization.

	New cases of hypertension during follow-up		Regression model[a] RR (95% CI)	*P* value
	No (n=142)	Yes (n=50)		
Intention-to-treat analysis				
Control	69 (70.4%)	29 (29.6%)	1 (Reference)	—
Intervention	73 (77.7%)	21 (22.3%)	0.76 (0.45-1.28)	.30
Dose-response analysis				
Control	69 (70.4%)	29 (29.6%)	1 (Reference)	—
<50% (<6 calls)	20 (66.7%)	10 (33.3%)	1.33 (0.64-2.75)	.44
≥50% (≥6 calls)	53 (82.8%)	11 (17.2%)	0.55 (0.29-1.04)	.07

[a]Models were adjusted by age and sex, as randomization was stratified for these variables.

Dose-Response Analysis

No changes in SBP (–4.05 mm Hg, 95% CI: –10.11 to 2.02) or DBP (1.48 mm Hg, 95% CI: –2.84 to 5.79; Table 3) were observed among participants receiving ≥50% of motivational calls. On the other hand, those receiving ≥50% of scheduled calls during the intervention had a 45% reduction (RR 0.55, 95% CI: 0.29-1.04) in the risk of developing hypertension 5 years after randomization; however, the estimates were not significant (Table 2).

Among secondary outcomes, participants in the intervention group who received ≥50% of calls had greater reductions in body weight (–6.23 kg, 95% CI: –11.47 to –0.99) and BMI (–2.81 kg/m^2, 95% CI: -4.77; -0.85); however, there were no changes in the other secondary outcomes (Table 3).

Table 3. Dose response to intervention in primary and secondary outcomes at 5-year follow-up.

	Mean difference (<50%/<6 calls)[b] (n=26)		Mean difference (≥50%/≥6 calls)[b] (n=54)	
	β (95% CI)	*P* value	β (95% CI)	*P* value
Primary outcomes				
Systolic blood pressure	0.42 (–5.89 to 6.72)	.90	–4.05 (–10.11 to 2.02)	.19
Diastolic blood pressure	7.13 (2.31 to 11.95)	.004	1.48 (–2.84 to 5.79)	.50
Secondary outcomes				
Body weight (kg)	–3.57 (–10.47 to 3.34)	.31	–6.23 (–11.47 to –0.99)	.02
BMI (kg/m^2)	–1.79 (–4.45 to 0.87)	.19	–2.81 (-4.77; -0.85)	.005
Physical activity (METs[a]/min per week)	–163.55 (–915.48 to 588.38)	.67	59.37 (–703.91 to 822.64)	.88
Daily intake of fruits and vegetables[b]	0.25 (–0.39 to 0.89)	.45	–0.04 (–0.56; 0.49)	.89
Daily intake of high-sodium food[b]	–0.05 (–0.47 to 0.37)	.82	–0.28 (–0.62; 0.06)	.11
Daily intake of high-fat and high-sugar foods[b]	1.74 (0.65 to 2.83)	.002	0.64 (–0.20; 1.49)	.14

[a]METs: metabolic equivalents.

[b]Mean differences were estimated using linear mixed models using information from the baseline, 6-month, 12-month, and 5-year follow-ups and were controlled by sex and age.

Discussion

Main Findings

Although no significant changes were observed in blood pressure levels 4 years after the original intervention was completed, our findings demonstrate important reductions in body weight and BMI. Moreover, participants who received ≥50% of motivational calls during the 1-year intervention potentially benefited most because greater reductions of body weight and BMI were observed; this may have further impact on hypertension incidence, as suggested by the risk estimates. Notably, none of the changes found in the behavioral factors could explain the reductions in body weight and BMI.

Comparison With Previous Studies

To our knowledge, this is one of the first randomized controlled trials assessing the long-term effects of an mHealth intervention created to promote healthy lifestyle behaviors among subjects at high risk of CVD (ie, with prehypertension) in Latin America. Two relatively recent systematic reviews highlighted the limited number of mHealth interventions in resource-constrained settings, especially from the prevention perspective [6,12].

Among existing studies, Green et al [22] demonstrated that web-delivered pharmacy team care resulted in greater reduction in SBP and improved blood pressure 6-18 months after completion of the interventions. Similarly, Margolis et al [23] showed that intensive intervention based on blood pressure telemonitoring with pharmacist management had sustained effects for up to 24 months (12 months after the intervention ended). Therefore, our study expands on current data by suggesting that mHealth intervention has a sustained effect on body weight 4 years after the intervention ended. On the other hand, Appel et al [24] reported a reduction of 4.6 kg among obese individuals receiving remote support (ie, telephone, website, and email) compared to a control group after 24 months of follow-up; however, no impact on CVD events or all-cause mortality was observed. On the other hand, Rubinstein et al [15] reported no change in blood pressure levels after 12 months of intervention in prehypertensive individuals, but participants in the intervention group had modest reductions in body weight and BMI and reported lower intake of high-fat and high-sugar foods. Recently, a meditation smartphone app appeared to decrease SBP in a 6-month dose-response feasibility trial; however, the adherence to this intervention declined over time [25].

The utility of mobile phone text messages has been reported mainly to support hypertension treatment and management, especially as reminders [26,27]. For example, Bobrow et al [26], using SMS text messages with hypertensive individuals, reported a small reduction in SBP compared to usual care after 12 months of intervention. Similarly, Hacking et al [27] reported that text messages only improved self-reported behavior changes. However, limited literature has assessed the long-term impact of mHealth interventions on the cardiometabolic profiles of prehypertensive individuals.

The GISMAL intervention had a marked long-term impact on body weight and BMI. Previous results of the original study at the Peruvian site showed reductions of 1.24 kg in body weight and 0.53 kg/m^2 in BMI among those in the intervention arm after 1 year of follow-up (Multimedia Appendix 2) [15]. These new results show that 4 years after the completion of the original study, both body weight and BMI are much lower among individuals who received the intervention than in those who did not. Thus, this 1-year intervention not only helped sustain previous weight loss but also helped ensure a greater weight reduction over time. Surprisingly, despite the clear reductions in weight and BMI at long-term follow-up, there was weak evidence of changes in blood pressure levels. The sample size of the original trial was calculated to detect a difference of 5 mm Hg in each country, and some decreases in SBP and hypertension incidence, especially among those receiving higher doses of the intervention, were noted.

On the other hand, there were no differences among evaluated behavioral factors. These findings suggest that the tools and questionnaires used during the evaluation were not accurate enough to assess selected lifestyles or that the intervention led to changes in unmeasured behaviors that were maintained beyond the period of the study intervention. Other studies have reported similar impacts of mHealth on bodyweight and BMI, but most of them were in the short term [28,29]. Therefore, our results support the fact that a short mHealth intervention comprising motivational interviewing calls and weekly text messages helps participants retain healthy habits and may help them maintain long-term effects.

Relevance to Public Health

Recently, the American Heart Association and the American College of Cardiology included individuals with prehypertension as having hypertension stage 1 (SBP=130-139 mm Hg or DBP=80-89 mm Hg) and proposed that these individuals need appropriate management [30]. However, although people with prehypertension are at a high risk of developing CVD, they do not receive treatment in resource-constrained settings. This highlights the need for prevention strategies to avoid further complications.

mHealth appears to be a promising way to reduce the risk of these individuals because participants in our intervention group only received monthly health counselling and weekly text messages for 12 months. Regular communication between patients and clinics or health posts may improve adherence to healthy behaviors, which in turn can prevent the onset of CVD later on and contribute to other positive health outcomes [31,32]. In addition, the effects of our intervention could have been greater if booster appointments were utilized, thus extending the behavior changes.

The perceived benefits of this mHealth intervention must outweigh the effort of receiving calls and text messages because self-management is an ongoing process that requires significant iteration. The introduction of apps to support calls and text messages even after the intervention period can help produce sustainable outcomes. However, evidence has demonstrated that in interventions based only on technology, when people are left alone with mobile self-help apps, participants are less adherent [33,34] and less motivated to engage in the proposed program than participants who are accompanied by health staff or coaches or who have other types of face-to-face interaction as part of the intervention [35]. The reductions in mean body weight and BMI indicate possible long-term success of the intervention, with a possible impact on hypertension; hence, this intervention can potentially be implemented to ensure prevention of CVD.

Limitations

Some limitations must be highlighted. First, this study included only data from participants in Peruvian settings, although the original intervention was conducted in three countries (Argentina, Guatemala, and Peru). However, the effects of the intervention were especially important in Peru, as shown by the 12-month results [15], and the sample size of the trial was calculated separately for each country. Second, the rate of attrition was over 20% after 5 years; thus, some bias may arise in the results. Despite this, the intention-to-treat principle was used in all the analyses. Third, the original intervention was based on the Transtheoretical Model, which is mainly used for smoking cessation [36,37]. Although recent literature shows the use of this model in other interventions [13], the intervention was adapted to be applied for cardiovascular prevention in

resource-constrained settings. Fourth, we did not assess differential exposition to other preventive interventions since study randomization. However, as the participants did not have hypertension or any other noncommunicable condition, the effects of this limitation may be negligible. Finally, recall and desirability bias may be present at the moment of evaluation, as is usual in these types of studies. However, validated scales and standardized procedures were used to reduce these biases as in the original study.

Conclusion

A 1-year mHealth intervention comprising motivational interview calls and text messages appears to have long-term effects on health 4 years after intervention completion. Although we detected no effects on blood pressure levels, important reductions of body weight and BMI were observed. Individuals receiving ≥50% of calls had greater reductions in body weight and BMI, and a potential effect on hypertension incidence was observed. Thus, mHealth appears to be a promising preventive strategy for noncommunicable diseases in resource-constrained settings.

Acknowledgments

AB-O was supported by a Research Training Fellowship in Public Health and Tropical Medicine (103994/Z/14/Z) funded by Wellcome Trust. The funder had no role in the study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics of the study population by intervention group.
[DOCX File , 15 KB - jmir_v22i4e14595_app1.docx]

Multimedia Appendix 2
Outcomes in Peru: comparison after 1-year intervention.
[DOCX File , 13 KB - jmir_v22i4e14595_app2.docx]

Multimedia Appendix 3
CONSORT e-HEALTH checklist (V. 1. 6. 1).
[PDF File (Adobe PDF File), 1683 KB - jmir_v22i4e14595_app3.pdf]

References

1. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. JAMA 2017 Dec 10;317(2):165-182. [doi: 10.1001/jama.2016.19043] [Medline: 28097354]
2. Lawes CMM, Vander Hoorn S, Rodgers A, International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. Lancet 2008 May 03;371(9623):1513-1518. [doi: 10.1016/S0140-6736(08)60655-8] [Medline: 18456100]
3. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003 May 21;289(19):2560-2572. [doi: 10.1001/jama.289.19.2560] [Medline: 12748199]
4. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). JAMA 2014 Feb 5;311(5):507-520. [doi: 10.1001/jama.2013.284427] [Medline: 24352797]
5. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
6. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
7. Devi BR, Syed-Abdul S, Kumar A, Iqbal U, Nguyen P, Li YJ, et al. mHealth: An updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Comput Methods Programs Biomed 2015 Nov;122(2):257-265. [doi: 10.1016/j.cmpb.2015.08.003] [Medline: 26304621]
8. Burns K, Keating P, Free C. A systematic review of randomised control trials of sexual health interventions delivered by mobile technologies. BMC Public Health 2016 Dec 12;16(1):778 [FREE Full text] [doi: 10.1186/s12889-016-3408-z] [Medline: 27514851]

9. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth Technologies to Influence Physical Activity and Sedentary Behaviors: Behavior Change Techniques, Systematic Review and Meta-Analysis of Randomized Controlled Trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
10. 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]
11. Kelly JT, Reidlinger DP, Hoffmann TC, Campbell KL. Telehealth methods to deliver dietary interventions in adults with chronic disease: a systematic review and meta-analysis. Am J Clin Nutr 2016 Dec;104(6):1693-1702. [doi: 10.3945/ajcn.116.136333] [Medline: 27935523]
12. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014;10:49 [FREE Full text] [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]
13. Carvalho de Menezes M, Bedeschi LB, Santos LCD, Lopes ACS. Interventions directed at eating habits and physical activity using the Transtheoretical Model: a systematic review. Nutr Hosp 2016 Sep 20;33(5):586 [FREE Full text] [doi: 10.20960/nh.586] [Medline: 27759990]
14. Patel ML, Wakayama LN, Bass MB, Breland JY. Motivational interviewing in eHealth and telehealth interventions for weight loss: A systematic review. Prev Med 2019 Sep;126:105738. [doi: 10.1016/j.ypmed.2019.05.026] [Medline: 31153917]
15. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, et al. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. The Lancet Diabetes & Endocrinology 2016 Jan;4(1):52-63. [doi: 10.1016/S2213-8587(15)00381-2] [Medline: 26653067]
16. Carrillo-Larco RM, Jiwani SS, Diez-Canseco F, Kanter R, Beratarrechea A, Irazola V, GISMAL Group. Implementation Tells Us More Beyond Pooled Estimates: Secondary Analysis of a Multicountry mHealth Trial to Reduce Blood Pressure. JMIR Mhealth Uhealth 2018 Nov 01;6(11):e10226 [FREE Full text] [doi: 10.2196/10226] [Medline: 30389646]
17. Beratarrechea A, Diez-Canseco F, Fernández A, Kanter R, Letona P, Martinez H, et al. [Acceptability of a mobile health based intervention to modify lifestyles in prehypertensive patients in Argentina, Guatemala and Peru: a pilot study]. Rev Peru Med Exp Salud Publica 2015;32(2):221-229 [FREE Full text] [Medline: 26338378]
18. Diez-Canseco F, Zavala-Loayza JA, Beratarrechea A, Kanter R, Ramirez-Zea M, Rubinstein A, et al. Design and Multi-Country Validation of Text Messages for an mHealth Intervention for Primary Prevention of Progression to Hypertension in Latin America. JMIR Mhealth Uhealth 2015 Feb 18;3(1):e19 [FREE Full text] [doi: 10.2196/mhealth.3874] [Medline: 25693595]
19. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
20. Kabagambe EK, Baylin A, Allan DA, Siles X, Spiegelman D, Campos H. Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. Am J Epidemiol 2001 Dec 15;154(12):1126-1135. [doi: 10.1093/aje/154.12.1126] [Medline: 11744518]
21. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: Toward an integrative model of change. Journal of Consulting and Clinical Psychology 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
22. Green BB, Anderson ML, Ralston JD, Catz SL, Cook AJ. Blood pressure 1 year after completion of web-based pharmacist care. JAMA Intern Med 2013 Jul 08;173(13):1250-1252 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1037] [Medline: 23689927]
23. Margolis KL, Asche SE, Dehmer SP, Bergdall AR, Green BB, Sperl-Hillen JM, et al. Long-term Outcomes of the Effects of Home Blood Pressure Telemonitoring and Pharmacist Management on Blood Pressure Among Adults With Uncontrolled Hypertension: Follow-up of a Cluster Randomized Clinical Trial. JAMA Netw Open 2018 Sep 07;1(5):e181617 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1617] [Medline: 30646139]
24. Appel LJ, Clark JM, Yeh H, Wang N, Coughlin JW, Daumit G, et al. Comparative effectiveness of weight-loss interventions in clinical practice. N Engl J Med 2011 Nov 24;365(21):1959-1968 [FREE Full text] [doi: 10.1056/NEJMoa1108660] [Medline: 22085317]
25. Adams ZW, Sieverdes JC, Brunner-Jackson B, Mueller M, Chandler J, Diaz V, et al. Meditation smartphone application effects on prehypertensive adults' blood pressure: Dose-response feasibility trial. Health Psychol 2018 Sep;37(9):850-860 [FREE Full text] [doi: 10.1037/hea0000584] [Medline: 30010353]
26. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile Phone Text Messages to Support Treatment Adherence in Adults With High Blood Pressure (SMS-Text Adherence Support [StAR]): A Single-Blind, Randomized Trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]

27. Hacking D, Haricharan HJ, Brittain K, Lau YK, Cassidy T, Heap M. Hypertension Health Promotion via Text Messaging at a Community Health Center in South Africa: A Mixed Methods Study. JMIR Mhealth Uhealth 2016 Mar 10;4(1):e22 [FREE Full text] [doi: 10.2196/mhealth.4569] [Medline: 26964505]
28. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
29. Shaw RJ, Bosworth HB, Silva SS, Lipkus IM, Davis LL, Sha RS, et al. Mobile health messages help sustain recent weight loss. Am J Med 2013 Nov;126(11):1002-1009 [FREE Full text] [doi: 10.1016/j.amjmed.2013.07.001] [Medline: 24050486]
30. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248. [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
31. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
32. Leon N, Surender R, Bobrow K, Muller J, Farmer A. Improving treatment adherence for blood pressure lowering via mobile phone SMS-messages in South Africa: a qualitative evaluation of the SMS-text Adherence SuppoRt (StAR) trial. BMC Fam Pract 2015;16:80 [FREE Full text] [doi: 10.1186/s12875-015-0289-7] [Medline: 26137844]
33. 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]
34. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
35. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
36. DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation: An analysis of precontemplation, contemplation, and preparation stages of change. Journal of Consulting and Clinical Psychology 1991 Apr;59(2):295-304. [doi: 10.1037//0022-006x.59.2.295] [Medline: 2030191]
37. Pallonen UE, Leskinen L, Prochaska JO, Willey CJ, Kääriäinen R, Salonen JT. A 2-year self-help smoking cessation manual intervention among middle-aged Finnish men: an application of the transtheoretical model. Prev Med 1994 Jul;23(4):507-514. [doi: 10.1006/pmed.1994.1069] [Medline: 7971879]

Abbreviations

CVD: cardiovascular disease
DBP: diastolic blood pressure
GISMAL: Grupo de Investigación en Salud Móvil en America Latina
mHealth: mobile health
METs: metabolic equivalents
RR: risk ratio
SBP: systolic blood pressure

Edited by G Eysenbach; submitted 14.05.19; peer-reviewed by S Omboni, PH Lin, A Carr; comments to author 03.10.19; revised version received 29.10.19; accepted 22.02.20; published 21.04.20.

Please cite as:
Bernabe-Ortiz A, Pauschardt J, Diez-Canseco F, Miranda JJ
Sustainability of mHealth Effects on Cardiometabolic Risk Factors: Five-Year Results of a Randomized Clinical Trial
J Med Internet Res 2020;22(4):e14595
URL: http://www.jmir.org/2020/4/e14595/
doi:10.2196/14595
PMID:32314970

Original Paper

Evaluating an Intervention Program Using WeChat for Patients With Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial

Yuyu Jiang[1], PhD[‡]; Fenglan Liu[1], MSc; Jianlan Guo[1], MSc; Pingping Sun[1], MSc; Zhongyi Chen[1], MSc; Jinping Li[2], MPH; Liming Cai[3], MB; Hongqing Zhao[4], MM; Ping Gao[5], MB; Zhaosheng Ding[6], BS; Xiaoliang Wu[7], MB

[1]Research Office of Chronic Disease Management and Rehabilitation, Wuxi School of Medicine, Jiangnan University, Wuxi, China

[2]Wuxi School of Medicine, Jiangnan University, Wuxi, China

[3]Affiliated Hospital of Jiangnan University, Wuxi, China

[4]Wuxi Second People's Hospital, Wuxi, China

[5]Wuxi Second Hospital of Traditional Chinese Medicine, Wuxi, China

[6]Rongjun Hospital of Jiangsu Province, Wuxi, China

[7]Huishan District Rehabilitation Hospital, Wuxi, China

[‡]Research Office of Chronic Disease Management and Rehabilitation

Corresponding Author:
Yuyu Jiang, PhD
Research Office of Chronic Disease Management and Rehabilitation
Wuxi School of Medicine
Jiangnan University
No 1800 Lihu Avenue
Wuxi
China
Phone: 86 0510 85328363
Fax: 86 0510 85328605
Email: doctoryuyu@jiangnan.edu.cn

Abstract

Background: The application of telemedicine in home pulmonary rehabilitation interventions for the management of patients with chronic obstructive pulmonary disease (COPD) has achieved promising results.

Objective: This study aimed to develop a WeChat official account (Pulmonary Internet Explorer Rehabilitation [PeR]) based on social media. It further evaluated the effect of PeR on the quality of life, symptoms, and exercise self-efficacy of patients with COPD.

Methods: The functional modules of PeR were developed by a multidisciplinary team according to the electronic health–enhanced chronic care model (eCCM) components. A total of 106 patients were randomly selected (53 in the PeR group and 53 in the outpatient face-to-face group [FtF]). Pulmonary rehabilitation intervention was conducted for 3 months, and the outcome was observed for 3 months. The primary outcome was patient quality of life measured with the COPD assessment test (CAT). The secondary outcomes were evaluated using the modified Medical Research Council scale (mMRC), exercise self-regulatory efficacy scale (Ex-SRES), and St George's Respiratory Questionnaire (SGRQ).

Results: The intention-to-treat analysis was used in the study. A total of 94 participants completed the 6-month pulmonary rehabilitation program. No statistically significant differences were observed in CAT ($F_{1,3}$=7.78, P=.001), Ex-SRES ($F_{1,3}$=21.91, P<.001), and mMRC scores ($F_{1,3}$=29.64, P<.001) between the two groups with the variation in time tendency. The Ex-SRES score had a significant effect on the CAT score (P=.03). The partial regression coefficient of Ex-SRES to CAT was 0.81, and Exp (B) was 2.24.

Conclusions: The telemedicine technology was effective using the eCCM combined with a behavioral intervention strategy centering on self-efficacy. Pulmonary rehabilitation at home through PeR and FtF could improve the sense of self-efficacy and quality of life and alleviate symptoms in patients with COPD.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900022770; https://tinyurl.com/tmmvpq3

(J Med Internet Res 2020;22(4):e17089) doi:10.2196/17089

KEYWORDS

chronic obstructive pulmonary disease; randomized controlled trial; self-efficacy; telemedicine; the eHealth enhanced chronic care model; WeChat

Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic and progressive respiratory disease [1]. In 2020, COPD is expected to become the third leading cause of death [2] and the fifth leading economic burden of disease [3].

Pulmonary rehabilitation is an important component of COPD treatment and management [4]. "Pulmonary rehabilitation is a comprehensive intervention based on a thorough patient assessment, followed by patient-tailored therapies that include, but are not limited to, exercise training, education, and behavioral change designed to improve the physical and psychological conditions of people with chronic respiratory disease and promote the long-term adherence to health-enhancing behaviors" [5]. Pulmonary rehabilitation could alleviate symptoms, enhance activity tolerance, improve quality of life, and reduce the burden of medical and health service system [6]. Despite the acknowledged benefits, participation and completion of pulmonary rehabilitation training have not lived up to expectations [7]. The standard duration of pulmonary rehabilitation training at home is 8 to 12 weeks. Patients needed long-term maintenance training to ensure the effect of pulmonary rehabilitation [4]. In the United Kingdom, less than 1.5% of patients with COPD receive pulmonary rehabilitation each year [8]. A clinical research showed that only 42% of patients with COPD successfully accomplished pulmonary rehabilitation training [7]. According to the American Thoracic Association and the European Respiratory Association, pulmonary rehabilitation has many limitations, such as insufficient resources for pulmonary rehabilitation, low proportion of medical insurance distribution, and lack of professional health care providers [9]. In addition, other elements (eg, transport, mobility of population, distance, and training location) also make patients with COPD incapable of requesting, participating, and persisting in pulmonary rehabilitation training [10].

Telemedicine service mode can be described as the use of electronic information and communication technology by professional health care providers to provide and support health care to patients in case of long distances [11]. The implementation of pulmonary rehabilitation through remote technology can not only reduce the medical service demand and expenses of patients with COPD but also improve the accessibility of service projects, solve the difficulty of transportation and distance during training, and expand the programs to remote areas [12]. Moreover, in response to the low adherence rate of patient pulmonary rehabilitation training, researchers intended to modify the treatment of chronic respiratory diseases by developing behavioral change interventions [13]. The systematic review by McCullough et al [13] included 46 studies; 19 of them applied 12 different behavioral change theories, among which self-efficacy theory and social cognition theory were used in several studies. Notably, few studies combined behavioral intervention strategies with remote technologies in COPD. Lorig et al [14] implemented internet-based chronic disease self-management projects in 958 patients with chronic diseases that included password protection, interactive network teaching, and health education. The content of health education included personalized sports design, cognitive symptom management, negative emotion management, drug overview, physician-patient communication, and healthy diet. The results showed that providing self-management support through the internet could effectively improve the health status of patients, and it was a feasible choice to replace face-to-face self-management [14].

The advantage of interventions based on theoretical models has been confirmed by researchers [15]. Theoretical models can help researchers conduct clinical research better by observing the relationship between telemedicine content, behavioral mechanisms, and expected outcomes from a holistic perspective [16]. The electronic health–enhanced chronic care model (eCCM) is a theoretical model of telemedicine intervention proposed by Gee and colleagues [17] based on the CCM (chronic care model). The CCM is currently considered the best comprehensive evidence for chronic disease prevention and management interventions [18]. The eCCM is the innovation of CCM and electronic health (eHealth) [17]. No transformation study about the eCCM or other theory based on the eCCM has been published to date.

This study aimed to construct functional modules of Pulmonary Internet Explorer Rehabilitation (PeR, mobile technology, a free social media WeChat official account) according to the features of eCCM components. Meanwhile, behavioral intervention strategies centered on self-efficacy were included to evaluate the effect of PeR's application. It was hypothesized that PeR could relieve symptoms in patients with COPD and improve their self-efficacy and quality of life.

Methods

Study Design

This study was a 6-month randomized controlled trial in which standardized pulmonary rehabilitation intervention lasted for 3 months, and the rehabilitation observation period was 3 months. The research assistant generated the random number sequence using a random assignment sequence table. Patients were randomly divided into intervention and control groups according to the ratio of 1:1. The intervention group received PeR management on the basis of evaluation, including respiratory training, sports training, diet guidance, medication knowledge, and so forth. Patients reported their training records at least once a week. The control group received the same content as the intervention group during an outpatient face-to-face (FtF) intervention. Patients reported their pulmonary rehabilitation

training records by mobile phone. The data of patients were collected by research assistants (those who did not know about the grouping) at baseline and in the third and sixth months. The study passed the ethical review of Wuxi Medical College of Jiangnan University (JNU20190318IRB61) and was registered with the Chinese Clinical Trial registry [ChiCTR1900022770]. The Consolidated Standards of Reporting Trials (CONSORT) checklist is in Multimedia Appendix 1.

Recruitment Processes

From October 1, 2018, to October 7, 2019, patients were recruited from the communities in Wuxi through leaflets, posters, and face-to-face communication. The pulmonary rehabilitation intervention was carried out. The inclusion criteria were aged 60 years and older, confirmed diagnosis of COPD according to the diagnosis and treatment guidelines for chronic obstructive pulmonary disease [19], forced expiratory volume in 1 second (FEV1)/forced vital capacity (FVC) ratio of <0.7, FEV1<80% predicted, and use of WeChat for effective communication. The exclusion criteria included patients with mental disorders, cognitive disorders, and limb dysfunction; patients with unstable heart disease or arrhythmia requiring drug intervention; patients with a history of myocardial infarction or cerebral infarction in the previous year; patients too weak to perform the muscle strength test; patients with hypertension that could not be controlled with drugs; and patients with a history of syncope after exercise. Patients who met the criteria signed the informed consent form. The baseline data measurement was organized by community doctors, nurses, and research assistants, including sex, age, education level, disease severity, and other demographic and sociological information. The St George's Respiratory Questionnaire (SGRQ), COPD assessment test (CAT), the modified Medical Research Council scale (mMRC), and exercise self-regulatory efficacy scale (Ex-SRES) assessments were completed.

Development of the Pulmonary Rehabilitation Intervention Program

Before the intervention, respiratory experts, clinicians, rehabilitation practitioners, nurses, research assistants, software engineers, user interface designers, and patients with COPD formed a multidisciplinary team. The team discussed and developed the pulmonary rehabilitation intervention program. The core part of the intervention program included respiratory training, diet guidance, medication knowledge, and exercise training. The implementation of the intervention program was divided into two types: PeR and outpatient FtF intervention. Respiratory training included lip contraction breathing and abdominal breathing. Diet guidance was to make a diet plan according to the standard weight, physical labor intensity, intake proportion of three major nutrients (proteins, carbohydrates, and lipids), and diet preference of patients. Medication knowledge was health education of routine drug use. Exercise training included upper extremity training, lower extremity training, and balance training. The intensity of exercise training was determined and adjusted according to patients' target heart rate and the score of conscious exertion. The target heart rate was calculated using the Karvonen formula, conscious exertion was calculated using level 13 on the rate of perceived exertion, and patients' slight fatigue was the best exertion. Training frequency was no less than 3 times a week, and training time was 20 to 30 minutes each time. Patients chose their own time of home-based pulmonary rehabilitation training. Within 3 months of pulmonary rehabilitation intervention, patients in the PeR group reported pulmonary rehabilitation training at least once a week through PeR, and patients in the traditional FtF group reported once a week by telephone. The program was adjusted according to patient reports of home-based pulmonary rehabilitation training. Patients in both groups were not required to report their condition for 3 months during the rehabilitation observation period.

Development of the Pulmonary Internet Explorer Rehabilitation App

The focus group method was applied in this study. The multidisciplinary team developed the functional modules of PeR according to eCCM components and the corresponding characteristics. The mapping table of PeR function module is shown in Table 1. PeR is a WeChat official account for patients with stable COPD. It uses the internet, mobile phones, computers, and WeChat to achieve community pulmonary rehabilitation for patients with COPD. The PeR includes two ports: the computer end and the WeChat end. The computer end is operated and managed by health professionals. Patients with COPD can see the functional modules of PeR on the WeChat terminal (Figure 1). The purpose of developing PeR was to allow more patients with COPD to receive pulmonary rehabilitation using free social media and help primary health institutions lacking rehabilitation resources realize the economic intervention and management of COPD pulmonary rehabilitation with the help of WeChat. PeR protects the privacy and data security of patients through access control and permission control, patient data transmission and anonymity, redundant storage, and data backup. PeR has obtained the computer software copyright registration certificate (Multimedia Appendix 2), and the intellectual property is protected by the State Copyright Administration of the People's Republic of China. A screenshot of the English version of the app is shown in Multimedia Appendix 3.

Table 1. Mapping table of the Pulmonary Internet Explorer Rehabilitation function module.

Components of the eCCM[a]	Characteristic	PeR[b] function module
eCommunity	• Participation, active, self-management	• PeR moments
Communication and the addition of the complete feedback loop	• Communication, complete feedback • Control over the timing	• Appointment for medical treatment • Medical guidance
Clinical information systems enhancements	• Register • Database • Support access	• Register log-in • Rehabilitation effect database • Self-report database
Self-management support enhancements	• Education (information) • Behavior support (an aid to behavioral change)	• Respiratory training • Diet guidance • Medication knowledge • Exercise training • Integral mall
Add self-efficacy resource component	• Mastery experience • Alternative experience • Verbal persuasion • Physiological and emotional state	• Respiratory training • Diet guidance • Medication knowledge • Exercise training • Integral mall

[a]eCCM: electronic health–enhanced chronic care model.

[b]PeR: Pulmonary Internet Explorer Rehabilitation.

Figure 1. Functional modules of Pulmonary Internet Explorer Rehabilitation on the WeChat terminal.

Pulmonary Internet Explorer Rehabilitation App Moments

The implementation process on WeChat Moments was designed according to the four resources of self-efficacy theory (mastery experience, verbal persuasion, physiological and emotional state, and alternative experience). In PeR Moments, patients with COPD can upload their rehabilitation training pictures or speeches. Other patients and health professionals can interact

with them by commenting or giving thumbs up, thus promoting not only peer support between patients but also communication between doctors and patients. Moreover, an incentive system was designed to encourage patients with COPD in their participation. To illustrate, patients obtained scores by posting PeR Moments, after which they could exchange prizes by accumulating scores to a certain amount. Meanwhile, the standardized pulmonary rehabilitation training patients shared with the PeR Moment was set as an example to encourage remaining patients to persist in training.

Appointment for Medical Treatment and Medical Guidance

Patient baseline situations were measured by health care personnel in community health centers, while professionals (medical personnel in the multidisciplinary team) developed the pulmonary rehabilitation program. After that, pulmonary rehabilitation training of patients with COPD was initiated and managed by professionals through PeR. Patients could get the electronic pulmonary rehabilitation prescription in the medical guidance module and contact nursing members online in the same module if the training program needed to be adjusted. When the disease worsened, patients could communicate with nursing members after making an appointment. Nursing members recommended appropriate professionals for further treatment according to their condition.

Register Log-in, Rehabilitation Effect Database, and Self-Report Database

PeR included a registration center, rehabilitation effect database, and self-report database. The rehabilitation effect database was uploaded by a research assistant through a computer at baseline and after 3 and 6 months of testing. The self-report database was completed and uploaded by patients themselves on the WeChat end. Two community nurses were assigned by the community hospital to manage all patient data. Patients could browse and obtain all personal data in the database on the WeChat end.

Respiratory Training, Diet Guidance, Medication Knowledge, Exercise Training, and Integral Mall

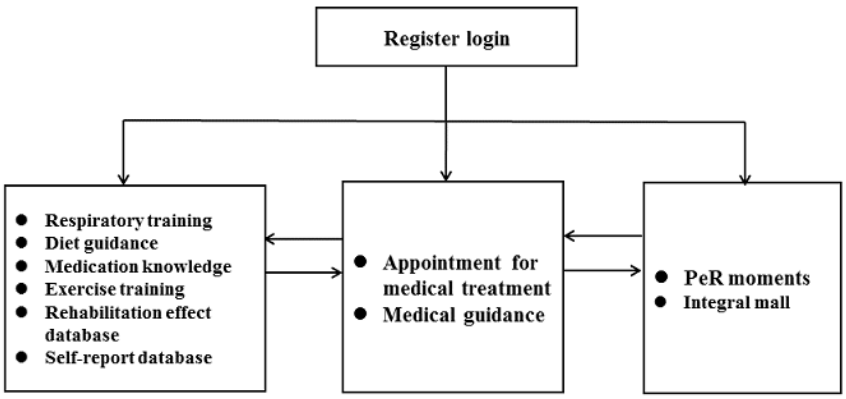
The audio and graphic versions of the modules on respiratory training, diet guidance, medication knowledge, and exercise training function were designed by a multidisciplinary team. Patients could choose different forms according to their own preferences. An integral mall was designed to encourage patients to generate network behavior so as to strengthen the function of the PeR Moments. Supervision of rehabilitation training behavior was increased to promote the maintenance of rehabilitation training behavior.

Pulmonary Rehabilitation Intervention

Intervention Process of the Pulmonary Rehabilitation Group

Before the start of pulmonary rehabilitation intervention, patients first received face-to-face training from medical and health care professionals. The training contents included how to obtain, understand, and apply the intervention program of pulmonary rehabilitation through PeR; how to complete self-report through PeR; how to share their own pulmonary rehabilitation training; and how to communicate and interact with peer patients and health care professionals through PeR. At the same time, patients received online training manuals. At the beginning, patients registered and logged on to PeR. Health care professionals evaluated patients, entered electronic pulmonary rehabilitation prescriptions, and distributed training aids such as resistance bands. Patients could view their electronic pulmonary rehabilitation prescription in the medical guidance module and the evaluation report in the rehabilitation effect database. Patients completed the home-based pulmonary rehabilitation training, completed the self-assessment report, and uploaded their training records in Moments. Health care professionals guided patients by reviewing the uploaded content. In the case of acute exacerbation of a patient's condition during this period, one could get in touch with health care professionals through the module of guidance and appointment for medical treatment, who would arrange the medical treatment schedule immediately. During the whole process of pulmonary rehabilitation intervention, patients could earn gifts in the integral mall using their own points and receive them in the third and sixth months of retest evaluation. The patient log-in process in the interface is shown in Figure 2.

Figure 2. Patient log-in process.



Intervention Process of the Face-to-Face Group in the Outpatient Department

Without using WeChat, the FtF group in the outpatient department received the same intervention as the PeR group, with the same pulmonary rehabilitation training equipment. According to the different training objectives, patients in the intervention and control groups all obtained one elastic band, and the elastic grade of the elastic band was consistent with the training objectives.

Follow-Up Process

The primary and secondary evaluation indexes were collected at baseline and in the third and sixth months. Patients in the intervention and control groups were tested separately face to face in the community hospital. Three days before the retest of each stage, the nurses in the community hospital informed patients to participate in the retest. Patients in the intervention group received an appointment reminder through PeR, and patients in the control group were informed by mobile phone. Patients reserved the right to withdraw from the study at any time during the study period.

Demographic and Sociological Information

The demographic and sociological information of patients in the intervention and control groups, including age, sex, education level, disease severity, and body mass index (BMI), were collected. A baseline comparison was then made.

Main Evaluation

CAT is a simple and easy-to-use health assessment tool for clinical practice to help patients and clinicians evaluate the symptoms and effects of diseases in a quantitative manner and to promote communication between patients and doctors. It includes 8 items in total. Each item evaluates a state, such as cough, expectoration, chest distress, energy, and so forth. Based on the state described by each item, the score range is from the best (score 0) to the worst (score 5). The total CAT score is the sum of the scores of all items. The higher the total score, the worse the patient's health [20].

Secondary Evaluation

SGRQ can be used as an important tool to evaluate the symptoms, pulmonary function, general well-being, and quality of life of patients with COPD and the effectiveness of medical services [21-23]. The scale contains 50 items, which are divided into 3 dimensions: symptom, activity, and influence. The score of the scale is from 0 to 100, which is calculated by a certain weight between the three subscales and the total scale. The higher the score, the worse the health status of patients with COPD. A score of 0 indicates that the disease caused no damage to a patient, and a score of 100 indicates maximum damage [20].

Ex-SRES was developed based on motor disorders to evaluate exercise self-efficacy [24]. Ex-SRES has 16 items reflecting patient confidence to continue to exercise under the conditions of bad weather, pain, exercise alone, busy, no support from others, lack of oxygen, vacation, fatigue, and no desire to exercise. Each item is measured with 1 to 10 points; 1 point means no self-confidence and 10 points mean very high self-confidence. The higher the score, the higher the confidence to keep exercising [24]. Ex-SRES is a single-factor structure with good internal consistency; the Cronbach coefficient is 0.917 [24]. Studies showed that Ex-SRES positively influenced adherence to regular exercise in patients with COPD [25].

mMRC is a questionnaire with 5 dimensions that provides a measurement method for perceived dyspnea. It is divided into 0 to 4 points: 0 points means do not feel dyspnea generally except for strenuous exercise; 1 point means short of breath when fast-walking on flat ground or when going uphill; 2 points means walking on flat ground slower than peers due to shortness of breath or having to stop walking to catch one's breath; 3 points means shortness of breath when walking on level ground for 100 meters or several minutes; and 4 points means inability to leave the room due to shortness of breath. The higher the score of the scale, the more severe the dyspnea [26].

Sample Size

The minimal clinically important difference in CAT score was 0.5 points. Previous studies showed that it could be improved by 2 points [19], and the estimated standard deviation was 4 points. With the test level of 0.05 on both sides and 80% test efficacy, the sample size estimation method was used for each group's measurement data. When the sample number of each group was equal, at least 31 cases were needed in each group. According to the estimation of 15% dropout rate in the study, at least 38 cases were needed in each group, and at least 76 cases were included in the sample size. The final sample size was 106.

Data Analysis

The intention-to-treat analysis was used in the study. The missing values of lost subjects used the last observation carried forward for analyzing. SPSS Statistics version 20.0 (IBM Corp) was used for data analysis, with the statistical significance set at $P<.05$. Sex, course of disease, education level, smoking status, and disease classification were analyzed using chi square tests. Age and BMI were analyzed by independent samples *t* tests. The changes in CAT, Ex-SRES, and mMRC scores in different intervention time periods (0, 3, and 6 months) were studied using the repeated measures analyses of variance. The paired samples *t* test was used to analyze the SGRQ at baseline and 6 months after the intervention. The effect comparison between the two groups in the sixth month used the difference between before and after the intervention for independent samples *t* test analysis. The intervention methods, intervention time, self-efficacy, and interaction among the aforementioned factors were analyzed using the regression modeling of the generalized estimating equation (GEE) of autocorrelation work structure to explore the significance of these factors in improving the quality of life of patients.

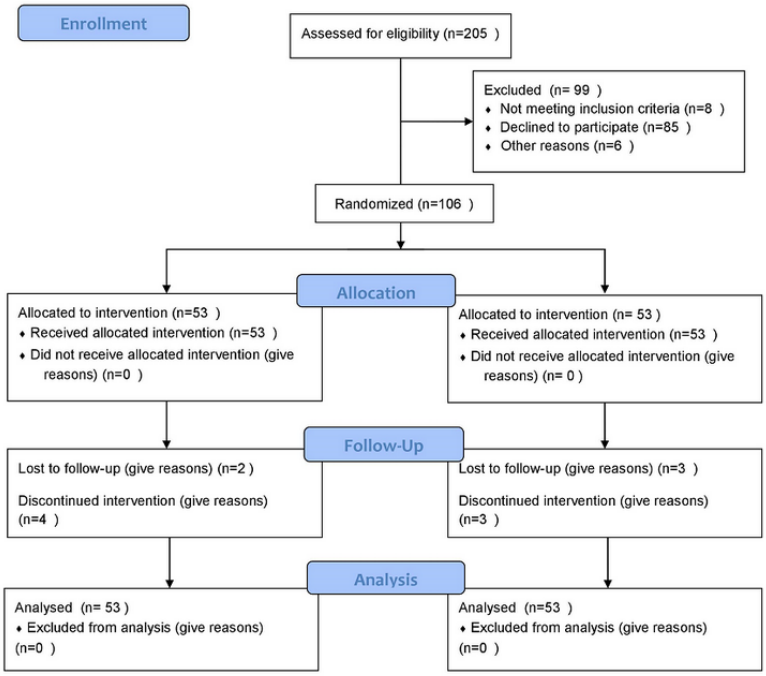
Results

Process of Recruitment and Research

The recruitment of patients with COPD started on October 1, 2018, and the process of follow-up was completed by October 7, 2019. Figure 3 shows the CONSORT flowchart of this study, which describes in detail the process of recruiting patients,

implementation of structured intervention program, observation effect, and loss to follow-up. During the study, 12 patients dropped out, and the dropout rate was 11.3%. In the PeR group, 6 patients dropped out, including 3 patients who felt that rehabilitation was of no use and dropped out immediately, 1 patient who got worse, and 2 who moved to other places. In the FtF group, 6 patients dropped out, including 2 patients who moved to other places, 2 patients who got worse, and 2 patients who felt rehabilitation was of no use.

Figure 3. Consolidated Standards of Reporting Trials flowchart.



Basic Demographic Characteristics and Measurements

No statistically significant difference was observed in socioeconomic information and measurement at baseline (P>.05; Table 2).

Main Evaluation Indexes

CAT was used to evaluate the health status of patients. A tendency to change in the CAT score was observed within 6 months (P=.001). The change trend of the CAT score over time is shown in Figure 4. No difference was found between the two groups (P=.53), and no interaction was observed between time and groups (P=.98). The results are shown in Table 3. The CAT score in the two groups changed from baseline to the third month, from the third month to the sixth month, and from baseline to the sixth month (P=.002, P=.70, and P=.001, respectively; Table 4).

Table 2. Basic demographic characteristics and measurements of the two groups.

Characteristics	PeR[a] group, (n=53)	FtF[b] group, (n=53)	Total, (n=106)	χ^2(δφ)	*t* (df)	*P* value
Sex, n (%)	—	—	—	0.06 (1)	—	.50
Male	44 (83)	43 (81)	87 (82)	—	—	—
Female	9 (17)	10 (19)	19 (18)	—	—	—
Age, mean (SD)	70.92 (6.38)	71.83 (7.60)	—	—	–0.67 (104)	.51
BMI[c], mean (SD)	22.21 (3.52)	21.27 (2.36)	—	—	1.62 (90.84)	.11
Disease duration, n (%)	—	—	—	0.05 (1)	—	>.99
<10 years	16 (30)	15 (28)	31 (29)	—	—	—
≥10 years	37 (70)	38 (72)	75 (71)	—	—	—
Education status, n (%)	—	—	—	3.58 (3)	—	.31
Primary school	12 (23)	13 (25)	25 (24)	—	—	—
Middle school	15 (28)	22 (42)	37 (35)	—	—	—
High school	17 (32)	14 (26)	31 (29)	—	—	—
Higher school	9 (17)	4 (8)	13 (12)	—	—	—
Smoking status, n (%)	—	—	—	0.00 (2)	—	>.99
No	12 (23)	12 (23)	24 (23)	—	—	—
Exsmoker	35 (66)	35 (66)	70 (66)	—	—	—
Current smoker	6 (11)	6 (11)	12 (11)	—	—	—
Disease classification[d], n (%)	—	—	—	6.69 (3)	—	.08
GOLD II	27 (50.9)	20 (38)	47 (44)	—	—	—
GOLD III	15 (28.3)	27 (51)	42 (40)	—	—	—
GOLD IV	11 (21)	6 (11)	17 (16)	—	—	—
Measurements, mean (SD)	—	—	—	—	—	—
CAT[e]	21.79 (6.85)	22.55 (6.48)	—	—	–0.58 (104)	.56
Ex-SRES[f]	72.25 (38.38)	71.48 (40.76)	—	—	0.10 (104)	.92
mMRC[g]	2.79 (0.66)	2.75 (0.71)	—	—	0.28 (104)	.78
SGRQ-S[h]	53.02 (19.90)	51.12 (18.63)	—	—	0.51 (104)	.61
SGRQ-A[i]	56.44 (23.96)	56.87 (22.47)	—	—	–0.09 (104)	.93
SGRQ-I[j]	45.82 (24.27)	44.92 (18.69)	—	—	0.22 (104)	.83
SGRQ-T[k]	50.24 (20.95)	49.57 (17.52)	—	—	0.18 (104)	.86

[a]PeR: Pulmonary Internet Explorer Rehabilitation.

[b]FtF: face to face.

[c]BMI: body mass index.

[d]GOLD: Global Initiative for Chronic Obstructive Lung Disease; GOLD II: FEV1/FVC<70%, 50%≤FEV1<80%; GOLD III: FEV1/FVC<70%, 30%≤FEV1<50%; GOLD IV: FEV1/FVC<70%, FEV1<30%.

[e]CAT: chronic obstructive pulmonary disease assessment test.

[f]Ex-SRES: exercise self-regulatory efficacy scale.

[g]mMRC: modified Medical Research Council scale.

[h]SGRQ-S: St George's Respiratory Questionnaire–System.

[i]SGRQ-A: St George's Respiratory Questionnaire–Activity.

[j]SGRQ-I: St George's Respiratory Questionnaire–Influence.

[k]SGRQ-T: St George's Respiratory Questionnaire–Total.

Figure 4. Chronic obstructive pulmonary disease assessment test score and its change over time.

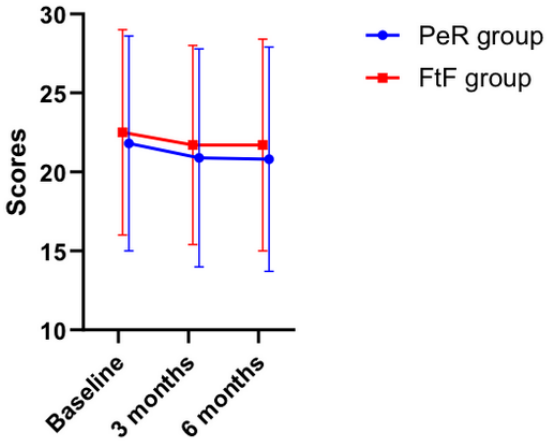


Table 3. Variation tendency of chronic obstructive pulmonary disease assessment test, exercise self-regulatory efficacy scale, and modified Medical Research Council scale scores in the two groups.

Item and group	CAT[a]		Ex-SRES[b]		mMRC[c]	
	$F_{1,3}$	*P* value	$F_{1,3}$	*P* value	$F_{1,3}$	*P* value
Time	7.78	.001	21.91	<.001	29.64	<.001
Group	0.39	.53	0.23	.63	0.00	>.99
Time×group	0.02	.98	2.02	.14	0.8	.43

[a]CAT: chronic obstructive pulmonary disease assessment test.

[b]Ex-SRES: exercise self-regulatory efficacy scale.

[c]mMRC: modified Medical Research Council scale.

Table 4. Comparison of chronic obstructive pulmonary disease assessment test, exercise self-regulatory efficacy scale, modified Medical Research Council scale, and St George's Respiratory Questionnaire scores between the two groups before and after intervention.

Outcomes	Baseline	After intervention (3 months)	After intervention (6 months)	$F_{1,3}$[a]/t_3[b] (*P* value)
CAT[c]				
PeR[d] group, mean (SD)	21.79 (6.85)	20.98 (6.99)	20.85 (7.11)	3.70 (.02)
FtF[e] group, mean (SD)	22.55 (6.48)	21.75 (6.25)	21.70 (6.69)	4.17 (.02)
F_1[f] (*P* value)	10.14 (.002)	—	—	—
F_2[g] (*P* value)	—	0.15 (.70)	—	—
F_3[h] (*P* value)	—	—	12.80 (.001)	—
Ex-SRES[i]				
PeR group, mean (SD)	72.25 (38.38)	85.36 (33.18)	80.53 (37.72)	17.22 (<.001)
FtF group, mean (SD)	71.48 (40.76)	78.49 (33.94)	78.25 (35.40)	6.47 (.008)
F_1 (*P* value)	54.10 (<.001)	—	—	—
F_2 (*P* value)	—	3.99 (.05)	—	—
F_3 (*P* value)	—	—	14.12 (<.001)	—
mMRC[j]				
PeR group, mean (SD)	2.79 (0.66)	2.51 (0.72)	2.40 (0.79)	14.73 (<.001)
FtF group, mean (SD)	2.75 (0.70)	2.58 (0.69)	2.36 (0.71)	15.77 (<.001)
F_1 (*P* value)	21.54 (<.001)	—	—	—
F_2 (*P* value)	—	14.06 (<.001)	—	—
F_3 (*P* value)	—	—	43.97 (<.001)	—
SGRQ-S[k]				
PeR group, mean (SD)	53.02 (19.90)	—	43.59 (23.63)	3.59 (.001)
FtF group, mean (SD)	51.12 (18.63)	—	45.33 (22.25)	2.89 (.006)
t[l] value (*P* value)	—	—	–1.10 (.27)	—
SGRQ-A[m]				
PeR group, mean (SD)	56.44 (23.96)	—	48.74 (24.28)	3.01 (.004)
FtF group, mean (SD)	56.87 (22.47)	—	53.46 (23.06)	2.89 (.006)
t value (*P* value)	—	—	–1.53 (.13)	—
SGRQ-I[n]				
PeR group, mean (SD)	45.83 (24.27)	—	33.27 (22.86)	3.89 (<.001)
FtF group, mean (SD)	44.92 (18.69)	—	38.63 (21.88)	3.65 (.001)
t value (*P* value)	—	—	–1.71 (.09)	—
SGRQ-T[o]				
PeR group, mean (SD)	50.24 (20.95)	—	39.66 (20.92)	4.06 (<.001)
FtF group, mean (SD)	49.57 (17.52)	—	44.24 (19.90)	3.84 (<.001)
t value (*P* value)	—	—	–1.78 (.08)	—

[a]$F_{1,3}$: repeated measures analysis of variance.

[b]$t_{3:}$ 6 months compared with baseline.

[c]CAT: chronic obstructive pulmonary disease assessment test.

[d]PeR: Pulmonary Internet Explorer Rehabilitation.

[e]FtF: face-to-face.

[f]F_1: 3 months compared with baseline.

[g]F_2: 6 months compared with 3 months.

[h]F_3: 6 months compared with baseline.

[i]Ex-SRES: exercise self-regulatory efficacy scale.

[j]mMRC: modified Medical Research Council scale.

[k]SGRQ-S: St George's Respiratory Questionnaire–System.

[l]*t*: comparison among groups.

[m]SGRQ-A: St George's Respiratory Questionnaire–Activity.

[n]SGRQ-I: St George's Respiratory Questionnaire–Influence.

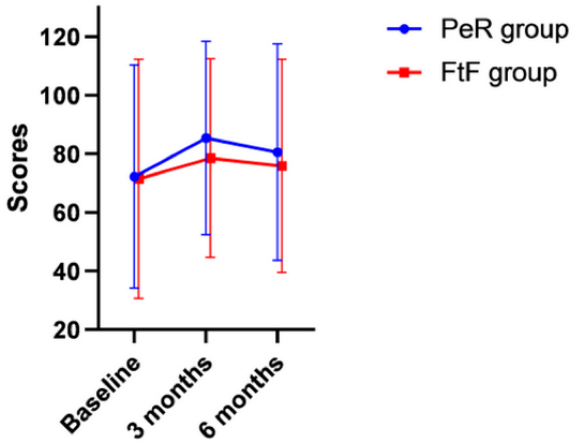
[o]SGRQ-T: St George's Respiratory Questionnaire–Total.

Secondary Evaluation Indexes

SGRQ was used to evaluate the symptom, activity, and influence of patients. Table 4 shows that the SGRQ-S, SGRQ-A, SGRQ-I, and SGRQ-T scores in the PeR group were improved from baseline to the sixth month (*P*=.001, *P*=.004, *P*<.001, and *P*<.001, respectively). The SGRQ-S, SGRQ-A, SGRQ-I, and SGRQ-T scores in the FtF group were improved (*P*=.006, *P*=.006, *P*=.001, and *P*<.001, respectively). In the sixth month, no significant difference was found between the measured value and the baseline value of SGRQ-S, SGRQ-A, SGRQ-I, and SGRQ-T in the two groups (*P*=.27, *P*=.13, *P*=.09, and *P*=.08, respectively).

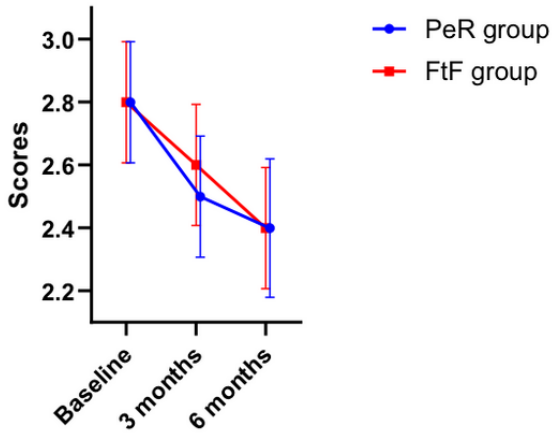
Ex-SRES was used to evaluate the patient sense of self-efficacy. A tendency to change in the CAT score was noted within 6 months (*P*<.001). The change trend of the Ex-SRES score over time is shown in Figure 5. No difference was found between the two groups (*P*=.63), and no interaction was observed between time and groups (*P*=.14). The results are shown in Table 3. The Ex-SRES score in the two groups changed from baseline to the third month, from the third month to the sixth month, and from baseline to the sixth month (*P*<.001, *P*=.05, and *P*<.001, respectively; Table 4).

Figure 5. Exercise self-regulatory efficacy scale score and its change over time.



mMRC was used to evaluate the status of dyspnea. A tendency to change in the CAT score was observed within 6 months (*P*<.001). The change trend of the mMRC score over time is shown in Figure 6. No difference was found between the two groups (*P*>.99), and no interaction was noted between time and groups (*P*=.43). The results are shown in Table 3. The mMRC score in the two groups changed from baseline to the third month, from the third month to the sixth month, and from baseline to the sixth month (*P*<.001, *P*<.001, and *P*<.001, respectively; Table 4).

Figure 6. Modified Medical Research Council scale score and its change over time.



Factors of Methods, Time, and Exercise Self-Regulatory Efficacy Scale Effects on Primary Outcome

Table 5 shows the analysis results using GEE. Only the Ex-SRES had a significant effect on the CAT score (*P*=.03), but the intervention method and time had no significant effect on the CAT score (*P*=.48 and *P*=.84, respectively). The partial regression coefficient of Ex-SRES to CAT was 0.81, and Exp (B) was 2.24. Interaction between intervention methods and Ex-SRES; interaction between intervention methods and intervention time; interaction between Ex-SRES and intervention time; and interaction among intervention methods, Ex-SRES, and intervention time had no statistically significant effect on the CAT score (*P*=.13, *P*=.85, *P*=.09, and *P*=.08, respectively).

Table 5. Analysis of the influence of intervention methods, intervention time, and self-efficacy on patients' quality of life using the generalized estimating equation.

Parameter	*B*	SE	Wald	*P* value	Exp *(B)*
Constant	–1.62	0.33	24.33	<.001	0.20
Intervention methods	–0.25	0.35	0.49	.48	0.78
Ex-SRES[a]	0.81	0.38	4.63	.03	2.24
Intervention time	–0.07	0.35	0.04	.84	0.93
Intervention methods×Ex-SRES	0.75	0.49	2.34	.13	2.12
Intervention methods×intervention time	0.10	0.51	0.03	.85	1.10
Ex-SRES×intervention time	0.97	0.56	2.96	.09	2.64
Intervention methods×Ex-SRES×intervention time	–1.86	1.05	3.16	.08	0.16

[a]Ex-SRES: exercise self-regulatory efficacy scale.

Discussion

Principal Findings

It is feasible and effective to implement PeR pulmonary rehabilitation based on the eCCM and self-efficacy theory to help patients with COPD relieve dyspnea symptoms, improve their self-efficacy, and improve their quality of life. The effect was the same as that of the face-to-face intervention.

Three months after the intervention, patient improved CAT scores maintained a subsequent effect and the mMRC scores continued to improve. Meanwhile, the Ex-SRES scores decreased after the peak in the third month instead of a subsequent effect, but it was still an improvement compared with before the intervention. All dimensions and total SGRQ scores of patients improved at the end of the intervention. Patient Ex-SRES scores correlated with CAT scores. Moreover, the improvement in health was 2.24 times higher in patients with high Ex-SRES scores than in those with low Ex-SRES scores.

Remote technology can be used for information support, information storage and management, telemonitoring, and teleconsultation. Common mobile health technologies include personal computers, tablets, smartphones, and so forth [27,28]. According to the review by McCabe et al [29], the self-management intervention of patients with COPD using mobile technology could improve the quality of life and activity level of patients, and the effect could be maintained for a period of time. However, research by Vorrink et al [30] showed that the evaluation and telemonitoring of physical activity, functional exercise ability, health-related quality of life, and other indicators of patients with COPD using smartphones and

websites could not improve the self-efficacy of patients. Thus, it is evident that telemonitoring alone cannot improve the health status of patients.

For patients with COPD, the increase in exercise ability and the change in adaptive behavior are the premises to improve patient health conditions [31]. In a previous study, patients were taught the skills of pulmonary rehabilitation training face to face, and health education content was provided to them. In addition, their training behaviors were evaluated, persuaded, and strengthened using short messaging service. The results showed that patients' self-efficacy, dyspnea symptoms, and quality of life were all improved [32]. Therefore, for patients with COPD, sports training should be considered as the core of pulmonary rehabilitation program and behavior intervention.

In this study, the incentive function of the integral mall was a way to promote behavioral change and maintenance. In addition, gamification is a good principle. The study by Tabak et al [33] showed that, through telemedicine, gamification strategies could increase patients' motivation for behavioral change and promote their active lives. This study applied the WeChat official account to carry out pulmonary rehabilitation intervention along with management and provided a pulmonary rehabilitation program centered on sports training for patients with COPD at home. During the process of implementation, the behavioral intervention aimed at enhancing self-efficacy could be realized through the design of WeChat's functional structure and the application of remote technology.

Many studies suggested that self-efficacy was a key cognitive (motivational) factor in adopting and maintaining self-management behaviors [34,35]. Zarski et al [36] proposed that the systematic development of planning skills and the maintenance of self-efficacy before or during internet-based interventions would help participants successfully complete treatments. Schwarzer's health action process approach points out that self-efficacy plays an important role in improving and maintaining health behavior [37]. Self-efficacy is a predictor of health and quality-of-life improvement in patients undergoing pulmonary rehabilitation [38].

This study was based on four kinds of information sources proposed by Bandura, mastery experience, alternative experience, verbal persuasion, and good physiological and emotional state, to promote the establishment of patient self-efficacy beliefs. Through the function modules of respiratory training, diet guidance, medication guidance, and sports training, patients could master the knowledge and skills of pulmonary rehabilitation, which could be defined as an information source related to mastery experience. Peer support and doctor-patient communication were promoted through comments and thumbs up interactions in Moments, which not only strengthened the frequency of information source stimulation of verbal persuasion but also built a good virtual social support system for patients' home-based rehabilitation training. The presentation of successful cases in Moments could construct an alternative experience. On account of the positive incentive function in the integral mall, patient confidence in completing the training could be strengthened and a good mood could be maintained. During the construction of the environment in the aforementioned individual action, social and psychological support environment realized a constantly strengthening behavior intervention process. Further, with the convenience of a mobile terminal across time and space, patients with COPD could continue to receive the stimulation of information sources in their daily life and gradually cultivate the confidence of persisting in long-time sports training. This study combined the strategy of behavior intervention in sociology with the content of telemedicine intervention to provide substantive medical services, making it a new telemedicine service model with warmth and humanistic care. Besides, such research production could provide inspiration for the application of remote technology combined with cognitive behavioral intervention strategies in COPD.

The systematic review by Cruz et al [39] showed that patient perceived difficulties was the reason for the failure to adhere to pulmonary rehabilitation during telemedicine intervention. Thus, organizing more training courses would help patients accept and use remote intervention technology. In this study, preintervention training was used to solve the problem—that is to say, patients were not proficient in technical operation skills. In addition, appointments and medical guidance could integrate medical resources. Patients who were far away from each other could access PeR through free social media and then join the pulmonary rehabilitation program. This was consistent with the concept of making pulmonary rehabilitation an available and affordable project proposed by the World Health Organization [40]. Thanks to the network advantage and easy operation of PeR, the response time of health care professionals was shortened, which helped promote complete feedback between professionals and patients.

The development of internet and mobile devices has promoted the development of telemedicine. Most of the telemedicine technologies applied in countries with developed medical resources intended to telemonitor and manage patients with COPD are expensive, such as remote monitoring systems [39], robots [41], and electronic health systems [42]. Remote technology applied in this study was a WeChat official account developed with certain functions. Due to a large number of patients with COPD in China, some technical and financial limitations still exist in realizing intelligent pulmonary rehabilitation management based on the internet. The number of active users of WeChat in China has exceeded 697 million [43]. WeChat is a widely used and free app on mobile terminals, and its function is similar to that of Facebook [43]. The WeChat app has a positive impact on health behavior. By acquiring information provided on WeChat, health knowledge can be increased, and the ability of public health decision making and action can be improved. The results of a survey indicated that 97.68% of respondents read health information through WeChat [44]. In this study, the network platform based on WeChat was used to implement pulmonary rehabilitation intervention. WeChat is an important information access port for users on mobile terminals. It can realize the interaction and communication between developers and subscribers by pushing graphics, video, and audio through multimedia [45]. At the same time, this study also guaranteed patient privacy and data security by conducting technical strategy, confidentiality agreements,

and patient coding. Using free social media to carry out pulmonary rehabilitation intervention and management can provide a reference for similar service projects in countries with resource shortages.

App Design

The development of telemedicine technology should be a process of interdisciplinary integration. A previous study showed that the continued participation of multiple stakeholders and users was critical to the design and development of successful eHealth solutions [46]. Participants with a multidisciplinary context have different resources and experience, and thus they can innovate and improve the existing service model. At the same time, the user-centered design concept enables the developers of telemedicine technology to give priority to patients' needs, which helps developers establish specific app functions and hardware as early as possible [47]. The existing methods for user-centered design include workshops and focus groups, paper prototyping, sketching, thinking aloud, scenarios, storytelling, interviews and field studies, questionnaires, and other methods [48]. In this study, respiratory experts, clinicians, rehabilitation practitioners, nurses, research assistants, and patients with COPD all participated in the process of development and application of PeR. Through the establishment of focus groups, the views of different stakeholders were absorbed and intervention content was determined along with the presentation form of PeR (graphic, audio, video, background color, font, font size, etc).

The results of the systematic review showed that the technology acceptance theory and its basic behavior theory could be applied to the acceptance research of telemedicine [49]. At present, technology acceptance models are commonly used. They evaluate the availability and acceptance of telemedicine technology using usefulness cognition, ease-of-use cognition, and use intention [50]. However, they do not mention how to carry out behavioral intervention strategies in the process of designing the functional structure of remote technology, which may obviously affect patients' acceptance of remote technology. The eCCM used in this study is a theoretical model of telemedicine intervention proposed by Gee and colleagues [17] based on the CCM. The model includes 8 interdependent and interactive parts: the eCommunity and an informatics framework, health system enhancements, delivery system design enhancements, self-management support enhancements, clinical decision support enhancements, clinical information systems enhancements, addition of eHealth education to the CCM, and communication and addition of the complete feedback loop [17]. A systematic review showed that the implementation of the CCM in primary care could significantly improve the medical effect on patients, improve the quality of life of patients, and reduce the social burden [51]. Evidence showed that eHealth tools could be used to enhance patient self-management behaviors, revise the CCM, and support effective interaction between patients and providers to improve health outcomes [52,53]. Through the components of the eCCM, this study expounded its core characteristics, constructed the functional structure of PeR, and combined the behavioral intervention strategy with the implementation process. The results confirmed the effectiveness of PeR in improving patient dyspnea symptoms, quality of life, and self-efficacy.

The evaluation design of the intervention effect is a necessary part of the health service project design. The multidimensional quality-of-life questionnaire is an efficient evaluation tool in the process of remote intervention. During the course of this study, 82% of patients were tired of using SGRQ and willing to use CAT. Although SGRQ is reliable and effective, it has many items and complex scoring methods, and therefore its widespread use in clinical work is difficult [54,55]. SGRQ is divided into 3 dimensions, with a total of 50 items and a long evaluation time. However, CAT has only 8 questions, with a short evaluation time. CAT is simple and easy to operate, providing a reliable measurement of the COPD health status. CAT and SGRQ have been shown to have a good correlation [56]. In this study, all patients were aged over 60 years. The numerous items of the questionnaire increased the difficulty of implementation. This suggested that patient personal experience of the assessment tool should be taken into account when choosing the online self-report effect assessment tool; a concise and effective assessment tool is the best choice. The combination of convenient, effective, and reliable effect evaluation methods and mobile health is a solution worth considering in telemedicine service.

Limitations

This study had some limitations. The sample size was small, and the main participants were patients with COPD in China. Hence, data on other races were lacking. Consultation during the implementation of the study may affect the use of PeR, such as the immediate response and extended response time of medical professionals. Despite the best efforts to ensure consistency, the implementation process can still be affected by inevitable factors. All outcomes of the study were self-reported. This is one of the limitations of this study, but it made sense because the study referred to a self-management tool. The clinical outcomes will be addressed in the next study.

Conclusions

This study used PeR to intervene and manage the pulmonary rehabilitation of patients with COPD at home, which could effectively improve their sense of self-efficacy and quality of life and also alleviate their symptoms. The implementation of PeR confirmed that the eCCM combined with behavioral strategy intervention, based on the self-efficacy theory, could be realized using remote technology. During the application of remote technology, it is worth considering how to construct a free platform that can integrate resources into daily life, especially in the field of COPD and other chronic diseases. Moreover, in the development process of remote technology, we should pay attention to the transformation of theoretical models and combine behavioral intervention at the same time.

Acknowledgments

This study was supported by the Wuxi Science and Technology Development Fund (grant/award number: WX0302B010507180074PB) and the Launching Fund of the Research Office of Chronic Disease Management and Rehabilitation, Wuxi School of Medicine, Jiangnan University. The authors thank Lei Chen and Hanjun Zhang, who gave us a lot of suggestions on the development of PeR. They also thank all patients who participated in this study. Finally, they thank MedSci for helping polish the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT - EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 505 KB - jmir_v22i4e17089_app1.pdf]

Multimedia Appendix 2
Computer software copyright registration certificate.
[PDF File (Adobe PDF File), 592 KB - jmir_v22i4e17089_app2.pdf]

Multimedia Appendix 3
English language version of the functional modules.
[PNG File , 322 KB - jmir_v22i4e17089_app3.png]

References

1. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease. Pneumologie 2017 Jan 23;71(01):9-14. [doi: 10.1055/s-0042-121903]
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196 [FREE Full text] [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
3. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 2017 Sep;5(9):691-706 [FREE Full text] [doi: 10.1016/S2213-2600(17)30293-X] [Medline: 28822787]
4. Alison JA, McKeough ZJ, Johnston K, McNamara RJ, Spencer LM, Jenkins SC, Lung Foundation Australiathe Thoracic Society of AustraliaNew Zealand. Australian and New Zealand Pulmonary Rehabilitation Guidelines. Respirology 2017 May;22(4):800-819. [doi: 10.1111/resp.13025] [Medline: 28339144]
5. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, ATS/ERS Task Force on Pulmonary Rehabilitation. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med 2013 Oct 15;188(8):e13-e64. [doi: 10.1164/rccm.201309-1634ST] [Medline: 24127811]
6. Puhan MA, Lareau SC. Evidence-based outcomes from pulmonary rehabilitation in the chronic obstructive pulmonary disease patient. Clin Chest Med 2014 Jun;35(2):295-301. [doi: 10.1016/j.ccm.2014.02.001] [Medline: 24874125]
7. Jones A, Taylor A, Gowler H, O'Kelly N, Ghosh S, Bridle C. Systematic review of interventions to improve patient uptake and completion of pulmonary rehabilitation in COPD. ERJ Open Res 2017 Jan;76(11-12):790 [FREE Full text] [doi: 10.1183/23120541.00089-2016] [Medline: 28154821]
8. Yohannes AM, Connolly MJ. Pulmonary rehabilitation programmes in the UK: a national representative survey. Clin Rehabil 2004 Jun;18(4):444-449. [doi: 10.1191/0269215504cr736oa] [Medline: 15180129]
9. Keating A, Lee A, Holland AE. What prevents people with chronic obstructive pulmonary disease from attending pulmonary rehabilitation? A systematic review. Chron Respir Dis 2011;8(2):89-99. [doi: 10.1177/1479972310393756] [Medline: 21596892]
10. Rochester CL, Vogiatzis I, Holland AE, Lareau SC, Marciniuk DD, Puhan MA, ATS/ERS Task Force on Policy in Pulmonary Rehabilitation. An Official American Thoracic Society/European Respiratory Society Policy Statement: enhancing implementation, use, and delivery of pulmonary rehabilitation. Am J Respir Crit Care Med 2015 Dec 01;192(11):1373-1386. [doi: 10.1164/rccm.201510-1966ST] [Medline: 26623686]
11. Field MJ. Telemedicine: a guide to assessing telecommunications in healthcare. J Digit Imaging 1997 Aug;10(3 Suppl 1):28 [FREE Full text] [doi: 10.1007/bf03168648] [Medline: 9268830]
12. Vitacca M, Montini A, Comini L. How will telemedicine change clinical practice in chronic obstructive pulmonary disease? Ther Adv Respir Dis 2018;12:1753465818754778 [FREE Full text] [doi: 10.1177/1753465818754778] [Medline: 29411700]

13. McCullough AR, Ryan C, Macindoe C, Yii N, Bradley JM, O'Neill B, et al. Behavior change theory, content and delivery of interventions to enhance adherence in chronic respiratory disease: a systematic review. Respir Med 2016 Jul;116:78-84 [FREE Full text] [doi: 10.1016/j.rmed.2016.05.021] [Medline: 27296825]
14. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: a randomized trial. Med Care 2006 Nov;44(11):964-971. [doi: 10.1097/01.mlr.0000233678.80203.c1] [Medline: 17063127]
15. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010 Mar;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
16. Gammon D, Johannessen LK, Sørensen T, Wynn R, Whitten P. An overview and analysis of theories employed in telemedicine studies. A field in search of an identity. Methods Inf Med 2008;47(3):260-269. [Medline: 18473093]
17. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LMS. The eHealth enhanced chronic care model: a theory derivation approach. J Med Internet Res 2015 Apr;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
18. Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the Chronic Care Model in the new millennium. Health Aff (Millwood) 2009 Jan;28(1):75-85 [FREE Full text] [doi: 10.1377/hlthaff.28.1.75] [Medline: 19124857]
19. Chronic obstructive pulmonary disease group, respiratory branch, Chinese Medical Association. Guidelines for diagnosis and treatment of chronic obstructive pulmonary disease (2013 Revision). Chinese Journal of medical frontier 2014(02):67-80.
20. Jones PW, Brusselle G, Dal Negro RW, Ferrer M, Kardos P, Levy ML, et al. Properties of the COPD assessment test in a cross-sectional European study. Eur Respir J 2011 May 12;38(1):29-35. [doi: 10.1183/09031936.00177210]
21. Peruzza S, Sergi G, Vianello A, Pisent C, Tiozzo F, Manzan A, et al. Chronic obstructive pulmonary disease (COPD) in elderly subjects: impact on functional status and quality of life. Respir Med 2003 Jun;97(6):612-617 [FREE Full text] [doi: 10.1053/rmed.2003.1488] [Medline: 12814144]
22. Doll H, Duprat-Lomon I, Ammerman E, Sagnier P. Validity of the St George's respiratory questionnaire at acute exacerbation of chronic bronchitis: comparison with the Nottingham health profile. Qual Life Res 2003 Mar;12(2):117-132. [doi: 10.1023/a:1022280532276] [Medline: 12639059]
23. Singh SJ, Sodergren SC, Hyland ME, Williams J, Morgan MD. A comparison of three disease-specific and two generic health-status measures to evaluate the outcome of pulmonary rehabilitation in COPD. Respir Med 2001 Jan;95(1):71-77 [FREE Full text] [doi: 10.1053/rmed.2000.0976] [Medline: 11207021]
24. Davis AH, Figueredo AJ, Fahy BF, Rawiworrakul T. Reliability and validity of the Exercise Self-Regulatory Efficacy Scale for individuals with chronic obstructive pulmonary disease. Heart Lung 2007;36(3):205-216. [doi: 10.1016/j.hrtlng.2006.08.007] [Medline: 17509427]
25. Tsai Y, Chen J, Davis AH, Yen W, Lin Y, Chiang L, et al. Development of the Chinese-version of the exercise self-regulatory efficacy scale for patients with chronic obstructive pulmonary disease. Heart Lung 2018;47(1):16-23. [doi: 10.1016/j.hrtlng.2017.10.007] [Medline: 29132846]
26. Munari AB, Gulart AA, Dos Santos K, Venâncio RS, Karloh M, Mayer AF. Modified Medical Research Council dyspnea scale in GOLD classification better reflects physical activities of daily living. Respir Care 2018 Jan;63(1):77-85 [FREE Full text] [doi: 10.4187/respcare.05636] [Medline: 28874609]
27. Gregersen TL, Green A, Frausing E, Ringbæk T, Brøndum E, Suppli Ulrik C. Do telemedical interventions improve quality of life in patients with COPD? A systematic review. Int J Chron Obstruct Pulmon Dis 2016;11:809-822 [FREE Full text] [doi: 10.2147/COPD.S96079] [Medline: 27143872]
28. Silva BMC, Rodrigues JJPC, de la Torre Diez I, López-Coronado M, Saleem K. Mobile health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
29. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Dec 23;5:CD011425. [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
30. Vorrink SNW, Kort HSM, Troosters T, Zanen P, Lammers JJ. Efficacy of an mHealth intervention to stimulate physical activity in COPD patients after pulmonary rehabilitation. Eur Respir J 2016 Oct;48(4):1019-1029. [doi: 10.1183/13993003.00083-2016] [Medline: 27587557]
31. Spruit MA, Pitta F, McAuley E, ZuWallack RL, Nici L. Pulmonary rehabilitation and physical activity in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2015 Oct 15;192(8):924-933. [doi: 10.1164/rccm.201505-0929CI] [Medline: 26161676]
32. Jiang YY, Zhou CX, Hao JF, Hua J. Application of short message service for community-based pulmonary rehabilitation. Chin J Rehabil Theory Pract 2016;22(05):596-600.
33. Tabak M, Dekker-van Weering M, van Dijk H, Vollenbroek-Hutten M. Promoting daily physical activity by means of mobile gaming: a review of the state of the art. Games Health J 2015 Dec;4(6):460-469. [doi: 10.1089/g4h.2015.0010] [Medline: 26397179]
34. Bourbeau J, Nault D, Dang-Tan T. Self-management and behaviour modification in COPD. Patient Educ Couns 2004 Mar;52(3):271-277. [doi: 10.1016/S0738-3991(03)00102-2] [Medline: 14998597]
35. Rice K, Bourbeau J, MacDonald R, Wilt TJ. Collaborative self-management and behavioral change. Clin Chest Med 2014 Jun;35(2):337-351. [doi: 10.1016/j.ccm.2014.02.004] [Medline: 24874129]

36. Zarski A, Berking M, Reis D, Lehr D, Buntrock C, Schwarzer R, et al. Turning good intentions into actions by using the health action process approach to predict adherence to internet-based depression prevention: secondary analysis of a randomized controlled trial. J Med Internet Res 2018 Jan 11;20(1):e9 [FREE Full text] [doi: 10.2196/jmir.8814] [Medline: 29326097]
37. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
38. Bentsen SB, Wentzel-Larsen T, Henriksen AH, Rokne B, Wahl AK. Self-efficacy as a predictor of improvement in health status and overall quality of life in pulmonary rehabilitation—an exploratory study. Patient Educ Couns 2010 Oct;81(1):5-13. [doi: 10.1016/j.pec.2009.11.019] [Medline: 20356700]
39. Cruz J, Brooks D, Marques A. Home telemonitoring in COPD: a systematic review of methodologies and patients' adherence. Int J Med Inform 2014 Apr;83(4):249-263. [doi: 10.1016/j.ijmedinf.2014.01.008] [Medline: 24529402]
40. Gimigliano F, Negrini S. The World Health Organization "Rehabilitation 2030: a call for action". Eur J Phys Rehabil Med 2017 Apr 04;53(2):155. [doi: 10.23736/S1973-9087.17.04746-3] [Medline: 28382807]
41. 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]
42. Farmer A, Williams V, Velardo C, Shah SA, Yu L, Rutter H, et al. Self-management support using a digital health system compared with usual care for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2017 May 03;19(5):e144 [FREE Full text] [doi: 10.2196/jmir.7116] [Medline: 28468749]
43. Feng S, Liang Z, Zhang R, Liao W, Chen Y, Fan Y, et al. Effects of mobile phone WeChat services improve adherence to corticosteroid nasal spray treatment for chronic rhinosinusitis after functional endoscopic sinus surgery: a 3-month follow-up study. Eur Arch Otorhinolaryngol 2017 Mar;274(3):1477-1485. [doi: 10.1007/s00405-016-4371-0] [Medline: 27822702]
44. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
45. Chuxin H, Dan W. The current situation, types and development trend of WeChat public address. News and Writing 2015(07):5-9.
46. 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. [doi: 10.2196/mhealth.5127] [Medline: 27282195]
47. Wang Z. Research on innovative design of mobile Internet products based on service design concept [Thesis]. Jiangxi: East China University of Technology; 2014.
48. Arsand E, Demiris G. User-centered methods for designing patient-centric self-help tools. Inform Health Soc Care 2008 Sep;33(3):158-169. [doi: 10.1080/17538150802457562] [Medline: 18850399]
49. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
50. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
51. Yeoh E, Wong M, Wong E, Yam C, Poon C, Chung R, et al. Benefits and limitations of implementing Chronic Care Model (CCM) in primary care programs: a systematic review. Int J Cardiol 2018 May 01;258:279-288. [doi: 10.1016/j.ijcard.2017.11.057] [Medline: 29544944]
52. Siminerio LM. The role of technology and the chronic care model. J Diabetes Sci Technol 2010 Mar 01;4(2):470-475 [FREE Full text] [doi: 10.1177/193229681000400229] [Medline: 20307409]
53. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
54. Weldam SW, Schuurmans MJ, Liu R, Lammers JJ. Evaluation of Quality of Life instruments for use in COPD care and research: a systematic review. Int J Nurs Stud 2013 May;50(5):688-707. [doi: 10.1016/j.ijnurstu.2012.07.017] [Medline: 22921317]
55. Karloh M, Fleig Mayer A, Maurici R, Pizzichini MM, Jones PW, Pizzichini E. The COPD assessment test: what do we know so far? A systematic review and meta-analysis about clinical outcomes prediction and classification of patients into gold stages. Chest 2016 Feb;149(2):413-425. [doi: 10.1378/chest.15-1752] [Medline: 26513112]
56. Sciriha A, Lungaro-Mifsud S, Scerri J, Magro R, Camilleri L, Montefort S. Health status of COPD patients undergoing pulmonary rehabilitation: a comparative responsiveness of the CAT and SGRQ. Chron Respir Dis 2017 Nov;14(4):352-359 [FREE Full text] [doi: 10.1177/1479972317694622] [Medline: 29081264]

Abbreviations

BMI: body mass index
CAT: chronic obstructive pulmonary disease assessment test
CCM: chronic care model
CONSORT: Consolidated Standards of Reporting Trials
COPD: chronic obstructive pulmonary disease
eCCM: electronic health–enhanced chronic care model
eHealth: electronic health
Ex-SRES: exercise self-regulatory efficacy scale
FtF: face-to-face
FEV1: forced expiratory volume in 1 second
FVC: forced vital capacity
GEE: generalized estimating equation
mMRC: modified Medical Research Council scale
PeR: Pulmonary Internet Explorer Rehabilitation
SGRQ: St George's Respiratory Questionnaire
SGRQ-A: St George's Respiratory Questionnaire–Activity
SGRQ-I: St George's Respiratory Questionnaire–Influence
SGRQ-S: St George's Respiratory Questionnaire–System
SGRQ-T: St George's Respiratory Questionnaire–Total

Edited by G Eysenbach; submitted 17.11.19; peer-reviewed by L Harst, J Lei; comments to author 09.12.19; revised version received 08.02.20; accepted 12.03.20; published 21.04.20.

Please cite as:
Jiang Y, Liu F, Guo J, Sun P, Chen Z, Li J, Cai L, Zhao H, Gao P, Ding Z, Wu X
Evaluating an Intervention Program Using WeChat for Patients With Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial
J Med Internet Res 2020;22(4):e17089
URL: http://www.jmir.org/2020/4/e17089/
doi:10.2196/17089
PMID:32314971

Original Paper

Online Interventions for the Selective Prevention of Illicit Drug Use in Young Drug Users: Exploratory Study

Tina Tomazic[1], PhD; Olivera Stanojevic Jerkovic[2,3], PhD

[1]Institute of Media Communications, Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[2]Medical Faculty, Department of Public Health, University of Maribor, Maribor, Slovenia

[3]National Institute of Public Health, Maribor, Slovenia

Corresponding Author:
Tina Tomazic, PhD
Institute of Media Communications
Faculty of Electrical Engineering and Computer Science
University of Maribor
Koroska 46
Maribor, 2000
Slovenia
Phone: 386 41704302
Fax: 386 22207272
Email: tina.tomazic@um.si

Abstract

Background: Digital technologies have a major impact on the daily lives of young people and are also used to seek information on and help with drug-related issues online.

Objective: The aim of this article was to analyze current online interventions for young drug users in Slovenia, with the purpose of contributing to the development of guidelines and key recommendations for effective online interventions.

Methods: This study was part of the project Click for Support. We performed a keyword search, received input from national experts in the field of drug prevention, and conducted an assessment of recognized national online interventions through workshop-based discussions with the target group of 20 young drug users.

Results: The current online intervention services in Slovenia are satisfactory but are still not sufficiently recognized. The most important issues for young drug users were the design and functionality of the online intervention, presence of a clear structure, possibility of using it on smartphones, comprehensive and quick professional feedback, and data security. Playful elements and the ability to share (experiences) with other or former users were also recognized as important.

Conclusions: With effective online interventions, we can include more young drug users, facilitate access to a more affordable service, provide quick professional feedback on patterns of consumption, increase knowledge about the effects and consequences of drugs, and support the reduction or cessation of drug use. From the public health perspective, it is challenging to provide drug interventions broadly to the target group and, hence, decrease inequities.

(J Med Internet Res 2020;22(4):e17688) doi:10.2196/17688

KEYWORDS

web-based; intervention; prevention; internet

Introduction

The growth of social media and increasing popularity of interactive social media platforms [1] have revolutionized methods of communication and influenced the way interactions are handled [2]. Most members of the younger generation are extremely quick to seize opportunities for participation in the new, freely accessible and low-cost model of new technologies [3]. Social networks have also changed with the development of the internet [4]. Furthermore, the increased popularity and capabilities of the internet have led to a revolution in the provision of health-related information and treatment [5].

Public health strategies and policies addressing issues of illicit drug use in Europe are heterogeneous [6]. Traditional public health and clinical interventions, with the central role of health professionals and passive role of patients [7], cannot fully

address this need due to resource constraints, limited access to a hard-to-reach population, and other factors. These conditions have spurred innovative approaches to health education and promotion [8]. Because digital technologies such as the internet and smartphone apps have a major impact on the daily lives of young people, there is a greater possibility that young people will search for drug-related assistance online. Approximately 95% of North Americans, 68.9% of the populations of Australia and Oceania, and 85.2% of Europeans use the internet [9]. The United Nations Office on Drugs and Crime describes the use of illicit drugs as a complex health condition that has social, psychological, and biological dimensions and consequences [10].

Online interventions, or web-based interventions (WBIs), are defined as a professional service for selective prevention that is delivered via the internet; includes interactive elements, computer-assisted behavior therapies, education, prevention, and information interventions; and provides individual feedback for young drug users [11,12]. It is possible to deliver interventions to large numbers of people at relatively low cost and ensure that the intervention is accessible 24 hours a day. Therefore, it is available at critical moments, enables anonymity, increases engagement through the use of interactive methods such as video streaming and sharing resources, is economic to run and maintain, and can effectively change health risk behaviors and their determinants [13-15]. Young people with a pattern of drug use can be considered experimental and problematic and are not reached currently. WBIs can offer specialized services in rural areas where distances are too great for easy access to drug treatment centers, which are often concentrated in urban areas. WBIs could also prove to be a cost-effective way of providing support to a larger number of clients than traditional treatment centers, which have limited capacity and human resources [16]. Interventions provided over the internet can overcome traditional barriers to accessing health services [17]. With wide reach and user engagement, social media tools offer a phenomenal opportunity to use social interactions to engage young people in behavior change interventions and to foster socially supportive communities for quitting [18] abuse of illicit substances. These techniques are perceived as reliable, efficient, and able to provide users with useful information and skills, although several aspects require further, in-depth assessment [7]. Online interventions should provide easy access to young people, especially those who would not seek help or advice in a conventional way and cannot be reached through traditional approaches. Another reason why online interventions should be used in day-to-day preventive work is the recent development of new psychoactive substances (NPS), the so-called “legal highs.” According to the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), there is a general lack of services in the European Union addressing young illicit drug users [11]. As Mounteney et al [19] reported, the purchase of a low threshold over the internet is typical for NPS, as the internet facilitates the movement of products, money, and information across global borders. Social media also plays a role in facilitating the interaction, advertising, and marketing of NPS.

According to Caudevilla [7], WBI also has limitations, as an online forum provides very limited information compared with genuine personal interviews and mediation. Many drug users do not want to enquire about traditional health services because of the possibility of being evaluated or the fear of a professional’s moral prejudices.

There is enough evidence on the effectiveness of online interventions in mental disorders [20-22], harmful alcohol consumption [15,23-34], and smoking [13,18,35-37], whereas the literature is not as cohesive in the field of illicit drugs. Meta-analyses have shown that online mediation has a clear positive effect on the reduction of cannabis use [38-40]. These findings are supported by the evaluation of the Quit the Shit project, which records very positive results in reducing the frequency and quantity of cannabis use [41]. Despite the potential for reaching drug users online, the number of quality intervention choices is still limited, and their effects have not been evaluated sufficiently [42]. In many parts of the world, services designed for drug users simply do not exist.

The aim of this article was to analyze the current online services for young drug users in Slovenia, with the purpose of contributing to the development of guidelines for effective online interventions and to increase the awareness of the importance of online counseling in preventive work with young drug users.

Methods

The study was a part of the project Click for Support, coordinated by the Landschaftsverband Westfalen-Lippe LWL-Coordination Office for Drug-Related Issues, Münster, Germany, and the National Institute of Public Health of RS as a partner. The project consisted of several phases. First, we reviewed and evaluated online interventions according to predetermined EMCDDA quality criteria, followed by workshops and surveys conducted with young drug users. Finally, key recommendations, or guidelines, for developing effective WBIs were designed. Based on the results of the questionnaire and group discussion and to reach consensus on the final key recommendations, we conducted two rounds [14] of an online Delphi study method to structure the group communication process [43], with the help of 90 external experts. The experts identified strategies and provided recommendations for effective WBIs.

The target group was 20 young people between the ages of 15 and 21 years (average, 17.1 years) with risky consumption of illicit drugs and who had problems with illegal drugs – either previously or at the time of the study [11] (Table 1). Most of the participants smoked cannabis, and some used cocaine, heroin, ecstasy, or amphetamines. Participants were users of the Program Centre for Drug Prevention in Maribor and had been sent to the Program Centre from different institutions including schools and organizations. Some of the participants were long-term Program Centre users, some were just experimenters, and some were sent from schools for a quick intervention.

Table 1. Age and gender distribution of the 20 workshop participants.

Characteristic	n (%)
Age group, years	
<14	0 (0)
14-15	4 (20)
16-17	8 (40)
18-21	8 (40)
>21	0 (0)
Gender	
Male	11 (55)
Female	9 (45)

To determine the current knowledge of WBIs and their use by young drug users, research was performed using a keyword search (intervention, prevention, offer, new psychoactive drugs, legal highs, illegal drugs, cannabis, cocaine, heroin, app, website, smart phone, internet, social media, chat, forum) and by inquiry among national experts in the field of drugs.

The study was prepared according to the following EMCDDA quality criteria:

1. Interventions should be web-based (eg, websites, apps, or social media app).

2. Interventions should include interactive elements that require the user to do something actively to receive individual feedback.

3. Interventions should be professional services (ie, not simple chats between consumers).

4. The target group should be young drug consumers.

5. The focus should be on illicit drugs, ideally NPS.

6. The effectiveness of the intervention should have been evaluated scientifically.

To show to which extent the criteria were met by the included services, the interventions were ranked according to the number of criteria they fulfilled (“A” for one criterion, “AAAAAA” for all 6 criteria) [11].

The next phase was the assessment of national online interventions with the inclusion of the target group. The first part of the survey addressed the general interest of young people to use WBIs and their prior knowledge and experience with these WBIs, including which elements they had already used. In addition, participants were asked whether anything was missing from the WBIs and which devices they would use to access online services. In the end, the participants were asked to state what features the WBI should have and what kind of information is important. In the section dealing with specific national WBIs, participants were asked whether they wanted to use the app and which specific parts they would use. They were invited to evaluate the attractiveness and usefulness of the service on a scale from 1 to 10, stating which aspects they liked. They were also asked which aspects needed improvement, what they thought the specific service was missing, and, finally, if they would recommend the app to their friends [11]. It is important to gather ideas to motivate users to remain on the website. We suggest incorporating interactive features, such as educational games, fun apps that attract the user to the website, and, of course, consultant feedback. Counselors are expected to have a high level of competence; be kind; give a sense of security, confidence, and motivation; foster a personal connection; and be open. Further, we suggest counseling without a moralizing attitude, but with an attitude that accepts and motivates users. The target group rejects services that promote strict abstinence as the only possible goal, and, consequently, they will not return to such a website.

Participation in the survey and workshop was free and anonymous. The minimum age was 15 years, so we did not request special permission from parents or guardians. Participants were provided a verbal explanation of the course and goal of the research within the project and were provided anonymity. Any data requiring identification of persons could not be included. Participants were offered the opportunity to participate in bowling as an incentive.

Results

Results of the Expert Rating of Existing Online Interventions

In Slovenia, 3 websites met the criteria, 1 of which already met the transitional criterion to include expert advice (Table 2).

Table 2. Rating of web-based interventions according to the European Monitoring Centre for Drugs and Drug Addiction criteria.

Intervention	Criteria						Rating[g]		
	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]		Online since	URL
DrogArt	A	A	A	A	A		AAAAA	2006 (Reduser app, 2013)	https://www.drogart.org/
Med.over.net	A	A	A				AAA	2000	https://med.over.net/
This is me	A	A	A				AAA	2001	http://www.tosemjaz.net/

[a]Interventions should be web-based (eg, websites, apps, or social media apps).

[b]Interventions should include interactive elements that require the user to do something actively to receive individual feedback.

[c]Interventions should be professional services (ie, not simple chats between consumers).

[d]The target group should be young drug consumers.

[e]The focus should be on illicit drugs, ideally new psychoactive substances.

[f]The effectiveness of the intervention should have been evaluated scientifically.

[g]A indicates it met one criterion, while AAAAAA means it met 6 criteria.

The website for the DrogArt program includes the associated Reduser app, which represents a rather new approach. It is an interactive tool for self-help that can be used to reduce use of or quit substances. In this app, users record their consumption patterns (consumer logs), feelings, desires, activities, and goals. They can also contact experts for help. With Reduser, providers are striving to help reduce user stress in the situations they face with drugs as well as the stress of their relatives and friends. Expert feedback is available on the DrogArt website via email, apps, forums, Facebook, Skype, and conventional consulting [44]. The target group includes all types of illicit drug and alcohol users, especially high school and university students and people who use club drugs and cocaine.

The website for the Med.Over.Net program answers questions about a healthy lifestyle, exercise, and nutrition; raises awareness on these topics; and offers advice. The website contains information and contacts of experts and institutions that can provide help and advice. Expert feedback is provided through a forum, where experts act as moderators, always encourage users to get help or initiate therapy, and remind them of the effects and consequences of drug use. They also enable communication with medical doctors and professionals [45]. The target group also includes drug users, regardless of their age.

The website for the To sem jaz (“This is me”) community deals not only with the topic of drugs but also all the issues that are important during adolescence, in particular, good self-esteem. Its purpose is to provide youngsters with anonymous, quick, and free access to expert advice from professionals in medicine, psychology, and social work. The website offers information primarily in the form of a forum for questions that are answered by experts. The issues discussed include the side effects of drugs, desires and fears to test substances, feelings under the influence of drugs, and the length of time some medicines may remain in the body or organism [46]. The target group includes adolescents 13-17 years old, especially high school students.

Reactions to Web-Based Interventions and the Rating of Existing Online Interventions by the Target Group

All 20 adolescents from the target group completed a questionnaire during the workshops. This evaluation of the online interventions revealed that a large percentage of the target group did not use online interventions, although 85% (17/20) were still interested in the use of the selected WBIs. There were positive responses to many parts of the selected WBIs, such as the quality and overall provision of information, design, and possibility of interacting with other users (Table 3).

Table 3. Results of the workshop discussion with the target group of adolescents.

Response or issue	DrogArt	Med.over.net	This is me
Number of participants who would like to try the website, n (%)	15 (75)	13 (65)	9 (45)
Number of participants interested in information about drugs, n (%)	18 (90)	17 (85)	16 (80)
Overall attractiveness of the website	Website is attractive	Website is not attractive (advertising is disturbing)	Website is attractive, especially for younger users
Most attractive feature of the website	Set up and structure	Quality of the information	Set up and structure
Effectiveness of the website	Seems to be effective	Does not seem to be effective	Seems to be effective
Number of participants who would recommend it to others, n (%)	15 (75)	8 (40)	7 (35)

The most important issues for the young people in the workshop discussions were the design; a clear structure; functionality, especially the possibility of using the service on smartphones; professional feedback; and data security. Comprehensive and objective information was essential for the target group that accepted the offer. Young people do not want to be lectured, but they need to be encouraged and motivated to change the pattern of drug use. Two of the 3 WBIs sought to incorporate some playful elements, such as quizzes and games, and the ability to share with other (former) users [11].

Results of the Delphi Study

The 90 international experts included project partners, individuals from the project's LinkedIn network, and other European experts. They voted anonymously about recommendations for the development and implementation of online interventions that had not yet gained consensus among the project partners. Thus, key recommendations were made regarding the development and implementation of online interventions [11].

Based on the questionnaire responses, feedback from the young people in the discussions, and the Delphi study (see also Jander et al [14] and Neale et al [47]), our key recommendations relate to technical aspects, interactive elements, reaching young drug users, motivation to use the intervention, and evaluation.

For the technical aspects, it is necessary to include experts, clearly lay down provisions, ensure anonymity and data protection in accordance with European Union legislation, and ensure that the setup of mobile versions is user-friendly.

Interactive elements should include entertainment elements, support elements that are as interactive as possible, quizzes, tests, apps, games, blogs, self-tests, animations, and video clips.

To reach young drug users, provide a recognizable offer or unique form; include special elements, access through Facebook, access through YouTube, and personal recommendations; and set up a promotion through the organization's links and links from other organizations, making sure to include the target group in the promotion.

Provide motivation to use the intervention through the intervention's structure and usability, content, mode of presentation and attitude, communication between the user and consultant, and transparency.

Make sure to evaluate the effectiveness of the intervention, by planning the process of evaluation from the beginning. Collect data and feedback regularly, and include the target group in the process.

Discussion

Principal Findings

Existing online interventions in Slovenia provide enough information to young people about illicit drugs, including the effects and risks of drug use, preventive measures, and professional advice to motivate drug users to reduce their consumption and seek help. However, the results of the questionnaire and feedback from the workshop discussions provide information on young people's needs and, at the same time, what should be avoided during the development of such interventions to increase their use among young drug users. Various interactive elements, such as self-testing, games, structured intervention programs, quizzes, chat functionality, forums, and email functionality, are recognized as very important [11]. The different participant characteristics in this study that could influence their answers, such as selection of young drug users from the ongoing program at the Centre for Addiction Prevention in Maribor, different statuses of drug use (former and current users), different ages, and the concern of some about being revealed as a drug user must be taken into account. There are practically no drug users that only use NPS; they typically use NPS in combination with traditional drugs, such as cannabis or cocaine. The legal statuses of these drugs are different.

Existing online interventions show significantly positive results in reducing the use of cannabis, tobacco. or alcohol. According to Sindinovic et al [48], over the past decade, treatment providers and policy makers have increasingly recognized the potential of web-based self-help. Therefore, it is recommended that online interventions are also explored for other illicit drug users. It seems that many factors drive change and the development of medicines on the internet. Most of the factors are connected with technology, globalization, and market innovations. Drug markets have become hybrid markets that combine traditional structures of social and economic opportunities with new opportunities offered by the internet [49]. The internet makes it possible to dramatically increase the number of people who can access health care, thereby achieving significant progress in health, among both the population without previous access to health care and general public [50]. Social media can also provide opportunities for creating online communities that support recovery from drug addiction [2].

Given the high level of internet use in Western societies, digital interventions have the potential to reach adolescents everywhere, especially in areas where physical facilities are rare, such as rural communities [16]. Digital interventions can be transmitted and promoted via various channels, such as email-based advertising campaigns aimed at risk groups, including graduates of higher education, military recruits, students in the first year of tertiary education, and unemployed young people in work centers; online advertising via social media, search engines, or music and video sharing sites; mobile advertising via SMS or MMS; mobile apps; and other sources of information (eg, press, television, radio) [39].

Online interventions generally face some criticism for using the internet as a health communication medium. Online communication cannot offer specific and complex forms of personal communication and assistance, which is particularly needed in the therapeutic environment and especially for problematic drug users. Lack of personal contact could increase existing problems because socially isolated people can participate in anonymous internet services potentially without improving their social position. In addition, it is difficult for participants to determine the reliability of a global consultant or online treatment in general.

People with a lack of computer skills or literacy would generally be excluded [9]. The cost-effectiveness of developing online services must be considered. Although apps are likely to be the most successful if they are targeted at a very specific group, the development of technology for small groups is not cost-effective and, therefore, unlikely to be a priority.

Considering the rapid technological process and new developments in the field of illicit drugs, it is very important to remain current in terms of new technologies and content. Funds should be reviewed and updated regularly, to ensure that they remain attractive and meet the needs of modern youth. The development and implementation of effective online interventions require considerable budget, staff, and network resources. In addition, the use of the guidelines depends on staff who have experience with modern technology and social media. A minimum of experience and, more importantly, a willingness to use these technologies are essential when planning to use a WBI [11]. Given these trends, it is not surprising that there has been significant interest in Europe devoted to the development and implementation of interventions and digital solutions delivered through computer and mobile technologies within European health care systems [51]. However, when we talk about reducing the use of illicit drugs, we cannot overlook the fact that it is necessary to reduce a number of factors that lead to drug use, such as poverty, lack of education, unemployment, poor parenting, and bad skills [52]. By understanding why individuals choose to consume the drugs, within the social and cultural context, we can begin to engage in truly helpful conversations about how to reduce drug-related harm [53]. Furthermore, governments should play an active role in strengthening the prevention of illicit drug use through the enforcement of legislative measures to deter access to illicit drugs; inclusion of illicit drug and health education in the curricula of secondary schools, higher secondary schools, and universities; financing of the evaluation of prevention programs and dissemination of piloted programs at the national level; promotion of community programs; and promotion of the exchange of good practices, guidelines, and quality standards at national and international levels. Long-term strategies should be implemented based on cooperation between government institutions, academic institutions, non-governmental organizations, and mass media [54].

Conclusions

With effective online methods, we could reach a large number of young drug users, facilitate access to more affordable services, provide a quick response or professional feedback on patterns of consumption, increase knowledge about the effects of drugs and their consequences, and support reduction or cessation of drug use. The guidelines, developed as a product of the results of our survey (and surveys from other partners), could improve and increase the effectiveness of existing online services or enable the development of new services designed especially for drug users.

It seems likely that WBIs will continue to develop and that their importance will increase. The speed at which the internet permits the transformation of the drug market will continue to be the main challenge for law enforcement and public health research and monitoring [55]. From the public health perspective, it is challenging to provide such interventions more broadly to the target group and, hence, decrease inequities.

Future Directions

Although technologies and platforms for accessing interventions via smartphones and tablets are evolving rapidly, the anonymity and protection of data related to the transmission of drug use data on the internet remain a real concern [56]. These issues should be addressed fully in the further development of these measures, because these measures are an excellent basis for developing practical and relevant guidelines for implementing effective online interventions for selective prevention.

Furthermore, based on these findings, there is a recognized need to extend the research regarding the effectiveness of WBI implementation in daily prevention work.

Acknowledgments

The study was performed as part of the project CLICK FOR SUPPORT, coordinated by the Landschaftsverband Westfalen-Lippe LWL-Coordination Office for Drug-Related Issues, Münster, Germany, with financial support from the Drug Prevention and Information Program of the European Union (JUST/2013/DPIP/AG/4802). The funding bodies had no role in the design of the study; collection, analysis, or interpretation of data; or writing the manuscript.

Conflicts of Interest

None declared.

References

1. Sznitman S, Taubman D, Lewis N. Analysing Twitter as an opportunity to understand substance use. In: Potter G, Wouters M, Fountain J. editors. Drugs in Europe: Change and Continuity. Lengerich: Pabst Science Publishers; 2014:103-122.
2. Thanki D, Frederick B. Social media and drug markets. In: Mounteney J, Bo A, Oteo A, editors. The internet and drug markets. Luxembourg: European Monitoring Centre for Drugs and Drug Addiction; 2016:E.
3. Tomazic T, Vilela NB. Ongoing Criminal Activities in Cyberspace: From the Protection of Minors to the Deep Web. Journal of Criminal Investigation and Criminology 2017;68:412-423.
4. Tomazic T. The Importance of Social Media from the Wine Marketing Perspective. LEX 2017 Oct 25;15(4):827-844. [doi: 10.4335/15.4.827-844(2017)]

5. Kay-Lambkin FJ, White A, Baker AL, Kavanagh DJ, Klein B, Proudfoot J, et al. Assessment of function and clinical utility of alcohol and other drug web sites: an observational, qualitative study. BMC Public Health 2011 May 05;11(1):277 [FREE Full text] [doi: 10.1186/1471-2458-11-277] [Medline: 21545748]
6. Pischke CR, Zeeb H, van Hal G, Vriesacker B, McAlaney J, Bewick BM, et al. A feasibility trial to examine the social norms approach for the prevention and reduction of licit and illicit drug use in European University and college students. BMC Public Health 2012 Oct 18;12(1):882 [FREE Full text] [doi: 10.1186/1471-2458-12-882] [Medline: 23075043]
7. Caudevilla F. In: Mounteney , Bo , Oteo , editors. The emergence of deep web marketplaces: a health perspective. In: Mounteney J, Bo A, Oteo A. editors. The internet and drug markets. Luxembourguropean Monitoring Centre for Drugs and Drug Addiction; 2016:69-76.
8. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]
9. Internet USTIBPWIUPS. Internet World Stats. internetworldstats URL: https://internetworldstats.com/ [accessed 2018-11-11]
10. United NOOD. Community based treatment and care for drug use and dependence. In: UNODC. Thailand: United Nations Office on Drugs and Crime; 2019.
11. Click FS. CLICK FOR SUPPORT. In: Guidelines for web-based interventions in selective drug prevention. Münster: Landschaftsverband Westfalen-Lippe, LWL-Coordination Office for Drug-Related Issues; 2015.
12. Quaglio G, Schellekens A, Blankers M, Hoch E, Karapiperis T, Esposito G, et al. A Brief Outline of the Use of New Technologies for Treating Substance Use Disorders in the European Union. Eur Addict Res 2017 Aug 12;23(4):177-181. [doi: 10.1159/000478904] [Medline: 28803249]
13. Epton T, Norman P, Dadzie A, Harris PR, Webb TL, Sheeran P, et al. A theory-based online health behaviour intervention for new university students (U@Uni): results from a randomised controlled trial. BMC Public Health 2014 Jun 05;14(1):563 [FREE Full text] [doi: 10.1186/1471-2458-14-563] [Medline: 24903620]
14. Jander A, Crutzen R, Mercken L, De Vries H. Web-based interventions to decrease alcohol use in adolescents: a Delphi study about increasing effectiveness and reducing drop-out. BMC Public Health 2015 Apr 09;15(1):340 [FREE Full text] [doi: 10.1186/s12889-015-1639-z] [Medline: 25881254]
15. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. Addiction 2007;103:227 [FREE Full text] [doi: 10.1111/j.1360-0443.2007.02063.x]
16. Internet-Based drug treatment interventions: Best practice and applications in EU Member States.: EMCDDA Accessed 14 January 2019 URL: http://www.emcdda.europa.eu/system/files/publications/544/EMCDDA-Insight10_155820.pdf [accessed 2019-01-14]
17. Tait RJ, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 01;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]
18. Ramo DE, Thrul J, Chavez K, Delucchi KL, Prochaska JJ. Feasibility and Quit Rates of the Tobacco Status Project: A Facebook Smoking Cessation Intervention for Young Adults. J Med Internet Res 2015 Dec 31;17(12):e291 [FREE Full text] [doi: 10.2196/jmir.5209] [Medline: 26721211]
19. Mounteney J, Griffiths P, Vandam L. What is the future for internet drug markets? In: Mounteney , Bo , Oteo , editors. The internet and drug markets. Luxembourg: EMCDDA; 2016:127-134.
20. Kok G, Burger H, Riper H, Cuijpers P, Dekker J, van Marwijk H, et al. The Three-Month Effect of Mobile Internet-Based Cognitive Therapy on the Course of Depressive Symptoms in Remitted Recurrently Depressed Patients: Results of a Randomized Controlled Trial. Psychother Psychosom 2015 Feb 21;84(2):90-99. [doi: 10.1159/000369469] [Medline: 25721915]
21. 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 20;38(5):776-783. [doi: 10.2337/dc14-1728] [Medline: 25710923]
22. Weiner MW, Nosheny R, Camacho M, Truran-Sacrey D, Mackin RS, Flenniken D, et al. The Brain Health Registry: An internet-based platform for recruitment, assessment, and longitudinal monitoring of participants for neuroscience studies. Alzheimers Dement 2018 Aug 10;14(8):1063-1076 [FREE Full text] [doi: 10.1016/j.jalz.2018.02.021] [Medline: 29754989]
23. Andersson C. Comparison of Automated Technologies for Brief Alcohol Interventions. European Addiction research 2015;21:240-252.
24. Arnaud N, Bröning S, Drechsel M, Thomasius R, Baldus C. Web-based screening and brief intervention for poly-drug use among teenagers: study protocol of a multicentre two-arm randomized controlled trial. BMC Public Health 2012 Sep 26;12(1):826 [FREE Full text] [doi: 10.1186/1471-2458-12-826] [Medline: 23013141]
25. Blankers M, Koeter M, Schippers GM. Evaluating real-time internet therapy and online self-help for problematic alcohol consumers: a three-arm RCT protocol. BMC Public Health 2009 Jan 14;9(1):16 [FREE Full text] [doi: 10.1186/1471-2458-9-16] [Medline: 19144162]
26. Khadjesari Z, Freemantle N, Linke S, Hunter R, Murray E. Correction: Health on the web: randomised controlled trial of online screening and brief alcohol intervention delivered in a workplace setting. PLoS One 2015 Apr 27;10(4):e0127371 [FREE Full text] [doi: 10.1371/journal.pone.0127371] [Medline: 25915505]

27. Kelly PJ, Kay-Lambkin FJ, Baker AL, Deane FP, Brooks AC, Mitchell A, et al. Study protocol: a randomized controlled trial of a computer-based depression and substance abuse intervention for people attending residential substance abuse treatment. BMC Public Health 2012 Feb 10;12(1):113 [FREE Full text] [doi: 10.1186/1471-2458-12-113] [Medline: 22325594]
28. Kuntsche E, Kuntsche S, Thrul J, Gmel G. Binge drinking: Health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug 17;32(8):976-1017. [doi: 10.1080/08870446.2017.1325889] [Medline: 28513195]
29. Penzenstadler L, Chatton A, Van Singer M, Khazaal Y. Quality of Smartphone Apps Related to Alcohol Use Disorder. Eur Addict Res 2016 Sep 7;22(6):329-338. [doi: 10.1159/000449097] [Medline: 27598779]
30. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
31. Tait RJ, Christensen H. Internet‐based interventions for young people with problematic substance use: a systematic review. Medical Journal of Australia 2010 Jun 07;192(S11):15-21. [doi: 10.5694/j.1326-5377.2010.tb03687.x]
32. Tebb KP, Erenrich RK, Jasik CB, Berna MS, Lester JC, Ozer EM. Use of theory in computer-based interventions to reduce alcohol use among adolescents and young adults: a systematic review. BMC Public Health 2016 Jun 17;16(1):517 [FREE Full text] [doi: 10.1186/s12889-016-3183-x] [Medline: 27317330]
33. 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 Dec 19;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]
34. Wilson SR, Lubman DI, Rodda S, Manning V, Yap MBH. The impact of problematic substance use on partners' interpersonal relationships: qualitative analysis of counselling transcripts from a national online service. Drugs: Education, Prevention and Policy 2018 May 31;26(5):429-436. [doi: 10.1080/09687637.2018.1472217]
35. Graham AL, Zhao K, Papandonatos GD, Erar B, Wang X, Amato MS, et al. A prospective examination of online social network dynamics and smoking cessation. PLoS One 2017 Aug 23;12(8):e0183655 [FREE Full text] [doi: 10.1371/journal.pone.0183655] [Medline: 28832621]
36. Thrul J, Bühler A, Ferguson SG. An Internet-based ecological momentary assessment study relying on participants' own mobile phones: insights from a study with young adult smokers. Eur Addict Res 2015 Oct 18;21(1):1-5. [doi: 10.1159/000363231] [Medline: 25342514]
37. Thrul J, Klein AB, Ramo DE. Smoking Cessation Intervention on Facebook: Which Content Generates the Best Engagement? J Med Internet Res 2015 Nov 11;17(11):e244 [FREE Full text] [doi: 10.2196/jmir.4575] [Medline: 26561529]
38. Gainsbury S, Blaszczynski A. A systematic review of Internet-based therapy for the treatment of addictions. Clin Psychol Rev 2011 Apr;31(3):490-498. [doi: 10.1016/j.cpr.2010.11.007] [Medline: 21146272]
39. Hoch E, Preuss UW, Ferri M, Simon R. Digital Interventions for Problematic Cannabis Users in Non-Clinical Settings: Findings from a Systematic Review and Meta-Analysis. Eur Addict Res 2016 May 4;22(5):233-242 [FREE Full text] [doi: 10.1159/000445716] [Medline: 27160333]
40. Jonas B, Tensil M, Tossmann P, Strüber E. Effects of Treatment Length and Chat-Based Counseling in a Web-Based Intervention for Cannabis Users: Randomized Factorial Trial. J Med Internet Res 2018 May 08;20(5):e166 [FREE Full text] [doi: 10.2196/jmir.9579] [Medline: 29739738]
41. Tossmann DH, Jonas B, Tensil M, Lang P, Strüber E. A controlled trial of an internet-based intervention program for cannabis users. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):673-679. [doi: 10.1089/cyber.2010.0506] [Medline: 21651419]
42. Tossman P, Leuschner F. EMCDDA insights. In: Internet-based drug treatment interventions. Best practice and applications in EU Member States. LisbonMCDDA. Lisbon: EMCDDA; 2009:E.
43. Linstone H, Turoff M. The Delphi Method: Techniques and Applications. The Delphi Method: Techniques and Applications 1975:A. [doi: 10.2307/3150755]
44. DrogArt. URL: http://www.drogart.org/ [accessed 2019-10-11]
45. Med.Over.Net. URL: http://med.over.net/ [accessed 2019-10-26]
46. To sem jaz. URL: http://www.tosemjaz.net/ [accessed 2019-10-26]
47. Neale J, Finch E, Marsden J, Mitcheson L, Rose D, Strang J, et al. How should we measure addiction recovery? Analysis of service provider perspectives using online Delphi groups. Drugs: Education, Prevention and Policy 2014 May 15;21(4):310-323. [doi: 10.3109/09687637.2014.918089]
48. Sinadinovic K, Wennberg P, Johansson M, Berman AH. Targeting individuals with problematic alcohol use via Web-based cognitive-behavioral self-help modules, personalized screening feedback or assessment only: a randomized controlled trial. Eur Addict Res 2014 Oct 4;20(6):305-318 [FREE Full text] [doi: 10.1159/000362406] [Medline: 25300885]
49. The internet and drug markets. Lisbon: EMCDDA; Feb 2016.
50. Curry SJ. eHealth research and healthcare delivery beyond intervention effectiveness. Am J Prev Med 2007 May;32(5 Suppl):S127-S130. [doi: 10.1016/j.amepre.2007.01.026] [Medline: 17466817]
51. Quaglio G, Pirona A, Esposito G, Karapiperis T, Brand H, Dom G, et al. Knowledge and utilization of technology-based interventions for substance use disorders: an exploratory study among health professionals in the European Union. Drugs: Education, Prevention and Policy 2018 Jun 06;26(5):437-446. [doi: 10.1080/09687637.2018.1475549]

52. Hechanova MRM, Alianan AS, Calleja MT, Melgar IE, Acosta A, Villasanta A, et al. The Development of a Community-Based Drug Intervention for Filipino Drug Users. Journal of Pacific Rim Psychology 2018 Feb 14;12:1-10. [doi: 10.1017/prp.2017.23]
53. Ivsins A, Yake K. Looking beyond harm: meaning and purpose of substance use in the lives of marginalized people who use drugs. Drugs: Education, Prevention and Policy 2018 Sep 11;27(1):27-36. [doi: 10.1080/09687637.2018.1497145]
54. Lotrean L, Santillan E, Thrasher J, Laza V. An Increasing Trend of Illicit Drug use among Romanian University Students from 1999 to 2011. Zdr Varst 2016 Mar;55(1):53-58 [FREE Full text] [doi: 10.1515/sjph-2016-0008] [Medline: 27647089]
55. Moore BA, Fazzino T, Garnet B, Cutter CJ, Barry DT. Computer-based interventions for drug use disorders: a systematic review. J Subst Abuse Treat 2011 Apr;40(3):215-223 [FREE Full text] [doi: 10.1016/j.jsat.2010.11.002] [Medline: 21185683]
56. Internet-based drug treatment. In: Perspective on drugs Internet-based drug treatment. Lisbon: EMCDDA; May 27, 2014.

Abbreviations

EMCDDA: European Monitoring Centre for Drugs and Drug Addiction.
NPS: new psychoactive substances.
WBI: Web-based intervention.

Edited by G Eysenbach; submitted 04.01.20; peer-reviewed by J Jurišić, G Shrivastava; comments to author 26.01.20; revised version received 31.01.20; accepted 01.02.20; published 22.04.20.

Please cite as:
Tomazic T, Jerkovic OS
Online Interventions for the Selective Prevention of Illicit Drug Use in Young Drug Users: Exploratory Study
J Med Internet Res 2020;22(4):e17688
URL: http://www.jmir.org/2020/4/e17688/
doi:10.2196/17688
PMID:32319962

Original Paper

Message Delivery for the Treatment of Posttraumatic Stress Disorder: Longitudinal Observational Study of Symptom Trajectories

Matteo Malgaroli[1*], PhD; Thomas Derrick Hull[2,3*], MS; Shannon Wiltsey Stirman[4,5], PhD; Patricia Resick[6], PhD

[1]NYU Grossman School of Medicine, New York, NY, United States
[2]Columbia University, New York, NY, United States
[3]Talkspace, New York, NY, United States
[4]Standford University, Stanford, CA, United States
[5]National Center for PTSD, Washington DC, DC, United States
[6]Duke University, Durham, NC, United States
[*]these authors contributed equally

Corresponding Author:
Matteo Malgaroli, PhD
NYU Grossman School of Medicine
One Park Avenue, 8th Floor
New York, NY, 10016
United States
Phone: 1 6467544030
Email: matteo.malgaroli@nyulangone.org

Abstract

Background: Individuals with posttraumatic stress disorder (PTSD) face symptoms that can hinder access to treatment, such as avoidance and guilt. Telemedicine offers a technological solution to increase access to mental health care and overcome barriers to treatment. Although an increasing body of literature focused on synchronous telehealth (eg, live video), no studies have examined the delivery of PTSD treatment via two-way multimedia messages (ie, texting or messaging).

Objective: The aim of this study was to conduct a longitudinal observation of treatment for PTSD delivered using two-way asynchronous messaging. We also sought to identify individual and treatment characteristics that could predict the observed outcome differences.

Methods: Outpatients diagnosed with PTSD (N=475) received interventions from licensed therapists, which were delivered via messaging once or more than once per day, 5 days a week for 12 weeks. PTSD symptoms were assessed every 3 weeks using the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-5. Trajectories of PTSD symptoms were identified using growth mixture modeling (GMM). Using logistic regression, the demographic, treatment, and messaging characteristics of patient groups that improved were compared with the characteristics of patient groups that did not improve.

Results: The GMM identified 4 trajectories of PTSD symptoms: *moderate improvement* (197/475, 41.4%), *high symptoms* (197/475, 41.4%), *chronic symptoms* (61/475, 12.9%), and *acute improvement* (20/475, 4.3%). Patients with a clinically significant reduction in PTSD symptoms (231/475, 48.6%) were more likely to communicate via video (odds ratio [OR] 1.01, 95% CI 1.01-1.05; *P*=.03), have a higher working alliance with their therapist (OR 1.03, 95% CI 1.01-1.05; *P*=.02), and be at their first treatment experience (OR 2.03, 95% CI 1.18-3.54; *P*=.01). Treatment adherence was associated with greater therapeutic alliance (OR 1.07, 95% CI 1.03-1.10; *P*<.001), education (OR 2.13, 95% CI 1.13-4.03; *P*=.02), and more patient-generated messages per week (OR 1.08, 95% CI 1.04-1.13; *P*<.001).

Conclusions: Multimedia message delivery for PTSD treatment showed symptom-reduction rates similar to traditional forms of treatment delivery, suggesting further study of messaging as a treatment medium. Most patients completed an 8-week course, reflecting the acceptability of messaging interventions. Delivering treatment via two-way messaging offers increased opportunities for widespread access to mental health care.

(J Med Internet Res 2020;22(4):e15587) doi:10.2196/15587

KEYWORDS

PTSD; telemedicine; messaging; textmessaging; psychotherapy; telehealth; digital health

Introduction

Over the course of a lifetime, most people are exposed to at least one potentially traumatic event [1], and approximately 9.7% of women and 3.6% of men [2] will develop posttraumatic stress disorder (PTSD) [3]. PTSD symptoms can be extremely debilitating and are associated with significant functional impairment in education, childbearing, relationships, and financial income [4]. Moreover, lifetime prevalence of PTSD comorbidity has been estimated at approximately 80% [5,6], with higher rates for substance abuse, depression, and anxiety [7]. Comorbidity, in turn, greatly increases day-to-day functional impairment and disability [8-10].

Although many evidence-based treatments for PTSD have been found to be efficacious [11,12], access to face-to-face treatment is not so straightforward. In addition to traditional barriers to treatment, including cost, insurance, stigma, and physical impairments [13-15], individuals with PTSD also face additional hindrances such as avoidance and shame that often result in isolation from their communities. Geographic remoteness can also prevent access to care, and in areas with ongoing violence (and consequently high incidence of PTSD), it can be dangerous for mental health professionals to practice [16]. These obstacles tend to reduce seeking and obtaining of treatment, perpetuating the economic and social costs of untreated illness [17]. Given the particularly debilitating impact of PTSD symptoms, it is imperative to improve rates of treatment access.

Technological delivery is a solution to removing socioeconomic barriers to treatment. Telemedicine enables synchronous (eg, videoconferencing) and asynchronous (eg, texts, images, audio recordings, and video recordings) interactions and is rapidly growing in popularity. Multiple studies have shown the equivalence or noninferiority of PTSD treatment via videoconferencing compared with in-person treatment [18-21]. Other types of technology have also shown encouraging results, such as virtual reality [22], web-based text-based interventions [16,23-25], and email [26].

Although mostly used as an adjunct to treatment [27], asynchronous two-way modalities such as multimedia message service (MMS) have emerged as a potentially primary means of treatment delivery. Owing to its wide population reach, MMS can provide accessible mental health treatment to large numbers of individuals [28], including those in challenging geographic contexts. Early studies suggested that treatment via two-way messaging (ie, texting) delivers outcome rates comparable with other methods for individuals with a wide range of diagnoses [29]. However, the effect of therapy via messaging has not been investigated in a large PTSD population.

In this study, we conducted a longitudinal observation of the delivery of PTSD treatment via two-way multimedia messaging (ie, text, audio, and video). Consistent with once-weekly face-to-face therapy standards, we examined a period of 12 weeks of treatment. Importantly, the study took place in an ecologically valid outpatient telemedicine setting, portraying how two-way messaging would be used in naturalistic contexts. To our knowledge, this is the first study to examine the course of messaging telemedicine treatment for PTSD. Our first aim was to examine overall rates of PTSD reduction and then identify heterogeneous trajectories of symptom changes during treatment. We also sought to identify individual and treatment characteristics that could predict the observed outcome differences.

Methods

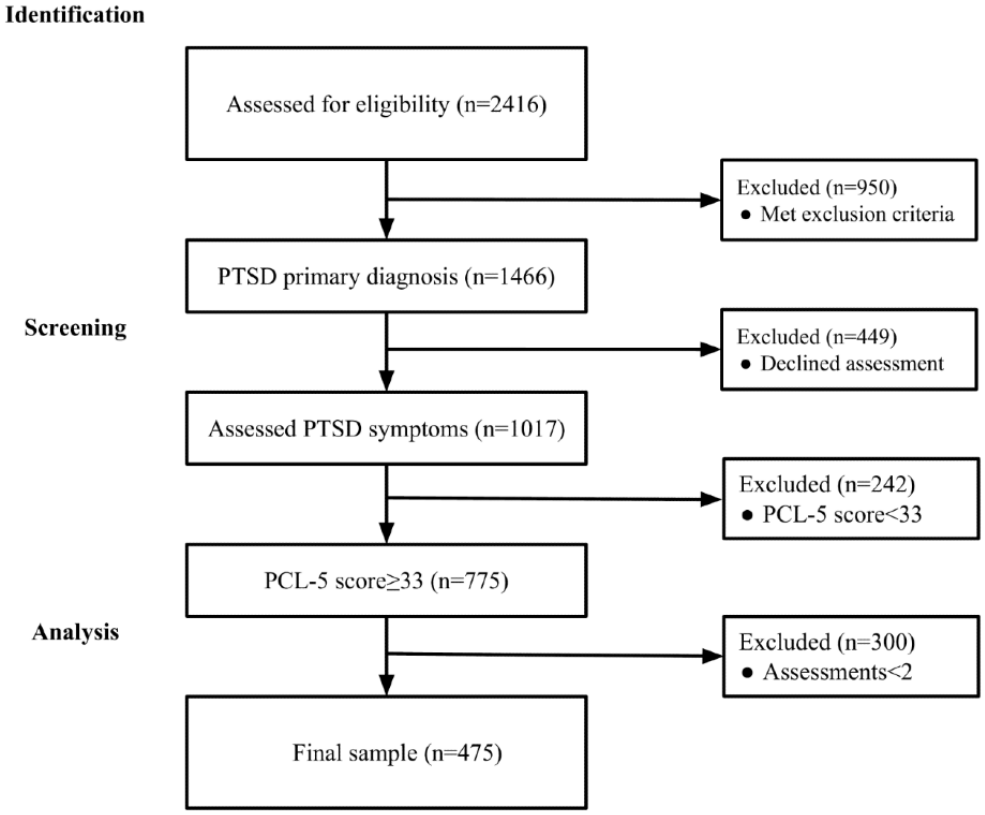
Participants and Setting

Our sample consisted of 18- to 65-year-old treatment-seeking individuals in the United States, who signed up for a web-based therapy platform (Talkspace). The platform is accessible through internet search, through employee assistance programs, and, as a behavioral health benefit, through some individual insurances. Patients first underwent a biopsychosocial intake interview with a licensed therapist through a live messaging system. Presenting complaints, diagnosis, treatment goals, treatment history, and provider preferences were determined by the clinician during the assessment. Patients then chose a therapist for their treatment from a selected assortment of clinicians, licensed in the state where the patients resided. The matching algorithm selection was based on intake information, therapists' previous treatment outcomes, and patients' preferences (eg, the therapist's gender). The study used archival data under the platform Terms of Use and was approved by the institutional review board at Columbia University Teachers College.

The inclusion criteria were as follows: PTSD diagnosis confirmed by clinical intake and working diagnosis with a licensed therapist, initial PTSD symptom score of 33 or above as measured by the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-5 (PCL-5), and regular internet and/or phone access. The exclusion criteria were as follows: comorbid bipolar or psychotic-spectrum disorders (established by clinical intake/interview); comorbid substance or alcohol abuse (established by clinical intake/interview); and suicidal ideation, intent, plan, and/or behavior requiring higher level of care, as measured by the Columbia Suicide Severity Rating Scale.

Of 1466 individuals who had a confirmed PTSD diagnosis from licensed providers, 1017 did not meet the exclusion criteria and agreed to complete measures before beginning treatment; of these, 775 had PTSD symptoms above the clinical threshold. The final sample consisted of 475 participants who completed two or more surveys (Figure 1).

Figure 1. Flowchart of patient selection for the study. PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders-5; PTSD: posttraumatic stress disorder.



Design and Procedure

Therapists in the study (n=173) were licensed in the same state as the patient and credentialed up to the National Committee for Quality Assurance standards. They were trained and experienced in PTSD (138/173, 79.7%), cognitive behavioral therapy (98/173, 56.9%), and third-wave behavioral (83/173, 47.7%) and psychodynamic (51/173, 29.4%) treatment modalities. They had at least three years of experience delivering mental health care post licensure. Modal experience of psychotherapy practice was 10+ years (69/173, 39.9%). Before providing treatment on the platform, therapists went through a 30-day orientation to the platform that included introduction to patients, completing informed consent, setting a frame for the treatment, appropriately pacing the treatment, handling crisis, making referrals, and other aspects related to telehealth competence.

Multimedia messages (similar in capability to commonly used *texting* apps) were used by the therapists to deliver interventions through a Health Insurance Portability and Accountability Act (HIPAA)–compliant interface for smartphones and computers. Messages consisted of two-way asynchronous communication, containing text, photo, audio, or video content. Therapists messaged patients a minimum of once (or more than once) per day, 5 days a week. Participants were able to send any number of multimedia messages at any time they wanted to their therapist, and the messages were stored for the clinicians to review. Therapist response times were scheduled and communicated to their patients at the start of treatment (eg, Monday to Friday from 9 AM to 11 AM and from 5 PM to 6 PM). All professional and ethical standards of the messaging treatment were observed just as in a face-to-face treatment, and higher levels of care referrals were provided when needed.

A period of 12 weeks of messaging was examined. Participants could discontinue treatment at any point as in conventional outpatient settings. Treatment engagement metrics included weekly average text, photo, audio, and video messaging communication for each week of the treatment. This information was automatically collected by the digital health platform. For each patient and therapist dyad, the average number of messages per week in treatment was assessed (text, audio, photo, and video) as well as the average number of weekly words used in text messages and the weekly average duration of audio and video content in minutes.

Measures

All participants had a primary diagnosis of PTSD from their licensed provider. In addition, the 20-item PCL-5 [30,31] was used to identify the presence of PTSD symptoms from baseline and then every 3 weeks for 12 weeks. A PCL-5 score of 33 or higher is considered indicative of clinically significant PTSD [32], and only participants initially meeting the cutoff were included. Post baseline survey completion was voluntary and described to patients as an important aspect of their care that facilitates goal setting and tracking progress. In addition, the 12-item Working Alliance Inventory (WAI)–revised short patient version [33] determined participants' treatment alliances after 3 weeks of treatment. WAI scores have been positively associated with good treatment outcomes [34].

Statistical Analyses

Posttraumatic Stress Disorder Symptom Trajectories

Growth mixture modeling (GMM) was performed using Mplus 8 [35] to identify trajectories of PTSD symptom scores over treatment (baseline to week 12). GMM has been extensively used to model PTSD symptom course [36], as it allows the identification of heterogeneous subpopulations (or classes) characterized by different longitudinal symptom trajectories, such as treatment responders and nonresponders. The GMM trajectories were estimated using the Muthén-Roy model [37], a more stringent approach to survey nonresponse that considers missing data as nonignorable and associated with treatment measures. Specifically, binary indicators of missingness for each observation were included in the model and described by classes of missing data patterns, which correlated with the PTSD trajectories predicted by growth factors (intercept, slope, and quadratic parameters) of PCL-5 score changes over time. Patients were then assigned to one missing data pattern and a PCL-5 class, conjointly representing their treatment outcomes. To determine the best fitting model, nested GMMs with an increasing number of trajectories and different missing data approaches to dropout were compared [37]. The optimal trajectory solution [38] was determined based on the lower Bayesian Information Criterion and higher classification entropy [39]. Theoretical and explanatory properties and theoretical coherence were also considered when determining the best fitting solution [40].

Predictors of Posttraumatic Stress Disorder Reduction and Treatment Adherence

After determining the best fitting trajectory solution, categorical class assignments as well as demographic, clinical, and treatment variables were added into a logistic regression in R [41] to predict clinically meaningful PTSD symptom improvement, as indicated by the PCL-5 score reduction of 10 or more points [30] at the last available observation or at 12 weeks, whichever came first. A second logistic regression examined predictors of patient adherence, measured as remaining in treatment for at least eight weeks.

Analyzed predictors consisted of demographics (gender, age, and education level), first experience in therapy, total number of treatment weeks, working alliance (WAI score), therapist-reported PTSD treatment expertise, and years of experience. Weekly communication characteristics (types of messages, average words sent, and duration of audio/video messages) of therapist interventions and patient responses were also analyzed. Examined predictor variables had 12% missingness or less, with the exception of the WAI (32%). Multiple imputations by iterative random forests (500 trees, 10 iterations) for missing values were performed in R using package missForest v 1.4 [42]. Clinical and outcome variables were masked during imputations to prevent information leakage. Categorical variables were then transformed into binary (eg, age ≤ or >35 years).

Results

Sample Characteristics

The final patient sample consisted of 475 individuals with both PTSD diagnosis and PCL-5 scores of 33 or above. They were predominantly female, 26- to 35-year-old, and college educated. Table 1 reports the full characteristics of the participants.

Table 1. Demographic, clinical, and treatment characteristics for full sample (N=475).

Variable	Total number of participants	Value
Age (years), n (%)	475	
18-25		125 (26.3)
26-35		234 (49.3)
36-49		103 (21.7)
≥50		13 (2.7)
Education, n (%)	422	
Bachelor's degree or higher		291 (69.0)
High school diploma		131 (31.0)
First time in therapy, n (%)		78 (17.8)
Gender, n (%)	475	
Female		412 (86.7)
Male		63 (13.3)
Patient's state, n (%)	475	
California		61 (14.2)
New York		52 (12.1)
Texas		28 (6.5)
Florida		28 (6.5)
Pennsylvania		20 (4.6)
Other US state		347 (56.1)
Posttraumatic stress disorder symptoms, mean (SD)		
PCL-5[a], baseline	475	50.64 (10.44)
PCL-5, week 3	475	43.53 (15.04)
PCL-5, week 6	282	39.71 (15.80)
PCL-5, week 9	168	37.65 (16.21)
PCL-5, week 12	108	36.03 (16.80)
Treatment duration (weeks), mean (SD)	475	10.28 (2.82)
Treatment focus, n (%)	68	
Traumatic memories		36 (53)
Challenges with daily living		32 (47)
Weekly engagement, patients, mean (SD)	423	
Number of messages		14.68 (18.04)
Text messages		
Count		13.78 (17.68)
Words		1039.86 (1191)
Audio messages		
Count		0.40 (1.46)
Duration (min)		83.32 (330.93)
Photo messages, count		0.42 (1.05)
Video messages		
Count		0.05 (0.37)
Duration (min)		5.20 (41.86)
Weekly interventions by therapists, mean (SD)	423	

Variable	Total number of participants	Value
Number of messages		8.75 (8.88)
Text messages		
Count		8.18 (8.84)
Words		679.78 (535.12)
Audio messages		
Count		0.33 (0.91)
Duration (min)		58.13 (180.48)
Photo messages, count		0.17 (0.49)
Video messages		
Count		0.07 (0.18)
Duration (min)		4.33 (17.33)
Working alliance, mean (SD)	322	45.44 (10.47)

[a]PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders-5.

Treatment Characteristics

Messaging Engagement

Therapists predominantly delivered therapy via text on a weekly basis (mean 8.18, SD 8.84), but audio, photo, and video messages were also used. Similar communication preferences were observed in the patients' weekly use of text (mean 13.78, SD 17.68), audio, photo, and video messages. A subset of therapists (68/173, 39.3%) described their treatment focus, with half reporting focusing on the memory of the trauma as the primary treatment goal (n=36) and the other on challenges with day-to-day living (n=32). Overall, patients reported good therapeutic alliance at 3 weeks of treatment (WAI, total scale: mean 45.44, SD 10.47; item average: mean 3.8, SD 0.9).

Dropout

The average treatment duration was 10.3 weeks (SD 2.7). A total of 58.7% (279/475) of patients completed the entire observed 12 weeks of treatment, with the remaining individuals discontinuing. Specifically, of the 475 patients, 11 (1.3% total) terminated treatment by week 3, 65 (13.7% total) terminated cumulatively by week 6, 127 (26.7%) terminated cumulatively by week 9, and an additional 69 discontinued before week 12, resulting in a total of 196 patients (41.3%) who discontinued treatment. Reasons for termination were reported in an exit survey of 108 individuals and included reaching personal treatment goals (n=52), considering the therapist not helpful or bad (n=23), money concerns (n=21), starting face-to-face therapy (n=9), lack of time (n=2), and technical difficulties with the platform (n=1). A therapeutic dose of PTSD treatment, consisting of 8 (or more) weeks of intervention, was achieved in 84.0% (399/475) of all patients.

Average Treatment Outcome

Mean PCL-5 scores averaged over the entire sample decreased from baseline (mean 50.64, SD 10.44) over the course of treatment: 3 weeks (mean 43.53, SD 15.04), 6 weeks (mean 39.71, SD 15.8), 9 weeks (mean 37.65, SD 16.21), and 12 weeks (mean 36.03, SD 16.8). There were 3.2 (SD 1.2) PCL-5 assessments available per patient. Scores falling below the established cutoff for probable PTSD (<33) were endorsed by 34.9% (166/475) of participants by their last observation; a more stringent threshold for remission based on the PCL-5 score of 19 and lower was reached by 14.3% (68/475) of the sample. The mean PCL-5 score reduction was 11 points (SD 14.47), with 48.6% (231/475) of patients reaching a clinically significant improvement of 10 or more points.

Posttraumatic Stress Disorder Symptom Trajectories

The best GMM fitting model of PCL-5 symptom trajectories over treatment is displayed in Figure 2. Table 2 reports fit indices for successfully estimated solutions of the GMM Muthén-Roy and other GMM models with progressively increasing numbers of classes.

GMM identified 4 subpopulations distinguished by their course of PCL-5 symptoms over treatment and their response patterns. The probability of distinct class membership was high, ranging from 0.78 to 0.92. The first trajectory was characterized by a steady reduction in symptom scores that eventually fell below the threshold established for probable PTSD (*moderate improvement*, 197/475, 41.4%). This class showed relatively lower levels of initial symptoms that decreased below the PTSD cutoff over the course of treatment. The second largest class (*high symptoms*, 197/475, 41.4%) described a population with PCL-5 scores that, although decreasing, remained above the clinical cutoff. The third identified class (*chronic symptoms*, 61/475, 12.9%) was a trajectory characterized by treatment nonresponse. The fourth trajectory (*acute improvement*, 20/475, 4.3%) was characterized by high initial PCL-5 scores, which decreased below the clinical cutoff through treatment. In terms of missing data patterns, the *moderate improvement* and *chronic symptoms* trajectories had the same latent profile, characterized by a higher risk of self-report measure noncompletion (44.0% of combined subpopulations completed only two measures) when compared with the *high symptoms* and *acute improvement* group (37.8%). There were no differences in dropout rates between the two missingness patterns (8+ weeks of treatment: 84.2% vs 83.8%). Consistent with their longitudinal course, the

moderate and *acute improvement* trajectories belonged to the same PTSD latent class, characterized by symptom reduction, whereas *high* and *chronic symptoms* were part of the same elevated-symptom PTSD class. Specifically, of the 166 patients reaching scores below probable PTSD (PCL-5≤32) by their last observation, 130 were assigned to the symptom-reduction trajectories and 36 to the elevated-symptom class.

Figure 2. Estimated means for Muthén-Roy growth mixture modeling trajectories of PCL-5 symptom scores over 12 weeks of treatment (N=475). PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders-5.

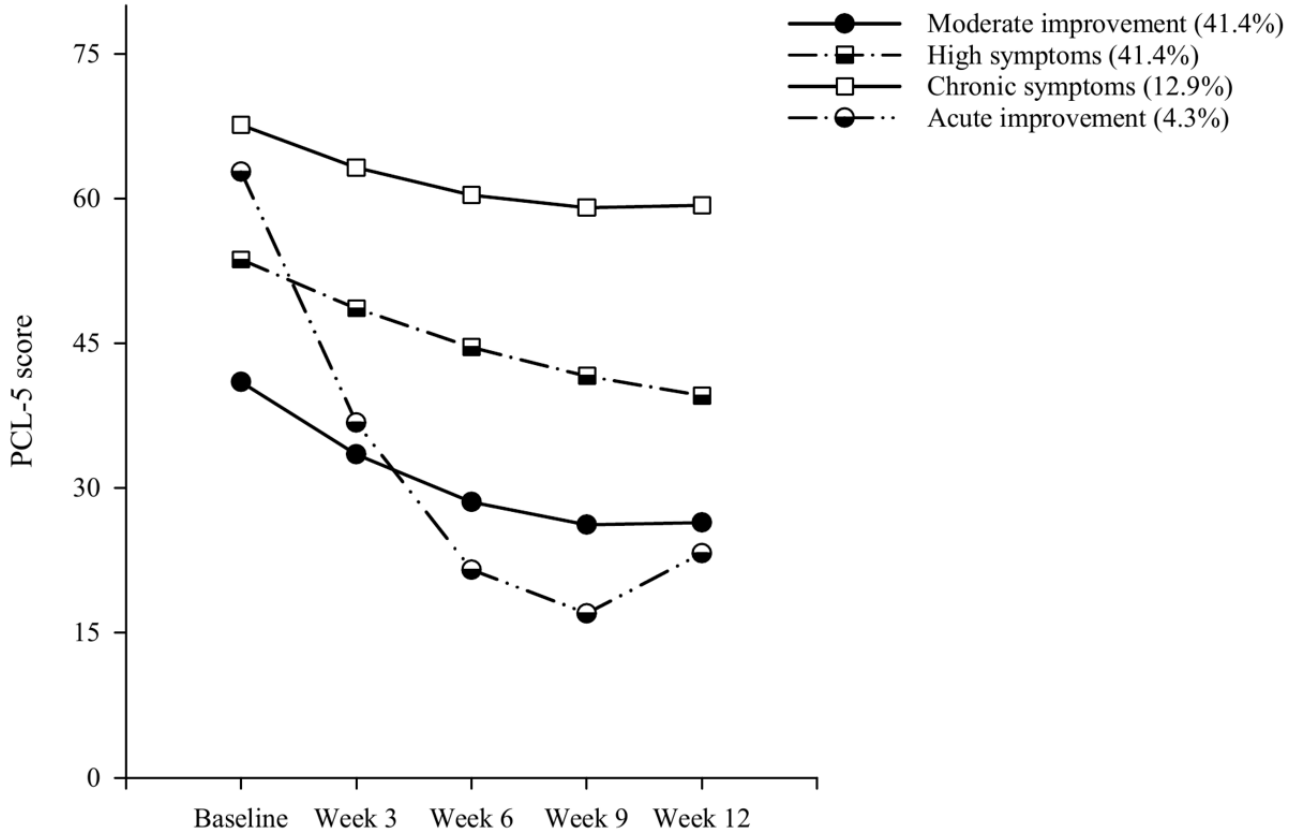


Table 2. Fit indices for Growth Mixture Models with increasing trajectory solutions.

Model and Number of classes	Bayesian Information Criterion	Entropy
Muthén–Roy		
2 (2)	11,508.56	0.80
1 (1)	11,556.90	0.70
Pattern mixture		
N/A[a]	11,583.12	N/A
Diggle-Kenward		
2	12,934.68	0.70
1	13,092.78	N/A
Missing at random GMM		
2	11,538.72	0.70
1	11,724.37	N/A

[a]N/A: not applicable.

Predictors of Posttraumatic Stress Disorder Reduction and Treatment Adherence

Logistic regression analyses were performed to determine the role of demographic, categorical GMM memberships, treatment, and communication characteristics as predictors of significant PTSD reduction and treatment adherence.

The threshold for clinically meaningful PTSD symptom improvement (PCL-5 score reduction ≥10) was reached by 48.6% (231/475) of patients by their last observation, with 130

patients assigned to the *acute* and *moderateimprovement* trajectories and 101 assigned to the *chronic* and *high symptom* trajectories. The likelihood ratio test of the full model against a constant-only model was significant ($\chi^2_{20,475}$=64.1, P<.001; McFadden R^2=0.10), and the predictor odds ratios (OR) are shown in Table 3. Results indicated that higher therapeutic alliance scores (OR 1.03, 95% CI 1.01-1.05; P=.02) and being at the first experience of psychotherapy (OR 2.03, 95% CI 1.18-3.54; P=.01) were significantly associated with PTSD symptom reduction. Consistent with the heterogeneous trajectories, the results also showed that participants assigned to the *moderate* and *acute* improvement trajectories were more likely to have significant symptom reduction (OR 5.19, 95% CI 2.83-10.0; P<.001) than those assigned to the elevated-symptom trajectories, while patients with less measure completion were also less likely to show improvements (OR 0.38, 95% CI 0.20-0.69; P=.002). In terms of average weekly engagement per week in treatment, patients sending more video messages per treatment week were more likely to show symptom improvements (OR 1.01, 95% CI 1.01-1.05; P=.03).

A second logistic regression examined predictors of staying in treatment until an adequate therapeutic dose (≥8 weeks), which was achieved by 84% (399/475) of patients. The analysis used all previously examined variables, with the exception of treatment duration. Results of the logistic regression ($\chi^2_{19,475}$=118, P<.001; McFadden R^2=0.28) indicated that patients with higher therapeutic alliances (OR 1.07, 95% CI 1.03-1.10; P<.001) and higher education level (OR 2.13, 95% CI 1.13-4.03; P=.02) were more likely to attain the treatment dose. Patients who were more engaged in treatment media who sent a higher number of messages per week of treatment were more likely to achieve treatment completion (OR 1.08, 95% CI 1.04-1.13; P<.001). Interestingly, therapists wrote more on average each week to patients who ended up discontinuing the therapy (OR 0.88, 95% CI 0.82-0.94; P<.001), possibly trying to keep them continuing the treatment. No other meaningful differences emerged in terms of covariates or quantitative intervention characteristics, and full estimates are reported in Table 3.

Table 3. Predictors of posttraumatic stress disorder symptoms clinically significant improvement and treatment adherence (N=475).

Variable	PTSD[a] symptoms improvement		Treatment dose 8+ weeks	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value
Demographics				
Age>35 years	0.79 (0.49-1.26)	.32	0.91 (0.45-1.90)	.80
Education: Bachelor's degree or higher	0.99 (0.64-1.54)	.97	2.13 (1.13-4.03)	.02
Gender - female	1.31 (0.73-2.36)	.37	0.69 (0.26-1.69)	.44
First therapy experience	2.03 (1.18-3.54)	.01	0.96 (0.44-2.23)	.92
Treatment characteristics				
Working alliance	1.03 (1.01-1.05)	.02	1.07 (1.03-1.10)	<.001
Duration of treatment (weeks)	1.07 (0.98-1.17)	.13	N/A[c]	N/A
Latent class: acute/moderate improvement	5.19 (2.83-10.0)	<.001	1.68 (0.72-3.92)	.23
Latent class: survey noncompletion	0.38 (0.20-0.69)	.002	0.58 (0.25-1.34)	.20
Therapist: years of experience (10+)	1.44 (0.92-2.27)	.11	0.68 (0.35-1.36)	.27
Therapist: PTSD expertise	0.76 (0.51-1.14)	.19	0.84 (0.44-1.55)	.57
Patient weekly engagement				
Number of messages	1.00 (0.98-1.01)	.64	1.08 (1.04-1.13)	<.001
Text messages: number of words	1.00 (1.00-1.00)	.67	1.00 (1.00-1.00)	.07
Audio messages: duration (min)	1.00 (1.00-1.00)	.52	1.00 (1.00-1.01)	.11
Photo messages: count	0.83 (0.65-1.03)	.11	1.22 (0.82-2.14)	.40
Video messages: duration (min)	1.02 (1.01-1.05)	.03	1.01 (0.99-1.04)	.47
Therapist weekly engagement				
Number of messages	1.01 (0.97-1.05)	.55	0.88 (0.82-0.94)	<.001
Text messages: number of words	1.00 (1.00-1.00)	.52	1.00 (1.00-1.00)	<.001
Audio messages: duration (min)	1.00 (1.00-1.00)	.92	1.00 (0.99-1.00)	.15
Photo messages: count	1.17 (0.74-1.99)	.51	1.43 (0.82-3.18)	.28
Video messages: duration (min)	0.98 (0.95-1.00)	.05	0.99 (0.96-1.02)	.43

[a]PTSD: posttraumatic stress disorder.

[b]OR: odds ratio.

[c]N/A: not applicable.

Discussion

This study analyzed two-way asynchronous messaging as a delivery modality for PTSD treatment. The research took place in a patient-centered naturalistic setting, where therapeutic dyads interacted daily through messaging. PTSD symptoms were assessed from baseline every 3 weeks over the course of 12 weeks of treatment. Messages were sent using a HIPAA-compliant online platform, which also automatically collected patient-therapist messaging engagement metrics. We first investigated whether PTSD treatment response patterns were influenced by the messaging delivery modality. We then divided patients into heterogeneous groups based on their symptom fluctuations to tease apart patterns of treatment response. We then assessed which individual, therapist, and messaging engagement differences could predict outcome differences.

The results indicated that 48.6% (231/475) of patients experienced clinically significant improvement of symptoms, with 34.9% (166/275) improving below the PTSD threshold. Similar remission rates are usually observed when delivering PTSD treatment through established modalities, including face-to-face [43,44] and live video [16,45]. In addition, we identified 4 heterogeneous trajectories of PTSD symptom changes. The majority of the sample showed a steady reduction in PTSD symptoms from a moderate baseline (*moderate improvement*, 198/475, 41.4%) and others more steeply from a higher baseline (*acute improvement*, 20/475, 4.3%). Nonetheless, one of the largest groups consisted of 198 individuals with symptoms that, while decreasing, remained above the clinical threshold (*high symptoms*, 198/475, 41.4%). A group of 12.9% (61/475) of patients with chronic high symptoms was also identified. The analytic approach allowed for a more accurate model of the heterogeneous PTSD course [46] and accounted for missing data in the course of treatment over time [37],

identifying symptom improvements for nearly half of the sample in the examined time frame. These trajectories raise important possibilities for assisting the two nonresponder groups. For example, individuals in these groups could be referred to face-to-face therapy, switched to another type of therapy, moved to live video sessions, or offered a psychiatric referral or other available resources. Further analysis of participants with clinically meaningful PTSD symptom reductions indicated that the improvement trajectories were significantly more likely to reach PTSD remission. A higher chance of meaningful symptom reduction was also associated with being in therapy for the first time, a higher working alliance, higher survey response rate, and more use of immersive communication modalities (ie, video messages). Therapists' expertise and the treatment focus were shown to be less important than other therapeutic factors (ie, working alliance). However, these findings may evolve, as therapeutic interventions were specifically identified for this medium. Of note, 84.0% (399/475) of the sample remained in treatment until reaching a therapeutic dose of 8+ weeks of therapy, a proportion higher than that observed in face-to-face PTSD treatment [43]. This finding is consistent with the 3 unique advantages of messaging therapy: (1) ease of access contributes to adherence and completion, (2) patients can write to therapists at any moment of the patient's choosing, and (3) there is greater frequency of therapeutic intervention (5 days a week). Logistic regression analysis showed that treatment adherence was associated with a higher therapeutic alliance and more messages sent by patients each week. The role of working alliances in predicting treatment adherence and symptom improvement through messaging was consistent with its importance during in-person and live video therapy delivery [16,47,48]. Moreover, the therapists of individuals who discontinued treatment prematurely were more likely to try to engage their patients, resulting in higher average messages generated per week.

Overall, PTSD treatment administered using two-way messaging resulted in symptom-reduction rates similar to those reported for face-to-face and videoconferencing therapy [16,45] and higher spontaneous remission rates [49]. To our knowledge, this study, which observes the PTSD symptom course and treatment delivery in a large sample of patient-therapist dyads interacting through two-way messaging, is the first of its kind. Although these initial observational results are promising, the lack of a control group and posttreatment assessment in this study hinders comparison with nondigital treatment as usual. It may be useful to conduct a randomized controlled study of PTSD treatment delivery via asynchronous two-way messaging to assess its effectiveness compared with established forms of treatment interventions. Another limitation of the study is that our research was focused on the delivery of treatment rather than its content. Accordingly, the treatment metrics examined were related to the frequency (ie, number of messages) and quantity (ie, number of words and length of audio/video messages) of the patient-therapist interactions. Although we investigated therapist experience and PTSD treatment expertise as well as reported on treatment focus, our level of analysis was not specific to the content of treatment interventions. Future research utilizing treatment protocols or otherwise quantifying the nature of the therapeutic intervention will also be important. There are additional limitations to consider when interpreting these findings. Despite the large sample size in this study, 449 potential participants chose to not complete any assessments. In addition, 300 participants did not complete the minimum of two assessments required to be included in the study. As such, self-selection could have influenced the findings. Moreover, although encouraged by the platform and therapist, survey responses were determined by the patient, with possible reasons for nonresponse ranging from dissatisfaction (eg, unhappiness with their progress, lack of motivation, or discomfort with the modality) to treatment success (eg, met goals). Although the trajectory modeling used in this study accounted for missing data and dropout attrition, future research should assess these characteristics in more detail, in addition to the outcome findings discussed, to strengthen the generalizability of these results. Finally, although therapists received telehealth orientation, no PTSD intervention manuals for messaging were available at the time of the study. Future research should examine the feasibility and effectiveness of specific treatment approaches and protocols for PTSD when delivered by messaging, such as cognitive processing therapy.

Despite these limitations, this study highlights new opportunities for telemedicine in the treatment of PTSD. Patients diagnosed with PTSD often face avoidance, guilt, fear, and alienation in addition to common barriers (wait lists, insurance coverage, affordability, and scheduling around work or other obligations) that prevent them from seeking treatment. Delivery of therapy through two-way messaging opens up opportunities for increased access to treatment with less fear of stigmatization in an immediate and convenient manner.

Acknowledgments

This study was supported by archival data provided by Talkspace. Talkspace had a minimal role in the design, data collection, and data management as the data provider. It had no role in the analysis, interpretation of the data, or decision to submit the manuscript for publication.

Authors' Contributions

All authors contributed to the study concept and design. MM, SW, and PR acquired, analyzed, and interpreted the data as well as supervised the study. All authors contributed to the drafting of the manuscript and its critical revision for important intellectual content. MM performed the statistical analysis. TH obtained the funding and provided the administrative, technical, and material support.

Conflicts of Interest

TH is an employee of the messaging platform used in this study. In the past 36 months, MM received minor consulting fees from Talkspace outside of the submitted work.

References

1. Ozer EJ, Best SR, Lipsey TL, Weiss DS. Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. Psychol Bull 2003 Jan;129(1):52-73. [doi: 10.1037/0033-2909.129.1.52] [Medline: 12555794]
2. PTSD: National Center for PTSD. Accessed February 13, 2017 URL: http://www.ptsd.va.gov/professional/PTSD-overview/epidemiological-facts-ptsd.asp [accessed 2019-08-01]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013.
4. Kessler RC. Posttraumatic stress disorder: the burden to the individual and to society. J Clin Psychiatry 2000;61(Suppl 5):4-12; discussion 13 [FREE Full text] [Medline: 10761674]
5. Helzer JE, Robins LN, McEvoy L. Post-traumatic stress disorder in the general population. Findings of the epidemiologic catchment area survey. N Engl J Med 1987 Dec 24;317(26):1630-1634. [doi: 10.1056/NEJM198712243172604] [Medline: 3683502]
6. Breslau N, Davis GC, Andreski P, Peterson E. Traumatic events and posttraumatic stress disorder in an urban population of young adults. Arch Gen Psychiatry 1991 Mar;48(3):216-222. [doi: 10.1001/archpsyc.1991.01810270028003] [Medline: 1996917]
7. Keane TM, Wolfe J. Comorbidity in post-traumatic stress disorder an analysis of community and clinical studies. J Appl Social Pyschol 1990;20(21):1776-1788. [doi: 10.1111/j.1559-1816.1990.tb01511.x]
8. Sareen J, Houlahan T, Cox BJ, Asmundson GJ. Anxiety disorders associated with suicidal ideation and suicide attempts in the National Comorbidity Survey. J Nerv Ment Dis 2005 Jul;193(7):450-454. [doi: 10.1097/01.nmd.0000168263.89652.6b] [Medline: 15985839]
9. Cougle JR, Resnick H, Kilpatrick DG. PTSD, depression, and their comorbidity in relation to suicidality: cross-sectional and prospective analyses of a national probability sample of women. Depress Anxiety 2009;26(12):1151-1157. [doi: 10.1002/da.20621] [Medline: 19842171]
10. Ikin JF, Creamer MC, Sim MR, McKenzie DP. Comorbidity of PTSD and depression in Korean War veterans: prevalence, predictors, and impairment. J Affect Disord 2010 Sep;125(1-3):279-286. [doi: 10.1016/j.jad.2009.12.005] [Medline: 20071032]
11. Rauch SA, Eftekhari A, Ruzek JI. Review of exposure therapy: a gold standard for PTSD treatment. J Rehabil Res Dev 2012;49(5):679-687 [FREE Full text] [doi: 10.1682/jrrd.2011.08.0152] [Medline: 23015579]
12. Bradley R, Greene J, Russ E, Dutra L, Westen D. A multidimensional meta-analysis of psychotherapy for PTSD. Am J Psychiatry 2005 Feb;162(2):214-227. [doi: 10.1176/appi.ajp.162.2.214] [Medline: 15677582]
13. Alleman JR. Online counseling: the internet and mental health treatment. Psychother Theory Res Pract Train 2002;39(2):199-209. [doi: 10.1037/0033-3204.39.2.199]
14. Hollon SD, Muñoz RF, Barlow DH, Beardslee WR, Bell CC, Bernal G, et al. Psychosocial intervention development for the prevention and treatment of depression: promoting innovation and increasing access. Biol Psychiatry 2002 Sep 15;52(6):610-630. [doi: 10.1016/s0006-3223(02)01384-7] [Medline: 12361671]
15. Nutting PA, Rost K, Dickinson M, Werner JJ, Dickinson P, Smith JL, et al. Barriers to initiating depression treatment in primary care practice. J Gen Intern Med 2002 Feb;17(2):103-111 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10128.x] [Medline: 11841525]
16. Knaevelsrud C, Maercker A. Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. BMC Psychiatry 2007 Apr 19;7:13 [FREE Full text] [doi: 10.1186/1471-244X-7-13] [Medline: 17442125]
17. Collins KA, Westra HA, Dozois DJ, Burns DD. Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care. Clin Psychol Rev 2004 Sep;24(5):583-616. [doi: 10.1016/j.cpr.2004.06.001] [Medline: 15325746]
18. Knaevelsrud C, Brand J, Lange A, Ruwaard J, Wagner B. Web-based psychotherapy for posttraumatic stress disorder in war-traumatized Arab patients: randomized controlled trial. J Med Internet Res 2015 Mar 20;17(3):e71 [FREE Full text] [doi: 10.2196/jmir.3582] [Medline: 25799024]
19. Maieritsch KP, Smith TL, Hessinger JD, Ahearn EP, Eickhoff JC, Zhao Q. Randomized controlled equivalence trial comparing videoconference and in person delivery of cognitive processing therapy for PTSD. J Telemed Telecare 2016 Jun;22(4):238-243. [doi: 10.1177/1357633X15596109] [Medline: 26231819]
20. Morland LA, Greene CJ, Grubbs K, Kloezeman K, Mackintosh M, Rosen C, et al. Therapist adherence to manualized cognitive-behavioral therapy for anger management delivered to veterans with PTSD via videoconferencing. J Clin Psychol 2011 Jun;67(6):629-638. [doi: 10.1002/jclp.20779] [Medline: 21360528]
21. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]

22. McLay RN, Wood DP, Webb-Murphy JA, Spira JL, Wiederhold MD, Pyne JM, et al. A randomized, controlled trial of virtual reality-graded exposure therapy for post-traumatic stress disorder in active duty service members with combat-related post-traumatic stress disorder. Cyberpsychol Behav Soc Netw 2011 Apr;14(4):223-229. [doi: 10.1089/cyber.2011.0003] [Medline: 21332375]
23. Ivarsson D, Blom M, Hesser H, Carlbring P, Enderby P, Nordberg R, et al. Guided internet-delivered cognitive behavior therapy for post-traumatic stress disorder: A randomized controlled trial. Internet Interv 2014 Mar;1(1):33-40. [doi: 10.1016/j.invent.2014.03.002]
24. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
25. Mouthaan J, Sijbrandij M, de Vries G, Reitsma JB, van de Schoot R, Goslings JC, et al. Internet-based early intervention to prevent posttraumatic stress disorder in injury patients: randomized controlled trial. J Med Internet Res 2013 Aug 13;15(8):e165 [FREE Full text] [doi: 10.2196/jmir.2460] [Medline: 23942480]
26. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-4:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
27. Godleski L, Cervone D, Vogel D, Rooney M. Home telemental health implementation and outcomes using electronic messaging. J Telemed Telecare 2012 Jan;18(1):17-19. [doi: 10.1258/jtt.2011.100919] [Medline: 22052966]
28. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
29. Hull TD, Mahan K. A study of asynchronous mobile-enabled SMS text psychotherapy. Telemed J E Health 2017 Mar;23(3):240-247. [doi: 10.1089/tmj.2016.0114] [Medline: 27797646]
30. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
31. PTSD: National Center for PTSD. PTSD Checklist for DSM-5 (PCL-5) URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2019-08-01]
32. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
33. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychother Res 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
34. Horvath AO, Symonds BD. Relation between working alliance and outcome in psychotherapy: A meta-analysis. J Couns Psychol 1991;38(2):139-149. [doi: 10.1037/0022-0167.38.2.139]
35. Muthén LK, Muthén BO. Muthén & Muthén, Mplus Home Page. Los Angeles, CA: Muthén & Muthén; 2017. Mplus User's Guide. Eighth Edition URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf [accessed 2020-02-18]
36. Galatzer-Levy IR, Huang SH, Bonanno GA. Trajectories of resilience and dysfunction following potential trauma: A review and statistical evaluation. Clin Psychol Rev 2018 Jul;63:41-55. [doi: 10.1016/j.cpr.2018.05.008] [Medline: 29902711]
37. Muthén B, Asparouhov T, Hunter AM, Leuchter AF. Growth modeling with nonignorable dropout: alternative analyses of the STAR*D antidepressant trial. Psychol Methods 2011 Mar;16(1):17-33 [FREE Full text] [doi: 10.1037/a0022634] [Medline: 21381817]
38. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Model A Multidiscip J 2007;14(4):535-569. [doi: 10.1080/10705510701575396]
39. Jung T, Wickrama KA. An introduction to latent class growth analysis and growth mixture modeling. Social Pers Psych Compass 2008;2(1):302-317. [doi: 10.1111/j.1751-9004.2007.00054.x]
40. Muthén BO. Latent variable analysis. In: Kaplan DW, editor. The SAGE Handbook of Quantitative Methodology for the Social Sciences. Thousand Oaks, CA: SAGE Publications Inc; 2004:345-368.
41. R Core Team. The R Project for Statistical Computing. Vienna, Austria: R Foundation for statistical computing URL: https://www.r-project.org/ [accessed 2020-02-18]
42. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 1;28(1):112-118. [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
43. Cusack K, Jonas DE, Forneris CA, Wines C, Sonis J, Middleton JC, et al. Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. Clin Psychol Rev 2016 Feb;43:128-141. [doi: 10.1016/j.cpr.2015.10.003] [Medline: 26574151]
44. Gutner CA, Gallagher MW, Baker AS, Sloan DM, Resick PA. Time course of treatment dropout in cognitive-behavioral therapies for posttraumatic stress disorder. Psychol Trauma 2016 Jan;8(1):115-121 [FREE Full text] [doi: 10.1037/tra0000062] [Medline: 26098737]

45. Morland LA, Mackintosh M, Greene CJ, Rosen CS, Chard KM, Resick P, et al. Cognitive processing therapy for posttraumatic stress disorder delivered to rural veterans via telemental health: a randomized noninferiority clinical trial. J Clin Psychiatry 2014 May;75(5):470-476. [doi: 10.4088/JCP.13m08842] [Medline: 24922484]
46. Galatzer-Levy IR, Bryant RA. 636,120 ways to have posttraumatic stress disorder. Perspect Psychol Sci 2013 Nov;8(6):651-662. [doi: 10.1177/1745691613504115] [Medline: 26173229]
47. Cloitre M, Stovall-McClough KC, Miranda R, Chemtob CM. Therapeutic alliance, negative mood regulation, and treatment outcome in child abuse-related posttraumatic stress disorder. J Consult Clin Psychol 2004 Jun;72(3):411-416. [doi: 10.1037/0022-006X.72.3.411] [Medline: 15279525]
48. Keller SM, Zoellner LA, Feeny NC. Understanding factors associated with early therapeutic alliance in PTSD treatment: adherence, childhood sexual abuse history, and social support. J Consult Clin Psychol 2010 Dec;78(6):974-979 [FREE Full text] [doi: 10.1037/a0020758] [Medline: 20873895]
49. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 1995 Dec;52(12):1048-1060. [doi: 10.1001/archpsyc.1995.03950240066012] [Medline: 7492257]

Abbreviations

GMM: growth mixture modeling
HIPAA: Health Insurance Portability and Accountability Act
MMS: multimedia message service
OR: odds ratio
PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders-5.
PTSD: posttraumatic stress disorder
WAI: Working Alliance Inventory

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by S Rauch, T Abdulai, T Aslanidis; comments to author 21.10.19; revised version received 16.12.19; accepted 24.01.20; published 29.04.20.

Please cite as:
Malgaroli M, Hull TD, Wiltsey Stirman S, Resick P
Message Delivery for the Treatment of Posttraumatic Stress Disorder: Longitudinal Observational Study of Symptom Trajectories
J Med Internet Res 2020;22(4):e15587
URL: http://www.jmir.org/2020/4/e15587/
doi:10.2196/15587
PMID:32347814

Original Paper

Effectiveness of One-Way Text Messaging on Attendance to Follow-Up Cervical Cancer Screening Among Human Papillomavirus–Positive Tanzanian Women (Connected2Care): Parallel-Group Randomized Controlled Trial

Ditte S Linde[1,2,3], MScPH, PhD; Marianne S Andersen[4], DMSc, MD, PhD; Julius Mwaiselage[5], MD, PhD; Rachel Manongi[6], MD, PhD; Susanne K Kjaer[7,8], DMSc, MD, PhD; Vibeke Rasch[1,2], DMSc, MD, PhD

[1]Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[2]Department of Obstetrics and Gynaecology, Odense University Hospital, Odense, Denmark
[3]Odense Patient Data Explorative Network, Odense University Hospital, Odense, Denmark
[4]Department of Medical Endocrinology, Odense University Hospital, Odense, Denmark
[5]Department for Cancer Prevention Services, Ocean Road Cancer Institute, Dar es Salaam, United Republic of Tanzania
[6]Department of Public Health, Kilimanjaro Christian Medical University College, Moshi, United Republic of Tanzania
[7]Department of Gynaecology, Rigshospitalet University Hospital, Copenhagen, Denmark
[8]Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark

Corresponding Author:
Ditte S Linde, MScPH, PhD
Department of Clinical Research
University of Southern Denmark
Kløvervænget 10, 10th Fl
Odense, 5000
Denmark
Phone: 45 61666564
Email: dsondergaard@health.sdu.dk

Abstract

Background: Rapid human papillomavirus (HPV) DNA testing is an emerging cervical cancer screening strategy in resource-limited countries, yet it requires follow-up of women who test HPV positive.

Objective: This study aimed to determine if one-way text messages improved attendance to a 14-month follow-up cervical cancer screening among HPV-positive women.

Methods: This multicenter, parallel-group randomized controlled trial was conducted at 3 hospitals in Tanzania. Eligible participants were aged between 25 and 60 years, had tested positive to a rapid HPV test during a patient-initiated screening, had been informed of their HPV result, and had a private mobile phone with a valid number. Participants were randomly assigned in a 1:1 ratio to the intervention or control group through an incorporated algorithm in the text message system. The intervention group received one-way text messages, and the control group received no text messages. The primary outcome was attendance at a 14-month health provider-initiated follow-up screening. Participants were not blinded, but outcome assessors were. The analysis was based on intention to treat.

Results: Between August 2015 and July 2017, 4080 women were screened for cervical cancer, of which 705 were included in this trial—358 women were allocated to the intervention group, and 347 women were allocated to the control group. Moreover, 16 women were excluded before the analysis because they developed cervical cancer or died (8 from each group). In the intervention group, 24.0% (84/350) women attended their follow-up screening, and in the control group, 23.8% (80/335) women attended their follow-up screening (risk ratio 1.02, 95% CI 0.79-1.33).

Conclusions: Attendance to a health provider-initiated follow-up cervical cancer screening among HPV-positive women was strikingly low, and one-way text messages did not improve the attendance rate. Implementation of rapid HPV testing as a primary screening method at the clinic level entails the challenge of ensuring a proper follow-up of women.

Trial Registration: ClinicalTrials.gov NCT02509702; https://clinicaltrials.gov/ct2/show/NCT02509702.

International Registered Report Identifier (IRRID): RR2-10.2196/10.2196/15863

(J Med Internet Res 2020;22(4):e15863) doi:10.2196/15863

KEYWORDS

telemedicine; cervical cancer; HPV; early detection of cancer; Africa

Introduction

Background

Despite the fact that cervical cancer can be prevented, the global age-standardized incidence and mortality rates of cervical cancer were 13.1 and 6.9, respectively, per 100,000 women in 2018 [1]. Cervical cancer is the fourth most common cancer among women worldwide [1], and the burden of disease is unequally distributed, with 70% of the cases occurring in low-resource countries [2]. In Tanzania, cervical cancer constitutes 38% of all new cancer cases among women, and the age-standardized incidence and mortality rates were 59.1 and 42.7, respectively, per 100,000 women in 2018 [3].

Primary prevention of cervical cancer includes vaccination against high-risk human papillomavirus (HPV) types, and secondary prevention involves early and accurate detection through screening and subsequent treatment of precancerous lesions [4]. In 2018, Tanzania launched HPV vaccination of girls aged 9 to 14 years [5]; however, older generations of women fully rely on screening to prevent the development of cervical cancer. Currently, Tanzania follows the World Health Organization's *screen-and-treat* guideline for resource-limited settings, which involves visual inspection with acetic acid (VIA) with immediate treatment if lesions appear [6]. Rapid HPV DNA testing is an emerging screening strategy in resource-limited countries, as it has proven to be an objective and more sensitive alternative to VIA [7,8]. However, the test only identifies women who are high-risk HPV positive and thus at increased risk of developing cervical cancer. Consequently, subsequent screening of these women is needed to identify those who have precancerous lesions. This leads to a risk of loss to follow-up compared with screening by the use of VIA. Yet, this issue is still largely unexplored in an African context.

In recent years, mobile phone access has expanded rapidly across Africa, and in 2018, there were 81 mobile subscriptions per 100 people in Tanzania [9]. The increasing mobile phone access has gone hand in hand with a growing number of mobile health (mHealth) interventions that aim to address global health issues in innovative ways. One-way text message interventions—also referred to as one-way SMS—involve sending text messages to a recipient who cannot respond to the text message. The advantages of text messages are that they are easy to use, can be sent to the receivers simultaneously, and require less staff. A recent systematic review and meta-analysis of one-way text message trials in Africa showed that one-way text messages have been tested within different clinical areas across Africa, although mainly in relation to medicine adherence and appointment attendance. The effect of one-way text messages varied across clinical areas, and overall, the highest effect was found in relation to increasing attendance to a childhood immunization appointment [10]. In relation to cervical cancer screening attendance, a systematic review from 2017 concerning text message interventions on cancer screening rates [11] found 1 trial from Malaysia with no effect of one-way text messages improving attendance to a repeat cervical smear compared with postal letters [12]. However, recent trials from Tanzania and Kenya have shown that one-way text messages increased attendance to cervical cancer screening among screening-naïve women [13] and a repeat cervical smear [14], compared with no text messages. Yet, it is still unknown whether or not one-way text messages can increase the attendance rate of HPV-positive women who have been appointed a follow-up screening.

Objectives and Study Context

The aim of this study (*Connected2Care)* was to assess the effect of one-way text messages on attendance to a provider-initiated follow-up screening appointment among women who had tested HPV positive during a patient-initiated opportunistic screening 14 months earlier. In addition, we examined factors associated with attendance regardless of group allocation. *Connected2Care* is a substudy of a larger research study, Comprehensive Cervical Cancer Prevention in Tanzania (CONCEPT). The CONCEPT study is linked to the existing national cervical cancer screening programs in Dar es Salaam and Kilimanjaro and will end in December 2021.

Methods

Study Design

Connected2Care is an unblinded, multicenter, parallel-group randomized controlled trial conducted at 3 hospitals in Tanzania: Ocean Road Cancer Institute (ORCI) in Dar es Salaam and Kilimanjaro Christian Medical Centre (KCMC) and Mawenzi Regional Hospital in Moshi. Originally, the study was designed as a double-site study (ORCI and KCMC); however, because of a slower-than-expected recruitment rate, a third study site (Mawenzi Regional Hospital) was added 6 months into the recruitment phase. To increase recruitment further, fliers informing about the screening were shared at mosques and churches close to the study sites, and 2 outreach screenings were conducted by the ORCI screening nurses in Dar es Salaam. The study protocol has been published elsewhere [15], and the joined ethical approval for all study sites was obtained from the National Institute for Medical Research in Tanzania.

Participants

Women aged 25 to 60 years, who attended a patient-initiated cervical cancer screening at the study sites, were assessed by a screening nurse for overall eligibility to participate in the CONCEPT study. Eligible women gave written informed consent. The exclusion criteria were pregnancy or menstruation on the day of enrollment, previous hysterectomy, cervical

cancer, or diagnosis of cervical precancerous lesions within the past 12 months. The subgroup of CONCEPT participants, who tested high-risk HPV positive at the enrollment screening, was assessed for eligibility to be further included in the *Connected2Care* study. Women were ineligible for inclusion in *Connected2Care* if they did not own a private mobile phone, had provided an invalid phone number (ie, digits missing), or were not informed of their positive HPV test result. Ineligible women were excluded before randomization.

Randomization and Blinding

Participants were randomly assigned to the intervention or control group in a 1:1 ratio. The randomization occurred through an incorporated algorithm in the text message system, which automatically assigned participants to the intervention or control group. The random sequence generator was developed by the external information technology consultant, who developed the text messaging system, and it was concealed to all investigators. Screening nurses enrolled participants and were concealed to the latter group allocation. The first author (DL) assigned participants to the trial by uploading the phone numbers of the eligible participants into the text messaging system and was not concealed to the group allocation. Due to the overt nature of the text message intervention, the study participants were not concealed to their group allocation. Yet, the outcome assessors—in the form of screening nurses registering the attendance date—were blinded.

Procedures

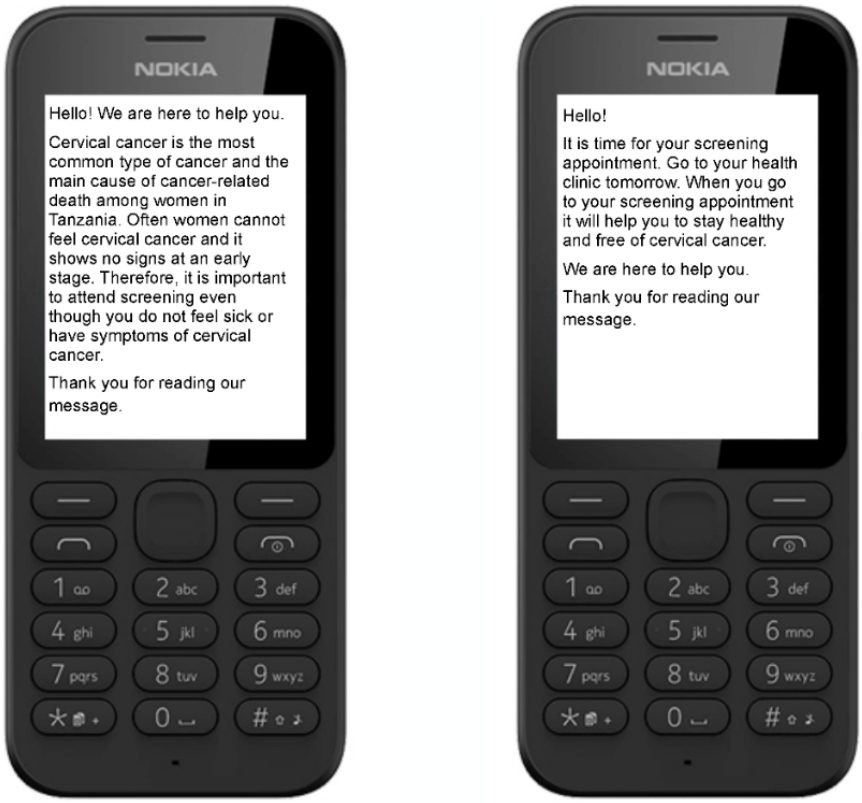
Eligible women were assigned an anonymous study ID and interviewed by a trained nurse using a structured questionnaire. The nurse also registered their home address and phone numbers as well as provided general cervical cancer screening education and counseling according to the national guidelines for cervical cancer screening in Tanzania. During the enrollment screening, 2 cervical specimens were taken using a careHPV brush [16] and a ThinPrep Pap Test plastic spatula [17] before VIA was conducted. Women who tested VIA positive were treated onsite with cryotherapy or loop electrosurgical excision procedure. VIA status and treatment did not influence inclusion into the trial, and the results from the ThinPrep-collected specimens (cytology and hybrid capture 2) were not available until postinclusion, hence did not influence enrollment either. The test performance of VIA, careHPV, and Hybrid Capture2 for the detection of cytologically diagnosed high-grade cervical lesions or cancer is not addressed in the *Connected2Care* trial, but in other parts of the CONCEPT project [18]. Once the screening had finished, all women were assigned a 14-month follow-up screening appointment, which was written on a physical appointment card together with their study ID. The appointment card was considered standard care.

Using the rapid careHPV test, the specimens were tested for 14 types of high-risk HPV (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) in a local laboratory at 2 of the study sites (ORCI and KCMC). Community nurses attempted to conduct a private home visit to inform all careHPV-positive women about their results. The women were informed that 1 of the extra tests they had had during their cervical cancer screening (HPV test) turned out positive, yet this did not mean that they had cervical cancer. Furthermore, they were informed that HPV is a common infection that most likely will clear on its own, but it was important that they attended their follow-up screening as it can develop into cervical cancer over time. Finally, the women were told that they may receive health educative and reminder text messages before their next screening appointment. Inclusion into *Connected2Care* did not occur until it was confirmed to the first author (DL) that a woman had been informed of her result, after which she was uploaded to the text message system and assigned to either the *text message plus standard care group* (intervention group) or the *standard care group* (control group).

Over a period of 10 months, 10 health educative messages, and 5 reminders were sent to the women in the intervention group. The health educative messages were sent once a month and contained information about screening, risk factors, and symptoms for cervical cancer (Figure 1). The reminders were sent 14, 7, and 1 day before the 14-month follow-up screening appointment as well as 1 and 7 days post the appointment date. The sender ID, content, timing, and number of text messages were pretested on screening clients as well as on screening nurses before starting the intervention. The rationale behind combining health education and reminders was that health education would enhance the perceived seriousness of the disease as well as the benefit of the screening, and combined with reminders, they would be cues to actions for screening attendance. The messages were not personalized out of privacy concerns, and they were developed in English and translated into Kiswahili using back-and-forth translation. A Web-based text message portal [19]—developed specifically for this study—automatically dispatched the messages and sent them with the ID *Elimu Ya Afya* (meaning *health education* in Kiswahili). The portal had a *delivery note feature*, which showed the discrepancy between the number of text messages sent and the number of text messages received. All data were entered and stored in Research Electronic Data Capture [20].

The women who attended their follow-up screening had a cervical specimen collected using the ThinPrep Pap Test plastic spatula and went through a standard VIA screening. Furthermore, they were reinterviewed using a structured questionnaire. Their attendance date was noted down in a registration book by the screening nurse. If a woman did not attend her follow-up screening within 30 days of their appointment date, the *Connected2Care* trial was considered finished, and an *active posttrial follow-up* started to trace the nonattending woman. First, a nurse called the woman to encourage her to attend her appointment at the clinic and offered to refund transportation costs. Second, if the woman did not attend, a nurse visited her at home and encouraged her to come to the clinic and attend her appointment. Third, if the woman still did not attend and she had consented to it, an outreach home visit was conducted, where the woman was interviewed, and a self-collected cervical swab was collected using the Evalyn brush [21]. This swab substituted the swab that would have been collected at the screening clinic by the use of the ThinPrep Pap Test plastic spatula, and these results are yet to be published as another part of the CONCEPT study.

Figure 1. Example of intervention text messages.



Outcomes

The primary outcome was attendance at a health provider–initiated follow-up screening at 14 months. It was measured as whether or not the participants attended the follow-up screening within 30 days of their screening appointment. Post hoc, we exploratively assessed factors that were associated with attendance, that is, the number of text messages received and sociodemographic characteristics, as well as how attendance had increased postintervention via phone calls and home visits. Prespecified secondary outcomes included cost-effectiveness of the intervention, the intervention's effect on the knowledge of cervical cancer and screening (16-item true-false questionnaire), and barriers against implementing a text message intervention in Tanzania (mixed methods study). The secondary outcomes, cost-effectiveness and knowledge, have not been assessed, and the barriers against implementation have been assessed in a qualitative postintervention study together with factors that influence screening attendance. The latter objective has been published elsewhere [22], whereas data on barriers related to the text message intervention are yet to be published.

Statistical Analysis

The sample size was calculated based on the methods described by Altman [23]. We estimated that the intervention would increase the attendance rate with 15 points and that 73% of women in the intervention group, and 58% of women in the control group would attend their follow-up appointment. With an allocation ratio of 1:1, a two-sided alpha of 5%, and 80% power, we required approximately 350 participants in each arm. At the time of design, no other text message interventions on cervical cancer had been conducted in Africa, and the estimation was based on the effect of mHealth interventions found within other clinical areas in Eastern Africa [24,25]. The analysis was based on intention to treat, and participants who developed cervical cancer or died during the intervention period were excluded before analysis, as attendance was not an option for them. We calculated a risk ratio (RR) to determine if the proportion of participants who attended their follow-up screening differed between groups and used a relative risk regression by the use of a generalized linear model with log-link function and binomial distribution as statistical family. In subgroup analyses, we assessed if the intervention had a differential effect across subgroups by including an interaction term between the intervention allocation and a subgroup-defining variable. Stata 15 [26] was used to analyze the data. A data monitoring committee did not oversee the study. The study has been registered at ClinicalTrials.gov (NCT02509702).

Results

Trial Results

Study participants were enrolled in the study between August 17, 2015, and July 6, 2017, and the follow-up ended on October 6, 2018. Altogether 4080 women were enrolled in the CONCEPT study. Of these women, 705 were included in *Connected2Care* (Figure 2). After randomization, 4 participants in each group developed cervical cancer, and 4 participants from each group died from the disease (n=16). These participants were excluded from the analysis, leaving 689 women for analysis.

The intervention and control group were comparable (Table 1), and the primary analysis showed that the intervention had no

effect on attendance to a follow-up screening appointment. Of the 350 women in the intervention group, 84 (24.0%) attended their follow-up screening, and of the 335 women in the control group, 80 (23.8%) attended their follow-up screening (RR 1.02, 95% CI 0.79-1.33).

Six months into the study—after enrollment of 38 participants—we discovered that the SMS system did not dispatch the text messages according to the study plan. We reassigned 38 participants into a new SMS system [19], and they stayed in the groups they were originally allocated to [15]. We conducted a post hoc sensitivity analysis excluding these women and another post hoc analysis excluding 16 participants who were misclassified as HPV positive and erroneously included in the study. The sensitivity analysis of the SMS system (RR 0.95, 95% CI 0.71-1.26) and misclassifications (RR 0.98, 95% CI 0.75-1.29) showed similar results as our main analysis. Moreover, 1 of 8 explorative subgroup analyses showed that HIV status was a potential effect modifier (not adjusted for multiplicity; Figure 3).

Figure 2. Flowchart of the trial. *Found out post inclusion. HPV: human papillomavirus.

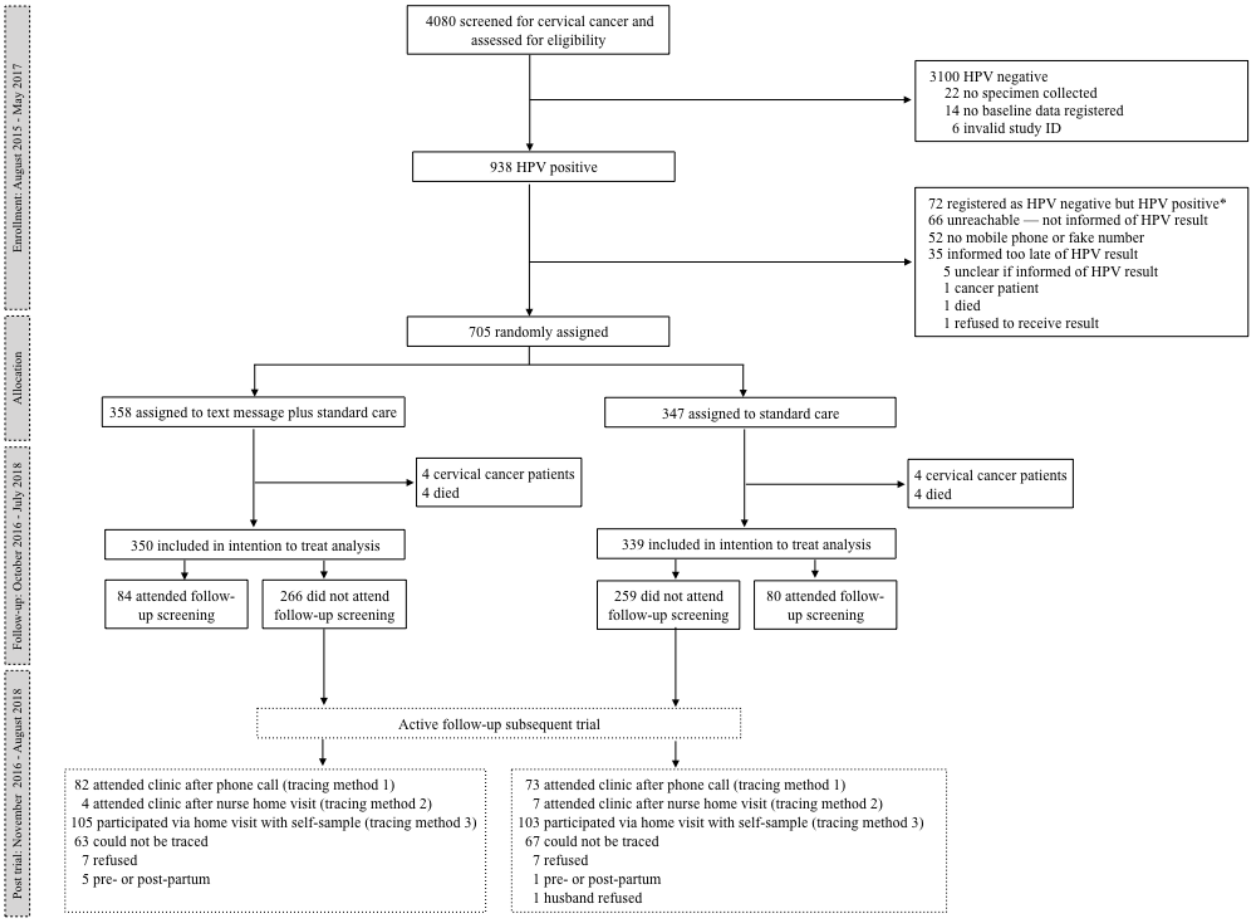
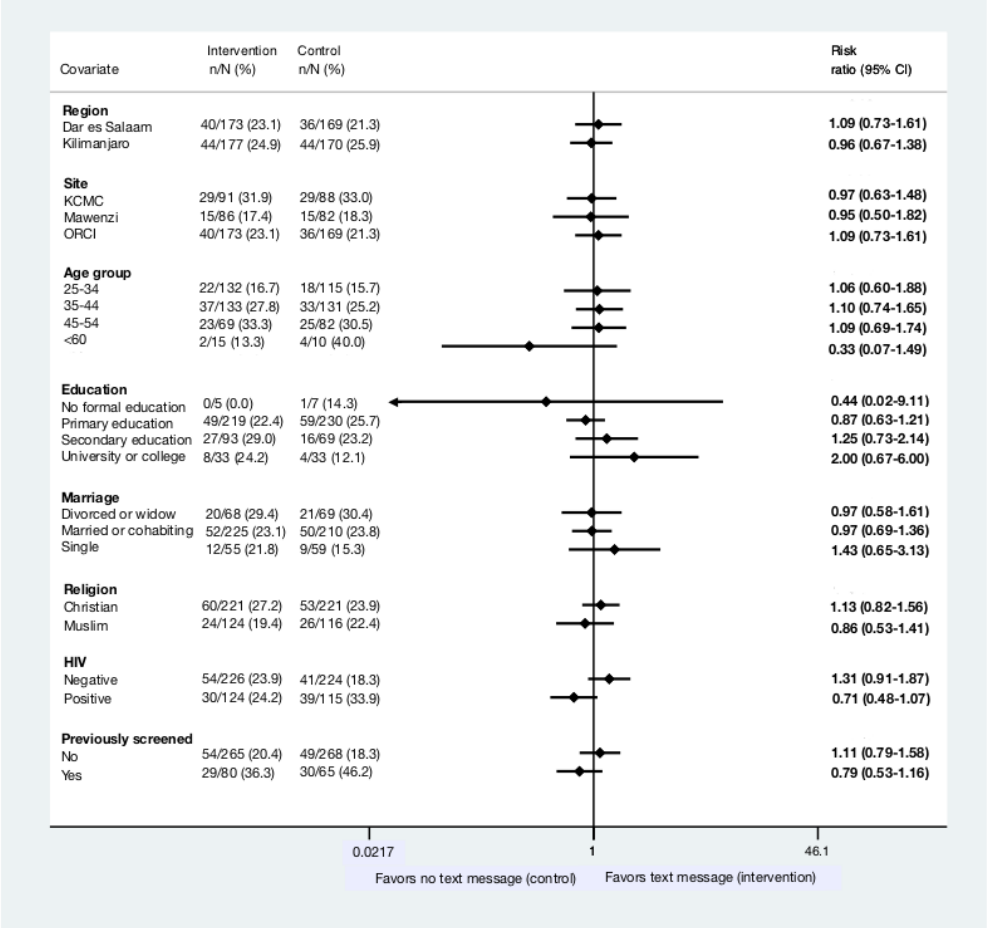


Table 1. Baseline characteristics.

Sociodemographic characteristics	Text message group (n=350)	Control group (n=339)
Region, n (%)		
Dar es Salaam	173 (49.4)	169 (49.8)
Kilimanjaro	177 (50.6)	170 (50.1)
Clinic, n (%)		
Ocean Road Cancer Institute	173 (49.4)	169 (59.8)
Kilimanjaro Christian Medical Centre	91 (26.0)	88 (25.9)
Mawenzi Regional Hospital	86 (24.6)	82 (24.2)
Age (years), mean (SD)	38.2 (8.8)	39.2 (8.4)
Education, n (%)		
No formal education	5 (1.4)	7 (2.1)
Primary education	219 (62.6)	230 (67.8)
Secondary education	93 (26.6)	69 (20.4)
University or college	33 (9.4)	33 (9.7)
Marital status[a], n (%)		
Married or cohabiting	226 (64.6)	210 (61.9)
Single	55 (15.7)	59 (17.4)
Divorced or widow	68 (19.4)	69 (20.1)
Religion[a], n (%)		
Christian	223 (63.7)	221 (65.2)
Muslim	124 (35.4)	116 (34.2)
HIV status, n (%)		
HIV positive	124 (35.4)	115 (33.9)
HIV negative	226 (64.6)	224 (66.1)
Previously screened for cervical cancer[a], n (%)		
Yes	81 (23.1)	65 (19.2)
No	268 (77.6)	271 (80.0)
Self-perceived health[a], n (%)		
Excellent or very good	62 (17.7)	77 (22.7)
Good	239 (68.3)	200 (59.0)
Bad or less good	48 (13.7)	60 (17.7)

[a]Missing under 1%.

Figure 3. Forrest plot of subgroup analysis. KCMC: Kilimanjaro Christian Medical Centre; ORCI: Ocean Road Cancer Institute.



The SMS dispatching system showed that all participants in the intervention group received at least one of the 15 text messages sent to them, 8% (2/26) received between 1 and 4 text messages, and 23.4% (26/111) received all 15 text messages (Table 2). When examining factors potentially associated with attendance to screening, we found that the number of text messages received did not appear to affect the attendance rate. In contrast, women aged 35 to 44 years (RR 1.64, 95% CI 1.16-2.33) and 45 to 54 years (RR 1.97, 95% CI 1.36-2.84) were more likely to attend their screening compared with women aged 25 to 34 years; HIV-positive women were more likely to attend their screening than HIV-negative women (RR 1.37, 95% CI 1.05-1.79); and women who had previously been screened were more likely to attend their screening than women who had not been screened before their screening at baseline (RR 2.09, 95% CI 1.61-2.72; Table 2).

Table 2. Factors associated with attendance.

Sociodemographic and SMS characteristics	Attendance, n (%)	Attendance (adjusted for SMS; linear)		
		Risk ratio (95% CI)	SE	z-statistic
SMS received[a]				
15 (all, n=111)	26 (23.4)	1.00	N/A[b]	N/A
10-14 (n=141)	37 (26.2)	1.12 (0.68-1.48)	0.25	0.51
9-5 (n=65)	18 (28)	1.18 (0.72-1.73)	0.31	0.63
1-4 (n=26)	2 (8)	0.32 (0.68-1.48)	0.23	−1.59
0 (control group, n=339)	80 (23.6)	1.00 (0.61-1.67)	0.2	0.04
Region				
Dar es Salaam (n=342)	76 (22.2)	1.00	N/A	N/A
Kilimanjaro (KCMC[c]/Mawenzi, n=347)	88 (25.3)	1.14 (0.87-1.49)	0.16	0.96
Clinic				
Ocean Road Cancer Institute (n=342)	76 (22.2)	1.00	N/A	N/A
KCMC (n=179)	58 (32.4)	1.46 (1.09-1.95)	0.22	2.55
Mawenzi Regional Hospital (n=168)	30 (17.9)	0.80 (0.55-1.18)	0.16	−1.13
Age (years)[d]				
25-34 (n=248)	40 (16.1)	1.00	N/A	N/A
35-44 (n=264)	70 (26.5)	1.64 (1.16-2.33)	0.29	2.8
45-54 (n=151)	48 (31.8)	1.97 (1.36-2.84)	1.97	0.37
55-60 (n=26)	6 (23)	1.47 (0.69-3.13)	1.47	0.57
Education				
No formal education (n=12)	1 (8)	0.35 (0.53-2.28)	0.33	−1.10
Primary education (n=449)	108 (24.1)	1.00	N/A	N/A
Secondary education (n=162)	43 (26.5)	1.10 (0.81-1.50)	0.17	0.63
University or college (n=66)	12 (18)	0.76 (0.44-1.29)	0.76	0.21
Marital status[e]				
Married or cohabiting (n=436)	102 (23.4)	1.00	N/A	N/A
Single (n=114)	21 (18.4)	0.79 (0.52-1.20)	0.17	−1.12
Divorced or widow (n=137)	41 (29.9)	1.28 (0.94-1.74)	0.2	1.56
Religion[e]				
Christian (n=444)	113 (25.5)	1.00	N/A	N/A
Muslim (n=240)	50 (20.8)	0.81 (0.61-1.09)	0.12	−1.37
HIV status				
HIV negative (n=450)	95 (21.1)	1.00	N/A	N/A
HIV positive (n=239)	69 (28.0)	1.37 (1.05-1.79)	0.19	2.29
Previously screened for cervical cancer[e]				
No (n=539)	104 (19.3)	1.00	N/A	N/A
Yes (n=146)	59 (40.4)	2.11 (1.62-2.75)	0.28	5.57
Self-perceived health[e]				
Excellent or very good (n=139)	31 (22.3)	1.00	N/A	N/A
Good (n=439)	108 (24.6)	1.10 (0.78-1.57)	0.2	0.56

Sociodemographic and SMS characteristics	Attendance, n (%)	Attendance (adjusted for SMS; linear)		
		Risk ratio (95% CI)	SE	z-statistic
Bad or less good (n=108)	25 (23.1)	1.04 (0.65-1.65)	0.25	0.16

[a]For 7 participants, it is unknown how many text messages were received. Not adjusted for SMS linear.

[b]Not applicable.

[c]KCMC: Kilimanjaro Christian Medical Centre.

[d]Risk Ratio, continuous (95% CI): 1.03 (1.01-1.04).

[e]Missing under 1%.

Posttrial Results

Once the trial had finished, the *active* posttrial follow-up started where women in the intervention and control groups were traced using the same methods. Overall, 24% of women attended their screening at the clinic within 30 days of their appointment (trial period). An additional 23% of women attended the follow-up screening at the clinic after being called (tracing method 1), and a further 2% of women attended the clinic after having a nurse home visit (tracing method 2). Finally, an additional 30% of women had their follow-up screening via a self-sample test during a home visit (tracing method 3). Hence, the total number of women who received a follow-up screening was 77.9% (537/689; Table 3).

Table 3. Post follow-up attendance rate.

Attendance	Text message group (n=350), n (%)	Control group (n=339), n (%)	Total (text message+control group combined, n=689), n (%)
During trial			
Attendance, primary analysis	84 (24.0)	80 (23.4)	164 (24.4)
Post trial—both groups receive the same tracing methods			
Attendance at clinic after a phone call (tracing method 1)	82 (23.4)	73 (21.5)	155 (22.5)
Attendance at clinic after a home visit (tracing method 2)	4 (1.1)	7 (2.1)	11 (1.6)
Participation via self-sampling at the home level (tracing method 3)	105 (30.0)	103 (30.3)	208 (30.1)
Total	275 (78.5)	262 (77.2)	537 (77.9)

Discussion

Principal Findings

This trial shows that one-way text messages did not improve the attendance rate of a health provider-initiated follow-up cervical cancer screening among HPV-positive women. Overall, one-fourth of the participants attended their follow-up screening regardless of whether or not they had received text messages. Hence, the barrier of getting women to return to the clinic was not one that one-way text messages appeared to overcome. Not all text messages dispatched were received among the intervention group, yet the number of messages received did not affect the attendance rate. Once the trial had finished, an additional 23% of women attended the clinic after a nurse had called them, and 2% of women attended after a nurse home visit. Of the remaining nonattendees, a further 30% of women participated via a home visit and self-sampling, which led to a total follow-up rate of 78%.

Limitations

We encountered several obstacles while conducting this trial. First, we had to switch to a new SMS system 6 months into the trial because of the first system being unreliable. However, our sensitivity analysis found no difference in effect among the participants who had been enrolled using the first system. Second, a number of women were misclassified as HPV negative or HPV positive in the process of transferring the laboratory reports to the CONCEPT investigators. This led to a number of women being incorrectly excluded from or included in the trial. The sensitivity analysis of the incorrectly included women showed no difference in results. Although these issues are study specific, they highlight a need for incorporating secure procedures when implementing a more complex screening method such as rapid HPV testing in a setting similar to Tanzania. Third, the attendance rate was much lower than what we had anticipated, which led us to part from some of our originally preplanned secondary outcomes and examine what affects attendance in more detail. However, we clearly specified which analyses and results were conducted post hoc so as to not hypothesize after the results were known (ie, avoid hypothesizing after the results are known [HARKing]) [27]. Furthermore, as the active posttrial follow-up could not be carried out over a short period, we cannot rule out that participants have been contaminated by this and, for example, have waited for a home visit and self-collected sample instead of having to go to the clinic. If this is the case, this could have had a negative effect on the attendance rate of the intervention and control group participants. Despite the fact that we pilot tested the text messages before starting the trial, it is plausible

that the use of a health behavior theoretical framework and further pilot testing of the intervention and the text messaging portal could have addressed some of the study challenges.

Owing to the lack of blinding of participants, there is a risk of performance bias and risk of contamination, which could have affected the internal validity of the trial [28]. To preserve privacy, we did not personalize the text messages, and we excluded women who did not own a private mobile phone. However, this exclusion criterion affects the external validity of the trial, as the participants may not represent the target population. Despite our effort to ensure privacy for the study participants, we cannot guarantee that the participants found the messages confidential enough. If this was an issue, it could have affected the acceptability and effectiveness of the messages. Our explorative subgroup analysis indicated that HIV status may be a potential effect modifier. However, our trial was not dimensioned to assess differential effects across subgroups, and it is likely an artifact of the data and a false discovery rather than HIV status modifying the relationship between text messages and follow-up.

Comparison With Prior Work

To our knowledge, this is the first trial from Africa that investigates if attendance to a health provider-initiated follow-up cervical cancer screening among HPV-positive women can be increased by the use of one-way text messages. No previous trials have addressed the follow-up of women who have tested HPV positive; hence, the actual comparability of our study results is limited. Nonetheless, the findings from the trial are somewhat in contrast to other cervical cancer one-way text message trials from East Africa. One trial from northern Tanzania found that one-way text messages increased the attendance rate among screening-naïve women (odds ratio [OR] 3.0, 95% CI 1.5-6.2) and that one-way text messages combined with an electronic voucher (eVoucher) increased it even further (OR 4.7, 95% CI 2.9-7.4). Still, the attendance rate was low, with 4.3% (12/281) of women attending in the control group, 12.9% (35/272) in the text message group, and 18.1% (54/298) in the text message plus eVoucher group, which indicates that one-way text messages did not manage to make the vast majority of women attend screening [13]. Furthermore, a trial from Kenya found that one-way text messages increased the attendance rate at a follow-up screening among women who had a normal pap smear at baseline (OR 8.0, 95% CI 4.7-13.7); 67.1% (96/143) of women attended in the intervention group and 20.3% (29/143) in the control group. However, the trial has been judged as a high-risk bias trial in a recent systematic review [10] and was published in a journal on Beall's list of potential predatory publishers [29].

The outcome of our trial is highly relevant in a larger mHealth context and in relation to how to address the cervical cancer burden in East Africa. Rapid HPV testing is an area that has the potential to improve the prevention of the disease; however, this trial shows that implementation of rapid HPV testing leads to the challenge of providing proper follow-up of the women who test HPV positive. This is an issue that policy makers and global health clinicians should be aware of in relation to implementing rapid HPV testing as a primary screening method in future cervical cancer screening programs in Africa. Our posttrial follow-up strategy indicates that phone calls where nurses engage with the women and emphasize the importance of reattendance may be more effective than a one-way text message intervention. This finding is supported by a Malaysian trial on screening attendance, which found that one-way text messages had no effect, but phone calls did [12]. Furthermore, our trial shows that there is a potential for addressing the issue of nonattendance by implementing HPV self-sample testing at the home level. A qualitative interview study, which we conducted in the postintervention period among 15 women from the intervention group, showed that both behavioral and structural barriers affected screening reattendance, including fear of the gynecologic examination as well as transport and waiting time at the clinic [22]. HPV self-sample testing has the potential to overcome some of these barriers, and it has also proven to be an acceptable and effective method in other settings in Africa [30-33]. However, women who test positive to an HPV self-sample test would still require a form of visual inspection of the cervix to assess the features that would make it amenable or not. How best to implement rapid HPV tests, the overall care pathway for HPV-positive women and the potential of phone call reminders for cervical cancer follow-up are areas worth exploring in future studies in Africa.

Conclusions

This study illustrates that the implementation of rapid HPV testing at cervical cancer screening clinics in Tanzania entails a number of challenges, including ensuring a proper follow-up of women who test HPV positive. One-way text messages had no effect on the attendance rate; however, it is plausible that phone calls or outreach services that involve HPV testing at home may be more promising methods than one-way text messages and repeat screenings at the clinic level. This should be explored in future studies.

Acknowledgments

The study is funded by the Danish International Development Agency (Danida; 14-P02-Tan/A26775) and the University of Southern Denmark (internal funds). The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. DSL, JM, and VR had access to all the data in the study, and DSL had the final responsibility for the decision to submit for publication. The authors would like to thank all partners involved in the CONCEPT study: screening nurses, laboratory personnel, outreach nurses, fellow researchers, and all the anonymous women who are part of the CONCEPT study. The authors would also like to thank Andreas Hover Lundh for commenting on the manuscript, Pia Veldt Larsen for statistical support, and Yusuph Kassim for developing the SMS system and providing technical support for the system throughout the trial.

Authors' Contributions

SK, VR, JM, and RM conceptualized the idea and designed the study in collaboration with MA and DSL. DSL analyzed the data, and SK, VR, and MA verified the analysis. DSL drafted the initial manuscript, and SK, VR, JM, RM, and MA critically revised it. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2658 KB - jmir_v22i4e15863_app1.pdf]

References

1. International Agency for Research on Cancer. Global Cancer Observatory - iarc.: The Global Cancer Observatory; 2019. Population Fact Sheets: World URL: http://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf [accessed 2019-07-20]
2. Stewart BW, Wild CP. World Cancer Report 2014. France: International Agency for Research on Cancer; 2014.
3. International Agency for Cancer Research. Global Cancer Observatory - iarc. 2019. Population Fact Sheets: Tanzania, United Republic of URL: http://gco.iarc.fr/today/fact-sheets-populations [accessed 2019-07-20]
4. Finocchario-Kessler S, Wexler C, Maloba M, Mabachi N, Ndikum-Moffor F, Bukusi E. Cervical cancer prevention and treatment research in Africa: a systematic review from a public health perspective. BMC Womens Health 2016 Jun 4;16:29 [FREE Full text] [doi: 10.1186/s12905-016-0306-6] [Medline: 27259656]
5. WHO | Regional Office for Africa. 2018. Tanzania Rolls Out Vaccination Against Cervical Cancer URL: https://www.afro.who.int/news/tanzania-rolls-out-vaccination-against-cervical-cancer [accessed 2019-07-15]
6. World Health Organization. WHO Guidelines for Screening and Treatment of Precancerous Lesions for Cervical Cancer Prevention. Geneva: World Health Organization; 2013.
7. Umulisa MC, Franceschi S, Baussano I, Tenet V, Uwimbabazi M, Rugwizangoga B, et al. Evaluation of human-papillomavirus testing and visual inspection for cervical cancer screening in Rwanda. BMC Womens Health 2018 Apr 24;18(1):59 [FREE Full text] [doi: 10.1186/s12905-018-0549-5] [Medline: 29699549]
8. Dartell MA, Rasch V, Iftner T, Kahesa C, Mwaiselage JD, Junge J, et al. Performance of visual inspection with acetic acid and human papillomavirus testing for detection of high-grade cervical lesions in HIV positive and HIV negative Tanzanian women. Int J Cancer 2014 Aug 15;135(4):896-904 [FREE Full text] [doi: 10.1002/ijc.28712] [Medline: 24391021]
9. Tanzania Communications Regulatory Authority. 2018. Quaterly Communications Statitics. October - December 2018 Operators' Submissions URL: https://www.tcra.go.tz/images/documents/telecommunication/TelCom_Statistics_December_2018.pdf [accessed 2020-01-28]
10. Linde DS, Korsholm M, Katanga J, Rasch V, Lundh A, Andersen MS. One-way SMS and healthcare outcomes in Africa: Systematic review of randomised trials with meta-analysis. PLoS One 2019;14(6):e0217485 [FREE Full text] [doi: 10.1371/journal.pone.0217485] [Medline: 31170176]
11. Uy C, Lopez J, Trinh-Shevrin C, Kwon S, Sherman S, Liang P. Text messaging interventions on cancer screening rates: a systematic review. J Med Internet Res 2017 Aug 24;19(8):e296 [FREE Full text] [doi: 10.2196/jmir.7893] [Medline: 28838885]
12. Rashid RM, Mohamed M, Hamid ZA, Dahlui M. Is the phone call the most effective method for recall in cervical cancer screening?--results from a randomised control trial. Asian Pac J Cancer Prev 2013;14(10):5901-5904 [FREE Full text] [doi: 10.7314/apjcp.2013.14.10.5901] [Medline: 24289597]
13. Erwin EM, Aronson KJ, Day A, Ginsburg O, Macheku G, Feksi A, et al. SMS behaviour change communication and eVoucher interventions to increase uptake of cervical cancer screening in the Kilimanjaro and Arusha regions of Tanzania: a randomised, double-blind, controlled trial of effectiveness. BMJ Innov 2019 Jan;5(1):28-34 [FREE Full text] [doi: 10.1136/bmjinnov-2018-000276] [Medline: 31645991]
14. Wanyoro AK, Kabiru EW. Use of mobile phone short text message service to enhance cervical cancer screening at Thika Level 5 Hospital, Kiambu County, Kenya: a randomised controlled trial. Res Gynecol Obstet 2017;5(1):10-20 [FREE Full text] [doi: 10.5923/j.rog.20170501.03]
15. Linde DS, Andersen MS, Mwaiselage JD, Manongi R, Kjaer SK, Rasch V. Text messages to increase attendance to follow-up cervical cancer screening appointments among HPV-positive Tanzanian women (Connected2Care): study protocol for a randomised controlled trial. Trials 2017 Nov 21;18(1):555 [FREE Full text] [doi: 10.1186/s13063-017-2215-x] [Medline: 29162148]
16. Qiagen. Netherlands: Qiagen careHPV Test Kit URL: https://www.qiagen.com/nl/products/diagnostics-and-clinical-research/sexual-reproductive-health/cervical-cancer-screening/carehpv-test-kit/#orderinginformation [accessed 2020-01-28]

17. Hologic. USA: Hologic Aptima+ThinPrep Cervical Health URL: https://www.hologic.com/hologic-products/diagnostic-solutions/aptimathinprep-cervical-health [accessed 2020-01-28]
18. Katanga J, Kjaer SK, Manongi R, Wu CS, Iftner T, Waldstrom M, et al. Performance of careHPV, hybrid capture 2 and visual inspection with acetic acid for detection of high-grade cervical lesion in Tanzania: A cross-sectional study. PLoS One 2019;14(6):e0218559 [FREE Full text] [doi: 10.1371/journal.pone.0218559] [Medline: 31216333]
19. ORCI Dashboard. URL: http://www.connected2care.org/ [accessed 2020-01-28]
20. REDCap. USA: Vanderbilt University; National Institutes of Health URL: https://www.project-redcap.org/ [accessed 2020-01-28]
21. Rovers Medical Devices. Netherlands: Rovers Evalyn Brush URL: https://www.roversmedicaldevices.com/cell-sampling-devices/evalyn-brush/ [accessed 2020-01-28]
22. Linde DS, Rasch V, Mwaiselage JD, Gammeltoft TM. Competing needs: a qualitative study of cervical cancer screening attendance among HPV-positive women in Tanzania. BMJ Open 2019 Feb 27;9(2):e024011 [FREE Full text] [doi: 10.1136/bmjopen-2018-024011] [Medline: 30819704]
23. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall; 1990.
24. Lester RT, 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]
25. Lund S, Nielsen BB, Hemed M, Boas IM, Said A, Said K, et al. Mobile phones improve antenatal care attendance in Zanzibar: a cluster randomized controlled trial. BMC Pregnancy Childbirth 2014 Jan 17;14:29 [FREE Full text] [doi: 10.1186/1471-2393-14-29] [Medline: 24438517]
26. Stata: Software for Statistics and Data Science. STATA Software USA: StataCorp URL: https://www.stata.com/ [accessed 2020-01-28]
27. Kerr NL. HARKing: hypothesizing after the results are known. Pers Soc Psychol Rev 1998;2(3):196-217. [doi: 10.1207/s15327957pspr0203_4] [Medline: 15647155]
28. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Hoboken, New Jersey: The Cochrane Collaboration; 2011.
29. Beall J. Beall's List of Predatory Journals and Publishers 2018. 2020 Jan 28. URL: https://beallslist.weebly.com/ [accessed 2019-07-01]
30. Untiet S, Vassilakos P, McCarey C, Tebeu P, Kengne-Fosso G, Menoud P, et al. HPV self-sampling as primary screening test in sub-Saharan Africa: implication for a triaging strategy. Int J Cancer 2014 Oct 15;135(8):1911-1917 [FREE Full text] [doi: 10.1002/ijc.28834] [Medline: 24615324]
31. Modibbo F, Iregbu KC, Okuma J, Leeman A, Kasius A, de Koning M, et al. Randomized trial evaluating self-sampling for HPV DNA based tests for cervical cancer screening in Nigeria. Infect Agent Cancer 2017;12:11 [FREE Full text] [doi: 10.1186/s13027-017-0123-z] [Medline: 28184239]
32. Mbatha JN, Galappaththi-Arachchige HN, Mtshali A, Taylor M, Ndhlovu PD, Kjetland EF, et al. Self-sampling for human papillomavirus testing among rural young women of KwaZulu-Natal, South Africa. BMC Res Notes 2017 Dec 6;10(1):702 [FREE Full text] [doi: 10.1186/s13104-017-3045-3] [Medline: 29208043]
33. Obiri-Yeboah D, Adu-Sarkodie Y, Djigma F, Hayfron-Benjamin A, Abdul L, Simpore J, et al. Self-collected vaginal sampling for the detection of genital human papillomavirus (HPV) using careHPV among Ghanaian women. BMC Womens Health 2017 Sep 26;17(1):86 [FREE Full text] [doi: 10.1186/s12905-017-0448-1] [Medline: 28950841]

Abbreviations

CONCEPT: Comprehensive Cervical Cancer Prevention in Tanzania
eVoucher: electronic voucher
HPV: human papillomavirus
KCMC: Kilimanjaro Christian Medical Centre
mHealth: mobile health
OR: odds ratio
ORCI: Ocean Road Cancer Institute
RR: risk ratio
VIA: visual inspection with acetic acid

Edited by G Eysenbach; submitted 15.08.19; peer-reviewed by Z Ma, K Bruno; comments to author 26.11.19; revised version received 16.12.19; accepted 19.12.19; published 02.04.20.

Please cite as:
Linde DS, Andersen MS, Mwaiselage J, Manongi R, Kjaer SK, Rasch V
Effectiveness of One-Way Text Messaging on Attendance to Follow-Up Cervical Cancer Screening Among Human Papillomavirus–Positive Tanzanian Women (Connected2Care): Parallel-Group Randomized Controlled Trial
J Med Internet Res 2020;22(4):e15863
URL: http://www.jmir.org/2020/4/e15863/
doi:10.2196/15863
PMID:32238335

Original Paper

Clinical Integration of a Smartphone App for Patients With Chronic Pain: Retrospective Analysis of Predictors of Benefits and Patient Engagement Between Clinic Visits

Edgar L Ross[1], MD; Robert N Jamison[2*], PhD; Lance Nicholls[3*], PharmD; Barbara M Perry[3*], PharmD; Kim D Nolen[3*], PharmD

[1]Brigham and Women's Hospital, Department of Anesthesiology, Pain Management Center, Chestnut Hill, MA, United States

[2]Brigham and Women's Hospital, Harvard Medical School, Pain Management Center, Chestnut Hill, MA, United States

[3]Pfizer, Inc, New York City, NY, United States

[*]these authors contributed equally

Corresponding Author:
Robert N Jamison, PhD
Brigham and Women's Hospital
Harvard Medical School
Pain Management Center
850 Boyston Street, Suite 320
Chestnut Hill, MA
United States
Phone: 1 3392211441
Email: rjamison@bwh.harvard.edu

Abstract

Background: Although many pain-related smartphone apps exist, little attention has been given to understanding how these apps are used over time and what factors contribute to greater compliance and patient engagement.

Objective: This retrospective analysis was designed to help identify factors that predicted the benefits and future use of a smartphone pain app among patients with chronic pain.

Methods: An app designed for both Android and iOS devices was developed by Brigham and Women's Hospital Pain Management Center (BWH-PMC) for users with chronic pain to assess and monitor pain and communicate with their providers. The pain app offered chronic pain assessment, push notification reminders and communication, personalized goal setting, relaxation sound files, topics of interest with psychological and medical pain management strategies, and line graphs from daily assessments. BWH-PMC recruited 253 patients with chronic pain over time to use the pain app. All subjects completed baseline measures and were asked to record their progress every day using push notification daily assessments. After 3 months, participants completed follow-up questionnaires and answered satisfaction questions. We defined the number of completed daily assessments as a measure of patient engagement with the pain app.

Results: The average age of participants was 51.5 years (SD 13.7, range 18-92), 72.8% (182/253) were female, and 36.8% (78/212) reported the low back as their primary pain site. The number of daily assessments ranged from 1 to 426 (average 62.0, SD 49.9). The app was easy to introduce among patients, and it was well accepted. Those who completed more daily assessments (greater patient engagement) throughout the study were more likely to report higher pain intensity, more activity interference, and greater disability and were generally overweight compared with others. Patients with higher engagement with the app rated the app as offering greater benefit in coping with their pain and expressed more willingness to use the app in the future (P<.05) compared with patients showing lower engagement. Patients completing a small number of daily assessments reported less pain intensity, less daily activity interference, and less pain-related disability on average and were less likely to use the two-way messaging than those who were more engaged with the pain app (P<.05).

Conclusions: Patients with chronic pain who appeared to manage their pain better were less likely to report benefits of a smartphone pain app designed for chronic pain management. They demonstrated lower patient engagement in reporting their daily progress, in part, owing to the perceived burden of regularly using an app without a perceived benefit. An intrinsically different pain app designed and targeted for individuals based on early identification of user characteristics and adapted for each individual would likely improve compliance and app-related patient engagement.

(J Med Internet Res 2020;22(4):e16939) doi:10.2196/16939

KEYWORDS

chronic pain; patient engagement; telemedicine; mHealth; pain measurement

Introduction

Background

Pain is a major reason that individuals seek health care treatment, and it is estimated that more than 25 million US adults are affected by daily pain [1]. Chronic pain is known to impose a tremendous burden on the quality of life of the affected individuals [2]. According to the Global Burden of Disease Study of the Institute for Health Metrics and Evaluation and the World Health Organization, chronic pain has consistently been ranked first in associated disability and overall burden between 1990 and 2017 [3]. It has been determined that chronic pain adversely affects individuals at a higher frequency than depression, substance abuse, and Alzheimer disease [4,5]. An influential report by the Institute of Medicine on *Relieving Pain in America* highlighted the urgent need for the development of better methods for tracking and treating pain because of the ever-increasing costs associated with this condition [6].

Innovative technology can be used by health care providers to track persons with chronic pain, engage the patients between clinic visits, and offer information and support to improve coping. There has been a rapid increase in smartphone apps used to monitor and record health data partly due to the increase in mobile device availability [7]. According to the Pew Research Center, about three-quarters of US adults (77%) stated that they owned a smartphone, and 46% of these owners said that their smartphone is something “they could not live without” [8,9]. Individuals living in both urban and rural communities are capable more than ever of monitoring their progress and sending information directly to their health care providers using sophisticated apps [10].

There is evidence that tracking real-time data using momentary ecological assessment is preferable to retrospective diary entry [11-14]. Apps using innovative time-stamped technology can be particularly helpful in tracking variations in pain intensity and other health-related symptoms between clinic visits [7,15,16]. Large datasets of daily pain assessment offer opportunities for the employment of computer-based classification and artificial intelligence [17]. Various available smartphone apps target people with both non–cancer- and cancer-related pain [18-21]. Although many of these apps are commercially accessible, most of them (approximately 86%) have been found to lack professional medical involvement in their development [22]. Lallo et al [23] reviewed 224 pain apps and found little evidence that health care professionals had been involved in creating the apps. The authors also found that only 2% of the apps they reviewed incorporated interactive social support and goal setting. None of the apps that were reviewed contained the recommended five main categories of functionality: the ability to self-monitor, set goals, build skills, educate, and provide social support.

In a more recent review, Bhatterai et al [22] examined 373 pain self-management apps; only 4 successfully met their inclusion criteria according to an established usability evaluation tool. In another recent review of 195 pain management apps, Portelli and Eldred [24] found only 6 apps that incorporated a specific psychological component. The authors concluded that existing pain apps were often constructed by software developers with little input from health care professionals and patients with pain. They also reported that the pain apps tend to contain minimal theoretical content for facilitating self-management or behavioral change. Unfortunately, the life expectancy of most smartphone apps is brief. Three-fourths (75%) of users discontinue using an app within 48 hours of downloading it, and 25% of apps are discarded after the first opening [25]. On the basis of anonymized data points from more than 125 million mobile phones, it is estimated that 80% of apps fade away in time frames as short as 72 hours, and 21% of users use an app only once [26,27].

Objective

The purpose of this analysis was to determine the long-term effects of using a smartphone pain app that offers pain management strategies and allows patients with chronic pain to assess, monitor, and communicate their condition to their health care providers. We were particularly interested in learning from quantitative and qualitative feedback from users about factors that might contribute to improved patient engagement and what might affect adherence to using a smartphone pain app between clinic visits. We were interested in identifying the type of user who would commit to continuing to use a pain app in the future. Finally, we examined qualitative feedback from the users to help identify ways to improve a smartphone pain app.

Methods

Design

This is a retrospective analysis of data gathered from a smartphone pain app designed by Brigham and Women’s Hospital Pain Management Center (BWH-PMC) to assess longitudinal combined information about satisfaction and compliance with the use of a smartphone pain app for persons with chronic pain over 3 months. The analysis plan was approved by the hospital’s internal review board. A team from the BWH-PMC helped develop and test multiple versions of a smartphone pain app used on iOS and Android devices. Initial input from 20 patients with chronic pain was obtained to assist in the development process of the first version of the app (PainApp Pilot; Figure 1). The pain app was tested for security, and all data were saved on a secure encrypted password-protected server. All versions of the pain app were tested and uploaded to the Apple iTunes Store and Google Play Store. A validated version of the app used for this analysis (BWH PainApp) could be downloaded for free and could be

used to monitor progress and provide feedback to the user through two-way messaging (Figure 2).

Data on the server were available only to BWH-PMC personnel through a secure password-protected administration portal. Components of the smartphone app included demographic and contact information, a comprehensive chronic pain assessment, 5-item daily assessments with push notification reminders (Figure 3), personalized goal setting, relaxation sound files, topics of interest with psychological and medical pain management strategies, and line graphs from the daily assessments that could be saved and placed on the patients' electronic medical record (EMR).

Data were collected by BWH-PMC from a series of studies using the third version of the smartphone pain app (BWH PainApp) between February 2015 and May 2018 among patients with noncancer-related chronic pain. Previously conducted study methods have been reported earlier by BWH-PMC [28-31]. Subjects were recruited either by invitation to participate in a randomized trial by their treating physician, or they responded to a flyer left in clinic waiting rooms. The trials were designed to investigate the efficacy of a pain app [28,29] and the efficacy of devices to help manage pain (eg, vibrating gloves and a transcutaneous electric nerve stimulation device) [30,31]. Interested and eligible participants signed a consent form and completed baseline measures through these previously conducted studies. Subjects were asked to complete a packet of questionnaires at baseline and again after a 3-month follow-up. The pre-post questionnaires were completed on paper, and data were transferred to an electronic database. Data were also captured from baseline and daily assessments from the pain app. All patients were asked to inform the investigators if any unforeseen medical changes occurred. Subjects were informed about how to find the app (either on the iTunes Store or Google Play Store) and were assisted in downloading the pain app if assistance was needed. Support was offered to address any technical problems that the subject might have encountered. Most of the on-boarding process was done live with a research assistant (RA), and additional assistance was offered with two-way messaging on the administrator portal. Subjects were also able to contact the RA if and when they encountered any technical problems.

All participants were encouraged to complete a 5-item daily assessment on the pain app about their pain, sleep, mood, activity interference, and whether they had gotten better or worse on a visual analog scale (Figure 3). Participants were instructed to complete the assessments around the same time each day for 3 months (although some participants continued to use the app beyond 3 months). Line graphs of the data were made available on the server for subjects. These reports could be copied and pasted onto the subject's hospital EMR. The app would sometimes have a lag in transmitting data to the server, and as such, subjects did not always get to see their summarized data on these line graphs. Participants involved in a specific study received US $25 after completing the baseline packet of questionnaires, and US $50 after completing the 3-month assessments. Compensation was based on the completion of the follow-up questionnaires regardless of app usage.

Figure 1. Key development highlights from each version of the pain app. BWH: Brigham and Women's Hospital; PMC: Pain Management Center.

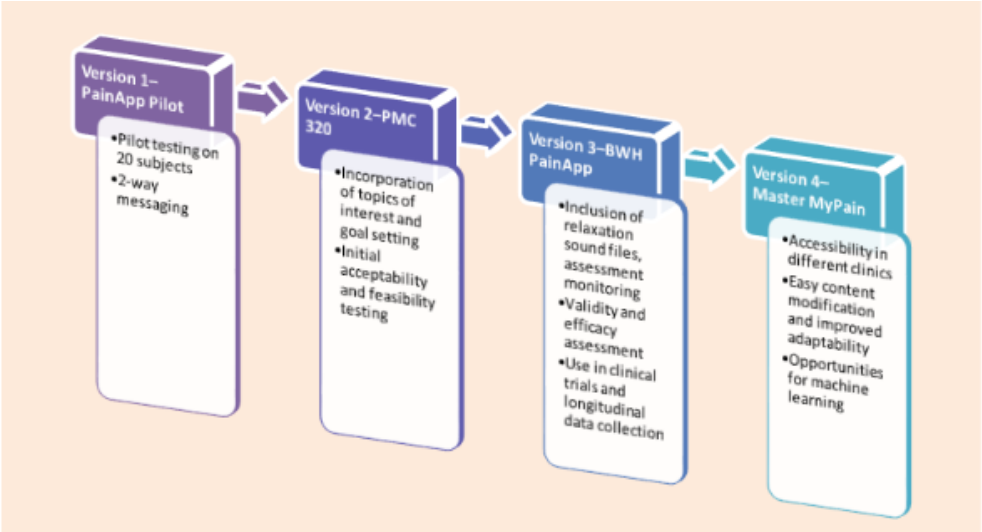


Figure 2. Pain app version 3 home page with links when scrolled down.

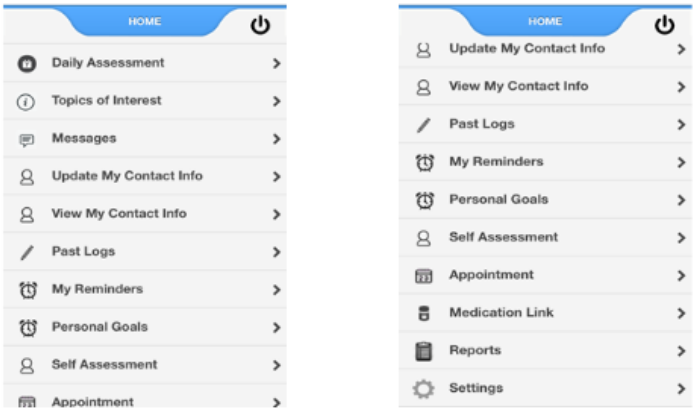
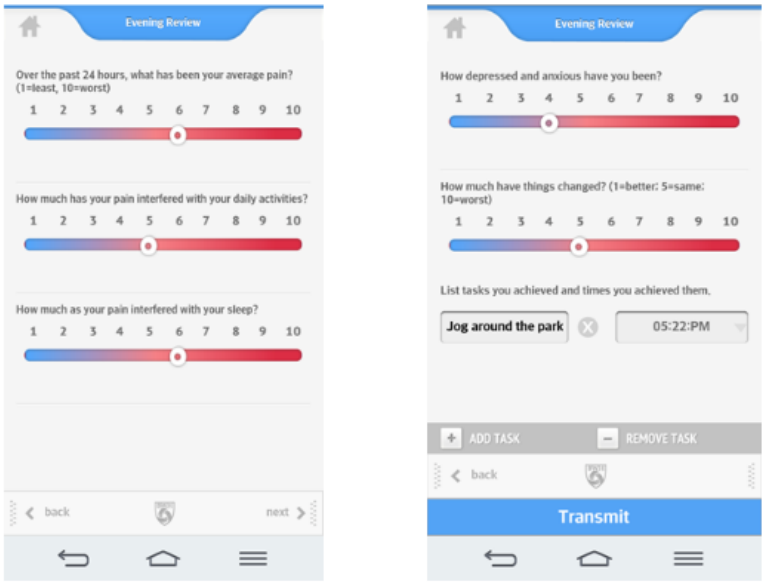


Figure 3. Pain app version 3 daily assessments and goal-setting tasks.



Participants

BWH-PMC recruited patients with chronic pain to participate in 1 of 4 published studies [28-31]. Participants needed to be 18 years or older and own a compatible smartphone (iOS or Android device). Other inclusion criteria included (1) having chronic pain for more than 6 months, (2) averaging 4 or greater on a pain intensity scale of 0 to 10, and (3) able to speak and understand English. Patients were excluded if they had (1) any cognitive impairment that would prevent them from understanding the consent, study measures, or procedures, (2) any clinically unstable medical condition judged to interfere with study participation, (3) a pain condition requiring urgent surgery, (4) a present psychiatric condition (eg, Diagnostic and Statistical Manual of Mental Disorders diagnosis of schizophrenia, delusional disorder, psychotic disorder or dissociative disorder) that was judged to interfere with the study, (5) visual or motor impairment that would interfere with the use of a smartphone, and (6) an active addiction disorder over the past 6 months (positive on the Mini International Neuropsychiatric Interview, v.5.0) [32] that would interfere with study participation.

Measures

Acceptability, tolerability, feasibility, and effectiveness of the third version of the pain app were assessed by examining the number and frequency of daily assessments, the number of subjects who continued to use the app after the initial download, and the numeric and qualitative satisfaction ratings. Any reported safety issues were also documented. Overall outcome efficacy was determined through standardized paper-based measures administered at baseline and again after 3 months from this baseline assessment [28-31].

Pain intensity and pain description were assessed using the Brief Pain Inventory (BPI) [33]. This self-report questionnaire, formerly the Wisconsin Brief Pain Questionnaire [34], has shown sufficient reliability and validity. Scale (rated from 0=no pain to 10=worst pain possible) indicates the intensity of pain at its worst, least, average, and pain now. A figure representing the body was used for the patient to shade the area corresponding to his or her pain. Test-retest reliability for the BPI ratings of pain revealed high correlations of 0.93 for worst pain, 0.78 for usual pain, and 0.59 for pain now [33].

Activity interference and disability was assessed with items from the BPI and the Pain Disability Inventory (PDI) [35]. The PDI is a 7-item questionnaire rated from 0 to 10 on the level of disability of 7 areas of activity interference including family or home responsibilities, recreation, social activity, occupation, sexual behavior, self-care, and life-supporting behaviors. It has shown to have excellent test-retest reliability and validity and is sensitive to high levels of disability [35].

Mood, negative affect, and emotional distress were assessed using the Hospital Anxiety and Depression Scale (HADS) [36,37] and certain initial baseline questions on the pain app. The HADS is a 14-item scale that helps determine the presence and severity of anxiety and depression. Each item is coded from 0 to 3 (eg, not at all, most of the time) with 7 items assessing anxiety and 7 items measuring depression. The HADS has adequate reliability (Cronbach alpha=.83) and validity [37]. We also examined the construct of catastrophizing using the Pain Catastrophizing Scale (PCS) [38,39]. The PCS is a 13-item self-report measure that examines pain rumination, magnification of symptoms, and general helplessness. The responses range from not at all (0) to all the time (4) on a 5-point scale. The PCS is found to predict levels of pain and distress among clinical patients and has good psychometric properties with excellent reliability (Cronbach alpha=.87-.95) and validity [17,40]. Total scores of 30 or greater represent a clinically relevant level of catastrophizing (75th percentile).

After 3 months, participants were asked to respond to a 5-item paper-based satisfaction questionnaire designed to investigate the perceived benefit of how easy the pain app was to use and navigate, how useful the daily reminders were, how much the program helped them cope with their pain, and how willing they would be to use the pain app in the future. All items, which were developed in a previous study [29] and adapted from a previously validated measure [41], were rated on a 0 to 10 scale (0=worse/not at all helpful to 10=best/very helpful).

Statistical Analysis

This retrospective analysis was conducted by BWH-PMC. Univariate and multivariate descriptive analyses were performed on all the dependent variables at baseline and at follow-up. Chi-square, *t* tests, interitem correlations, exploratory factor analyses, and canonical discriminant function analyses were conducted as appropriate using SPSS (version 25.0, IBM Corporation) [42,43]. Subjective comments about the use of the pain app were also collected and summarized.

Results

Patient Demographic Characteristics

A total of 253 patients with chronic pain were engaged by BWH-PMC to use a revised third version of the smartphone pain app. The average age of patients was 51.4 years (SD 13.7, range 18-92); 73.1% (185/253) of patients were female and 82.9% (209/252) of patients were white (Table 1). Pain duration averaged 11.8 years (SD 10.7), and 36.8% (78/212) of patients reported having primary low back pain. Most of the patients (n=243) were taking prescription medication at the time of the study, and 39.9% (97/243) were prescribed opioids for pain. Most subjects were overweight with a BMI averaging 30.1 kg/m^2 (SD 7.4). Most of the subjects (171/253, 67.6%) had iPhones, whereas 32.4% (82/253) of the subjects had Android smartphones.

Table 1. Patient demographic characteristics (N=253).

Variable	Value	Range
Age (years), mean (SD)	51.5 (13.8)	18-92
Gender, female, n (%)	182 (72.7)	N/A[a]
Ethnicity, n (%)		
White	206 (82.7)	N/A
African American	16 (6.4)	N/A
Hispanic	17 (6.8)	N/A
Other	10 (4.0)	N/A
Pain duration (years), mean (SD)	11.8 (10.7)	0.5-50
Pain site, n (%)[b]		
Low back	78 (36.8)	N/A
Multiple sites	77 (36.3)	N/A
Cervical/upper extremity	31 (14.6)	N/A
Lower extremity	11 (5.2)	N/A
Abdominal/pelvic	13 (6.1)	N/A
Head/face	2 (0.9)	N/A
Pain intensity[c], mean (SD)		
Worst pain	7.7 (2.1)	1-10
Least pain	3.3 (2.3)	0-10
Average pain	5.4 (1.8)	1-10
Depth of pain, mean (SD)	203.3 (48.0)	0-270
Take opioid medication, n (%)[d]	97 (39.9)	N/A
BMI (kg/m^2), mean (SD)	30.1 (7.4)	12.2-54.7
Number of times wake during night, mean (SD)	2.6 (2.1)	0-10
Sleep hours, mean (SD)	6.3 (1.8)	1-12
Pain interference (total)[e], mean (SD)	4.9 (2.7)	0-10
Pain Disability Index, mean (SD)	31.5 (17.7)	0-70
Hospital Anxiety and Depression Scale total, mean (SD)	14.9 (7.7)	0-36
Pain Catastrophizing Scale, mean (SD)	17.3 (12.2))	0-50
Number of symptoms[f] (present or absent), mean (SD)	1.6 (2.4)	0-13
Number of pain descriptors[g] (present or absent), mean (SD)	4.1 (1.9)	1-9
Number of daily assessments, mean (SD)	62.0 (49.9)	1-426

[a]N/A: not applicable.

[b]N=212.

[c]0=no pain; 10=pain as bad as you can imagine.

[d]N=243.

[e]During the past 24 hours, how much has your pain interfered with (1) general activity, (2) mood, (3) walking ability, (4) normal work, (5) relations with others, (6) sleep, and (7) enjoyment of life? 0=has not interfered; 10=completely interfered.

[f]Side effect symptoms: constipation, dizziness, dry mouth, headache, itching, memory lapse, confusion, nausea, nightmares, sneezing, sweating, visual problems, weakness, and other.

[g]Pain descriptors: throbbing, stabbing, aching, burning, pricking, pulling, shooting, numbing, and other.

Patient Engagement Results

Of the 253 subjects considered for the analysis, 43 (18.1%) reported some type of technical problem with the app during the study period that briefly restricted their daily assessments. This did not significantly affect their engagement with the pain app. The total number of daily assessments from the pain app averaged 62.0 (SD 49.9). Comparisons between baseline measures and repeat measures at 3 months showed an overall decrease in average pain intensity on the BPI (5.3, SD 1.8 vs 4.9, SD 2.3; t_{185}=4.0; *P*<.001) and a decrease in disability on the PDI (30.6, SD 17.7 vs 27.2, SD 18.2; t_{156}=3.9; *P*<.001), but no differences in mood (mean HADS score 14.4, SD 8.1 vs 14.6, SD 8.3) and pain catastrophizing (mean PCS score 15.7, SD 1.3 vs 16.3, SD 12.8).

A total of 72.3% (183/253) users completed the satisfaction questionnaire after approximately 3 months. No significant differences in demographic characteristics were found between those who completed the satisfaction questionnaire and those who did not complete this questionnaire. Most users found the app easy to use (mean 8.7, SD 2.2) and easy to navigate (mean 8.5, SD 2.4; 0=not at all easy; 10=very easy). The majority of users also found the daily reminders to be useful (mean 6.7, SD 3.9; 0=not at all useful, 10=very useful). Some of the users, primarily the Android users, reported that the push notification reminders did not consistently work on their phone, and they were more likely to rate lower perceived usefulness of the daily reminders because they did not work. The users felt that the app offered some help in coping with their pain (mean 4.5, SD 3.7; 0=not at all helpful, 10=very helpful), whereas the majority of the users felt that they would be willing to use the app in the future (mean 7.1, SD 3.3; 0=not at all willing, 10=very willing).

No significant differences were found on demographic variables of age, gender, ethnicity, or pain duration on all outcome variables. Those who reported liking the pain app were more likely to use it often to submit more daily reports and reported greater pain intensity and more disability. Pearson product-moment correlations between the 5 satisfaction questions ranged between 0.21 and 0.58.

Factor analysis of the satisfaction questionnaire responses using principal component analysis with Varimax rotation found two factors above an eigenvalue of greater than 1.0: (1) easy to use, easy to navigate, useful reminders (correlation *r*=0.48; eigenvalue=2.47) and (2) helped to cope with pain and would use the app in the future (*r*=0.51; eigenvalue=1.01). The first factor, containing 3 of the satisfaction questions, was labeled *easy to use*. The second factor, containing the other two satisfaction questions, was labeled *help with coping and future use.* Those who felt that the app helped them in coping also were more likely to report that they would use the pain app in the future. Combined, these 2 factors accounted for 70.0% of the variance.

Pearson product-moment correlations were run between the combined satisfaction ratings of *easy to use* (satisfaction questions 1, 2 and 3) and *help with coping and future use* (questions 4 and 5) and the number of daily assessments, BPI activity interference, and PDI scores at 3 months (Table 2).

Discriminant function analyses were run using those variables, which revealed significant differences between those with generally higher ratings on *help with coping and future use*, compared with those with lower ratings on these items. Three items were identified using stepwise entry: (1) total number of messages sent and received, (2) total PDI baseline scores, and (3) BMI score (Wilks Lambda=0.89; *P*<.001) and correctly classifying 69.4% of the cases entered. This means that, overall, these 3 variables would correctly classify someone approximately 70% of the time as to the app being helpful to cope with pain and used in the future.

Differences were examined on the baseline and outcome variables between those selected patients with pain who felt that the pain app both helped them cope with their pain and were willing to use the app in the future (n=84; >7/10) and those who reported that the pain app both did not help them cope and were less inclined to use it in the future (n=81; <7/10) based on the 3-month satisfaction questions (Table 3). No differences were found between groups in age, gender, pain site, ethnicity, or pain duration. Those who rated the pain app more favorably reported higher pain ratings (baseline and follow-up), more activity interference, more pain-related disability, used more words to describe their pain, reported more side effects, had higher BMI scores, and were more often taking opioids for their pain than those who were less favorable about the pain app (*P*<.05). Those who felt that the app helped them cope with their pain and would use the app in the future completed more daily assessments and used the two-way messaging service on the app more than those who felt that the app was not as helpful (*P*<.05).

Most of those who responded to the follow-up question *Is there anything about the pain app that you would change?* had no comments (eg, no, none, not really, not sure), and there were a number of positive comments (eg, I found it easy to use, Thanks for the opportunity to use the app). Examples of negative feedback and specific comments about difficulties encountered with the push notification and recommendations for improvements are included in Multimedia Appendix 1. Some users encountered a number of difficulties with the functionality of the app (slow, not accurate, problems with deleting password) and expressed difficulties in seeing past logs and concerns about being constantly reminded about their pain. Some pointed out the challenges they experienced when updating their phones or changing their phone carriers. Many expressed problems they encountered with daily push notification reminders. This was found to be particularly prevalent among Android device owners. Other requests included making the pain app more adaptive to each user's specific condition, adding more instruction when starting to use the app, giving a clear indication when a daily assessment was completed, being able to continue to listen to the relaxation sound files even when the app is closed, and incorporating clearly designated areas to type in free text that can be sent to providers.

Table 2. Pearson product-moment correlations among patient satisfaction questionnaire responses between those who found the pain app easy to use, and those who felt that the app helped them to cope and would be willing to use the pain app in the future (0=very satisfied; 10=very unsatisfied).

Variable	Pearson product-moment correlations for *easy to use*	Pearson product-moment correlations for *helped cope and future use*
Age (years)	0.20[a]	−0.08
Brief Pain Inventory (0-10)		
Worse pain	−0.06	0.18[a]
Least pain	0.15	0.22[b]
Average pain	0.10	0.24[b]
BMI (kg/m^2)	0.11	0.26[b]
Brief Pain Inventory activity interference (0-10)	−0.03	0.24[b]
Pain Disability Inventory total (0-70)	0.00	0.19[a]
Side effect list total (0-14)[a,c]	−0.18[a]	0.19[a]
Pain description total (0-9)[b,d]	−0.23[a]	0.15
Number of daily assessments entered	0.09	0.15
Total number of messages sent and received	−0.01	0.18[a]
Opioids (yes/no)	0.11	0.19[a]

[a]*P*<.05.

[b]*P*<.01.

[c]Side effect symptoms: constipation, dizziness, dry mouth, headache, itching, memory lapse, confusion, nausea, nightmares, sneezing, sweating, visual problems, weakness, and other.

[d]Pain descriptors: throbbing, stabbing, aching, burning, pricking, pulling, shooting, numbing, and other.

Table 3. Differences between patients with pain who felt that the pain app helped them cope with their pain and were willing to use the app in the future (n=84) and those who reported that the pain app did not help them cope and were less inclined to use it in the future (n=81).

Variable[a]	Yes (n=84)	No (n=81)	*t* test (*df*)	Chi-square (*df*)
BPI[b] pain (baseline, range 0-10), mean (SD)	5.7 (1.9)	5.1 (1.8)	2.3 (157)[c]	N/A[d]
BPI pain (3-month follow-up, range 0-10), mean (SD)	5.2 (2.4)	4.4 (2.0)	2.3 (155)[c]	N/A
BPI activity interference (range 0-10), mean (SD)	5.0 (2.4)	3.9 (2.6)	2.7 (157)[c]	N/A
BPI activity interference (3-month follow-up, range 0-10), mean (SD)	4.7 (2.8)	3.8 (2.8)	Not significant	N/A
PDI[e] total (baseline), mean (SD)	34.6 (16.9)	27.2 (17.0)	2.7 (145)[f]	N/A
PDI total (3-month follow-up), mean (SD)	31.8 (18.3)	25.0 (17.0)	2.4 (150)[c]	N/A
Pain description (range 0-9)[g], mean (SD)	4.5 (2.1)	3.9 (1.6)	2.1 (160)[c]	N/A
Side effects total (yes, range 0-14)[h], mean (SD)	7.8 (15.6)	4.1 (8.6)	Not significant	N/A
BMI (kg/m^2), mean (SD)	31.6 (7.8)	28.1 (6.5)	3.2 (160)[f]	N/A
Opioids (% yes of total)	22.6	16.4	N/A	4.3 (1)[c]
Number of daily assessments entered, mean (SD)	83.6 (62.3)	65.9 (37.9)	2.2 (162)[c]	N/A
Total messages, mean (SD)	13.1 (12.3)	8.9 (7.9)	2.6 (162)[c]	N/A

[a]No differences were found between groups on age, gender, pain site, ethnicity, or pain duration.

[b]BPI: Brief Pain Inventory.

[c]$P<.05$.

[d]N/A: not applicable.

[e]PDI: Pain Disability Inventory.

[f]$P<.01$.

[g]Pain descriptors: throbbing, stabbing, aching, burning, pricking, pulling, shooting, numbing, and other.

[h]Side effect symptoms: constipation, dizziness, dry mouth, headache, itching, memory lapse, confusion, nausea, nightmares, sneezing, sweating, visual problems, weakness, and other.

Discussion

Principal Findings

Although many pain-related apps exist, attention has been given recently to understanding how these apps are used over time and what factors contribute to greater compliance and patient engagement [44,45]. This study examined factors that contributed to increased patient engagement in using a smartphone pain app. Overall, the pain app was found to be usable and easily accepted among most of the users and, based on the 3-month follow-up assessments, most of the users reported improvement in pain intensity and activity interference. However, those rating the app as easy to use did not necessarily report that the app improved their ability to cope with the pain or that they would necessarily continue to use the app in the future. Demographic variables such as age, gender, or ethnicity were not found to play a role in predicting overall improvement, compliance or satisfaction with use of the pain app. Throughout the study, those persons with chronic pain who reported higher pain intensity and greater pain-related disability were found to like the app more, use the app more, and express greater willingness to use the app in the future compared with those who were less disabled because of their pain. This suggests that the ability to tailor the app to meet the needs of each user could have an important effect on improving compliance. These analyses also suggest that apps should be selectively assigned to those who may present with certain indicators signaling a greater likelihood of benefiting from a pain app to cope with their pain.

The challenge with mobile health (mHealth) technology is to encourage and motivate participants to continue to use an app to track behavior, maintain contact with their provider, and make improvements in their condition. This is particularly important among individuals with chronic illnesses. The goal of innovative mHealth technology is to offer medical and psychological assistance remotely to reduce health care utilization by reducing clinic and emergency room visits and unnecessary expensive tests. This is a future direction for health care technology, but engaging individuals in ways that increase use of this technology continues to represent a challenge among app developers. It may be no surprise that those patients with pain who used the app more were more satisfied with the pain-related software program. It is interesting to speculate why those with more pain, greater self-reported disability, greater weight, more use of opioids, and more pain descriptors were more satisfied with the smartphone pain app. Quite possibly, those who were busy throughout the day found the app to be more bothersome. Subjective feedback suggests that some preferred not to focus on their pain and found the frequent monitoring to be more of a bother than helpful. Those who

reported more limitations owing to their pain might have been more focused on their pain and welcomed the opportunity to share their experience with their providers. Some may have also wanted to verify their disability and document their limitations for others.

There are many challenges with pain apps going forward. Few physicians recommend pain apps because of lack of time, lack of information about which apps are reliable, concerns of liability, and insufficient evidence that the use of an app will improve outcomes [46]. Even if a physician recommends an app to a patient, there is no guarantee that the patient will download it, use it, and continue to use it. Patients need to have the desire to self-manage, and the role of patient engagement is vitally important. Monitoring data with stand-alone apps that collect data but make it difficult to share with providers will reduce the chance that the apps will be used. In addition, physicians do not have time to wade through raw data, so analytics are needed to help make the data digestible. Providers will also not spend time to open a website to view data; thus, there needs to be an easy way to incorporate pain app data into an EMR system.

There are a number of limitations of this analysis that should be highlighted. As with any new technology, we encountered some software and hardware difficulties that may have adversely affected the use of the app and consequently affected the outcome data. Some subjects did not receive reminders or push notifications to complete their daily assessments, which seemed to be reported mostly by Android smartphone owners. In addition, some encountered difficulties when they upgraded their smartphones, including problems downloading the program to their new device. They also reported minor problems with the app when software updates were made to either the iPhone or Android devices. Corresponding changes were needed in the software code of the pain app every time these changes were made to the iOS and Android platforms. The BWH-PMC staff also needed to make periodic changes to the administrative portal and server, which caused delays in capturing patient data. Thus, factors other than patient noncompliance, including technical difficulties with the software and the devices, may have accounted for the perceived benefit from the pain app. Not all users were able to participate owing to the limitations of their phone capabilities or them not owning a smartphone. Thus, these results may have been affected by selection bias. Patients were encouraged to use the app as part of a study, which may have influenced the use of the app more than what might have been done if patients were not involved in a study. We also could not determine how the availability of RA support was an influencing factor in engagement. It should also be pointed out that the results are correlational in nature, and no causal relationships can be assumed.

Conclusions

This retrospective analysis demonstrates that a smartphone pain app for persons with chronic pain can be perceived to be easy to use, but certain factors, including greater pain and disability, might have an increased influence in motivating individuals to use the app. It also highlights potential challenges in using mHealth technology. Future improvements are needed to make pain apps more adaptive and engaging and directly tailored to the individual user. This would likely have a positive impact on adherence and may lead to increased improvements among persons with chronic pain.

Acknowledgments

A portion of the results of this analysis was presented at the American Medical Informatics Association’s (AMIA) annual symposium, San Francisco, California, November 6, 2018. The AMIA poster and this manuscript were funded by Pfizer to understand the importance of utilizing a pain app to communicate between patients and providers between clinic visits. The authors gratefully recognize Joseph C Cappelleri, Walter J McClain, Elizabeth Scanlan, Margarita Udall, Limeng Wan, and the patients and staff of BWH-PMC for their assistance and participation related to the development of this manuscript. Special thanks are also extended to the staff of Technogrounds Inc for helping to develop the different versions of the app.

This manuscript development was funded by Pfizer Inc. ER and RJ are employees of BWH-PMC, which received financial support from Pfizer for participation in this manuscript. LN, BP, and KN are Medical Affairs employees of Pfizer Inc.

Authors' Contributions

ER and RJ conducted the retrospective analysis without any support or input from Pfizer Inc. RJ developed the initial draft of this manuscript. All other authors and contributors provided edits, suggestions, and revisions to the final manuscript.

Conflicts of Interest

LN, BP, and KN are employees of Pfizer Inc. ER and RJ declare no conflicts of interest.

Multimedia Appendix 1
Comments regarding use of the pain app after a 3 month trial "Is there anything about the pain app that you would change.".
[DOCX File , 17 KB - jmir_v22i4e16939_app1.docx]

References

1. Nahin RL. Estimates of pain prevalence and severity in adults: United States, 2012. J Pain 2015 Aug;16(8):769-780 [FREE Full text] [doi: 10.1016/j.jpain.2015.05.002] [Medline: 26028573]
2. Koes BW, van Tulder MW, Ostelo R, Burton AK, Waddell G. Clinical guidelines for the management of low back pain in primary care: an international comparison. Spine (Phila Pa 1976) 2001 Nov 15;26(22):2504-13; discussion 2513. [doi: 10.1097/00007632-200111150-00022] [Medline: 11707719]
3. Institute for Health Metrics and Evaluation. Seattle: IHME; 2019. Findings from the Global Burden of Disease Study 2017 URL: http://www.healthdata.org/policy-report/findings-global-burden-disease-study-2017 [accessed 2020-02-22]
4. GBD 2013 DALYsHALE Collaborators, Murray CJ, Barber RM, Foreman KJ, Ozgoren AA, Abd-Allah F, et al. . [doi: 10.1016/S0140-6736(15)61340-X] [Medline: 26321261]
5. Murray CJ, Atkinson C, Bhalla K, Birbeck G, Burstein R, Chou D, Foreman, Lopez, Murray, Dahodwala, Jarlais, Fahami, Murray, Jarlais, Foreman, Lopez, Murray, US Burden of Disease Collaborators. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. J Am Med Assoc 2013 Aug 14;310(6):591-608 [FREE Full text] [doi: 10.1001/jama.2013.13805] [Medline: 23842577]
6. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. NCBI - NIH. Washington, DC: National Academies Press (US); 2011. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research URL: https://www.ncbi.nlm.nih.gov/pubmed/22553896 [accessed 2020-02-26]
7. Sundararaman LV, Edwards RR, Ross EL, Jamison RN. Integration of mobile health technology in the treatment of chronic pain: a critical review. Reg Anesth Pain Med 2017;42(4):488-498. [doi: 10.1097/AAP.0000000000000621] [Medline: 28570436]
8. Smith A. Pew Research Center. 2015 Apr 1. US Smartphone Use in 2015 URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2020-02-02]
9. Perrin A. The Pew Research Center, June 28. 2017 Jun 28. 10 Facts About Smartphones as the iPhone Turns 10 URL: http://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones [accessed 2020-02-22]
10. Singh K, Bates D. The Commonwealth Fund. 2016 Feb 18. Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications URL: http://www.commonwealthfund.org/publications/issue-briefs/2016/feb/evaluating-mobile-health-apps [accessed 2018-08-09]
11. Blumenthal S, Somashekar G. The Huffington Post. 2015 Aug 10. Advancing Health With Information Technology in the 21st Century URL: http://www.huffingtonpost.com/susan-blumenthal/advancing-health-with-inf_b_7968190.html [accessed 2019-08-09]
12. Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, Katz NP. Electronic diaries for monitoring chronic pain: 1-year validation study. Pain 2001 Apr;91(3):277-285. [doi: 10.1016/s0304-3959(00)00450-4] [Medline: 11275385]
13. Jamison RN, Gracely RH, Raymond SA, Levine JG, Marino B, Herrmann TJ, et al. Comparative study of electronic vs paper VAS ratings: a randomized, crossover trial using healthy volunteers. Pain 2002 Sep;99(1-2):341-347. [doi: 10.1016/s0304-3959(02)00178-1] [Medline: 12237213]
14. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004 Feb;107(3):213-219. [doi: 10.1016/j.pain.2003.10.005] [Medline: 14736583]
15. Stone AA, Broderick JE. Real-time data collection for pain: appraisal and current status. Pain Med 2007 Oct;8(Suppl 3):S85-S93. [doi: 10.1111/j.1526-4637.2007.00372.x] [Medline: 17877531]
16. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
17. Jamison RN, Xu X, Wan L, Edwards RR, Ross EL. Determining pain catastrophizing from daily pain app assessment data: role of computer-based classification. J Pain 2019 Mar;20(3):278-287. [doi: 10.1016/j.jpain.2018.09.005] [Medline: 30273687]
18. Hundert AS, Huguet A, McGrath PJ, Stinson JN, Wheaton M. Commercially available mobile phone headache diary apps: a systematic review. JMIR Mhealth Uhealth 2014 Aug 19;2(3):e36 [FREE Full text] [doi: 10.2196/mhealth.3452] [Medline: 25138438]
19. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar 8;15(3):e51. [doi: 10.2196/jmir.2350] [Medline: 23475457]
20. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909. [doi: 10.1111/pme.12327] [Medline: 24422990]
21. de la Vega R, Roset R, Castarlenas E, Sánchez-Rodríguez E, Solé E, Miró J. Development and testing of painometer: a smartphone app to assess pain intensity. J Pain 2014 Oct;15(10):1001-1007. [doi: 10.1016/j.jpain.2014.04.009] [Medline: 24854065]
22. Bhattarai P, Newton-John TR, Phillips JL. Quality and usability of arthritic pain self-management apps for older adults: a systematic review. Pain Med 2018 Mar 1;19(3):471-484. [doi: 10.1093/pm/pnx090] [Medline: 28541464]

23. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. 'There's a Pain App for That': review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
24. Portelli P, Eldred C. A quality review of smartphone applications for the management of pain. Br J Pain 2016 Aug;10(3):135-140 [FREE Full text] [doi: 10.1177/2049463716638700] [Medline: 27583140]
25. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
26. Perro J. Localytics. 2018 Mar 23. Mobile Apps: What's A Good Retention Rate? URL: http://info.localytics.com/blog/mobile-apps-whats-a-good-retention-rate?_ga=2.90317082.1607914945.1575073756-556096695.1575073756 [accessed 2020-02-22]
27. Chen A. AndrewChen. New Data Shows Losing 80% of Mobile Users is Normal, and Why the Best Apps Do Better URL: http://andrewchen.co/new-data-shows-why-losing-80-of-your-mobile-users-is-normal-and-that-the-best-apps-do [accessed 2019-08-08]
28. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: Who benefits from a pain app? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
29. Jamison RN, Mei A, Ross EL. Longitudinal trial of a smartphone pain application for chronic pain patients: predictors of compliance and satisfaction. J Telemed Telecare 2018 Feb;24(2):93-100. [doi: 10.1177/1357633X16679049] [Medline: 27831496]
30. Jamison RN, Mei A, Edwards RR, Ross EL. Efficacy of vibrating gloves for chronic hand pain due to osteoarthritis. Pain Med 2018 May 1;19(5):1044-1057. [doi: 10.1093/pm/pnx230] [Medline: 29025096]
31. Jamison RN, Wan L, Edwards RR, Mei A, Ross EL. Outcome of a high-frequency transcutaneous electrical nerve stimulator (hfTENS) device for low back pain: a randomized controlled trial. Pain Pract 2019 Jun;19(5):466-475. [doi: 10.1111/papr.12764] [Medline: 30636101]
32. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34 [FREE Full text] [Medline: 9881538]
33. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994 Mar;23(2):129-138. [Medline: 8080219]
34. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. Pain 1983 Oct;17(2):197-210. [doi: 10.1016/0304-3959(83)90143-4] [Medline: 6646795]
35. Tait RC, Pollard CA, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: psychometric and validity data. Arch Phys Med Rehabil 1987 Jul;68(7):438-441. [Medline: 3606368]
36. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
37. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
38. Sullivan MJ, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
39. Sullivan MJ, Stanish W, Waite H, Sullivan M, Tripp DA. Catastrophizing, pain, and disability in patients with soft-tissue injuries. Pain 1998 Sep;77(3):253-260. [doi: 10.1016/s0304-3959(98)00097-9] [Medline: 9808350]
40. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. J Behav Med 2000 Aug;23(4):351-365. [doi: 10.1023/a:1005548801037] [Medline: 10984864]
41. Chapman SL, Jamison RN, Sanders SH. Treatment Helpfulness Questionnaire: a measure of patient satisfaction with treatment modalities provided in chronic pain management programs. Pain 1996 Dec;68(2-3):349-361. [doi: 10.1016/s0304-3959(96)03217-4] [Medline: 9121824]
42. Rosner B. Fundamentals of Biostatistics. Boston, MA: Cengage Learning; 2016.
43. Stevens JP. Applied Multivariate Statistics for the Social Sciences. Fifth Edition. New York, NY: Taylor and Francis Group; 2009.
44. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
45. Leung K, Lu-McLean D, Kuziemsky C, Booth RG, Rossetti SC, Borycki E, et al. Using patient and family engagement strategies to improve outcomes of health information technology initiatives: scoping review. J Med Internet Res 2019 Oct 8;21(10):e14683 [FREE Full text] [doi: 10.2196/14683] [Medline: 31596241]
46. Gruessner V. mHealthIntelligence. 2015 Jun 4. Mobile Health Apps Lack Physician Trust Due to Few Regulations URL: https://mhealthintelligence.com/news/mobile-health-apps-lack-physician-trust-due-to-few-regulations [accessed 2020-02-26]

Abbreviations

AMIA: American Medical Informatics Association
BPI: Brief Pain Inventory
BWH-PMC: Brigham and Women's Hospital Pain Management Center
EMR: electronic medical record
HADS: Hospital Anxiety and Depression Scale
mHealth: mobile health
PCS: Pain Catastrophizing Scale
PDI: Pain Disability Inventory
RA: research assistant

Edited by G Eysenbach; submitted 06.11.19; peer-reviewed by T Janmohamed, L Whitehead, P Seaton; comments to author 18.12.19; revised version received 15.01.20; accepted 26.01.20; published 16.04.20.

Please cite as:
Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD
Clinical Integration of a Smartphone App for Patients With Chronic Pain: Retrospective Analysis of Predictors of Benefits and Patient Engagement Between Clinic Visits
J Med Internet Res 2020;22(4):e16939
URL: http://www.jmir.org/2020/4/e16939/
doi:10.2196/16939
PMID:32297871

Original Paper

Social Media Strategies for Health Promotion by Nonprofit Organizations: Multiple Case Study Design

Isabelle Vedel[1,2*], MD, PhD; Jui Ramaprasad[3*], BS, PhD; Liette Lapointe[3*], BA, MSc, PhD

[1]Department of Family Medicine, McGill University, Montreal, QC, Canada

[2]Lady Davis Institute, Jewish General Hospital, Montreal, QC, Canada

[3]Desautels Faculty of Management, McGill University, Montreal, QC, Canada

[*]all authors contributed equally

Corresponding Author:
Isabelle Vedel, MD, PhD
Department of Family Medicine
McGill University
5858 Chemin de la Côte-des-Neiges, 3rd Fl
Montreal, QC, H3S 1Z1
Canada
Phone: 1 5143999107
Email: isabelle.vedel@mcgill.ca

Abstract

Background: Nonprofit organizations have always played an important role in health promotion. Social media is widely used in health promotion efforts. However, there is a lack of evidence on how decisions regarding the use of social media are undertaken by nonprofit organizations that want to increase their impact in terms of health promotion.

Objective: The aim of this study was to understand why and how nonprofit health care organizations put forth social media strategies to achieve health promotion goals.

Methods: A multiple case study design, using in-depth interviews and a content analysis of each social media strategy, was employed to analyze the use of social media tools by six North American nonprofit organizations dedicated to cancer prevention and management.

Results: The resulting process model demonstrates how social media strategies are enacted by nonprofit organizations to achieve health promotion goals. They put forth three types of social media strategies relative to their use of existing information and communication technologies (ICT)—replicate, transform, or innovate—each affecting the content, format, and delivery of the message differently. Organizations make sense of the social media innovation in complementarity with existing ICT.

Conclusions: For nonprofit organizations, implementing a social media strategy can help achieve health promotion goals. The process of social media strategy implementation could benefit from understanding the rationale, the opportunities, the challenges, and the potentially complementary role of existing ICT strategies.

(J Med Internet Res 2020;22(4):e15586) doi:10.2196/15586

KEYWORDS

neoplasm; social media; information technology; organizations, nonprofit; cancer

Introduction

Background

Nonprofit organizations have always played an important role in health promotion, such as advertising campaigns using billboards [1], radio [2], or television [3]. However, many health promotion programs run by nonprofit organizations have difficulty achieving success. This can be attributed to challenges with disseminating information to the appropriate target group, often because the target audience is not easily identifiable [4], or individuals ignoring information and not feeling engaged [1].

As a complement to more traditional information and communication technologies (ICT), social media is creating opportunities to address these challenges. Social media "encompasses a wide range of online, word-of-mouth forums" [5] and is characterized by its interactive and digital nature [6]. Nonprofit organizations are increasingly relying on social media

to effectively design health promotion strategies [7-9] and to facilitate the reach of word of mouth [10], although some such organizations are not necessarily leveraging all the power social media can offer [11].

To date, research has mainly examined patients' and professionals' motives, barriers, and facilitators to the use of social media [12-15], as well as its impacts, both positive and negative [16]. On the one hand, social media has positive impacts for patients, such as enabling them to share experiences, seek information and opinions, engage with peers and providers, and belong to a community [14,16-19]. This, in turn, can improve patients' sense of participation, motivation, autonomy, empowerment, perceived self-efficacy engagement in decision making, emotional support, and self-care [14,16-18]. These factors associated with social media can contribute to a positive impact on patient health: if social media enables patients to be more engaged in their health, they will change their behavior more easily [17]. However, there is also the risk of unreliable and incorrect health information provided by the community for the community [20].

Objectives

What is not clear from this literature is how decisions regarding the use of social media are undertaken by nonprofit organizations that want to increase their impact in terms of health promotion. Our study, conducted in the context of cancer, aims at understanding why and how nonprofit organizations develop social media strategies, with the goal of eliciting how such organizations can successfully leverage social media. Looking at the use of social media from the organizational perspective allows us to understand the characteristics of the social media strategies that are utilized by nonprofit organizations and to identify how social media may help organizations attain their goals of health promotion. This understanding is critical in providing guidance on how such organizations can leverage social media and manipulate the factors or change the conditions of their social media use to ultimately increase their impact on health promotion.

Methods

Design

We conducted a multiple case study to examine how six North American nonprofit cancer organizations engage in the use of social media for health promotion.

Theoretical Framework

Our study is based on the organizing vision theoretical lens [21], which leverages the concept of mindfulness. In a learning organization, there is a commitment on learning and communication. The leadership of such organizations associate learning to organizational success and to sustaining a supportive learning culture [22]. Organizational mindfulness is "a combination of ongoing scrutiny of existing expectations, continuous refinement and differentiation of expectations based on newer experiences, willingness and capability to invent new expectations that make sense of unprecedented events" [23]. Hence, although a learning organization is focused on ensuring *organizational memory*, the construct of mindfulness embeds, in addition, a prospective and innovative perspective. The concept of mindfulness has proven to be useful to shed light not only on the organizational adoption of ICT innovations but also to inform how organizations can chart a successful course for ICT implementations, by remaining vigilant vis-à-vis ICT evolution [21,24-27]. To the best of our knowledge, this lens has not been used to examine social media.

Mindful behaviors of organizations mean openness to new information and awareness of multiple perspectives [28]. Mindful organizations are described as those that make appropriate interpretations of their nature and needs and respond adaptively to changes in their environment [29]. Rooted in this perspective, the organizing vision is a lens that helps explain how organizations can implement ICT innovations mindfully [30]. It shows how mindful organizations can become increasingly attentive to their idiosyncrasies and environment, to make the most of their ICT investments [31]. Mindfully innovating with ICT means that the organization "attends to an IT [Information Technology] innovation with reasoning grounded in its own organizational facts and specifics" [30], whereas innovating mindlessly with ICT refers to the instance where "a firm's actions betray an absence of such attention and grounding" [30].

Leveraging on the organizing vision lens, we adopted a theory-building approach, based on a multiple case study design [32,33].

Cases

The six cases in this study were selected based on a maximum variation sampling strategy [34] and focused on organizations using social media for cancer prevention and management (Table 1), a major public health issue in our society [35]. A detailed description of the key characteristics of each case is provided in Multimedia Appendix 1, including the rationale of social media use and the ICT and social media tools used.

Table 1. Case characteristics and social media tools used.

Cases	Characteristics	Social media tools
Case 1: Breast Cancer Action	Country: United States, disease type: breast cancer, year founded: 1990, and number of employees: 8	Facebook, Twitter, YouTube, and blog
Case 2: Breast Cancer Society	Country: Canada, disease type: breast cancer, year founded: 1991, and number of employees: 5	Facebook, Twitter, LinkedIn, and YouTube
Case 3: Breast Cancer Foundation	Country: Canada, disease type: breast cancer, year founded: 1986, and number of employees: 197	Facebook, Twitter, LinkedIn, YouTube, Flickr, and blog
Case 4: Us Too International	Country: United States, disease type: prostate cancer, year founded: 1990, and number of employees: 5	Facebook, Twitter, LinkedIn, YouTube, Wikis, Groupon, and blog
Case 5: Prostate Cancer Foundation	Country: United States, disease type: prostate cancer, year founded: 1993, and number of employees: 30	Facebook, Twitter, LinkedIn, YouTube, and blog
Case 6: Pints for Prostate	Country: United States, disease type: prostate cancer, year founded: 2008, and number of employees: 2	Facebook, Twitter, Flickr, and Vimeo

Data Sources and Data Collection

We triangulated our data sources: semistructured interviews with key informants, analysis of the documentation (eg, documentation describing the organization, reports, and newsletters), and qualitative content analysis of the websites and the social media tools used (eg, Facebook, Twitter, and YouTube). In each organization, we conducted semistructured interviews with the chief executive officer or the person responsible for the social media development and use (ie, the key informants) in winter 2008-2009 [34]. These respondents had a thorough knowledge of the origins, implementation, use, barriers, and enabling factors of ICT and social media usage in their respective organizations. Our interview guide (Multimedia Appendix 2) was validated and refined using four pilot interviews. The interviews lasted 1 hour on average and were recorded and transcribed verbatim in their entirety. In addition, we asked our participants to provide relevant documentation. We also collected data from the social media tools across 1 calendar year (2012), to minimize biases. In the end, for each organization, we created a data dossier that provides a structured summary of the characteristics of the organization, content of the website, and social media tools (Multimedia Appendix 2). The overall data collection process resulted in several hundred pages of transcripts and social media content data dossiers.

Analysis

Analytic induction was deemed to provide the best analytic strategy for this study [34,36-38]. Indeed, analytic induction begins with a deductive phase [34,39], which allows for the use of existing theory, and is followed by an inductive phase that allows for new insights to emerge from the data. Following the data collection process, we proceeded with the first round of coding of the social media data dossier and interview transcripts. Our initial codes were deductively based on the categories derived from our organizing vision theoretical lens to understand how organizations learned to best exploit social media through comprehension, adoption, implementation, and assimilation. Next, we proceeded to a round of open coding and identified new themes (eg, actions, tools, and practices put in place). Afterward, following axial coding, codes with the same content and meaning were grouped in higher-level categories (eg, rationale for using social media tools, complementarity with existing ICT, and challenges). Finally, through selective coding, we linked the resulting categories to the main category (eg, strategies). The analysis of the documentation was used to provide additional information and to corroborate and validate the information gathered via the interviews and the social media data dossier. During the overall process of data coding, as a team, we reviewed and discussed the codification of data until we had reached a consensus; this helped eliminate any potential discrepancy. Examples of codes are provided in Multimedia Appendix 3. N'Vivo 9 (QRS International Pty Ltd) was used to support the coding and analysis of the transcripts.

The analysis followed an iterative process, from reading the data to the data analysis multiple times. This iteration allowed a progressive theory development process with an increasing level of abstraction [40], that is, the creation of a shared understanding that forms a coherent structure, a unified whole. This was repeated until theoretical saturation (ie, the point at which additional analysis repeatedly confirmed the interpretations already made) [41]. Following this iterative analysis process, we developed our process model of social media strategies for health promotion by nonprofit organizations.

Results

Overall Findings

Overall, the analysis allowed us to build upon the four pillars of our organizing vision theoretical lens. First, we saw how organizations need to *comprehend* how social media can—or cannot—apply to their needs and reality in terms of health promotion. Second, mindful ICT *adoption* signifies the ability "to anchor the decision in local particulars, rather than simply follow the lead and public rationales or prior adopters" [31]. Third, in *implementing* social media, organizations have to be sensitive to their reality and idiosyncrasies. Finally, the mindfulness challenge in *assimilation* is to decide how to optimally integrate social media into everyday operations to have a better impact on health promotion. We provide illustrative quotes in Multimedia Appendix 4 and examples from the data dossier in Multimedia Appendix 5.

The cross-case analysis—of the ICT and social media tools, interviews, and documents—revealed no major variation in the

results among cases based on the cancer type they were concerned with, the country the organization is based in, the nature of the social media tools the organization employed, or the organization size. Although some of the larger organizations were able to assign some nonspecialized personnel to their social media activities, these activities mainly consisted of feeding the social media platforms, not developing the social media strategy. The analysis of the data dossiers did not reveal any major differences in why and how nonprofit organizations develop social media strategies.

Comprehension

Organizations tend to have one or several of the five following *rationales* for the adoption of social media in health promotion:

1. Creating awareness: Organizations use social media tools to advertise about the disease and to promote healthy behaviors (eg, screening). Social media can be particularly useful to provide information that can be tailored to a specific audience and to reach people who are not voluntarily seeking the information (see quotes 1-3 in Multimedia Appendix 4).
2. Educating: Social media tools can provide up-to-date information on the disease (eg, risk factors) and can enable end users (patients, families, and significant others) to make better informed decisions (eg, about treatment options—see quotes 4 and 5).
3. Providing a forum to interact and support: Social media tools such as blogs, forums, or tweets allow users to get advice from the organization and to facilitate user interactions among themselves for support (see quotes 6-8).
4. Advocating: Social media tools are also, at times, used to play an activist role in relation to the organizations' missions (see quotes 9 and 10).
5. Raising funds: Social media could be a way to facilitate communications and connections with donors (see quotes 11-13). Organizations may also track and report on social media metrics (eg, number of tweets and retweets), for the purposes of board and donor accountability.

In addition, six important *opportunities* associated with the use of the social media tools were identified:

1. Ease-of-use: Social media tools are perceived to be easy to use and provide the opportunity to easily reach a large number of individuals, as evidenced by the number of fans, followers, posts, and blogs (see quote 14 and Multimedia Appendix 5).
2. Low cost: Social media is seen as a low-cost tool compared with traditional marketing tools. For small organizations with limited budgets, such low-cost tools provide new opportunities to communicate and provide information (see quotes 15 and 16).
3. Interactivity with end users: Social media provides a forum for individuals to connect with each other and to engage in more personalized discussions in a timely manner (see quotes 17 and 18). Data show active participation of users (Multimedia Appendix 5) and better effectiveness. For example, end users can follow links and choose the path of information that they would like to explore deeper (see quotes 19 and 20).
4. Flexibility: Social media tools do not impose a strict structure on how the tools are used, how individuals choose to interact and access information using these tools, and how they are integrated with other media (see quotes 21 and 22). This was further evidenced by the links for YouTube videos that were found on many Facebook pages (Multimedia Appendix 5).
5. Status: The use of social media tools was associated with a desire for status differentiation and perceptions of popularity, trendiness, reputation, efficiency, etc (see quotes 23 and 24).
6. Virability: Social media's increased ease in spreading information compared with more traditional ICT—what we call virability—was evidenced by the ability to repost information on Facebook and Twitter (Multimedia Appendix 5), sometimes through mobile devices (see quotes 25 and 26).

Adoption

To maximize the impact, all six organizations used social media tools in addition to some ICT tools (eg, webpages and electronic newsletters) and even more traditional communication tools (eg, posters, magazine, and television advertisements; see quotes 27, 28, and 29 and Multimedia Appendix 5). They saw social media as a way to add to what they were already doing, to give more strength to their activities, and to augment and expand the capabilities of the ICT tools (see quotes 30-32). Concretely, analysis revealed three specific social media *strategies*:

1. Replicate: Organizations essentially imitate their existing use of ICTs, but through a different channel to reach a different and broader audience (see quotes 33 and 34).
2. Transform: Organizations use social media for the same purpose as it uses ICT tools, but the message is transformed in the way it is formatted and delivered, to better engage end users (see quotes 35 and 36).
3. Innovate: To truly tap in the soul of social media, organizations modify the message or action for a new purpose, seeking different results. Such a strategy entails, for example, reposting a message, taking advantage of the virability of the media, and using blogs for press conferences or virtual billboards for advertising. Altogether such a strategy may ultimately enable the development of a community (see quotes 37 and 38).

Implementation

To better take into account the reality of their usage and context, organizations have had to deal with several *challenges*:

- Lack of control: Managing the openness in communication that is enabled through social media (Multimedia Appendix 5) and appropriately monitor the quality, quantity, and format of conversations individuals were having (see quotes 39 and 40). This difficulty concerns both the user contribution and the information that the organization and partners themselves provided (see quotes 41 and 42).
- Technology-related issues: Although user friendly, technology usage introduces challenges such as forced upon updates and characteristics that create limitations (see quotes 43 and 44).

- Diversity of audience: Reaching a wider audience creates challenges in tailoring the message to different communities (eg, an older population and less educated individuals; see quotes 45 and 46 and Multimedia Appendix 5).
- Availability of resources: Finding the resources to develop and manage social media was considered challenging, given the need to find individuals with the expertise in both the content (cancer) and the social media tool. Moreover, there is a need to maintain a social media presence at a high level of interactivity, which requires an extensive amount of time (see quotes 47-50).
- Difficulty in measuring impacts: It is difficult to define relevant indicators of success and objectively assess whether social media use truly helps meet goals (see quotes 51 and 52).

Assimilation

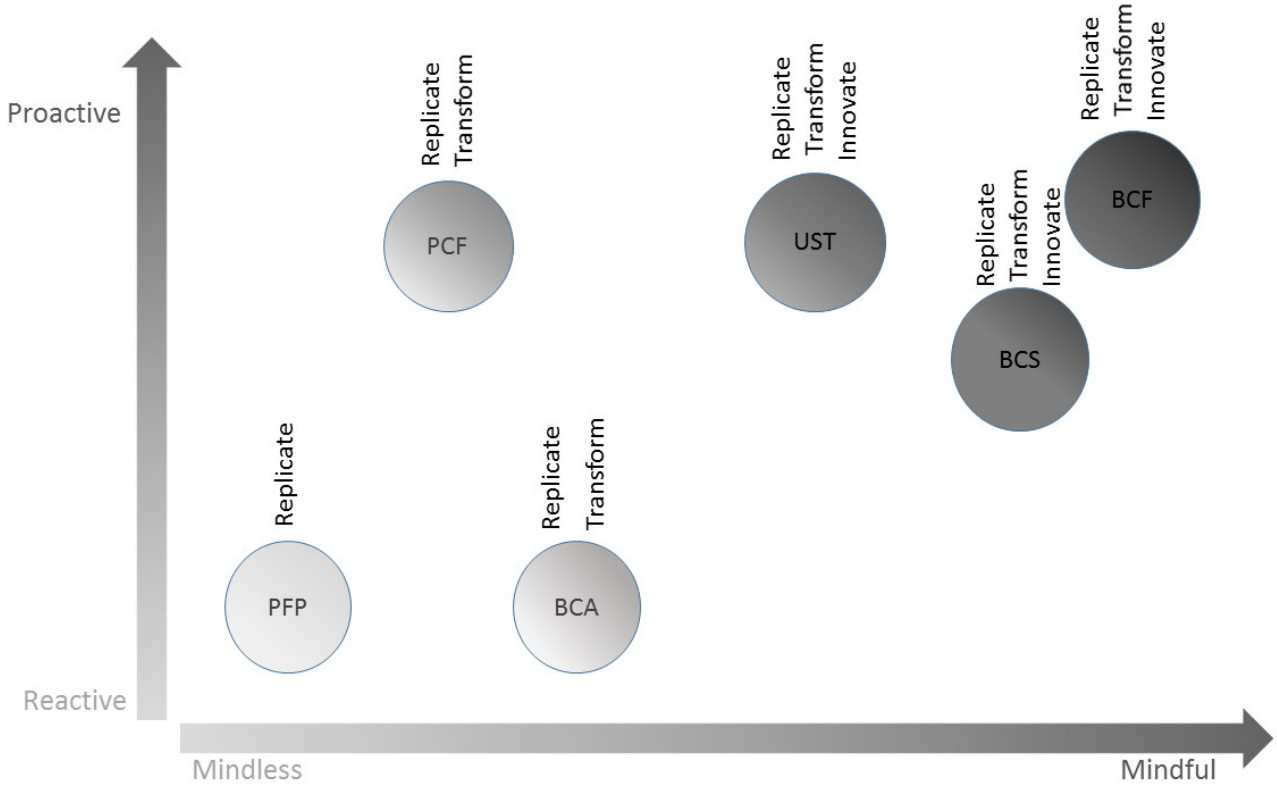
In assimilation, organizations decide how to optimally integrate the new social media tools into everyday operations.

1. Mindless/mindful: At the onset, organizations did not necessarily adopt or use social media in a well thought-out manner, with clear objectives in mind. Actually, the initial use of social media in most of the organizations was primarily mindless. This was particularly noticeable in the case of two organizations where the decision to use social media was not a planned event and where social media strategies were enacted to seize emergent opportunities (see quotes 53 and 54). The level of mindfulness of social media use by the organizations we studied evolved. With time, some organizations were beginning to reflect more about social media (see quotes 55 and 56). Interestingly, in the organization that was most mindful at the onset, social media usage continued to evolve in the same manner, maintaining a mindful stance (see quote 57).
2. Reactive/proactive: Above and beyond the mindful/mindless stance of the process, our results show that the social media strategies were at times enacted in a reactive manner and at other times in a proactive manner. Social media strategies were initially implemented mainly in a reactive manner (ie, in response to users' explicit needs; see quote 58). Only one organization exhibited goal-directed behavior and demonstrated anticipation—a proactive orientation—that is, enabling change before such needs are overtly expressed (see quote 59).

Connecting the Dots

In summary, our data revealed that in addition to considering the level of mindfulness, it was important to consider the proactiveness, or lack thereof, exhibited by the organizations. We linked the strategies put forth by organizations to their overall level of mindfulness and proactive orientation (Figure 1).

Figure 1. Mindfulness and proactive orientation of the six cases. BCA: Breast Cancer Action; BCF: Breast Cancer Foundation; PCF: Prostate Cancer Foundation; PFP: Pints for Prostate; UsT: Us Too International.



We identified three clusters:

1. Cluster 1: The organization exhibits a low level of mindfulness and little proactiveness. The only strategy that was mobilized is this case was replicate. Hence, this organization mostly used social media to carry on the same activities but using social media (see quotes 60-62).
2. Cluster 2: One organization exhibited a fairly low level of mindfulness but a high proactive orientation; another

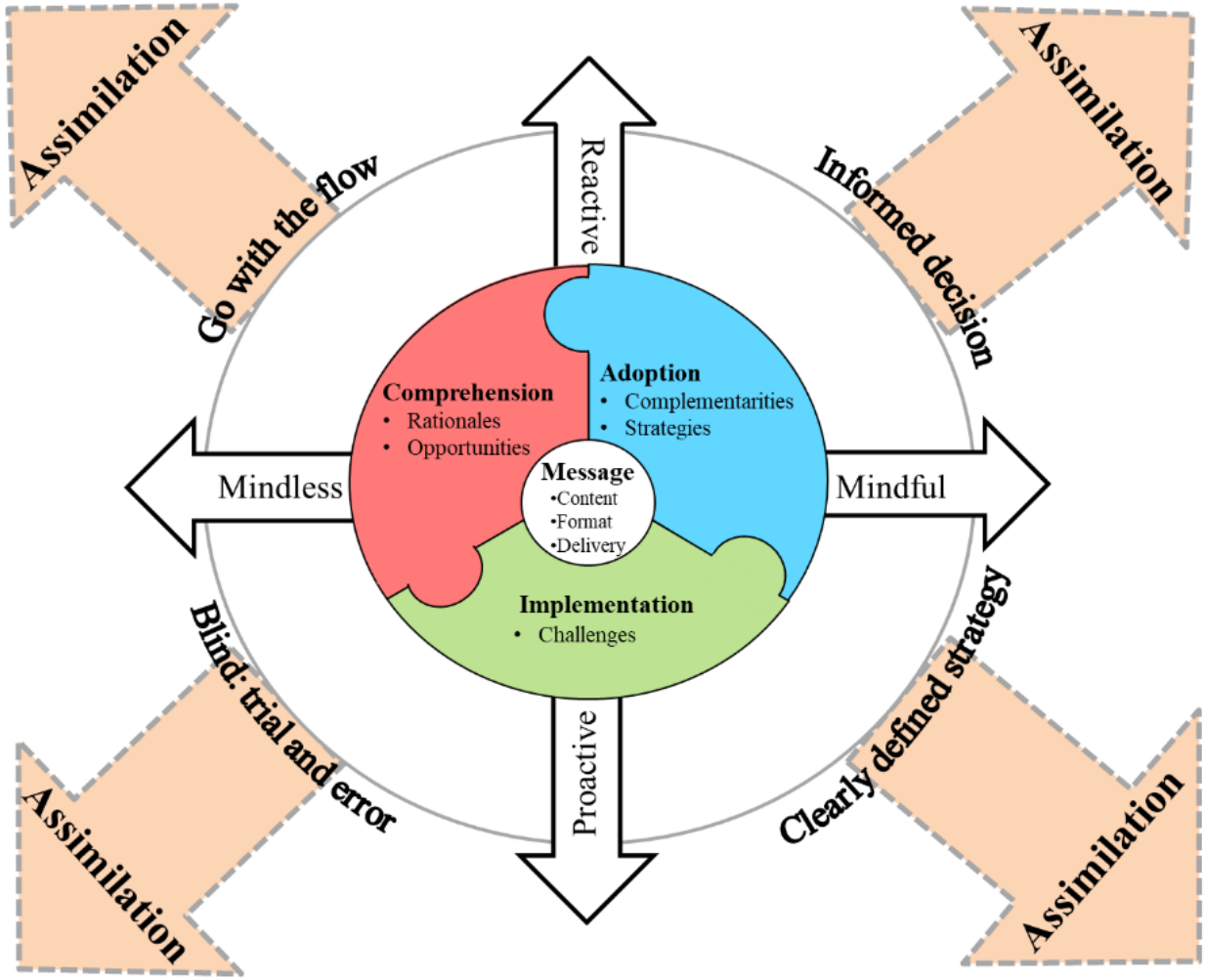
organization exhibited a low proactive orientation but a higher level of mindfulness. In both cases, these organizations leverage social media to transform their message, using the particularities of social media to better engage users (see quotes 63 and 64). Despite the fact that these organizations are not both proactive and mindful, they do appear to derive higher value from their social media strategies in terms of health promotion (see quotes 65 and 66) than organizations exhibiting a low level of mindfulness and little proactiveness (ie, cluster 1).

3. Cluster 3: Organizations exhibit a higher level of mindfulness compared with the other clusters. In all, three organizations did not use social media simply to replicate or to transform their message but most importantly to innovate by leveraging the potential offered by social media (see quote 67 and 68). Not surprisingly, these organizations appear to derive the most value from their involvement in social media (see quotes 69 and 70).

The Process Model of Social Media Strategies for Health Promotion by Nonprofit Organizations

On the basis of our data analysis and the organizing vision theoretical lens, we developed a process model that reveals the elements and patterns of relationships that underlie the enactment of social media strategies by organizations for health promotion (Figure 2).

Figure 2. Process model of social media strategies for health promotion by nonprofit organizations.



It first shows that the four pillars of social media strategy enactment—*comprehension*, *adoption*, *implementation*, and *assimilation*—are not necessarily observed sequentially. Instead, they are intertwined, can occur in any order, and are often iterative. As such, assimilation can occur anywhere in the social media enactment process.

Our model also shows that the organizations need to comprehend the *rationales* and *opportunities* linked with social media tools. They develop their social media *strategies* (replicate, transform, innovate) based on the *complementarities* they seek between existing ICT and social media, which will affect the content, the format, and the delivery of the message (Table 2). Our model also shows that to leverage their social media *strategies*, organizations also need to balance opportunities with the inherent *challenges* of social media.

This social media enactment process is also embedded in the orientation—*proactive* vs *reactive*—and the level of *mindfulness* vs *mindlessness* in which social media strategies are put in place, as illustrated in Table 3.

Table 2. Social media strategies: key message characteristics in the synergistic use of information and communication technologies and social media tools.

Strategies	Replicate	Transform	Innovate
Content	Same	Same	Different
Format	Same	Different	Different
Delivery	Different	Different	Different

Table 3. Reactive/proactive and mindless/mindful social media strategies enactment.

Orientation	Reactive	Proactive
Mindless	Type 1—go with the flow	Type 2—blind: trial and error
Mindful	Type 3—informed decision	Type 4—clearly defined strategy

When organizations are mindless and reactive (type 1), they generally *go with the flow*, that is, they observe and follow what is happening in the field. When organizations are more proactive, although still mindless (type 2), they do not have a clear plan for their social media strategy. Regardless, they attempt to stay in the forefront of their social media use and iteratively adjust their subsequent social media decisions on a *trial-and-error* basis. When organizations are mindful and reactive (type 3), they are observing others' usage of social media and assessing its potential value. They then decide whether and how to engage in implementing their social media strategy, thus making an *informed decision* but without a clear and definite plan of action. The final category (type 4) is when organizations are self-aware, staying on the edge, and create a *clearly defined strategy*. They then act with foresight, in a strategic and rational manner, which occurs when organizations are proactive and mindful.

Discussion

Principal Findings

Understanding how social media strategies are enacted and how social media can be strategically leveraged at the organizational level is an understudied area of research in health care. Recent work has established the importance of social media for patients and professionals to enable interactions and to access information [14]. We complement this work by looking at social media adoption by nonprofit cancer organizations—institutions that are central in health promotion. The goals of this study were to understand why and how six organizations put forth and enact social media strategies to achieve health promotion goals. Our analysis revealed five main rationales for adoption of social media, as described above, and a process of organizational adoption that we visualize in Figure 2. A key aspect of the all the rationales identified is that they have the common goal of enabling interaction with patients, families, and members of the community for reasons ranging from creating awareness and educating individuals to raising funds for the organizations.

This study adds to the existing literature around patient and professional use of social media [14,42] and extends it by delving deep into the process of adoption of social media by nonprofit organizations. In doing this, we not only look at social media by itself but also its use alongside other ICT tools [43]. To the best of our knowledge, no prior work has taken this approach, which provides an overarching view of the social media adoption process by organizations, a comprehensive understanding of opportunities and challenges associated with adoption of social media, and practical implications for managers who seek to use social media.

One of our key findings in this study is that to leverage their social media strategies, organizations need to balance *opportunities* with the inherent *challenges* of social media, such as lack of control [44], risk of misinformation, lack of privacy, limited audience, usability of social media programs, and the manipulation of identity [17]. With the recent attention to the spread of misinformation on the Web, organizations must understand and implement mechanisms to combat the risks associated with misinformation and privacy. It is critical that information is disseminated from credible sources, such as the organizations that we studied, using tools and technologies that end users, such as patients and their families, can access.

Furthermore, when studying organizational social media use, the question of *how* organizations should communicate with stakeholders is vital [45]. Results from our study suggest that it is imperative to consider the existing ICT when adopting a social media strategy. Our results shows that depending on the *complementarity* sought by the concomitant use of ICT and social media [46], organizations will seek to create the optimal *synergy* between the two strategies when interacting with users, which is consistent with current research findings that suggest that ICT provides most value when combined with other existing resources in the organization [46]. In developing social media *strategies* that take this complementarity into account, organizations must consider the capabilities of the tools along three dimensions: the content, the format, and the delivery [47,48]. Indeed, "...strategies do not need to be drastically overhauled to incorporate social media but merely retooled in framing messages and targeting audiences using the new media" [49].

Overall, although some organizations embrace social media to be at the forefront of innovation to provide health promotion, for others, social media adoption appears to be more of a bandwagon effect. Organizations feel pressure to use social media as they see their competitors and peers using it. In making decisions about social media, organizations face a highly

ambiguous environment because of its novelty. Indeed, at the organizational level, the impacts of social media strategies, and their benefits and risks, are still uncertain. Previous research indicates that under high-ambiguity conditions, bandwagon pressures tend to increase [28]. In addition, it has been said that the idea of "mindlessness in innovating with IT [Information Technologies] can reasonably be entertained whenever and wherever its likely rewards outweigh its risks" [30]. However, with time, as the understanding of social media and its role at the organizational level becomes clearer, it is to be expected that organizations would move toward enacting more mindful and proactive social media strategies. Indeed, "mindfulness is not something that an organization possesses: Instead, it is something that emerges in a process of becoming" [50]. Our results suggest that a proactive/mindful stance contributes to improve health promotion.

These results also pave the way for future research, such as testing the model using a larger sample to understand how this process may change depending on the type of organizations (eg, public health agencies, hospitals, private health care organizations, and bigger organization with dedicated staff for the social media activities). Moreover, it would be interesting to take into account the material properties of the social media tools themselves [51-53]. In that perspective, a study of the affordances of each social media tool could be insightful.

Conclusions

Our process model of social media strategies for health promotion by nonprofit organizations provides a means for managers of nonprofit organizations to understand the rationale of social media strategies and the role that social media can play in health promotion. Our process model can also be used as a guiding framework for nonprofit organizations engaging in social media use for health promotion. These organizations often face the challenge of effectively disseminating information to and engaging with the correct target group, all at low cost. This study provides these organizations with a mechanism for assessing how they can best exploit social media, taking into consideration the opportunities and challenges they face and the complementarities with their existing ICT. Using and understanding these mechanisms can help them create a well-defined strategy that will permit synergies between the existing ICT and social media, so that the use of both sets of tools together will bring in benefits that will surpass the simple sum of each.

Acknowledgments

Funding for this research was provided by a William Dawson Scholar Award (McGill University). Isabelle Vedel also received a New Investigator Salary Award from the Canadian Institutes of Health Research. The authors would like to thank Marine Hardouin and Andrea Zdyb for their help in editing this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed presentation of each case.
[DOCX File , 30 KB - jmir_v22i4e15586_app1.docx]

Multimedia Appendix 2
Data collection tools.
[DOCX File , 31 KB - jmir_v22i4e15586_app2.docx]

Multimedia Appendix 3
Coding scheme.
[DOCX File , 26 KB - jmir_v22i4e15586_app3.docx]

Multimedia Appendix 4
Quotes from key informants.
[DOCX File , 32 KB - jmir_v22i4e15586_app4.docx]

Multimedia Appendix 5
Data dossier.
[DOCX File , 26 KB - jmir_v22i4e15586_app5.docx]

References

1. Fox KF, Kotler P. The marketing of social causes: the first 10 years. J Mark 1980;44:24-33. [Medline: 12284730]

2. PSA Research Center. Mass Media and Health Practices Project: PSA Bibliography URL: http://www1.psaresearch.com/success8.html [accessed 2018-02-22]
3. Evans WD. Social marketing campaigns and children's media use. Future Child 2008;18(1):181-203. [doi: 10.1353/foc.0.0009] [Medline: 21338010]
4. Bloom PN, Novelli WD. Problems and challenges in social marketing. J Mark 1981;45(2):79-88. [Medline: 12280283]
5. Mangold WG, Faulds DJ. Social media: the new hybrid element of the promotion mix. Bus Horiz 2009;52(4):357-365. [doi: 10.1016/j.bushor.2009.03.002]
6. Winer RS. New communications approaches in marketing: issues and research directions. J Interact Mark 2009;23(2):108-117. [doi: 10.1016/j.intmar.2009.02.004]
7. 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 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
8. Thackeray R, Neiger BL, Smith AK, van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012 Mar 26;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
9. Ramanadhan S, Mendez SR, Rao M, Viswanath K. Social media use by community-based organizations conducting health promotion: a content analysis. BMC Public Health 2013 Dec 5;13:1129 [FREE Full text] [doi: 10.1186/1471-2458-13-1129] [Medline: 24313999]
10. Thackeray R, Neiger BL, Hanson CL, McKenzie JF. Enhancing promotional strategies within social marketing programs: use of web 2.0 social media. Health Promot Pract 2008 Oct;9(4):338-343. [doi: 10.1177/1524839908325335] [Medline: 18936268]
11. Park H, Rodgers S, Stemmle J. Health organizations' use of Facebook for health advertising and promotion. J Interact Advert 2011;12(1):62-77 [FREE Full text] [doi: 10.1080/15252019.2011.10722191]
12. 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]
13. Rupert DJ, Moultrie RR, Read JG, Amoozegar JB, Bornkessel AS, O'Donoghue AC, et al. Perceived healthcare provider reactions to patient and caregiver use of online health communities. Patient Educ Couns 2014 Sep;96(3):320-326. [doi: 10.1016/j.pec.2014.05.015] [Medline: 24923652]
14. Benetoli A, Chen TF, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Educ Couns 2018 Mar;101(3):439-444. [doi: 10.1016/j.pec.2017.08.015] [Medline: 28882545]
15. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
16. 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]
17. 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]
18. Litchman ML, Rothwell E, Edelman LS. The diabetes online community: older adults supporting self-care through peer health. Patient Educ Couns 2018 Mar;101(3):518-523. [doi: 10.1016/j.pec.2017.08.023] [Medline: 28947360]
19. Liang B, Scammon D. E-Word-of-Mouth on health social networking sites: an opportunity for tailored health communication. J Consumer Behav 2011;10(6):322-331 [FREE Full text] [doi: 10.1002/cb.378]
20. Ralph LJ, Berglas NF, Schwartz SL, Brindis CD. Finding teens in TheirSpace: using social networking sites to connect youth to sexual health services. Sex Res Soc Policy 2011;8(1):38-49. [doi: 10.1007/s13178-011-0043-4]
21. Swanson EB, Ramiller NC. The organizing vision in information systems innovation. Organ Sci 1997;8(5):458-474. [doi: 10.1287/orsc.8.5.458]
22. Garwin D. Building a learning organization. Harv Bus Rev 1993;71(4):73-91 [FREE Full text]
23. Weick KE, Sutcliffe KM. Managing the Unexpected: Resilient Performance in an Age of Uncertainty. New York City, USA: Jossey-bass; 2001.
24. Ramiller NC, Swanson EB. Organizing visions for information technology and the information systems executive response. J Manag Inf Syst 2003;20(1):13-50 [FREE Full text] [doi: 10.1080/07421222.2003.11045760]
25. Sun H, Fang Y, Zou HM. Choosing a fit technology: understanding mindfulness in technology adoption and continuance. J Assoc Inf Syst 2016;17(6):377-412. [doi: 10.17705/1jais.00431]
26. Thatcher JB, Wright RT, Sun H, Zagenczyk TJ, Klein R. Mindfulness in information technology use: definitions, distinctions, and a new measure. Manag Inf Syst Q 2018;42(3):831-847. [doi: 10.25300/misq/2018/11881]
27. Dernbecher S, Beck R. The concept of mindfulness in information systems research: a multi-dimensional analysis. Eur J Inf Syst 2017;26(2):121-142 [FREE Full text] [doi: 10.1057/s41303-016-0032-z]
28. Fiol CM, O'Connor EJ. Waking up! Mindfulness in the face of bandwagons. Acad Manag Rev 2003;28(1):54-70 [FREE Full text] [doi: 10.2307/30040689]
29. Valorinta M. Information technology and mindfulness in organizations. Ind Corp Change 2009;18(5):963-997. [doi: 10.1093/icc/dtp027]

30. Swanson EB, Ramiller NC. Innovating mindfully with information technology. Manag Inf Syst Q 2004;28(4):553-583. [doi: 10.2307/25148655]
31. Ramiller NC, Swanson EB. Mindfulness routines for innovating with information technology. J Decis Syst 2009;18(1):13-26. [doi: 10.3166/Jds.18.13-26]
32. Eisenhardt KM, Graebner ME. Theory building from cases: opportunities and challenges. Acad Manag J 2007;50(1):25-32. [doi: 10.5465/Amj.2007.24160888]
33. Jaccard J, Jacoby J. Theory Construction and Model-Building Skills: A Practical Guide for Social Scientists. New York City, USA: Guilford Publications, Inc; 2010.
34. Patton MQ. Qualitative Research & Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2002.
35. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015;65(1):5-29 [FREE Full text] [doi: 10.3322/caac.21254] [Medline: 25559415]
36. Edmondson AC, Mcmanus SE. Methodological fit in management field research. Acad Manag Rev 2007;32(4):1246-1264 [FREE Full text] [doi: 10.5465/amr.2007.26586086]
37. Gilgun JF. We shared something special: the moral discourse of incest perpetrators. J Marriage Fam 1995;57(2):265-281. [doi: 10.2307/353682]
38. Gilgun JF. Grounded theory, other inductive methods, and social work. In: Thyer BA, editor. The Handbook of Social Work Research Methods. Thousand Oaks, CA: Sage Publications Inc; 2001:345-364.
39. Rivard S, Lapointe L, Kappos A. An organizational culture-based theory of clinical information systems implementation in hospitals. J Assoc Inf Syst 2011;12(2):123-162. [doi: 10.17705/1jais.00258]
40. Urquhart C, Lehmann H, Myers MD. Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems. Inf Syst J 2010;20(4):357-381 [FREE Full text] [doi: 10.1111/j.1365-2575.2009.00328.x]
41. Strauss AL. Qualitative Analysis for Social Scientists. Cambridge, UK: Cambridge University Press; 1987.
42. Mazanderani F, O'Neill B, Powell J. 'People power' or 'pester power'? YouTube as a forum for the generation of evidence and patient advocacy. Patient Educ Couns 2013 Dec;93(3):420-425 [FREE Full text] [doi: 10.1016/j.pec.2013.06.006] [Medline: 23830239]
43. Gentry S, van-Velthoven MH, Car LT, Car J. Telephone delivered interventions for reducing morbidity and mortality in people with HIV infection. Cochrane Database Syst Rev 2013 May 31(5):CD009189 [FREE Full text] [doi: 10.1002/14651858.CD009189.pub2] [Medline: 23728687]
44. Macnamara J, Zerfass A. Social media communication in organizations: the challenges of balancing openness, strategy, and management. Int J Strat Commun 2012;6(4):287-308 [FREE Full text] [doi: 10.1080/1553118x.2012.711402]
45. Aral S, Dellarocas C, Godes D. Introduction to the special issue—social media and business transformation: a framework for research. Inf Syst Res 2013;24(1):3-13 [FREE Full text] [doi: 10.1287/isre.1120.0470]
46. Trainor KJ. Relating social media technologies to performance: a capabilities-based perspective. J Pers Selling Sales Manag 2012;32(3):317-331. [doi: 10.2753/Pss0885-3134320303]
47. Welch M, Jackson PR. Rethinking internal communication: a stakeholder approach. Corp Commun 2007;12(2):177-198. [doi: 10.1108/13563280710744847]
48. Fetscherin M, Knolmayer G. Business models for content delivery: an empirical analysis of the newspaper and magazine industry. Int J Media Manag 2004;6(1-2):4-11 [FREE Full text] [doi: 10.1080/14241277.2004.9669377]
49. Holmes W. Trace. 2011. Crisis Communications and Social Media: Advantages, Disadvantages and Best Practices URL: https://trace.tennessee.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1003&context=ccisymposium [accessed 2020-01-23]
50. Carlo JL, Lyytinen K, Boland Jr RJ. Dialectics of collective minding: Contradictory appropriations of information technology in a high-risk project. Manag Inf Syst Q 2012;36(4):1081-1108. [doi: 10.2307/41703499]
51. Leonardi PM, Barley SR. Materiality and change: challenges to building better theory about technology and organizing. Inf Organ 2008;18(3):159-176. [doi: 10.1016/j.infoandorg.2008.03.001]
52. Zammuto RF, Griffith TL, Majchrzak A, Dougherty DJ, Faraj S. Information technology and the changing fabric of organization. Organ Sci 2007;18(5):749-762. [doi: 10.1287/orsc.1070.0307]
53. Leonardi PM. When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. Manag Inf Syst Q 2011;35(1):147-167. [doi: 10.2307/23043493]

Abbreviations

ICT: information and communication technologies

Edited by G Eysenbach; submitted 23.07.19; peer-reviewed by S Smith, M Jones, A Benetoli; comments to author 01.09.19; revised version received 05.12.19; accepted 15.12.19; published 06.04.20.

Please cite as:
Vedel I, Ramaprasad J, Lapointe L
Social Media Strategies for Health Promotion by Nonprofit Organizations: Multiple Case Study Design
J Med Internet Res 2020;22(4):e15586
URL: https://www.jmir.org/2020/4/e15586
doi:10.2196/15586
PMID:32250282

Original Paper

The Impact of a Social Networking Service–Enhanced Smart Care Model on Stage 5 Chronic Kidney Disease: Quasi-Experimental Study

Feng-Jung Yang[1,2,3,4,5,6], MD, MMSc, MSci, MBA; Ying-Hui Hou[7*], PhD; Ray-E Chang[1*], PhD

[1]Institute of Health Policy and Management, College of Public Health, National Taiwan University, Taipei, Taiwan

[2]Department of Internal Medicine, National Taiwan University Hospital Yun Lin Branch, Douliu, Taiwan

[3]School of Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

[4]Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

[5]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[6]Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan

[7]Department Health Industry Management, School of Healthcare Management, Kainan University, Taoyuan, Taiwan

[*]these authors contributed equally

Corresponding Author:
Ray-E Chang, PhD
Institute of Health Policy and Management
College of Public Health
National Taiwan University
No 17, Xu-Zhou Road, Room 639
Taipei, 10055
Taiwan
Phone: 886 2 3366 8069
Fax: 886 2 2343 4200
Email: rchang@ntu.edu.tw

Abstract

Background: Stage 5 chronic kidney disease (CKD) presents a high risk for dialysis initiation and for complications such as uremic encephalopathy, uremic symptoms, gastrointestinal bleeding, and infection. One of the most common barriers to health care for patients with stage 5 CKD is poor continuity of care due to unresolved communication gaps.

Objective: Our aim was to establish a powerful care model that includes the use of a social networking service (SNS) to improve care quality for patients with CKD and safely delay dialysis initiation.

Methods: We used a retrospective cohort of CKD patients aged 20-85 years who received care between 2007 and 2017 to evaluate the efficacy of incorporating an SNS into the health care system. In 2014, author F-JY, a nephrologist at the National Taiwan University Hospital Yunlin Branch, started to use an SNS app to connect with stage 5 CKD patients and their families. In cases of emergency, patients and families could quickly report any condition to F-JY. Using this app, F-JY helped facilitate productive interactions between these patients and the health care system. The intention was to safely delay the initiation of dialysis therapy. We divided patients into four groups: group 1 (G1) included patients at the study hospital during the 2007-2014 period who had contact only with nephrologists other than F-JY; group 2 (G2) included patients who visited F-JY during the 2007-2014 period before he began using the SNS app; group 3 (G3) included patients who visited nephrologists other than F-JY during the 2014-2017 period and had no interactions using the SNS; and group 4 (G4) included patients who visited F-JY during the 2014-2017 period and interacted with him using the SNS app.

Results: We recruited 209 patients with stage 5 CKD who had been enrolled in the study hospital's CKD program between 2007 and 2017. Each of the four groups initiated dialysis at different times. Before adjusting for baseline estimated glomerular filtration rate (eGFR), the G4 patients had a longer time to dialysis (mean 761.7 days, SD 616.2 days) than the other groups (G1: mean 403.6 days, SD 409.4 days, *P*=.011 for G4 vs G1; G2: 394.8 days, SD 318.8 days, *P*=.04; G3: 369.1 days, SD 330.8 days, *P*=.049). After adjusting for baseline eGFR, G4 had a longer duration for each eGFR drop (mean 84.8 days, SD 65.1 days) than the other groups (G1: mean 43.5 days, SD 45.4 days, *P*=.005; G2: mean 42.5 days, SD 26.5 days, *P*=.03; G3: mean 3.8.7 days, SD 33.5 days, *P*=.002).

Conclusions: The use of an SNS app between patients with stage 5 CKD and their physicians can reduce the communication gap between them and create benefits such as prolonging time-to-dialysis initiation. The role of SNSs and associated care models should be further investigated in a larger population.

(J Med Internet Res 2020;22(4):e15565) doi:10.2196/15565

KEYWORDS

chronic kidney disease; stage 5 chronic kidney disease; chronic care model; dialysis initiation; social networking services; social networking; healthcare

Introduction

Background

The incidence of end-stage renal disease (ESRD) in Taiwan is very high [1], with an estimated prevalence of 15.46% for all stages of chronic kidney disease (CKD) and 9.06% for CKD stages 3-5 [2]. Research has shown that patients with advanced CKD enrolled in a pre-ESRD pay-for-performance program had lower risks of progressing to ESRD as well as lower risks of mortality and a significantly longer time to initiation of dialysis therapy (430 versus 347 days; $P<.001$) [3]. In a review of 32 articles reporting on 19 trials, earlier initiation of dialysis therapy was not found to reduce mortality compared to later initiation [4]. Economic analyses based on findings from the Initiating Dialysis Early And Late (IDEAL) trial and the US Renal Data System suggest that significant cost savings could be achieved by reversing the trend toward early initiation [5].

The most common barriers to health care reported by patients with CKD include poor continuity of care (eg, seeing a different specialist each visit; 49.3%), inadequate understanding and education about CKD (43.5%), feeling unwell (42.6%), and having trouble maintaining dietary and fluid restrictions (40.1%) [6]. The Chronic Care Model (CCM) is a well-developed and validated framework that illustrates a comprehensive approach to caring for people with chronic illness in a way that supports improved functional and clinical outcomes [7]. The most important aspect of CCM for CKD patients is self-management support, because successful self-management requires CKD patients to know how to monitor their disease, to manage symptoms, to interpret the results of home-monitoring therapies, and to carry out daily treatment plans, including adhering to medication regimens and dietary and fluid restrictions and dealing with side effects.

Surprisingly, clinicians tend to overlook patient access to and use of information and communication technologies (ICTs) to manage their health [8,9]. One study reported that less than 25% of CKD patients obtained information about renal health care from the internet. The ICTs most preferred by their renal health care teams were telephone (56.5%), internet (50%), email (48.3%), and text messages (46%) [9].

However, health care systems and patients are increasingly turning to the internet—including websites as well as social media platforms—for health-related information and support. The US-based National Kidney Foundation, for example, designed a comprehensive and user-friendly digital ecosystem that contains content relevant to each audience and helps promote prompt interactions between CKD patients and their health care providers as envisioned by the CCM [10]. The ecosystem received high satisfaction scores (88%) on the ForeSee survey, a customer satisfaction survey administered on the US National Institute of Diabetes and Digestive and Kidney Diseases websites [10]. Results from a paper by Gee et al [11] showed that chronic care needs to reform to incorporate ICT tools. They concluded that (1) eHealth (electronic health) education has a critical role in self-care; (2) eHealth support should be put into the community, and patients should be empowered with the benefits of the e-community (electronic community) or virtual communities; and (3) productive technology-based interactions ensure feedback loop between the patient and the provider [11].

A randomized controlled trial found no difference in mean estimated glomerular filtration rate (eGFR) at baseline and the number of patients who progressed renal replacement regardless of use of a standardized self-management program (where patient education, telephone-based support, and support groups were delivered by a multidisciplinary team of management nurses, dietitians, peers, and volunteer) [12]. In another study, the University of Pittsburgh Medical Center instituted an impressive patient- and family-initiated rapid response system called Condition Help based out of a hospital [13]. Safety issues could be identified and prevented through the Condition Help system, although the majority (83.4%) of calls involved nonsafety issues. In one study looking at a cohort of Hispanic patients with CKD, lower patient-physician interaction scores were independently associated with a higher risk of hospitalization but not with incidence of ESRD or death [14].

Further testing of ICT interventions to improve self-management is necessary. Despite the potential benefits of ICTs for health care, few studies have addressed the usage and preferences regarding these technologies among patients with chronic diseases such as CKD and ESRD [10].

Objectives

A new model of physician-patient interaction, the eHealth Enhanced Chronic Care Model, was discussed by Gee et al [11] (Multimedia Appendix 1). Following this model, our study aimed to establish the value of a new kind of connection between patients and the hospital where they are receiving care—a connection that uses a social networking service (SNS) to encourage proactive action. Our study explores a smart care model with SNS to determine whether it improves how patients and providers connect across digital platforms and to advance our understanding of how an SNS might be able to improve health-related outcomes for patients with CKD.

Methods

Study Population

In 2002, Taiwan's health care system, known as National Health Insurance, launched the Project of Integrated Care for CKD, a nationwide pre-ESRD pay-for-performance program providing more comprehensive care to patients with CKD (stages 3-5). The National Taiwan University Hospital (NTUH) Yunlin Branch, a regional teaching hospital in southern Taiwan, joined this program in 2004 and prospectively enrolled patients with CKD (stages 3-5) who were willing to participate. Patients were diagnosed with CKD according to the criteria of the National Kidney Foundation's Kidney Disease Outcomes Quality Initiative (KDOQI) clinical practice guidelines [15] and received follow-up care at an outpatient department. Biochemical profiles including serum creatinine, blood urea nitrogen (BUN), and the spot urine protein to creatinine ratio (UPCR) were measured at least every 12 weeks. Our study followed all enrolled patients until initiation of long-term renal replacement therapy (hemodialysis, peritoneal dialysis, or transplantation) or until December 31, 2017, whichever occurred first.

Patient Selection

For our study, we selected a retrospective cohort of patients who had stage 5 CKD when they enrolled in the Project of Integrated Care for CKD. To further ensure that study patients were regularly treated for CKD, eligible patients were required to have had at least two outpatient visits within 3 months of their first date of diagnosis. In our final analysis, we included 209 patients aged 20 to 85 years with stage 5 CKD who had enrolled in the program and remained in it between 2007 and 2017.

Patient Grouping

Author F-JY is a nephrologist at the teaching hospital where this study took place. In 2015, he started using an SNS app to connect with his patients with stage 5 CKD and their families. In cases of emergency, patients and families could quickly report any condition to F-JY through this app. In this way, F-JY helped promote productive interactions and prompt responses between patients with stage 5 CKD and the health care system. In addition, family members and patients could exchange messages and photos with their health care providers via this platform. No medical decisions were made on this platform; if any risks emerged, clinic appointments were made. All nephrologists at the hospital acted in accordance with the guidelines of the KDOQI and the pay-for-performance program.

To examine the effect of the SNS intervention on care for patients with stage 5 CKD, we employed a quasi-experimental (single group pre-post) study design. We divided patients into four groups. Group 1 (G1) included patients who had contact only with hospital nephrologists other than F-JY during the 2007-2014 period; group 2 (G2) included patients who visited F-JY during the 2007-2014 period before he began using the SNS app; group 3 (G3) included patients who visited nephrologists other than F-JY during the 2014-2017 period and had no SNS interactions; and group 4 (G4) included patients who visited F-JY during the 2014-2017 period and used the SNS. The number of patients per group is shown in Multimedia Appendix 2.

Social Networking Service

The SNS used by F-JY was Line, a mobile app operated by Naver Corporation. Users can use texts, images, video, and audio for contact, and have free voice conversations and video conferences at any time. In Taiwan, the Line app has become increasingly popular since 2014 [16].

Measurement of Kidney Function

A serum creatinine level of <15 ml/min/1.73 m2 was used to define the baseline eGFR and establish the patient's CKD stage as 5 at enrollment. eGFR was calculated with the Modification of Diet in Renal Disease equation:

$$\text{eGFR} = 175 \times \text{creatinine}^{-1.154} \times \text{age}^{-0.203} \times 1.212 \text{ (if black)} \times 0.742 \text{ (if female)}$$

The value of eGFR was recorded by every 3 months until dialysis. The rate of decline in daily eGFR was defined as baseline eGFR divided by time to dialysis in days.

Data Collection

Blood samples were collected before every clinic visit. Hematological and biochemical tests were conducted in the central laboratory of the hospital. Baseline comorbidities and clinical laboratory data were recorded by taking medical histories and conducting detailed chart reviews. All medical histories were recorded, including diabetes, hypertension, and cardiovascular disease (defined as coronary artery disease, myocardial infarction, stroke, and heart failure), after reviewing the electronic medical records.

Statistical Analyses

Continuous variables were reported as means and standard deviations, or medians with interquartile ranges if distributions were skewed. Categorical variables, such as cardiovascular risk factors (eg, hypercholesterolemia, hypertension, and smoking), were reported as frequencies and percentages.

Baseline characteristics, including demographics (age and gender), laboratory data (sodium, potassium, calcium, phosphate, BUN, creatinine, and albumin), and presence of comorbid disease (coronary artery disease, congestive heart failure, hypertension, diabetes mellitus, and malignancy) were compared between the groups. Baseline patient characteristics were compared using chi-square tests for categorical variables and the Mann-Whitney *U* test for continuous variables.

We performed multiple linear regression analysis to evaluate the associations between time to dialysis, duration of each eGFR drop, and daily eGFR decline rate with baseline covariables. We combined difference in differences with matching on pretreatment outcomes to address nonparallel trends between the treatment and control groups. The covariables included care model, age, diabetes mellitus (yes vs no), albumin, and hemoglobin. A multiple logistic regression analysis was performed to determine the factors associated with time to dialysis, duration of each eGFR drop, and daily eGFR decline rate. Statistical analyses were performed using SPSS Version 25 (IBM) and Prism 7.0d (GraphPad).

Ethics Statement

The study was approved by the NTUH ethical review board (NTUH 201901030RINB and 201903005RINA). To maintain confidentiality, all data sets in the study were pseudonymized, and personal IDs, birth dates, and names were encrypted. This deidentification process was supervised by NTUH's Institutional Review Board (IRB), which verified the anonymity of data analysis performed in this study. Because the data were analyzed anonymously and in accordance with IRB guidelines, informed consent was not obtained from the study participants. All research procedures followed the directives of the Declaration of Helsinki.

Results

Baseline Measurements

We investigated the differences in demographics and clinical and laboratory data among the four study groups. There were no differences between groups in terms of gender, age, diabetes, hypertension, hemoglobin (g/dL), baseline eGFR, BUN (mg/dL), creatinine (mg/dL), sodium (mmol/L), potassium (mmol/L), calcium (mg/dL), phosphorus (mg/dL), or uric acid (mg/dL). As shown in Table 1, patients in the four groups did show differences in time to dialysis, albumin, and calcium levels, all of which were statistically significant.

Table 1. Baseline demographic, clinical, and laboratory measurements by group.

Measurement[a]	Group 1[b]		Group 2[b]		Group 3[b]		Group 4[b]		Total		*P* value[c]
	Value	n	Value	n	Value	n	Value	n	Value	N	
Male, n (%)	39 (44)	88	18 (55)	33	40 (56)	71	8 (47)	17	105 (50)	209	.46
Age (years), mean (SD)	66.8 (13.0)	88	67.2 (12.7)	33	66.9 (13.9)	71	63.1 (10.2)	17	66.6 (13.0)	209	.72
Diabetes mellitus (yes), n (%)	45 (51)	88	18 (48)	33	37 (52)	71	10 (59)	17	109 (52)	209	.92
Hypertension (yes), n (%)	58 (66)	88	20 (61)	33	51 (72)	71	12 (71)	17	140 (67)	209	.69
Albumin (g/dL), mean (SD)	3.7 (0.4)	86	4.0 (0.5)	33	3.6 (0.5)	71	3.8 (0.5)	17	3.7 (0.5)	207	*.005*
Hemoglobin (g/dL), mean (SD)	9.0 (1.5)	87	9.5 (1.6)	33	9.4 (1.4)	71	8.8 (1.5)	17	9.2 (1.5)	208	.26
eGFR (mL/min/1.73 m^2), mean (SD)	9.2 (2.9)	88	9.1 (3.1)	33	9.8 (3.0)	71	9.1 (2.0)	17	9.4 (2.9)	209	.44
Time to dialysis (days), mean (SD)	403.6 (409.4)	88	394.8 (318.8)	33	369.1 (330.8)	71	761.7 (616.2)	17	419.6 (403.0)	209	*.003*
BUN (mg/dL), mean (SD)	86.4 (24.3)	87	93.7 (39.8)	33	94.9 (35.4)	71	88.6 (31.5)	17	90.6 (31.7)	208	.37
Creatinine (mg/dL), mean (SD)	9.4 (3.9)	88	9.7 (3.9)	33	8.8 (3.5)	71	9.7 (3.8)	17	9.3 (3.8)	209	.66
Sodium (mmol/L), mean (SD)	137.5 (4.3)	77	137.7 (4.6)	31	136.9 (4.7)	45	137.1 (3.8)	13	137.3 (4.4)	166	.84
Potassium (mmol/L), mean (SD)	4.7 (0.7)	88	4.5 (0.7)	33	4.6 (0.7)	71	4.9 (0.6)	17	4.6 (0.7)	209	.29
Calcium (mg/dL), mean (SD)	8.8 (0.9)	87	9.1 (1.0)	33	8.3 (1.4)	71	8.7 (0.7)	17	8.7 (1.1)	208	*.001*
Phosphorus (mg/dL), mean (SD)	5.9 (1.7)	87	5.9 (1.6)	33	5.8 (1.5)	71	6.2 (1.4)	17	5.9 (1.6)	208	.80
Uric acid (mg/dL), mean (SD)	8.4 (1.9)	80	8.4 (2.1)	32	7.5 (2.8)	47	7.1 (2.2)	13	8.1 (2.2)	172	.05

[a]Measurements are for patients with stage 5 CKD at time of enrollment.

[b]Groups are stratified by physician and care model. Group 1: patients who had contact only with hospital nephrologists other than author F-JY, 2007-2014; group 2: patients who visited F-JY before he began using the SNS app, 2007-2014; group 3: patients who visited nephrologists other than F-JY, 2014-2017 (no SNS interactions); group 4: patients who visited F-JY and used the SNS app, 2014-2017.

[c]*P* values express differences in data between groups. *P* values less than .05 are marked in italics.

Age and Time to Dialysis

Aging—but not age—presented a risk for dialysis initiation. As shown in Table 2, age did not increase with daily eGFR decline rate in late-enrolled patients with stage 5 CKD. Decline in daily eGFR, albumin, hemoglobin, care model, and eGFR at enrollment were not associated with age, but age showed a positive association with eGFR at the last outpatient visit (r=.136, P=.049). In addition, time to dialysis was associated positively with age, albumin, and eGFR at enrollment but negatively with diabetes.

Table 2. Pearson correlations (*r*) between age and other risk factors, and time to dialysis and other risk factors, among enrolled patients with stage 5 chronic kidney disease.

Variable	*r*	*P* value[a]
Age		
Daily eGFR[b] decline rate	−.065	.35
Albumin	.063	.37
Hemoglobin	−.026	.71
Diabetes mellitus (yes)	−.046	.51
Care model	−.058	.41
eGFR at enrollment	.081	.25
eGFR at last outpatient visit	.136	*.049*
Time to dialysis		
Age	*.166*	*.02*
Diabetes mellitus (yes)	*−.183*	*.008*
Hemoglobin	.016	.82
Albumin	*.254*	*<.001*
Blood urea nitrogen	.094	.18
eGFR at enrollment	*.268*	*<.001*
Creatinine	.104	.14
Sodium	.080	.31
Potassium	−.081	.24
Calcium	.113	.11
Phosphorus	−.015	.83

[a]*P* values less than .05 are marked in italics.

[b]eGRF: estimated glomerular filtration rate.

Time to Dialysis and Duration of Each eGFR Drop

Time to dialysis differed among the four groups of patients. Before adjusting for baseline eGFR, patients in G4 had a longer time to dialysis initiation (mean 761.7 days, SD 616.2 days) than patients in the other groups (G1: mean 403.6 days, SD 409.4 days, *P*=.011; G2: mean 394.8 days, SD 318.8 days, *P*=.04; G3: mean 369.1 days, SD 330.8 days, *P*=.049; Figure 1). Each patient showed differences in renal function at baseline. As shown in Table 3, we adjusted time to dialysis according to baseline eGFR. The duration of each eGFR drop is defined by time to dialysis (days) divided by the baseline eGFR (mL/min/1.73 m^2). G4 had longer durations for each eGFR drop (mean 84.8 days, SD 65.1 days) compared to the other groups (G1: mean 43.5 days, SD 45.4 days, *P*=.005; G2: mean 42.5 days, SD 26.5 days, *P*=.03; G3: mean 38.7 days, SD 33.5 days, *P*=.002; Figures 2 and 3).

Figure 1. Trends in time to dialysis were compared between the four groups. The lines indicate median values. Statistical calculation of *P* values was performed using the nonparametric Mann-Whitney *U* test.

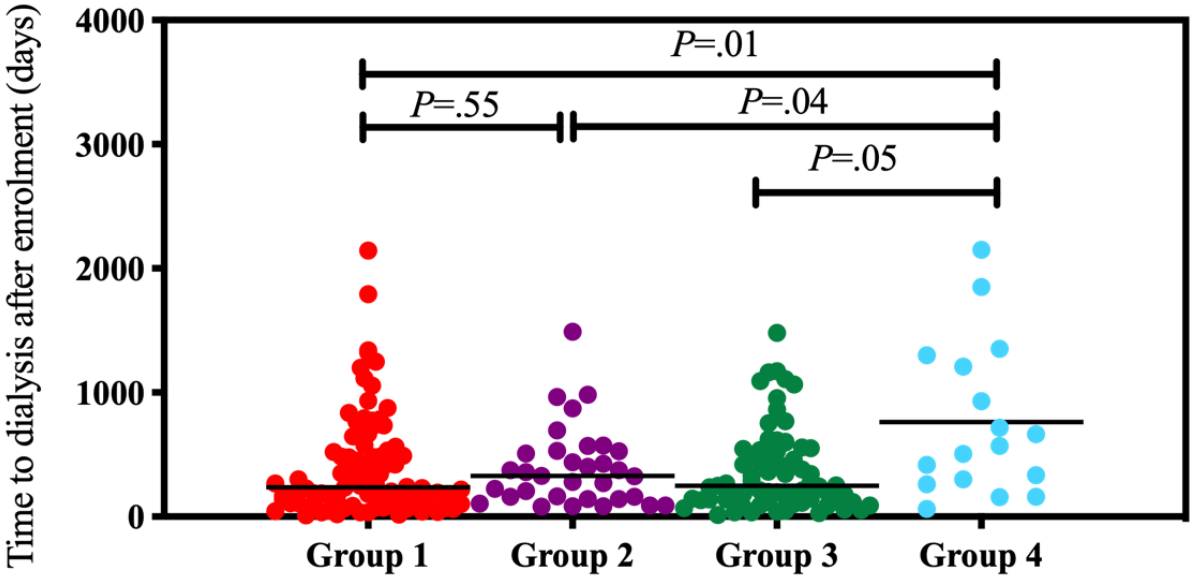


Table 3. Comparison of clinical and laboratory measurements by group, adjusted for baseline estimated glomerular filtration rate (eGFR).

Measurement[a]	Group 1	Group 2	Group 3	Group 4	*P* value[b]
Age (years), mean (SD)	66.8 (13.0)	67.2 (12.7)	66.9 (13.9)	63.1 (10.2)	.72
Diabetes mellitus (yes), n (%)	45 (51)	16 (48)	37 (52)	10 (59)	.92
Albumin (g/dL), mean (SD)	3.7 (0.4)	4.0 (0.5)	3.6 (0.5)	3.8 (0.5)	*.005*
Hemoglobin (g/dL), mean (SD)	9.0 (1.5)	9.5 (1.6)	9.4 (1.4)	8.8 (1.5)	.26
eGFR (mL/min/1.73 m^2), mean (SD)	9.2 (2.9)	9.1 (3.1)	9.8 (3.0)	9.1 (2.0)	.44
Time to dialysis (days), mean (SD)	403.6 (409.4)	394.8 (318.8)	369.1 (330.8)	761.7 (616.2)	*.003*
Duration of each eGFR drop, (days/[mL/min/1.73 m^2]), mean (SD)	43.5 (45.4)	42.5 (26.5)	38.7 (33.5)	84.8 (65.1)	*.010*

[a]Measurements are for patients with stage 5 CKD at time of enrollment.

[b]*P* values less than .05 are marked in italics.

Figure 2. Trends in the decline of daily estimated glomerular filtration rate (eGFR; [mL/min/1.73 m^2]/day) were compared between the four groups. The lines indicate median values. Statistical calculation of *P* values was performed using the nonparametric Mann-Whitney *U* test.

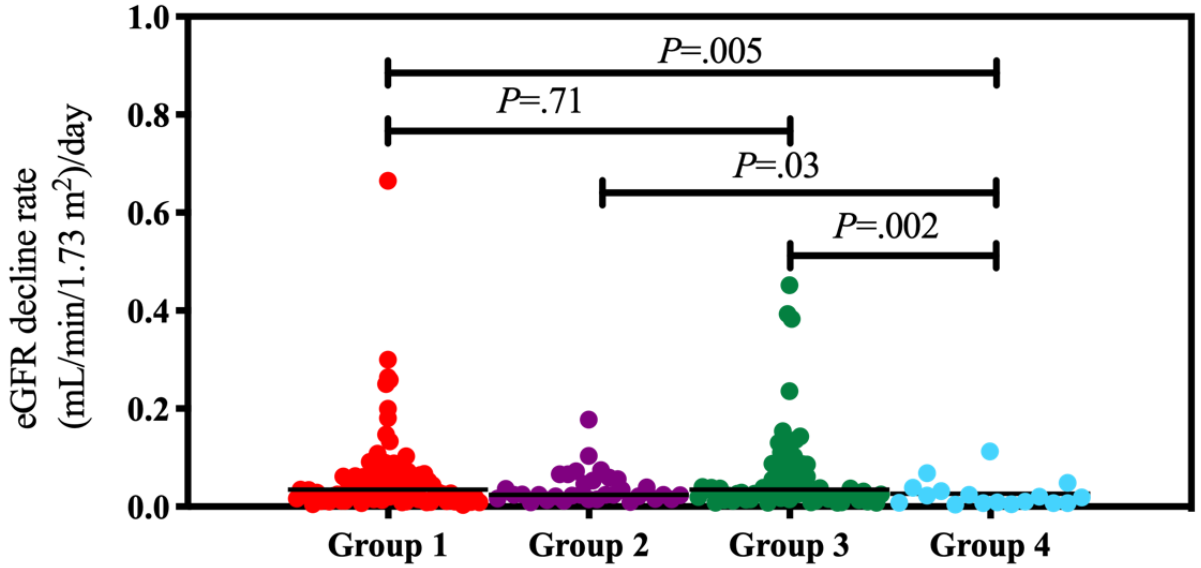
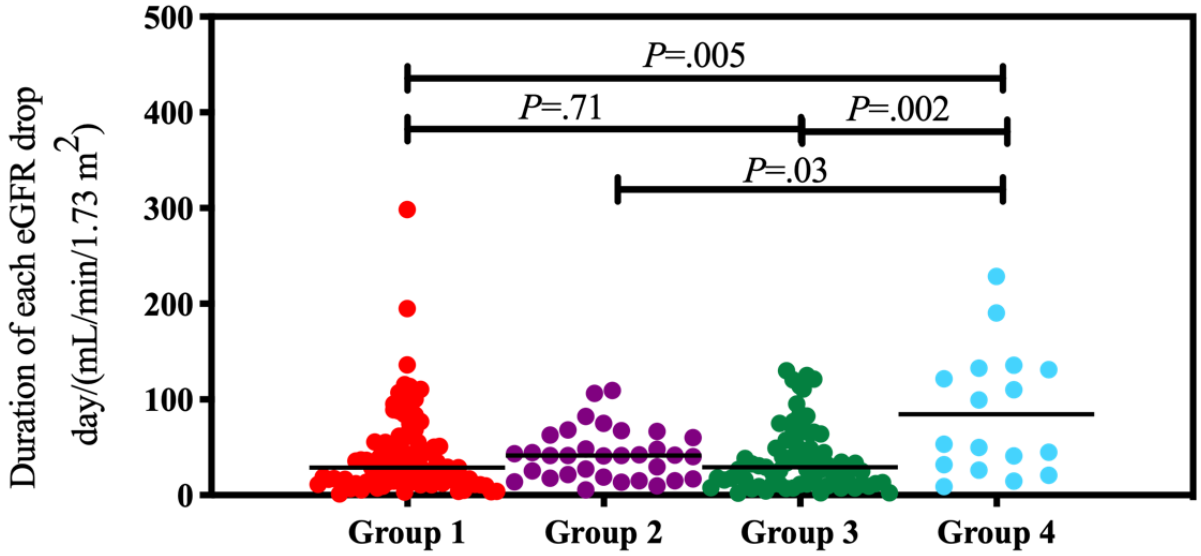


Figure 3. Durations of each drop in estimated glomerular filtration rate (EGFR; day/[mL/min/1.73 m^2]) were compared between the four groups. The lines indicate median values. Statistical calculation of *P* values was performed using the nonparametric Mann-Whitney *U* test.



Difference in Differences in the Care Model

Finally, we tested the independent association of the new care model and time to dialysis, and combined difference in differences with matching on pretreatment outcomes to address nonparallel trends between groups. We measured two factors: factor A was the individual physician factor (ie, "team" versus F-JY, where "team" refers to the involvement of any nephrologist other than F-JY); and factor B refers to the intervention phase (ie, the control phase, 2007-2014, versus the experimental phase, 2014-2017). Multivariable-adjusted logistic regression models, including age, diabetes mellitus, albumin, and hemoglobin, were used to investigate the independent association between factor A, factor B, and time to dialysis.

As shown in Table 4, the interaction of factor A with factor B was independently associated with time to dialysis in regression models after adjustment for age, diabetes mellitus, albumin, and hemoglobin. We found that the new care model driven by SNS had increased time to dialysis to 459.38 days. Similar results were found when the duration of each eGFR drop was used in the regression model instead of time to dialysis (Table 5). The SNS care model increased the duration of each eGFR drop to 52.7 days. In addition, in Table 6, we showed results in comparison with a reference group of adult patients less than 50 years of age. Multivariable-adjusted logistic regression models, including three other age groups (50-59 years, 60-69 years, and >70 years), diabetes mellitus, gender, albumin, and hemoglobin, were used to investigate the independent association between factor A (team vs F-JY), factor B (control vs experimental), and time to dialysis. Using difference in differences, our new care model had an increased time to dialysis of approximately 417.6 days.

Table 4. Associations between time to dialysis and care model using difference in differences[a].

Variable	Unstandardized coefficients		Standardized coefficients			
	B	SE	Beta	*t*	*P* value[b]	95% CI
(Constant)	−569.77	292.84	—[c]	−1.95	.053	−1147.27 to 7.72
Age	5.60	2.02	0.18	2.77	*.006*	1.62 to 9.58
Diabetes	−117.53	54.44	−0.15	−2.16	*.03*	−224.89 to −10.17
Hemoglobin	10.51	17.57	0.039	0.60	.55	−24.14 to 45.16
Albumin	155.07	56.87	0.19	2.73	*.007*	42.92 to 267.22
Factor A: team vs author F-JY	−67.08	77.94	−0.07	−0.86	.39	−220.78 to 86.63
Factor B: control vs experimental	−30.32	60.79	−0.04	−0.50	.62	−150.19 to 89.56
Factor A × Factor B	459.38	128.31	0.31	3.58	*<.001*	206.36 to 712.40

[a]Multivariable-adjusted logistic regression models, including age, diabetes mellitus, albumin, and hemoglobin, were used to investigate the independent association between factor A, factor B, and time to dialysis.

[b]*P* values less than .05 are marked in italics.

[c]Not applicable.

Table 5. Associations between the duration of each drop in estimated glomerular filtration rate and the care model using difference in differences[a].

Variable	Unstandardized coefficients		Standardized coefficients			
	B	SE	Beta	*t*	*P* value[b]	95% CI
(Constant)	−50.55	30.81	—[c]	−1.64	.10	−111.30 to 10.21
Age	0.43	0.21	0.13	2.03	*.044*	0.01 to 0.85
Diabetes	−14.54	5.73	−0.17	−2.54	*.012*	−25.83 to −3.24
Hemoglobin	1.18	1.85	0.04	0.64	.53	−2.47 to 4.82
Albumin	16.98	5.98	0.20	2.84	*.005*	5.18 to 28.78
Factor A: team vs author F-JY	−7.17	8.20	−0.07	−0.88	.38	−23.34 to 9.00
Factor B: control vs experimental	−4.16	6.40	−0.05	−0.65	.52	−16.77 to 8.45
Factor A × factor B	52.71	13.50	0.34	3.91	*<.001*	26.09 to 79.33

[a]Multivariable-adjusted logistic regression models, including age, diabetes mellitus, albumin, and hemoglobin, were used to investigate the independent association between factor A, factor B, and duration of each eGFR drop.

[b]*P* values less than .05 are marked in italics.

[c]Not applicable.

Table 6. Associations between time to dialysis and care model using difference in differences compared with the reference group (age <50 years)[a].

Variable	Unstandardized coefficients		Standardized coefficients			
	B	SE	Beta	*t*	*P* value[b]	95% CI
(Constant)	−158.05	268.87	—[c]	−0.59	.56	−688.32 to 372.22
Age (50-59.9 years)	−30.66	95.51	−0.03	−0.32	.75	−219.03 to 157.70
Age (60-69.9 years)	−14.67	92.97	−0.02	−0.16	.88	−198.02 to 168.69
Age (≥70 years)	88.57	84.02	0.11	1.05	.29	−77.15 to 254.28
Gender	−176.07	52.94	−0.22	−3.33	*.001*	−280.47 to −71.67
Diabetes	−107.57	53.68	−0.13	−2.00	*.046*	−213.44 to −1.70
Hemoglobin	9.94	17.22	0.04	0.58	.56	−24.02 to 43.90
Albumin	155.03	55.78	0.19	2.78	*.006*	45.03 to 265.03
Factor A: team vs author F-JY	−32.56	77.01	−0.04	−0.42	.67	−184.44 to 119.32
Factor B: control vs experimental	−5.31	59.98	−0.007	−0.09	.93	−123.60 to 112.98
Factor A × factor B	417.61	126.71	0.29	3.30	*.001*	167.72 to 667.50

[a]Multivariable-adjusted logistic regression models, including age, diabetes mellitus, albumin, and hemoglobin, were used to investigate the independent association between factor A (team vs F-JY), factor B (control vs experimental), and duration of each eGFR drop. The dependent variable was time to dialysis.

[b]*P* values less than .05 are marked in italics.

[c]Not applicable.

Discussion

Principal Findings

This study is significant in that it is the first study to combine an SNS with standard care for patients with stage 5 CKD. Other studies of SNSs and CKD recruited patients with CKD stages 3-4 only or patients with an unknown CKD stage. In this study, the physician, not the case manager, played the central role in integrated care.

Our study demonstrates the effectiveness of a new care model that includes use of an SNS for patients with stage 5 CKD and their physicians. After introducing the model, patients using the SNS care model saw benefits such as a longer time to dialysis, a longer duration of each eGFR drop, and a daily decline in eGFR, perhaps because they gained more real-time mental health–related and other support from their physicians.

Comparison With Prior Work

The IDEAL trial, a randomized controlled trial, showed a median time to initiation of dialysis of 1.80 months (95% CI 1.60 to 2.23) in its early-start group and 7.40 months (95% CI 6.23 to 8.27) in its late-start group, but showed no significant difference between the groups in terms of adverse events (cardiovascular events, infections, or complications of dialysis) [17]. This trial has demonstrated it can be safe to wait for lower eGFR levels or specific symptoms before beginning dialysis.

Razzaghi and Afshar [18] reported on the four key components of the physician-patient healing relationship: (1) valuing the patient as a person, (2) effective management of the power imbalance between the physician and the patient, (3) commitment, and (4) competence and character of the physician. They additionally stated that the three necessary relational elements of physician-patient relationship are trust, peace and hope, and acknowledgment [18].

Patient adherence to treatment is highly influenced by the quality of communication patients receive during medical care [19]. Physicians must be trained in efficient communication to enhance this adherence. The physician-patient relationship, therefore, plays a central role in patient health-related outcomes. The components of this relationship may impact the patient's experience of the health system.

Our study addressed the potential of a primary physician–led care model with an SNS in delaying dialysis initiation for patients with stage 5 CKD. This SNS-integrated care practice showed a significant effect in the intervention group compared with the control group in slowing the daily reduction in eGFR, and there was significant improvement in time to initiation of dialysis among patients in the intervention group compared with patients in the control group (with an imbalance between groups resulting from their relatively small sample size). This is a quasi-experimental design that suggests that SNSs in primary care can increase patient awareness, delay dialysis initiation, and provide patients with more mental health support from their physician.

Future Directions

To our knowledge, this is the first study of SNS that took place in a real-world primary care practice that measured outcomes for patients with stage 5 CKD longitudinally. To confirm the usefulness of the SNS care model, more data and objective results regarding efficacy are needed. The role of SNSs and associated care models should be further investigated in a larger population. Observational studies, clinical trials, systematic reviews, and meta-analyses should aim to further establish the role of digital health technologies in patient care.

Limitations

This study is retrospective to review the quality of care with SNS in a teaching hospital. The sample size is relatively small and only few physicians are involved in the study.

Conclusions

Use of SNSs by patients with stage 5 CKD and their physicians could resolve important communication gaps and create better conditions for treatment of CKD. In our study, we demonstrated that use of an SNS in a physician-led care model was associated with a significant delay in dialysis initiation for stage 5 CKD patients. SNSs can act as a communication bridge and assist in improving the connection between the health care system and the community. Patient adherence is highly correlated to communication between the medical caregiver and the patient. Better communication skills improve patient adherence and outcomes [19]. Consistent with other works, we found that high-quality communication and interpersonal support can improve care quality.

Acknowledgments

This work was supported by the National Taiwan University Hospital Yunlin Branch (grants NTUHYL105.X010 and NTUHYL106.X009). The authors thank Yi-Chen Wang and Pei-Hsuan Chung for their assistance in data collection.

Authors' Contributions

F-JY conceived the study, obtained the funding, performed the statistical analyses, and wrote the first draft. All authors contributed to the analysis and the writing of the final manuscript. R-EC and Y-HH provided expert opinions for the study design and execution and edited the manuscript; F-JY, Y-HH, and R-EC participated in discussion, interpretation, and final preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The eHealth enhanced Chronic Care Model.
[DOCX File , 548 KB - jmir_v22i4e15565_app1.docx]

Multimedia Appendix 2
Grouping of patients with stage 5 chronic kidney disease by physician and care model (with or without social networking service).
[DOCX File , 24 KB - jmir_v22i4e15565_app2.docx]

References

1. United States Renal Data System. Chapter 11: International Comparisons. American Journal of Kidney Diseases 2019 Mar;73(3):S549-S594. [doi: 10.1053/j.ajkd.2019.01.004]
2. Tsai M, Hsu C, Lin M, Yen M, Chen H, Chiu Y, et al. Incidence, Prevalence, and Duration of Chronic Kidney Disease in Taiwan: Results from a Community-Based Screening Program of 106,094 Individuals. Nephron 2018;140(3):175-184. [doi: 10.1159/000491708] [Medline: 30138926]
3. Hsieh H, Lin M, Chiu Y, Wu P, Cheng L, Jian F, et al. Economic evaluation of a pre-ESRD pay-for-performance programme in advanced chronic kidney disease patients. Nephrol Dial Transplant 2017 Jul 01;32(7):1184-1194. [doi: 10.1093/ndt/gfw372] [Medline: 28486670]
4. Slinin Y, Greer N, Ishani A, MacDonald R, Olson C, Rutks I, et al. Timing of dialysis initiation, duration and frequency of hemodialysis sessions, and membrane flux: a systematic review for a KDOQI clinical practice guideline. Am J Kidney Dis 2015 Nov;66(5):823-836. [doi: 10.1053/j.ajkd.2014.11.031] [Medline: 26498415]
5. Abra G, Kurella Tamura M. Timing of initiation of dialysis: time for a new direction? Curr Opin Nephrol Hypertens 2012 May;21(3):329-333 [FREE Full text] [doi: 10.1097/MNH.0b013e328351c244] [Medline: 22388556]
6. Lo C, Teede H, Fulcher G, Gallagher M, Kerr PG, Ranasinha S, et al. Gaps and barriers in health-care provision for co-morbid diabetes and chronic kidney disease: a cross-sectional study. BMC Nephrol 2017 Mar 28;18(1):80 [FREE Full text] [doi: 10.1186/s12882-017-0493-x] [Medline: 28245800]
7. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [Medline: 8941260]
8. Romano MF, Sardella MV, Alboni F, Russo L, Mariotti R, Nicastro I, et al. Is the digital divide an obstacle to e-health? An analysis of the situation in Europe and in Italy. Telemed J E Health 2015 Jan;21(1):24-35. [doi: 10.1089/tmj.2014.0010] [Medline: 25495564]
9. Bonner A, Gillespie K, Campbell KL, Corones-Watkins K, Hayes B, Harvie B, et al. Evaluating the prevalence and opportunity for technology use in chronic kidney disease patients: a cross-sectional study. BMC Nephrol 2018 Dec 02;19(1):28 [FREE Full text] [doi: 10.1186/s12882-018-0830-8] [Medline: 29394930]

10. Goldstein K, Briggs M, Oleynik V, Cullen M, Jones J, Newman E, et al. Using digital media to promote kidney disease education. Adv Chronic Kidney Dis 2013 Jul;20(4):364-369 [FREE Full text] [doi: 10.1053/j.ackd.2013.04.001] [Medline: 23809289]
11. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LMS. The eHealth Enhanced Chronic Care Model: a theory derivation approach. J Med Internet Res 2015;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
12. Chen S, Tsai Y, Sun C, Wu I, Lee C, Wu M. The impact of self-management support on the progression of chronic kidney disease--a prospective randomized controlled trial. Nephrol Dial Transplant 2011 Nov;26(11):3560-3566. [doi: 10.1093/ndt/gfr047] [Medline: 21414969]
13. Eden EL, Rack LL, Chen L, Bump GM. Condition Help: A Patient- and Family-Initiated Rapid Response System. J Hosp Med 2017 Mar;12(3):157-161. [doi: 10.12788/jhm.2697] [Medline: 28272591]
14. Cedillo-Couvert EA, Hsu JY, Ricardo AC, Fischer MJ, Gerber BS, Horwitz EJ, CRIC Study Investigators. Patient Experience with Primary Care Physician and Risk for Hospitalization in Hispanics with CKD. Clin J Am Soc Nephrol 2018 Nov 07;13(11):1659-1667 [FREE Full text] [doi: 10.2215/CJN.03170318] [Medline: 30337326]
15. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002 Feb;39(2 Suppl 1):S1-266. [Medline: 11904577]
16. Taipei Times. 2014 Jan 30. Messaging app Line calls Taiwan a major market URL: http://www.taipeitimes.com/News/biz/archives/2014/01/30/2003582448 [accessed 2020-03-31]
17. Cooper BA, Branley P, Bulfone L, Collins JF, Craig JC, Fraenkel MB, IDEAL Study. A randomized, controlled trial of early versus late initiation of dialysis. N Engl J Med 2010 Aug 12;363(7):609-619. [doi: 10.1056/NEJMoa1000552] [Medline: 20581422]
18. Razzaghi MR, Afshar L. A conceptual model of physician-patient relationships: a qualitative study. J Med Ethics Hist Med 2016;9:14 [FREE Full text] [Medline: 28050244]
19. Zolnierek KBH, 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]

Abbreviations

BUN: blood urea nitrogen
CCM: Chronic Care Model
CKD: chronic kidney disease
e-community: electronic community
eGFR: estimated glomerular filtration rate
eHealth: electronic health
ESRD: end-stage renal disease
G1-4: groups 1 to 4
ICT: information and communication technology
IDEAL: Initiating Dialysis Early And Late
IRB: Institutional Review Board
KDOQI: Kidney Disease Outcomes Quality Initiative
NTUH: National Taiwan University Hospital
SNS: social networking service
UPCR: urine protein to creatinine ratio

Edited by G Eysenbach; submitted 20.07.19; peer-reviewed by IC Chang, D Schatell, D O'Donoghue, H Barahimi, T Kothari, M Singh; comments to author 21.08.19; revised version received 30.01.20; accepted 22.03.20; published 14.04.20.

Please cite as:
Yang FJ, Hou YH, Chang RE
The Impact of a Social Networking Service–Enhanced Smart Care Model on Stage 5 Chronic Kidney Disease: Quasi-Experimental Study
J Med Internet Res 2020;22(4):e15565
URL: http://www.jmir.org/2020/4/e15565/
doi:10.2196/15565
PMID:32200348

Review

Patients' Use of Social Media for Diabetes Self-Care: Systematic Review

Abdelaziz Elnaggar[1], MBBS, MPH; Van Ta Park[1], PhD, MPH; Sei J Lee[2], MD, MAS; Melinda Bender[3], PhD, PNP-BC; Lee Anne Siegmund[4], PhD, RN, ACSM-CEP; Linda G Park[1], PhD, MS, FNP-BC, FAAN

[1]Department of Community Health Systems, School of Nursing, University of California, San Francisco, San Francisco, CA, United States

[2]Division of Geriatrics, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

[3]Department of Family Health Care Nursing, School of Nursing, University of California, San Francisco, San Francisco, CA, United States

[4]Office of Nursing Research and Innovation, Cleveland Clinic, Cleveland, OH, United States

Corresponding Author:
Linda G Park, PhD, MS, FNP-BC, FAAN
Department of Community Health Systems
School of Nursing
University of California, San Francisco
San Francisco, CA
United States
Phone: 1 415 502 6616
Email: linda.park@ucsf.edu

Abstract

Background: Patient engagement with diabetes self-care is critical to reducing morbidity and mortality. Social media is one form of digital health that is available for diabetes self-care, although its use for peer-to-peer communication has not been systematically described, and its potential to support patient self-care is unclear.

Objective: The primary aim of this systematic review was to describe the use of social media among patients (peer-to-peer) to manage diabetes and cardiovascular disease (CVD). The secondary aim was to assess patients' clinical outcomes, behavioral outcomes, quality of life, and self-efficacy resulting from peer-to-peer social media use.

Methods: We conducted a literature search in the following databases: PubMed, EMBASE, Web of Science, CINAHL, and PsycINFO (January 2008 through April 2019). The inclusion criteria were quantitative studies that included peer-to-peer use of social media for self-care of diabetes mellitus (with all subtypes) and CVD, including stroke.

Results: After an initial yield of 3066 citations, we selected 91 articles for a full-text review and identified 7 papers that met our inclusion criteria. Of these, 4 studies focused on type 1 diabetes, 1 study included both type 1 and 2 diabetes, and 2 studies included multiple chronic conditions (eg, CVD, diabetes, depression, etc). Our search did not yield any individual studies on CVD alone. Among the selected papers, 2 studies used commercial platforms (Facebook and I Seek You), 3 studies used discussion forums developed specifically for each study, and 2 surveyed patients through different platforms or blogs. There was significant heterogeneity in the study designs, methodologies, and outcomes applied, but all studies showed favorable results on either primary or secondary outcomes. The quality of studies was highly variable.

Conclusions: The future landscape of social media use for patient self-care is promising. However, current use is nascent. Our extensive search yielded only 7 studies, all of which included diabetes, indicating the most interest and demand for peer-to-peer interaction on diabetes self-care. Future research is needed to establish efficacy and safety in recommending social media use among peers for diabetes self-care and other conditions.

(J Med Internet Res 2020;22(4):e14209) doi:10.2196/14209

KEYWORDS

social media; diabetes mellitus; peer group; self-care; systematic review

Introduction

Background

Diabetes is one of the most prevalent chronic conditions in the United States and worldwide [1-3], associated with high morbidity and mortality, mainly as a result of complications from cardiovascular disease (CVD) [4-8]. In 2016, the World Health Organization estimated that diabetes was the seventh leading cause of death [9]. Evidence indicates that managing blood glucose and diabetes risk factors (including CVD) can prevent or delay mortality because of CVD by 33% [10-13]. Patient engagement is critical to successfully managing diabetes and thereby reducing morbidity and mortality [14,15].

Self-care has been described as a vital component in diabetes prevention and management in addition to other chronic conditions such as CVD [16-19]. Defined as a "naturalistic decision - making process addressing both the prevention and management of chronic illness" [16], self-care for chronic disease is a complex, multi-factorial endeavor with few effective intervention strategies to help patients manage their conditions [20]. Patients spend very little time each year with their providers and therefore need to independently build skills, knowledge, and motivation to improve individual outcomes. Several meta-analyses and reviews of multiple self-care intervention trials found lifestyle modification programs were more effective than usual care in improving clinical outcomes for diabetes and CVD [21-23].

Despite the known benefits, patients face many barriers in meeting the necessary lifestyle changes involved in self-care, including depression, poor self-efficacy, and cognitive decline [16]. Given the exponential rise in digital technology use among all age groups in the United States [24], mobile technologies are now frequently employed with lifestyle interventions to promote prevention, management, and self-care of chronic diseases [25-27]. Other technology-based programs such as telehealth and home-based rehabilitation have been successful for older patients and reflect their ability to adapt to the use of technology to support their health [28-30].

Peer-to-peer engagement [31], which is communicating with other people experiencing the same chronic condition to learn more about controlling and managing their condition, was found helpful to overcome some of these barriers [16] and has been shown to facilitate self-care, resulting in improved health behaviors [32]. Peer-to-peer communication through engagement on social media offers a convenient venue that is easily accessible for addressing patients' educational needs and providing real-time interaction with others who share many of the challenges in disease management [33]. In a scoping review of social media use between patients and caregivers, researchers found that social media was used to facilitate self-care in 77.1% (219/284) studies identified. Among these studies, the majority of conclusions were positive about social media use [34]. Although younger age and ease with technology use have been shown to affect the likelihood of using social media for disease-related support [35], the number of older adults who engage in social media has continued to climb and offers significant potential to affect self-care [24]. In addition, more capable social media users have recognized the potential for providing support to others who are managing chronic conditions [35].

Innovative strategies and effective interventions are required to improve self-care and health outcomes for patients with diabetes and CVD. A recent systematic review found supplementing usual health care services using social media platforms can satisfy patients' social support needs with managing their CVD, which health providers cannot easily accommodate [36]. Therefore, leveraging social media may be a viable strategy to help improve self-care for diabetes. Understanding how patients use social media to manage their chronic disease is a first step in validating social media platforms as a potentially effective intervention strategy to provide peer-to-peer support and improve diabetes self-care.

Study Aims

The primary aim of this systematic review was to summarize the available evidence on the peer-to-peer use of social media for managing diabetes. A secondary aim was to assess patients' clinical outcomes, behavioral outcomes (ie, self-care and patient activation), quality of life, and self-efficacy resulting from patients' social media use.

Methods

Overview

In this systematic review, we conducted a comprehensive search to capture all of the relevant quantitative studies that were published on the use of a social media platform as a communication tool between patients (peer-to-peer) on health-related topics pertaining to diabetes and CVD self-care. The outcome of interest included any change in clinical outcomes, behavioral outcomes, quality of life, and self-efficacy in participating individuals who used social media for peer-to-peer communication. This systematic review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines [37]. The protocol of this review was registered on the International Prospective Register of Systematic Reviews on November 13, 2018, using the same name as this study's review title.

The inclusion criteria included quantitative studies that addressed the use of social media as a communication tool between patients (ie, not between patients and providers). An 11-year interval (January 2008 to April 2019) was used to search for eligible studies as most studies with social media began in the late 2000s [38]. All US and international studies were included if they were available in the English language. We included studies that provided blogs, chats, and discussion forums from their Web-based platforms, but we excluded studies that were solely Web-based interventions (eg, education-based without interactions between participants). We limited our paper to describe the peer-to-peer use of social media and did not include studies describing the effect of health care provider-to-patient interactions on social media. We also excluded articles that did not mention which disease was studied. We excluded studies that were duplicates, book chapters, systematic or meta-analysis reviews, qualitative studies,

editorials, and meeting abstracts. No studies were excluded on the basis of quality.

A systematic methodology was developed to capture all the relevant data from the selected articles. We ensured our included studies had a clear research question on the basis of population, intervention, comparator, outcomes, and study design criteria (Textbox 1) [39]. This paper presents a narrative synthesis as it was not possible to pool results for a meta-analysis.

Textbox 1. Outline of research questions on the basis of the population, intervention, comparator, outcomes, and study design criteria (PICOS framework).

Population:

- Patients with diabetes

Intervention:

- Use of all social media platforms (eg, discussion forum, blogs, microblogs, and group chatting) for peer-to-peer communication for health-related reasons including support, advice, and education

Comparator:

- Patients receiving the same sort of treatment without social media exposure
- No comparator

Outcome:

- Clinical outcomes (eg, biological measures)
- Behavioral outcomes (eg, self-care and patient activation)
- Quality of life and self-efficacy

Study design:

- Randomized controlled trials
- Cohort
- Cross-sectional

Search Strategy

The search terms were developed on the basis of our research question with the assistance of a health sciences librarian. The selected terms were intended to capture studies that used the most popular social media platforms in all major languages. These terms were adjusted to fit each database to avoid missing any articles (Multimedia Appendix 1). The literature search was conducted in PubMed, EMBASE, Web of Science (including all the databases included in it), CINAHL, and PsycINFO to identify potential articles. We then conducted a manual review of published articles and their bibliographies to assess eligibility for inclusion. In addition, we conducted a hand search of possible relevant articles in the *Journal of Medical Internet Research* and *JMIR Diabetes*.

Study Selection

Initial screening of the studies was done by 2 independent reviewers (AE and MB). Primary screening and data extraction were done using the Cochrane Covidence primary screening and data extraction tool to import all the search results from databases followed by preliminary screening, which included titles and abstracts. If the preliminary screening of the abstract was not conclusive, the full text was screened (AE, MB, and VP). On the basis of the abovementioned criteria, studies were selected for a full-text review, with disagreements resolved by 2 other reviewers (LP and VP) who assessed the eligibility of the studies and approved the final selection of all included studies.

Data Extraction and Analysis

We developed data extraction guidelines. One reviewer (AE) performed data extraction for each eligible article, which was subsequently verified by a second reviewer (MB). The following variables were extracted from the selected studies: name of the first author, year of publication, country, target condition and age of participants, study design and sample size, exposure or intervention, form of social media and purpose, outcome measures, and results. We conducted a descriptive analysis with a summary of the studies.

Results

Study Characteristics

The initial database search applying our terms yielded 3066 citations. After removing duplicates, the remaining 1923 titles and abstracts were screened. On the basis of the inclusion and exclusion criteria, 91 articles were identified as eligible for a full-text review. Of these 91 articles, 84 did not meet the criteria and were eliminated as displayed in Figure 1, leaving 7 studies for inclusion that were related to diabetes and multiple chronic conditions, including CVD. We did not identify any studies focused on CVD alone.

The 7 selected studies for this review included 1 pilot randomized controlled trial (RCT) [40], 1 prospective cohort study [41], 3 cross-sectional studies [42-44], and 2 hybrid cross-sectional/cohort studies [45,46]. Of these, 1 study used

Facebook [46], 1 study used a chat line platform [41], 3 studies used discussion forums that were developed specifically for each study [40,43,45], and 2 studies used surveys to assess the use of social networking sites/blogs [42,44]. As presented in Table 1, a total of 2 studies included patients with multiple chronic diseases (including diabetes) and the other 5 studies focused solely on diabetes—4 studies focused on type 1 diabetes (T1D) [40,41,44,46], whereas 1 study included both T1D and type 2 diabetes [42]. The other studies included all adults, but some did not specify the mean age of those who participated, as shown in Table 1. With regard to the country of origin, 3 studies were conducted in the United States, with 4 out of the 7 studies originating from Israel, Macedonia, and Italy. The studies were published between 2011 and 2019.

Figure 1. Flow chart of study selection process.

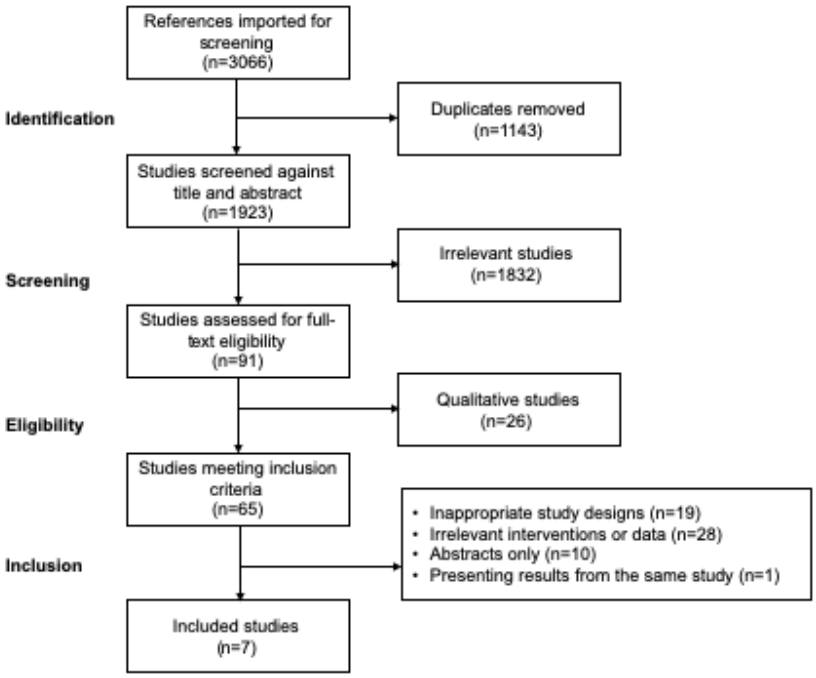


Table 1. Studies on the use of social media among patients for self-care.

References and country	Target condition and age group (years) of participants	Study design and sample size	Exposure/intervention of experimental groups	Form of social media used and purpose	Outcome measures	Results
Grosberg et al [45]; Israel	DM[a], chronic pain, hypertension, and depression (15-≥60)	Cross-sectional and prospective cohort (3 months), N=686	Active participation in a Hebrew-only website designed for chronic conditions	Discussions and blogs: *Camoni* (a Web-based social health network) for advice, consults with experts, and chats with other patients	Personal Involvement in Health Care Related to Site Use, PAM[b]	At baseline, experienced users had higher PAM scores (mean 69.3, SD 19.1; PAM level 4; *P*<.001) than new users (mean 62.8, SD 18.7; PAM level 3). At follow-up, there was a positive correlation between the frequency of visits or time spent and 3 indices of health empowerment (confidence from knowledge acquired about the disease, a sense of shared support, and personal involvement in treatment). PAM scores were higher among experienced users compared with new users (mean 62.8 vs 69.3, respectively; Z=−4.197; *P*<.001)
Iafusco et al [41]; Italy	T1D[c] (10-18), mean age: 13.6 (SD 2.7) chat group, 14.1 (SD 2.3) control	Prospective cohort (2 years), N=396	Online group messaging once a week for 90 min	Group chatting: I Seek You program for educational purposes and social support	DQOLY[d] and HbA_{1c}[e]	The intervention group showed significant improvements in all 3 subscales of DQOLY compared with the control: impact of diabetes (mean 75, SD 7 vs mean 81, SD 14; *P*<.001), worries about diabetes (mean 27, SD 3 vs mean 49, SD 2; *P*=.001), and satisfaction with life (mean 68, SD 13 vs mean 35, SD 13; *P*<.001). No statistically significant difference (*P*=.06) was observed in HbA_{1c} values between the chat and nonchat groups
Magnezi et al [43]; Israel	DM, CVD[f], kidney disease, spinal cord injury, depression/anxiety (20-≥65)	Cross-sectional, N=296	Active participation in a Hebrew-only website designed for chronic conditions	Discussions and blogs: *Camoni* for advice, consults with experts, and chats with other patients	Perceived Usefulness of Online Groups, PAM-13	Perceived usefulness was significantly higher in the 20-29 age group (mean 2.26, SD 1.24) than 50-64 age group (mean 1.43, SD 1.18; *P*=.04) and ≥65 age group (mean 1.38, SD 1.00; *P*<.05). PAM-13 was significantly lower in the 20-29 age group (mean 48.44, SD 21.25) compared with the 30-39 age group (mean 62.28, SD 19.78; *P*=.01) and the 50-64 age group (mean 57.50, SD 17.66; *P*<.05)
Nelakurthi et al [42]; United States	Type 1, type 2, and unspecified type DM, ≥18 (mean age 57, SD 14)	Cross-sectional, N=212	Visiting DM—specific social networking websites	DM—specific social networking websites	Following advice regarding eating habits, exercise habits, and lifestyle changes related to diabetes	Website users showed a significant correlation between offering advice and applying it to their own eating habits (*r*=0.29; *P*=.005), exercise (*r*=0.41; *P*=.001), and lifestyle modification (*r*=0.38; *P*=.001)

References and country	Target condition and age group (years) of participants	Study design and sample size	Exposure/intervention of experimental groups	Form of social media used and purpose	Outcome measures	Results
Newton et al [40]; United States	T1D (13-18)	RCT[g], N=50	Standard medical care plus website participation (7 weeks)	Discussion, blogs, and group chatting: *Diabetes Teen Talk*, to discuss solutions to psychosocial problems that make compliance difficult	DQOLY, Self-Efficacy of Diabetes Self-Management, and Outcome Expectations of Diabetes Self-Management	No significant differences between the control and intervention group on Quality of Life (*P*=.63), Self-Efficacy (*P*=.53), or Negative Outcome Expectation (*P*=.31) scores. Higher positive outcome expectations on treatment conditions was in the control group compared with the intervention group (mean 44.5, SD 6.9, *P*=.03)
Petrovski et al [46]; Macedonia	T1D (11-25), mean age: noninternet group 15.2 (SD 2.9), internet group 16.4 (SD 1.9)	Cross-sectional and retrospective cohort, N=728	Participating members in a national closed Facebook group	Discussion and blogs: Facebook, better blood glucose control, and social support	HbA_{1c} (%), HbA_{1c} (mmol/mol), diabetes ketoacidosis per patient/year, severe hypoglycemia per patient/year, and total daily insulin	Significant differences in the Facebook group between HbA_{1c} (%) and HbA_{1c} (mmol/mol; mean 7.1, SD 3.2 and mean 54, SD 35, respectively) compared with the control (mean 7.6, SD 2.8 and mean 60, SD 31, respectively; *P*<.05 in both). No significant differences in other measures
Oser et al [44]; United States	T1D (≥18)	Cross-sectional, N=282	Only passive readers of T1D-related blogs with no active contribution, insulin pump use, and CGM[h]	T1D-related blogs	HbA_{1c} (%)	HbA_{1c} levels of blog users were significantly lower than nonusers (7.0% vs 7.5%; *P*=.006), blog readers on insulin pump vs blog nonusers and those not on insulin pump (7.0% vs 8.0%), and blog users using CGM vs blog nonusers not using CGM (6.9% vs 7.5%)

[a]DM: diabetes mellitus.

[b]PAM: Patient Activation Measure.

[c]T1D: type 1 diabetes.

[d]DQOLY: Diabetes Quality of Life for Youth Inventory.

[e]HbA_{1c}: glycated hemoglobin.

[f]CVD:cardiovascular disease

[g]RCT: randomized controlled trial.

[h]CGM: continuous glucose monitor.

Form of Social Media Used and Purpose

Out of the 7 studies, 4 used discussion forums or blogs, either through websites developed especially for the targeted population [40,43-45] or through commercially available platforms (eg, Facebook and I Seek You) [41,46]; 2 studies did not use a platform or website but instead evaluated respondents' social networking site behaviors [42,44]; 2 studies used social media as a form of social support [40,46]; 3 studies assessed the usefulness of the platforms for educational purposes [41,43,45]; 2 studies used social media as a tool to improve blood sugar control through educating participants on the technicalities of blood glucose measurement and management, especially for the youth [40,46]; and 1 study assessed the accessibility and usefulness of Web-based medical information [42].

A Description of Social Media Use and Intended Outcomes

Clinical Outcomes

Concerning clinical outcomes, 3 of the 7 studies reviewed reported glycated hemoglobin (HbA_{1c}). Petrovski and Zivkovic [46] and Iafusco et al [41] focused on the adolescent age group, whereas Oser et al [44] targeted adults (≥18 years) with T1D. Petrovski and Zivkovic [46] sought to evaluate a Facebook group as a communication tool to interact with questions, answers, and comments to improve glucose control among adolescents and young people with T1D. Using a retrospective cohort design, Petrovski and Zivkovic [46] reported on data that were collected about Facebook users via electronic medical records and a cross-sectional analysis via social media (both

Facebook and Viber). Patients from the Facebook group had a mean of 1.5 (SD 3.5) posts per day [46]. Among 728 members in their diabetes center, they found significantly lower levels of HbA_{1c} among Facebook group users compared with nonusers 1 year after joining the closed Facebook group (users mean 7.1, SD 3.2; nonusers mean 7.6, SD 2.8; $P<.05$; N=728) [46].

Iafusco et al [41] evaluated the effectiveness of a chat line for T1D education among the youth using a prospective cohort design. In contrast to the study above, Iafusco et al [41] did not find a statistically significant difference in HbA_{1c} levels between 2 groups after adjusting for therapy choice, although the differences approached significance ($P=.05$). HbA_{1c} was assessed on each participant (N=396) at baseline, year 1, and year 2 (N=193) [41]. One important consideration of this study is that children mature physically, mentally, and emotionally over the course of 2 years. It is possible that HbA_{1c} changed similarly for both groups because blood glucose control was an issue of maturity and not necessarily related to the chat line.

Oser et al [44] focused on adults with T1D to assess HbA_{1c} differences between blog readers and blog nonusers [44]. This cross-sectional study also looked at differences in technology use (insulin pump and continuous glucose monitors) in these 2 groups and self-reported HbA_{1c} differences in blog use and technology subgroups [44]. Among 214 blog readers and 68 blog nonusers who completed their survey, the authors found HbA_{1c} was lower for blog readers (7.0%) compared with blog nonusers (7.5%; $P=.006$) [44]. The difference between blog users vs blog nonusers was compared with the clinically significant difference in HbA_{1c} seen among those who used continuous glucose monitors (compared with nonusers) and insulin pump use (compared with multiple daily injections) [44]. These results show that reading and communicating through blogs with other individuals with diabetes leads to learning pertinent information and thereby is associated with lower HbA_{1c} values [44].

Behavioral Outcomes

Magnezi et al [43] and Grosberg et al [45], in 2 separate studies, evaluated *patient activation* (defined as a patient's level of active participation in his or her health care) with chronic care management as a result of using social media. In particular, they examined the use of an online health-related social network called *Camoni*, a platform that was developed for individuals with a variety of chronic diseases to assist them in finding others with similar conditions [43,45]. The website provided advice about their common condition through blogs, discussion forums, online support groups, chats, and a secure channel to communicate with experts. Magnezi et al [43] included individuals with 5 chronic conditions: diabetes mellitus, CVD, renal disease, and depression/anxiety (N=296), whereas Grosberg et al [45] focused on individuals with diabetes, chronic pain, hypertension, and depression (N=696). The purpose of the studies was to evaluate the effects and benefits of participating in an online health-related social network on patient activation and to determine which variables predict the perceived usefulness of the site [43,45]. They found that the usefulness of the website was negatively correlated with age, and it was perceived as being more useful among participants who were less involved in their own care [43]. In addition, the level of activity on the website correlated with the perceived usefulness [43], and those with at least six months experience on the site had the highest patient activation scores (level 4) compared with new visitors ($P<.001$) [45]. There was a significant positive association among experienced users between both the frequency and duration of website visits and self-reported personal empowerment in health [45]. Gender differences were documented as men browsed the website for more than 30 min, whereas the average time for women was 10 to 30 min [45].

Using a cross-sectional study design, 2 separate studies conducted by Nelakurthi and colleagues [42] and Iafusco and colleagues [41] sought to evaluate the reasons behind the use of social networking sites among patients with diabetes and its impact on self-care. Nelakurthi et al [42] used surveys distributed through clinics and websites, whereas Iafusco et al [41] used a chat line moderated by a supervised physician, although it was unclear in the paper by Nelakurthi et al [42] which clinics and health websites were used and accessed by the patients. The top 2 reported reasons for the use of social networking sites were either to offer support or to share personal experiences [42]. Self-reported insulin therapy was significantly higher among users of social media ($P=.01$) [42]. Respondents were more likely to follow the advice received from the website about lifestyle changes and diabetes care compared with advice that was received from their health care provider, 69% and 65% of the time, respectively [42]. However, Iafusco et al [41] revealed that most of the patients thought that sharing HbA_{1c} readings on the group page was motivational for the other members of the group (64%) with the use of both Facebook and Viber.

Quality of Life and Self-Efficacy Factors

We found 2 other themes among 2 of the studies in this review: self-efficacy and quality of life. In a pilot RCT, Newton and Ashley [40] recruited adolescents (13-18 years of age) with T1D to assess the efficacy of a website, *DiabetesTeenTalk.com*, which provided blogs, chat rooms, and discussion forums to improve adherence to treatment protocols. All of the components were designed using Bandura's self-efficacy theory [47]. Although 81 participants were recruited, 59 completed the pretests, and 50 (85%) completed the posttests at 7 weeks [40]. In addition to standard medical care, the experimental group participated in the intervention through logging into the website at least three times weekly over 7 weeks, updated their blogs, and participated in the discussion forums and chats; the control group received standard medical care only [40]. Blinding of subjects was not feasible considering the intervention. However, the assessors of outcomes were not blinded. Differences in characteristics between experimental and control patients were not compared with statistical analyses, although there appeared to be differences in age groups and gender between the intervention and control groups.

Newton and Ashley [40] assessed the effectiveness of the intervention using Diabetes Quality of Life for Youths (DQOLY), Self-Efficacy of Diabetes Self-Management, and Outcome Expectations of Diabetes Self-Management.

Comparatively, Iafusco et al [41] examined DQOLY (N=396) at baseline, year 1, and year 2 (N=193). Newton and Ashley [40] found no significant differences between treatment groups on quality of life scores (P=.63), self-efficacy scores (P=.53), or negative outcome expectations (P=.31). However, the control group had higher positive outcome expectations (mean 48.1, SD 6.3) than those in the experimental group (mean 44.5, SD 6.9; P=.03) [40]. A large majority (78%) of the participants in the intervention group indicated that social support was the most helpful component of the website [40]. Iafusco et al [41] identified significant positive improvements in all subscales of DQOLY in the intervention (chat) group compared with controls who were randomly selected because they refused to participate in chat sessions [41]. At year 2, these included impact of diabetes (chat: mean 75, SD 7; nonchat: mean 81, SD 14; P<.001), worries about diabetes (chat: mean 27, SD 3; nonchat: mean 49, SD 2; P=.001), and satisfaction with life (chat: mean 68, SD 3; nonchat: mean 35, SD 13; P<.001) [41].

Discussion

Principal Findings

To our knowledge, this paper is the first to systematically review the literature for quantitative studies on the use of social media by patients with diabetes to communicate with peers for self-care. We identified 7 studies that examined the use of social media in managing various types of diabetes and reported on participants' change in clinical outcomes, behavioral outcomes, quality of life, and self-efficacy factors as the study outcomes. The studies were diverse, utilizing various social media platforms (eg, discussion forums, blogs, and group chats), research designs and methodologies (eg, RCT, feasibility, prospective and retrospective cohort, and cross-sectional), outcomes (eg, questionnaires and clinical/laboratory measures), and patient populations (eg, adolescents, young adults). Although there is no consensus among experts on the best form of social media platform to connect patients with each other, there is a promising benefit of using Facebook groups, blogs, and mobile phone apps for connecting patients with chronic conditions to their peers.

Both commercially available and customized social media platforms were used by patients in our review. Facebook groups have been found to be a useful tool as they provide a multimodal platform to access content, deliver skills, monitor progress, and organize online and live groups [48,49]. In addition, these groups could be a useful tool for patients and their caregivers to learn about blood glucose devices and receive technological assistance. Through closed private groups, members provided assistance to the community by spreading awareness, technical assistance, and emotional support. Furthermore, members put a high level of trust in their peers and followed their advice in many health situations about lifestyle changes for their chronic conditions, although almost all patients reported no harm using Facebook [46,50]. Similarly, establishing online connections with other individuals experiencing a similar chronic condition through blogging was shown to decrease the sense of isolation and increase the sense of purpose. In addition, active engagement in blogs was shown to be associated with a higher sense of self-accountability and provided a greater opportunity for patients to gain knowledge about their conditions [51,52].

Among the studies we included in this review, users' interactions with one another in the platforms were structured by 4 elements: (1) seeking support or encouragement from individuals with similar conditions, (2) seeking information and advice about clinical diabetes care, (3) obtaining advice about lifestyle changes, and (4) providing a sense of companionship [42,53]. Although obtaining information was the primary motive behind using these platforms rather than seeking relationships, several studies demonstrated that social support and motivation were the most helpful components of these platforms. For instance, a few studies demonstrated that most of the patients shared their last HbA_{1c} level with a social media group, which was used as a motivational and supportive tool for other patients [40,42,46]. Similarly, some were motivated to make other contributions in various forms, such as informational, technical, emotional, or financial support [51].

Our findings are consistent with a recent scoping review by Litchman et al [54] who assessed the potential or actual benefits and consequences of using a diabetes online community (DOC) by analyzing different study designs (cohort, cross-sectional, social network analysis, and text mining). They found that DOC use was highly beneficial with minimal risk or negative consequences [54]. Our review updates this earlier review, which analyzed patients' communication with each other by focusing on quantitative studies. In addition, unlike our study that focused on peer-to-peer interactions, previous reviews have reported on studies between patients and health care providers and showed positive outcomes with using social media and improvement in patient care to provide social, emotional, or experiential support in chronic diseases [48,49].

Potential Impact of Social Media in Diabetes

The benefit of peer-to-peer use is that social ties formed on online platforms provide support for self-care activities that can improve an individual's perceived illness experience, a particularly difficult area to address otherwise [55-60]. Social media platforms provide social support with practical options for facilitating self-care and emotional support to those living with chronic conditions [61-63], which is preferred by patients except when information on prescription medications is needed. In addition, there is no liability to the health care provider with peer-to-peer communication. Health care providers need to assess their capacity to monitor and any potential risks before encouraging widespread use of social media as a communication tool for patients and families [46] and include the communication as a part of the patient's health record. The American Association of Diabetes Educators emphasized in their most recent guidelines about the various benefits of online peer support, which included clinical, behavioral, psychosocial, and educational support [64]. This adds to the potential benefit of incorporating social media use for the management of chronic conditions such as diabetes mellitus.

Consideration of Potential Risks

Accuracy and creditability of medical information obtained from social media platforms remains to be one of the primary

concerns to patients. A number of studies have found that DOCs have beneficial effects with minimal risk [50,65-67]. Although there were positive results in this review with social media use overall, one should consider the risks that may emerge from using these platforms. These risks include access to misinformation, difficulty interpreting medical or scientific outcomes for the average reader, threats to individuals' privacy, and distraction by advertisements on the blogs [57,67-69]. There are limited data on the potential negative outcomes resulting from such activities to warn against using social media with chronic conditions. In addition, there are currently no rigorous quantitative or qualitative data to support the use of social media within the domains of diagnosis or education.

Limitations

There are limitations to be considered in our study. A systematic approach was used to select the relevant articles in the literature; however, we were unable to assess the methodological quality across studies because of the various study designs and some studies using a hybrid approach of cross-sectional surveys with cohort studies. A noted limitation is the small number of studies that fit the inclusion criteria of peer-to-peer communication for this systematic review paper. However, this strengthens the argument that many more clinical research opportunities exist in this area. In addition, because an inclusion criterion for this review paper included the specific mention of a chronic condition (ie, CVD, stroke, or diabetes), it is plausible that there may have been papers that were inadvertently excluded that did include these chronic conditions. Although some of the studies did not include a mean age, the majority of participants were adolescents or young adults, thus our conclusions cannot be generalized to older populations. Finally, this review only included studies published in the English language. Therefore, it is possible some relevant studies may have been excluded.

Future Research

Future research opportunities and current gaps have been identified in this review. There is a clear need to conduct more rigorous RCTs on patients using social media to manage their chronic disease through peer-to-peer communication as we only identified 1 pilot study. By providing a strong evidence base for applying social media for self-care, we will be able to determine the efficacy of using these platforms. We must also improve our outreach to diverse populations (ie, age, types of chronic disease, and race/ethnicity) and geographic locations to establish generalizability. Social media interventions need to be tested with the overall goal of engaging patients, caregivers, and providers to improve health and psychosocial outcomes. Given the limited studies that were included in this systematic review paper, some questions require future research: What type of social media platforms are the most effective and feasible? Which is better in the self-care of chronic conditions: commercially available or customized social media platforms? Which populations benefit the most from the use of social media for the self-care of chronic conditions?

Conclusions

This review contributes to our limited understanding of the impact of using contemporary social media platforms as a peer-to-peer communication tool among patients with diabetes to enhance self-care. Findings from this review may serve as a resource for researchers and clinicians to tailor their interventions in the way social media is currently used between patients and/or diversify their social media platforms according to the communities that they serve. There is a paucity of published research on social media use for peer-to-peer communication among patients with diabetes, which provides a ripe opportunity for clinicians and scientists to explore this digital means of communication among patients with chronic diseases. Social media platforms provide a cost-effective tool that may improve patient self-care and knowledge [54], thereby increasing patient activation, improving problem solving, and providing social support.

Acknowledgments

The authors would like to thank Julia von Oppenfeld, BA, for her editorial assistance with this manuscript. Also, Evans Whitaker, MD, MLIS, for his help developing the search terms.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search terms used to generate this review.
[DOCX File , 15 KB - jmir_v22i4e14209_app1.docx]

References

1. Centers for Disease Control and Prevention (CDC). 2017. National Diabetes Statistics Report, 2017 URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2020-02-14]
2. Yach D, Leeder SR, Bell J, Kistnasamy B. Global chronic diseases. Science 2005 Jan 21;307(5708):317. [doi: 10.1126/science.307.5708.317] [Medline: 15661976]

3. Joszt L. The American Journal of Managed Care. 2017 Nov 28. Identifying the Most Prevalent and Costly Chronic Conditions in Medicaid URL: https://www.ajmc.com/newsroom/identifying-the-most-prevalent-and-costly-chronic-conditions-in-medicaid [accessed 2020-02-14]
4. Szuszkiewicz-Garcia MM, Davidson JA. Cardiovascular disease in diabetes mellitus: risk factors and medical therapy. Endocrinol Metab Clin North Am 2014 Mar;43(1):25-40. [doi: 10.1016/j.ecl.2013.09.001] [Medline: 24582090]
5. Yu OH, Suissa S. Identifying causes for excess mortality in patients with diabetes: closer but not there yet. Diabetes Care 2016 Nov;39(11):1851-1853. [doi: 10.2337/dci16-0026] [Medline: 27926885]
6. Bertoni AG, Krop JS, Anderson GF, Brancati FL. Diabetes-related morbidity and mortality in a national sample of US elders. Diabetes Care 2002 Mar;25(3):471-475. [doi: 10.2337/diacare.25.3.471] [Medline: 11874932]
7. Zhou C, Byard RW. An analysis of the morbidity and mortality of diabetes mellitus in a forensic context. J Forensic Sci 2018 Jul;63(4):1149-1154. [doi: 10.1111/1556-4029.13674] [Medline: 29228513]
8. Emerging Risk Factors Collaboration, Sarwar N, Gao P, Seshasai SR, Gobin R, Kaptoge S, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010 Jun 26;375(9733):2215-2222 [FREE Full text] [doi: 10.1016/S0140-6736(10)60484-9] [Medline: 20609967]
9. World Health Organization. 2018 Oct 30. Diabetes URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2003-10-19]
10. Kottke TE, Faith DA, Jordan CO, Pronk NP, Thomas RJ, Capewell S. The comparative effectiveness of heart disease prevention and treatment strategies. Am J Prev Med 2009 Jan;36(1):82-88. [doi: 10.1016/j.amepre.2008.09.010] [Medline: 19095166]
11. Fonseca VA. Defining and characterizing the progression of type 2 diabetes. Diabetes Care 2009 Nov;32(Suppl 2):S151-S156 [FREE Full text] [doi: 10.2337/dc09-S301] [Medline: 19875543]
12. Nesmith JD. Type 2 diabetes mellitus in children and adolescents. Pediatr Rev 2001 May;22(5):147-152. [doi: 10.1542/pir.22-5-147] [Medline: 11331736]
13. DeFronzo RA. Current issues in the treatment of type 2 diabetes. Overview of newer agents: where treatment is going. Am J Med 2010 Mar;123(3 Suppl):S38-S48. [doi: 10.1016/j.amjmed.2009.12.008] [Medline: 20206731]
14. Peters AE, Keeley EC. Patient engagement following acute myocardial infarction and its influence on outcomes. Am J Cardiol 2017 Nov 1;120(9):1467-1471. [doi: 10.1016/j.amjcard.2017.07.037] [Medline: 28843395]
15. Sacks R, Greene J, Hibbard J, Overton V, Parrotta C. Does patient activation predict the course of type 2 diabetes? A longitudinal study. Patient Educ Couns 2017 Jul;100(7):1268-1275. [doi: 10.1016/j.pec.2017.01.014] [Medline: 28159442]
16. Riegel B, Moser DK, Buck HG, Dickson VV, Dunbar SB, Lee CS, American Heart Association Council on Cardiovascular and Stroke Nursing; Council on Peripheral Vascular Disease;Council on Quality of Care and Outcomes Research. Self-care for the prevention and management of cardiovascular disease and stroke: a scientific statement for healthcare professionals from The American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9):pii: e006997 [FREE Full text] [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
17. Beck J, Greenwood DA, Blanton L, Bollinger ST, Butcher MK, Condon JE, 2017 Standards Revision Task Force. 2017 National standards for diabetes self-management education and support. Diabetes Care 2017 Oct;40(10):1409-1419. [doi: 10.2337/dci17-0025] [Medline: 28754780]
18. White N, Manning ML, Brawer R, Plumb J. Using accreditation standards as a framework to evaluate and improve a community-based diabetes self-management education program. Popul Health Manag 2014 Feb;17(1):8-12. [doi: 10.1089/pop.2013.0044] [Medline: 24138340]
19. Wahowiak L. Providing lifelong education and support: updates in the 2017 National Standards for Diabetes Self-Management Education and Support. Diabetes Spectr 2017 Nov;30(4):298-300 [FREE Full text] [doi: 10.2337/ds17-0060] [Medline: 29151720]
20. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
21. Janssen V, de Gucht V, Dusseldorp E, Maes S. Lifestyle modification programmes for patients with coronary heart disease: a systematic review and meta-analysis of randomized controlled trials. Eur J Prev Cardiol 2013 Aug;20(4):620-640. [doi: 10.1177/2047487312462824] [Medline: 23022703]
22. Cardona-Morrell M, Rychetnik L, Morrell SL, Espinel PT, Bauman A. Reduction of diabetes risk in routine clinical practice: are physical activity and nutrition interventions feasible and are the outcomes from reference trials replicable? A systematic review and meta-analysis. BMC Public Health 2010 Oct 29;10:653 [FREE Full text] [doi: 10.1186/1471-2458-10-653] [Medline: 21029469]
23. Alharbi M, Gallagher R, Kirkness A, Sibbritt D, Tofler G. Long-term outcomes from Healthy Eating and Exercise Lifestyle Program for overweight people with heart disease and diabetes. Eur J Cardiovasc Nurs 2016 Feb;15(1):91-99. [doi: 10.1177/1474515114557222] [Medline: 25344059]
24. Pew Research Center. 2019 Jun 12. Mobile Fact Sheet URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2019-07-01]

25. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
26. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
27. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
28. Bernocchi P, Vitacca M, La Rovere MT, Volterrani M, Galli T, Baratti D, et al. Home-based telerehabilitation in older patients with chronic obstructive pulmonary disease and heart failure: a randomised controlled trial. Age Ageing 2018 Jan 1;47(1):82-88. [doi: 10.1093/ageing/afx146] [Medline: 28985325]
29. Inglis SC, Conway A, Cleland JG, Clark RA. Is age a factor in the success or failure of remote monitoring in heart failure? Telemonitoring and structured telephone support in elderly heart failure patients. Eur J Cardiovasc Nurs 2015 Jun;14(3):248-255. [doi: 10.1177/1474515114530611] [Medline: 24681423]
30. Wolkanin-Bartnik J, Pogorzelska H, Bartnik A. Patient education and quality of home-based rehabilitation in patients older than 60 years after acute myocardial infarction. J Cardiopulm Rehabil Prev 2011;31(4):249-253. [doi: 10.1097/HCR.0b013e31821c1391] [Medline: 21623216]
31. Lauckner H, Hutchinson SL. Peer support for people with chronic conditions in rural areas: a scoping review. Rural Remote Health 2016;16(1):3601 [FREE Full text] [Medline: 26943760]
32. 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]
33. Roblin DW. The potential of cellular technology to mediate social networks for support of chronic disease self-management. J Health Commun 2011;16(Suppl 1):59-76. [doi: 10.1080/10810730.2011.596610] [Medline: 21843096]
34. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. BMJ Open 2013 May 9;3(5):pii: e002819 [FREE Full text] [doi: 10.1136/bmjopen-2013-002819] [Medline: 23667163]
35. Partridge SR, Grunseit AC, Gallagher P, Freeman B, O'Hara BJ, Neubeck L, et al. Cardiac patients' experiences and perceptions of social media: mixed-methods study. J Med Internet Res 2017 Sep 15;19(9):e323 [FREE Full text] [doi: 10.2196/jmir.8081] [Medline: 28916507]
36. 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]
37. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
38. Giustini D, Ali SM, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. Online J Public Health Inform 2018;10(2):e215 [FREE Full text] [doi: 10.5210/ojphi.v10i2.8270] [Medline: 30349633]
39. Higgins J, Green S. Cochrane Training. 2011 Mar. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 URL: https://training.cochrane.org/handbook/archive/v5.1/ [accessed 2003-02-19]
40. Newton KT, Ashley A. Pilot study of a web-based intervention for adolescents with type 1 diabetes. J Telemed Telecare 2013 Dec;19(8):443-449. [doi: 10.1177/1357633X13512069] [Medline: 24197399]
41. Iafusco D, Galderisi A, Nocerino I, Cocca A, Zuccotti G, Prisco F, et al. Chat line for adolescents with type 1 diabetes: a useful tool to improve coping with diabetes: a 2-year follow-up study. Diabetes Technol Ther 2011 May;13(5):551-555. [doi: 10.1089/dia.2010.0188] [Medline: 21406010]
42. Nelakurthi AR, Pinto AM, Cook CB, Jones L, Boyle M, Ye J, et al. Should patients with diabetes be encouraged to integrate social media into their care plan? Future Sci OA 2018 Jul;4(7):FSO323 [FREE Full text] [doi: 10.4155/fsoa-2018-0021] [Medline: 30112191]
43. Magnezi R, Bergman YS, Grosberg D. Online activity and participation in treatment affects the perceived efficacy of social health networks among patients with chronic illness. J Med Internet Res 2014 Jan 10;16(1):e12 [FREE Full text] [doi: 10.2196/jmir.2630] [Medline: 24413148]
44. Oser SM, Stuckey HL, Parascando JA, McGinley EL, Berg A, Oser TK. Glycated hemoglobin differences among blog-reading adults with type 1 diabetes compared with those who do not read blogs: cross-sectional study. JMIR Diabetes 2019 Apr 2;4(2):e13634 [FREE Full text] [doi: 10.2196/13634] [Medline: 30938693]
45. Grosberg D, Grinvald H, Reuveni H, Magnezi R. Frequent surfing on social health networks is associated with increased knowledge and patient health activation. J Med Internet Res 2016 Aug 10;18(8):e212 [FREE Full text] [doi: 10.2196/jmir.5832] [Medline: 27511272]

46. Petrovski G, Zivkovic M. Are we ready to treat our diabetes patients using social media? Yes, we are. J Diabetes Sci Technol 2019 Mar;13(2):171-175 [FREE Full text] [doi: 10.1177/1932296818795441] [Medline: 30132685]
47. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
48. 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]
49. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
50. White K, Gebremariam A, Lewis D, Nordgren W, Wedding J, Pasek J, et al. Motivations for participation in an online social media community for diabetes. J Diabetes Sci Technol 2018 May;12(3):712-718 [FREE Full text] [doi: 10.1177/1932296817749611] [Medline: 29310448]
51. Vaala SE, Lee JM, Hood KK, Mulvaney SA. Sharing and helping: predictors of adolescents' willingness to share diabetes personal health information with peers. J Am Med Inform Assoc 2018 Feb 1;25(2):135-141 [FREE Full text] [doi: 10.1093/jamia/ocx051] [Medline: 28575462]
52. Ressler PK, Bradshaw YS, Gualtieri L, Chui KK. Communicating the experience of chronic pain and illness through blogging. J Med Internet Res 2012 Oct 23;14(5):e143 [FREE Full text] [doi: 10.2196/jmir.2002] [Medline: 23092747]
53. Zhang J, Brackbill D, Yang S, Centola D. Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. Prev Med Rep 2015;2:651-657 [FREE Full text] [doi: 10.1016/j.pmedr.2015.08.005] [Medline: 26844132]
54. Litchman ML, Walker HR, Ng AH, Wawrzynski SE, Oser SM, Greenwood DA, et al. State of the science: a scoping review and gap analysis of diabetes online communities. J Diabetes Sci Technol 2019 May;13(3):466-492. [doi: 10.1177/1932296819831042] [Medline: 30854884]
55. 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]
56. Litchman ML, Rothwell E, Edelman LS. The diabetes online community: older adults supporting self-care through peer health. Patient Educ Couns 2018 Mar;101(3):518-523. [doi: 10.1016/j.pec.2017.08.023] [Medline: 28947360]
57. Bissonnette-Maheux V, Provencher V, Lapointe A, Dugrenier M, Dumas A, Pluye P, et al. Exploring women's beliefs and perceptions about healthy eating blogs: a qualitative study. J Med Internet Res 2015 Apr 8;17(4):e87 [FREE Full text] [doi: 10.2196/jmir.3504] [Medline: 25858777]
58. Malik FS, Panlasigui N, Gritton J, Gill H, Yi-Frazier JP, Moreno MA. Adolescent perspectives on the use of social media to support type 1 diabetes management: focus group study. J Med Internet Res 2019 May 30;21(6):e12149 [FREE Full text] [doi: 10.2196/12149] [Medline: 31199310]
59. Litchman ML, Snider C, Edelman LS, Wawrzynski SE, Gee PM. Diabetes online community user perceptions of successful aging with diabetes: analysis of a #DSMA Tweet Chat. JMIR Aging 2018 Jun 22;1(1):e10176 [FREE Full text] [doi: 10.2196/10176] [Medline: 31518231]
60. Gavrila V, Garrity A, Hirschfeld E, Edwards B, Lee JM. Peer support through a diabetes social media community. J Diabetes Sci Technol 2019 May;13(3):493-497 [FREE Full text] [doi: 10.1177/1932296818818828] [Medline: 30600704]
61. 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 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
62. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
63. Nordqvist C, Hanberger L, Timpka T, Nordfeldt S. Health professionals' attitudes towards using a web 2.0 portal for child and adolescent diabetes care: qualitative study. J Med Internet Res 2009 Apr 6;11(2):e12 [FREE Full text] [doi: 10.2196/jmir.1152] [Medline: 19403464]
64. Association of Diabetes Care & Education Specialists. The Importance of Peer Support: Online and In-Person Support for People Living with Diabetes URL: https://www.diabeteseducator.org/living-with-diabetes/Tools-and-Resources/peer-support [accessed 2003-05-19]
65. Kovic I, Lulic I, Brumini G. Examining the medical blogosphere: an online survey of medical bloggers. J Med Internet Res 2008 Sep 23;10(3):e28 [FREE Full text] [doi: 10.2196/jmir.1118] [Medline: 18812312]
66. Litchman ML, Edelman LS. Perceptions of the Diabetes Online Community's Credibility, Social Capital, and Help and Harm: Cross-Sectional Comparison Between Baby Boomers and Younger Adults. JMIR Aging 2019 Sep 26;2(2):e10857 [FREE Full text] [doi: 10.2196/10857] [Medline: 31573907]
67. Ho Y, O'Connor BH, Mulvaney SA. Features of online health communities for adolescents with type 1 diabetes. West J Nurs Res 2014 Oct;36(9):1183-1198 [FREE Full text] [doi: 10.1177/0193945913520414] [Medline: 24473058]

68. Keogh CJ, McHugh SM, Moloney MC, Hannigan A, Healy DA, Burke PE, et al. Assessing the quality of online information for patients with carotid disease. Int J Surg 2014;12(3):205-208 [FREE Full text] [doi: 10.1016/j.ijsu.2013.12.011] [Medline: 24380751]
69. Sharma N, Tridimas A, Fitzsimmons PR. A readability assessment of online stroke information. J Stroke Cerebrovasc Dis 2014 Jul;23(6):1362-1367. [doi: 10.1016/j.jstrokecerebrovasdis.2013.11.017] [Medline: 24389374]

Abbreviations

CVD: cardiovascular disease
CGM: continuous glucose monitor
DOC: diabetes online community
DQOLY: Diabetes Quality of Life for Youth
HbA_{1c}: glycated hemoglobin
PAM: Patient Activation Measure
RCT: randomized controlled trial
T1D: type 1 diabetes

Edited by G Eysenbach; submitted 29.03.19; peer-reviewed by M Litchman, E Da Silva; comments to author 13.08.19; revised version received 08.01.20; accepted 26.01.20; published 24.04.20.

Please cite as:
Elnaggar A, Ta Park V, Lee SJ, Bender M, Siegmund LA, Park LG
Patients' Use of Social Media for Diabetes Self-Care: Systematic Review
J Med Internet Res 2020;22(4):e14209
URL: http://www.jmir.org/2020/4/e14209/
doi:10.2196/14209
PMID:32329745

Original Paper

A Comparison of the Psycholinguistic Styles of Schizophrenia-Related Stigma and Depression-Related Stigma on Social Media: Content Analysis

Ang Li[1,2], PhD; Dongdong Jiao[3], MSc; Xiaoqian Liu[2], PhD; Tingshao Zhu[2], PhD

[1]Department of Psychology, Beijing Forestry University, Beijing, China

[2]Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[3]National Computer System Engineering Research Institute of China, Beijing, China

Corresponding Author:
Tingshao Zhu, PhD
Institute of Psychology
Chinese Academy of Sciences
No 16 Lincui Rd
Chaoyang District
Beijing, 100101
China
Phone: 86 150 1096 5509
Email: tszhu@psych.ac.cn

Abstract

Background: Stigma related to schizophrenia is considered to be the primary focus of antistigma campaigns. Accurate and efficient detection of stigma toward schizophrenia in mass media is essential for the development of targeted antistigma interventions at the population level.

Objective: The purpose of this study was to examine the psycholinguistic characteristics of schizophrenia-related stigma on social media (ie, Sina Weibo, a Chinese microblogging website), and then to explore whether schizophrenia-related stigma can be distinguished from stigma toward other mental illnesses (ie, depression-related stigma) in terms of psycholinguistic style.

Methods: A total of 19,224 schizophrenia- and 15,879 depression-related Weibo posts were collected and analyzed. First, a human-based content analysis was performed on collected posts to determine whether they reflected stigma or not. Second, by using Linguistic Inquiry and Word Count software (Simplified Chinese version), a number of psycholinguistic features were automatically extracted from each post. Third, based on selected key features, four groups of classification models were established for different purposes: (a) differentiating schizophrenia-related stigma from nonstigma, (b) differentiating a certain subcategory of schizophrenia-related stigma from other subcategories, (c) differentiating schizophrenia-related stigma from depression-related stigma, and (d) differentiating a certain subcategory of schizophrenia-related stigma from the corresponding subcategory of depression-related stigma.

Results: In total, 26.22% of schizophrenia-related posts were labeled as stigmatizing posts. The proportion of posts indicating depression-related stigma was significantly lower than that indicating schizophrenia-related stigma (χ^2_1=2484.64, P<.001). The classification performance of the models in the four groups ranged from .71 to .92 (F measure).

Conclusions: The findings of this study have implications for the detection and reduction of stigma toward schizophrenia on social media.

(J Med Internet Res 2020;22(4):e16470) doi:10.2196/16470

KEYWORDS

stigma; schizophrenia; depression; psycholinguistic analysis; social media

Introduction

Stigma is a destructive phenomenon that undermines efforts to improve mental health and well-being in people with mental illness. A high perception of stigma is associated with reduced self-disclosure to psychotherapists and others, leading to delayed treatment. Schizophrenia is among the most stigmatized mental illnesses; it can be considered a paradigm for mental illness [1,2]. Therefore, it is essential to maintain schizophrenia-related stigma as the main focus of antistigma campaigns [3-5].

Enhancing public knowledge and beliefs about mental illnesses is foundational for stigma reduction. Because information distributed through mass media is an important contributor to the dissemination of materials that may increase mental illness stigma [6-11], mass media campaigns, especially social media campaigns, are effective in raising public awareness of mental health literacy. Social media enables users to bring personal experience into the public domain with the potential to influence public perceptions of mental illness [12]. It also allows users to create their own social networks, which can be leveraged to facilitate the acceptance of received knowledge and then accelerate changes in individual attitudes and behaviors [13-15]. However, such campaigns should be carefully planned and target knowledge deficits, suggesting a need for monitoring mental illness–related stigma on social media. To address this concern, a number of content analysis studies have been performed by human coders on social media to examine stigma related to different mental illnesses, including schizophrenia, depression, suicide, eating disorder, and obsessive-compulsive disorder [2,16-19]. However, the sheer volume of information available on social media makes it difficult for human coders to keep track of all information. For example, in China, Sina Weibo (a free social media site that is similar to Twitter) has over 500 million registered users and produces more than 100 million microblogs (Weibo posts) per day. Therefore, in order to develop more effective strategies to challenge stigma, there is a need for accurate and efficient detection of mental illness–related stigma on social media.

The way people use words provides insight into their psychological profiles [20]. By using computerized text analysis tools (eg, Linguistic Inquiry and Word Count [LIWC]), psycholinguistic analysis can be automatically performed on texts to characterize language use patterns in terms of psychologically meaningful categories, making it easier to keep track of social media data. Recent studies have concluded that psycholinguistic analysis methods can be used to discover characteristics of stigmatizing expressions in social media posts (eg, suicide- and depression-related stigma) [21,22]. However, to date, no research has investigated psycholinguistic characteristics of stigma related to schizophrenia. To our knowledge, there is no convincing evidence for the lack of psycholinguistic differences in the expression of stigma across different mental illnesses. Therefore, additional analysis is necessary to examine the ways in which stigma associated with schizophrenia is presented on social media.

To address this concern, this study aims to investigate psycholinguistic characteristics of schizophrenia-related stigma on social media (ie, Sina Weibo), and then attempts to explore whether schizophrenia-related stigma can be distinguished from stigma related to other mental illnesses (ie, depression-related stigma). Stigma associated with depression was selected for the comparison since it is considered to be one of the most stigmatized mental illnesses, particularly among social media users [2,21].

Methods

Overview

This protocol was reviewed and approved by the Institutional Review Board at the Institute of Psychology, Chinese Academy of Sciences. Participant consent was not obtained, as it is not required for analyzing publicly available data [17,18,21,22]. To protect the privacy of participants, personally identifiable information (including usernames, real names, and other personal information) was excluded from data analysis.

The research process included the following three steps: (a) data collection, (b) data preprocessing, and (c) data modeling.

Data Collection

First, a participant pool was created for data collection. Similar to Twitter, Sina Weibo is a free social media site that enables users to communicate and interact with others by posting digital messages (Weibo posts). Although some users opt to privatize their accounts, the majority of Weibo posts are publicly available for viewing and downloading. According to a previous study [23], a total of 1,953,485 Sina Weibo users were identified as potential participants. However, to due restrictions on data access imposed by Sina Weibo, of these potential participants, Weibo posts from only 1.06 million users were available for download.

Second, data from these participants were downloaded to construct a database of Weibo posts. By using an application programming interface (API), on April 2012, Weibo posts were downloaded automatically from 1.06 million users since the beginning of their registration. After that, the database of Weibo posts was maintained and updated regularly, with the latest update occurring in June 2017.

Third, relevant Weibo posts were identified from the database. To identify posts that are highly relevant to the topics of schizophrenia and depression, all downloaded posts should be searched using schizophrenia- and depression-related keywords, respectively. In this study, in order to make sure that identified posts refer to schizophrenia or depression as mental illnesses, two sets of keywords were selected, including “depressive disorder” ( ) and “schizophrenia” ( ). Therefore, posts with such keywords were included for further analysis. A total of 15,879 depression-related posts were obtained from 10,130 distinct users (time period: September 21, 2009, to June 9, 2017); a total of 19,224 schizophrenia-related posts were obtained from 15,676 distinct users (time period: September 24, 2009 to June 11, 2017). More details about the gender and location of these users can be found in Table 1.

Table 1. Demographics of included users.

Characteristic	Users, depression-related posts (n=10,130), n (%)	Users, schizophrenia-related posts (n=15,676), n (%)
Gender		
Male	3208 (31.67)	7152 (45.62)
Female	6457 (63.74)	8425 (53.74)
Not specified	465 (4.59)	99 (0.63)
Location (provinces)		
Anhui	97 (0.96)	187 (1.19)
Aomen	13 (0.13)	24 (0.15)
Beijing	1138 (11.23)	2039 (13.01)
Chongqing	143 (1.41)	205 (1.31)
Fujian	283 (2.79)	458 (2.92)
Gansu	27 (0.27)	66 (0.42)
Guangdong	1732 (17.10)	2789 (17.79)
Guangxi	134 (1.32)	201 (1.28)
Guizhou	62 (0.61)	115 (0.73)
Hainan	31 (0.31)	78 (0.50)
Hebei	122 (1.20)	217 (1.38)
Henan	208 (2.05)	285 (1.82)
Heilongjiang	86 (0.85)	124 (0.79)
Hong Kong	46 (0.45)	121 (0.77)
Hubei	206 (2.03)	323 (2.06)
Hunan	157 (1.55)	241 (1.54)
Inner Mongolia	40 (0.39)	74 (0.47)
Jilin	64 (0.63)	86 (0.55)
Jiangsu	421 (4.16)	680 (4.34)
Jiangxi	68 (0.67)	111 (0.71)
Liaoning	185 (1.83)	273 (1.74)
Ningxia	19 (0.19)	38 (0.24)
Qinghai	6 (0.06)	8 (0.05)
Shandong	752 (7.42)	1146 (7.31)
Shanxi	63 (0.62)	122 (0.78)
Shaanxi	134 (1.32)	235 (1.50)
Shanghai	1073 (10.59)	1524 (9.72)
Sichuan	271 (2.68)	489 (3.12)
Taiwan	18 (0.18)	55 (0.35)
Tianjin	125 (1.23)	180 (1.15)
Tibet	11 (0.11)	16 (0.10)
Xinjiang	49 (0.48)	61 (0.39)
Yunnan	80 (0.79)	161 (1.03)
Zhejiang	555 (5.48)	740 (4.72)
International	634 (6.26)	1112 (7.09)
Not specified	1077 (10.63)	1092 (6.97)

Data Preprocessing

After data collection, data preprocessing was performed on raw data to prepare it for data modeling.

First, to obtain predicted class labels for data modeling, a content analysis was performed on collected Weibo posts to determine whether they reflected stigma or not. The coding framework was developed on the basis of expert consensus and available evidence. Specifically, a researcher reviewed relevant studies [17,24-27], and conducted an inductive analysis of all collected posts to construct an initial framework. After that, two human coders were recruited and received training on the content of the initial framework and gave suggestions for its amendment. Finally, the initial framework was amended accordingly, and the formal framework was established. As a result, the formal framework of schizophrenia-related stigma included 11 subcategories while the formal framework of depression-related stigma included 9 subcategories [21]. There was a high degree of overlap in stigma subcategories between the two frameworks. Specifically, the framework of schizophrenia-related stigma was quite similar to that of depression-related stigma, but two additional subcategories (ie, weird and stupid) were added to make a set of eleven. By using the formal frameworks, two independent human coders were instructed to analyze all Weibo posts with keywords (15,879 depression-related posts and 19,224 schizophrenia-related posts). Levels of consistency of coding were evaluated by computing Cohen kappa coefficients. If cases of inconsistency, input from a third researcher was used to resolve the issue. The coding results were considered as the ground truth for data modeling validation.

Second, to obtain predictors for data modeling, LIWC software (Simplified Chinese version) was used to automatically extract psycholinguistic features from each post. This is a reliable and valid text analysis tool for the automatic estimation of word frequency in multiple psychologically meaningful categories, including linguistic processes (eg, personal pronouns), psychological processes (eg, affective processes), personal concerns (eg, achievement), and spoken categories (eg, assent) [28]. For example, values of features indicating personal pronouns, affective processes, achievement, and assent refer to frequencies of words associated with personal pronouns (eg, them), affective processes (eg, abandon), achievement (eg, hero), and assent (eg, agree), respectively. To remove the effects of keywords on data modeling, keywords, including “depressive disorder” and “schizophrenia”, were deleted from each post before feature extraction. Therefore, a number of psycholinguistic features can be obtained for each post. Finally, standardized values of psycholinguistic features were computed as potential predictors.

Data Modeling

In this study, the Waikato Environment for Knowledge Analysis software (Weka, version 3.8.1) was used to build four groups of classification models.

The first group of classification models was built to differentiate schizophrenia-related stigma from nonstigma. To solve the class imbalance problem, a certain number of posts were randomly selected from the majority class to obtain a well-balanced data set. After that, to improve the performance of data modeling, psycholinguistic features that were valid for differentiating between posts with and without stigma toward schizophrenia were selected as key features. Specifically, for each psycholinguistic feature, an independent sample *t* test was performed to compare values between two groups (schizophrenia-related stigma and nonstigma), and then the effect size value (Cohen *d*, which is one of the most common ways to measure effect size and can be used to indicate the standardized difference between two means) was calculated using the estimated *t* value. In this study, features that were statistically significant at .05 and had a Cohen $d > 0.20$ or < -0.20 were considered as key features. Finally, by using four different algorithms (support vector machine [SVM]; naïve Bayes [NB]; multilayer perceptron neural network [MPNN]; logistic model trees [LMT]), four classification models (SVM, NB, MPNN, and LMT models) were established based on selected key features. Each model was tested by 10-fold cross-validation. Specifically, the data set was randomly divided into ten subgroups with the same sample size. Each subgroup was used to test the model that was built on the other nine subgroups. After 10 rounds of model training, the modeling results were integrated into a final model. The classification performance was evaluated by three indicators: precision (number of true positives / number of instances predicted to be positive), recall (number of true positives / number of positive instances), and F measure (a tradeoff between precision and recall).

The second group of classification models was built to differentiate a certain subcategory of schizophrenia-related stigma from other subcategories. In this study, two major subcategories of stigma related to schizophrenia (unpredictable and dangerous stigma) were examined (unpredictable stigma / other subcategories; dangerous stigma / other subcategories). The balanced classification data sets, key features, and classification models were obtained using the same method outlined above.

The third group of classification models was built to differentiate schizophrenia-related stigma from depression-related stigma. The balanced classification data set, key features, and classification models were obtained using the same method outlined in the section on the first group of classification models.

For the third group of classification models, the reason for good classification performance may be attributed to marked differences in amount and distribution of stigma subcategories between schizophrenia and depression rather than actual existence of differences in psycholinguistic style between schizophrenia- and depression-related stigma. To clarify this issue, the fourth group of classification models was built to differentiate a certain subcategory of schizophrenia-related stigma from the corresponding subcategory of depression-related stigma. To obtain enough data for further analysis in this study, two subcategories of stigma (unpredictable and glorified stigma) were examined (unpredictable stigma related to schizophrenia / unpredictable stigma related to depression; glorified stigma related to schizophrenia / glorified stigma related to depression). The balanced classification data sets, key features, and classification models were obtained using the same method outlined in the section on the first group of classification models.

Results

Coding

The coding results for stigma related to schizophrenia are shown in Table 2 (see Multimedia Appendix 1 for posts in Chinese). The Cohen kappa coefficients for schizophrenia-related stigma and its subcategories reached .77 and .78, respectively, reflecting a satisfying level of agreement [29]. Of all schizophrenia-related posts, 26.22% (5041/19,224) were labeled as stigmatizing posts. Of these posts, 41.14% (n=2074) and 26.86% (n=1354) reflected the view that "people with schizophrenia are unpredictable" (unpredictable stigma) and "people with schizophrenia are dangerous" (dangerous stigma), respectively.

Table 2. Coding framework for schizophrenia-related stigma.

Subcategory	Definition	Representative Weibo post (English translation)	Posts, n (%)
It is best to avoid people with schizophrenia so that you do not become vulnerable to schizophrenia	Beliefs that schizophrenia is an infectious disease	"It is said that schizophrenia is highly infectious..."	26 (0.52)
Schizophrenia is a sign of personal weakness	Beliefs that people with schizophrenia show a lack of strength and cannot sustain pressure	"...A mentally weak person is so vulnerable to schizophrenia..."	39 (0.77)
People with schizophrenia are dangerous	Beliefs that people with schizophrenia are likely to cause harm or injury	"When you need to talk to a person with schizophrenia, it is very important to pay attention to your safety!"	1354 (26.86)
People with schizophrenia are unpredictable	Beliefs that people with schizophrenia behave in a way that cannot be not easily predicted	"People with schizophrenia can suddenly turn crying into laughing (still with snots and tears of sadness)..."	2074 (41.14)
Schizophrenia is not a real medical illness	Beliefs that schizophrenia is a made up, rather than a medical disease	"Schizophrenia is really just an excuse to get a lesser sentence or to get out of prison"	21 (0.42)
People with schizophrenia could snap out of it if they wanted	Beliefs that people with schizophrenia can recover from their illness at will	"As you pray to the Buddha, your schizophrenia will be recovered soon..."	13 (0.26)
People would not tell anyone if they had schizophrenia	Beliefs that people should be ashamed of their own schizophrenia	"...Schizophrenia is God's punishment for family sins..."	115 (2.28)
People with schizophrenia are glorified	Beliefs that schizophrenia is a sign of noble souls or a quality of being graceful	"I think those with schizophrenia are charming..."	281 (5.57)
People with schizophrenia are self-centered	Beliefs that people with schizophrenia only think of their own advantage	"...Schizophrenia is the same as narcissism."	48 (0.95)
People with schizophrenia are weird	Beliefs that people with schizophrenia behave in an unsettling way that is strikingly odd or unusual	"...Someone is throwing money like paper towels. Could this person have schizophrenia?"	929 (18.43)
People with schizophrenia are stupid	Beliefs that people with schizophrenia are silly or unwise	"The schizophrenia patients, the idiot people"	141 (2.80)

The coding results for stigma associated with depression showed that 6.09% (967/15,879) of depression-related posts were labeled as stigmatizing posts. Further details on coding results can be found in a previous study [21].

The proportion of posts indicating depression-related stigma (967/15,879) was significantly lower than that indicating schizophrenia-related stigma (5041/19,224) (χ^2_1=2484.64, P<.001).

Differentiating Schizophrenia-Related Stigma From Nonstigma

A total of 13 key features were selected for data modeling (Multimedia Appendix 2). Within each predicted class (schizophrenia-related stigma and nonstigma), there existed 5041 posts. The LMT model had the best classification performance (precision=.89, recall=.89, F measure=.89) (Table 3).

Table 3. Performance of classification models.

Models	Stigma / nonstigma	Unpredictable / other subcategories	Dangerous / other subcategories	Depression / schizophrenia	Unpredictable (depression) / unpredictable (schizophrenia)	Glorified (depression) / glorified (schizophrenia)
Support vector machine						
Precision	.70	.72	.76	.80	.93	.65
Recall	.70	.72	.76	.79	.92	.65
F measure	.70	.72	.76	.79	.92	.65
Naïve Bayes						
Precision	.65	.67	.69	.67	.88	.67
Recall	.60	.64	.66	.65	.88	.65
F measure	.57	.62	.64	.64	.88	.64
Multilayer perceptron neural network						
Precision	.67	.69	.71	.75	.88	.71
Recall	.67	.69	.71	.75	.88	.71
F measure	.67	.69	.71	.75	.88	.71
Logistic model trees						
Precision	.89	.70	.75	.78	.91	.65
Recall	.89	.70	.75	.77	.91	.65
F measure	.89	.70	.75	.77	.91	.65

Differentiating a Certain Subcategory of Schizophrenia-Related Stigma From Other Subcategories

A total of 28 and 27 key features were selected for building two subgroups of classification models respectively: one for differentiating between unpredictable stigma and other subcategories, and one for differentiating between dangerous stigma and other subcategories (Multimedia Appendix 2). For the first subgroup, within each predicted class (unpredictable stigma and other subcategories), there existed 2074 posts, while for the second subgroup, within each predicted class (dangerous stigma and other subcategories), there existed 1354 posts. For both of the two subgroups, the SVM model had the best classification performance (subgroup 1: precision=.72, recall=.72, F measure=.72; subgroup 2: precision=.76, recall=.76, F measure=.76) (Table 3).

Differentiating Schizophrenia-Related Stigma From Depression-Related Stigma

A total of 30 key features were selected for data modeling (Multimedia Appendix 2). Within each predicted class (depression- and schizophrenia-related stigma), there existed 967 posts. The SVM model had the best classification performance (precision=.80, recall=.79, F measure=.79) (Table 3).

Differentiating a Certain Subcategory of Schizophrenia-Related Stigma From the Corresponding Subcategory of Depression-Related Stigma

A total of 52 and 19 key features were selected for building two subgroups of classification models, respectively: one for differentiating the expression of unpredictable stigma between depression and schizophrenia, and one for differentiating the expression of glorified stigma between depression and schizophrenia (Multimedia Appendix 2). For the first subgroup, within each predicted class (unpredictable stigma related to depression and that related to schizophrenia), there existed 380 posts, while for the second subgroup, within each predicted class (glorified stigma related to depression and that related to schizophrenia), there existed 114 posts. Furthermore, for the first subgroup, the SVM model had the best classification performance (precision=.93, recall=.92, F measure=.92), while for the second subgroup, the MPNN model had the best classification performance (precision=.71, recall=.71, F measure=.71) (Table 3).

Discussion

Principal Findings

According to our knowledge, this is the first study to investigate the psycholinguistic characteristics of schizophrenia-related stigma on social media and characterize psycholinguistic differences in the expression of stigma between two different mental illnesses (depression and schizophrenia). The findings of this study have a number of implications for the detection and reduction of stigma associated with schizophrenia on social media.

First, it is necessary for campaigns to reduce stigma associated with schizophrenia on social media. Results showed that schizophrenia-related stigma was prevalent on Chinese social media. According to this study, 26.22% of relevant Weibo posts indicated stigmatizing attitudes toward schizophrenia, which is higher than that reported by previous studies on Twitter (5%-9.7%) [17,30]. Such inconsistency may be partly due to the role of anonymity in increasing the likelihood of users to express stigmatizing attitudes on Sina Weibo (an anonymous social media site) rather than on Twitter (an open public forum), and partly due to a low level of schizophrenia literacy among people in China [31-33]. In addition, stigmatization of schizophrenia was significantly higher than stigmatization of depression (χ^2_1=2484.64, P<.001), suggesting the essential role of schizophrenia-related stigma in antistigma campaigns on social media. Moreover, people with schizophrenia were more frequently perceived as unpredictable and dangerous (41.14% and 26.86%), which should be the targets of stigma reduction campaigns. It is worth noting that these results are largely consistent with relevant studies [2,24,26,34-38], supporting the use of social media to monitor mental illness–related stigma.

Second, the use of psycholinguistic analysis methods facilitates automatic detection of stigma toward schizophrenia on social media. Results showed that, by using psycholinguistic analysis methods, the best performance for detecting schizophrenia-related stigma and its subcategories ranged from .72 to .89 (F measure). Compared with the results of other studies (F measure=.66 to .86) [21,22], this performance is satisfactory. It is worth noting that, in this study, no single algorithm achieved the best performance in all six classification tasks (Table 3). Although it is still unclear why the performance of algorithms varied considerably across tasks, the results of this study imply that it is necessary to select appropriate algorithms for solving different tasks.

More importantly, the use of psycholinguistic analysis methods may provide insight into the ways in which schizophrenia-related stigma is presented on social media. Multimedia Appendix 2 showed that the expression of schizophrenia-related stigma was associated with increased use of words related to social processes (eg, mate), humans (eg, adult), death (eg, kill), and anger (eg, hate). Such language use patterns indicate a preference for more social comparisons and a higher level of negative emotion, which may fit into two elements of stigma processes, including cognitive separation (comparisons between people with and without a stigmatizing label) and emotional reactions (negative emotional reactions to people with a stigmatizing label) [39]. In addition, there existed psycholinguistic differences between subcategories of schizophrenia-related stigma as well. For example, compared with other subcategories, the expression of unpredictable stigma was associated with more frequent use of words related to cognitive processes (eg, ought) and personal pronouns (eg, them), which may be due to sustained confusion about patient behavior and the misbelief that people with schizophrenia have multiple personalities. Unlike unpredictable stigma, the expression of dangerous stigma was associated with an increased use in words related to death (eg, kill) and health (eg, clinic), which may be because of the expectation that patients with schizophrenia are likely to cause harm or injury.

Third, the development of an accurate tool for measuring stigma should be disease-specific. Results showed that stigma associated with schizophrenia can be distinguished from depression-related stigma in terms of psycholinguistic style. These significant differences existed not only at the level of general stigma (depression- or schizophrenia-related stigma as a whole), but also at the level of stigma subcategories (eg, unpredictable and glorified stigma) (Multimedia Appendix 2). The classification performance of the corresponding models ranged from .71 to .92 (F measure) (Table 3). These results can be explained by the fact that the reason for good classification performance may not be solely attributed to differences in amount and distribution of stigma subcategories between schizophrenia and depression, but also to the actual existence of differences in psycholinguistic style between stigma related to schizophrenia and depression. Therefore, in order to improve stigma detection performance, it is necessary to develop disease-specific measurement tools.

Limitations

There are a number of limitations. First, it is unknown whether all relevant posts can be searched by keywords that were used in this study. Therefore, it is unclear whether any additional language use patterns are associated with stigmatizing expressions. In addition, it is uncertain whether the current keywords selection may bias estimations of the number of stigmatizing posts. Second, social media users are not representative of the general population. For example, in China, Sina Weibo users are more likely to be aged between 20 to 29 years old, well educated, and located in urban areas [40]. Therefore, the findings of this study might have limited generalizability. Third, in this study, all analyzed social media posts were written in Chinese. Therefore, it is uncertain whether the findings of this study will be applicable to other languages. Fourth, although all established models were validated using the method of cross-validation, they should be further tested on other data sets in the future.

Conclusions

In this study, a nonintrusive method was used to collect and analyze data under nonexperimental conditions. As a result, the current research should have high ecological validity and could investigate actual attitudes toward people with schizophrenia. The results of this study may facilitate the automatic detection of stigma on social media and improve social media campaigns related to stigma reduction.

Acknowledgments

The authors thank RY Li for inspiring this study. This study was supported by the National Social Science Fund of China (grant number: 16AZD058) and the National Natural Science Foundation of China (grant number: 31700984). The funders played no role in this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coding framework for schizophrenia-related stigma.
[PDF File (Adobe PDF File), 91 KB - jmir_v22i4e16470_app1.pdf]

Multimedia Appendix 2
Key features selected for data modeling.
[PDF File (Adobe PDF File), 240 KB - jmir_v22i4e16470_app2.pdf]

References

1. Stuart H. Fighting the stigma caused by mental disorders: past perspectives, present activities, and future directions. World Psychiatry 2008 Oct;7(3):185-188 [FREE Full text] [Medline: 18836546]
2. 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. [doi: 10.1007/s00127-018-1571-5] [Medline: 30069754]
3. Sartorius N, Schulze H. Reducing the Stigma of Mental Illness: A Report from a Global Association. Cambridge, England: Cambridge University Press; 2005:152-171.
4. Gaebel W, Zäske H, Baumann AE, Klosterkötter J, Maier W, Decker P, et al. Evaluation of the German WPA "Program against stigma and discrimination because of schizophrenia — Open the Doors": results from representative telephone surveys before and after three years of antistigma interventions. Schizophr Res 2008 Jan;98(1-3):184-193. [doi: 10.1016/j.schres.2007.09.013] [Medline: 17961985]
5. Razali SM. Prevention of relapse and interventions for enhancing medication adherence in schizophrenia: an East Asian perspective. Asia Pac Psychiatry 2010 May;2(2):68-76. [doi: 10.1111/j.1758-5872.2010.00067.x]
6. Corrigan PW, Watson AC, Gracia G, Slopen N, Rasinski K, Hall LL. Newspaper stories as measures of structural stigma. Psychiatr Serv 2005 May;56(5):551-556. [doi: 10.1176/appi.ps.56.5.551] [Medline: 15872163]
7. Dietrich S, Heider D, Matschinger H, Angermeyer MC. Influence of newspaper reporting on adolescents' attitudes toward people with mental illness. Soc Psychiatry Psychiatr Epidemiol 2006 Apr;41(4):318-322. [doi: 10.1007/s00127-005-0026-y] [Medline: 16424966]
8. Morgan AJ, Jorm AF. Recall of news stories about mental illness by Australian youth: associations with help-seeking attitudes and stigma. Aust N Z J Psychiatry 2009 Sep;43(9):866-872. [doi: 10.1080/00048670903107567] [Medline: 19670060]
9. McGinty EE, Webster DW, Jarlenski M, Barry CL. News media framing of serious mental illness and gun violence in the United States, 1997-2012. Am J Public Health 2014 Mar;104(3):406-413. [doi: 10.2105/AJPH.2013.301557] [Medline: 24432874]
10. Niederkrotenthaler T, Reidenberg DJ, Till B, Gould MS. Increasing help-seeking and referrals for individuals at risk for suicide by decreasing stigma: the role of mass media. Am J Prev Med 2014 Sep;47(3 Suppl 2):S235-S243. [doi: 10.1016/j.amepre.2014.06.010] [Medline: 25145745]
11. Koike S, Yamaguchi S, Ojio Y, Ohta K, Ando S. Effect of name change of schizophrenia on mass media between 1985 and 2013 in Japan: a text data mining analysis. Schizophr Bull 2016 May;42(3):552-559 [FREE Full text] [doi: 10.1093/schbul/sbv159] [Medline: 26614786]
12. Betton V, Borschmann R, Docherty M, Coleman S, Brown M, Henderson C. The role of social media in reducing stigma and discrimination. Br J Psychiatry 2015 Jun;206(6):443-444 [FREE Full text] [doi: 10.1192/bjp.bp.114.152835] [Medline: 26034176]
13. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
14. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, et al. A 61-million-person experiment in social influence and political mobilization. Nature 2012 Sep;489(7415):295-298 [FREE Full text] [doi: 10.1038/nature11421] [Medline: 22972300]
15. Valente TW. Network interventions. Science 2012 Jul;337(6090):49-53. [doi: 10.1126/science.1217330] [Medline: 22767921]

16. Fu K, Cheng Q, Wong PWC, Yip PSF. Responses to a self-presented suicide attempt in social media: a social network analysis. Crisis 2013 Jan;34(6):406-412 [FREE Full text] [doi: 10.1027/0227-5910/a000221] [Medline: 23871954]
17. Reavley NJ, Pilkington PD. Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. PeerJ 2014 Oct;2:e647 [FREE Full text] [doi: 10.7717/peerj.647] [Medline: 25374786]
18. Li A, Huang X, Hao B, O'Dea B, Christensen H, Zhu T. Attitudes towards suicide attempts broadcast on social media: an exploratory study of Chinese microblogs. PeerJ 2015 Sep;3:e1209 [FREE Full text] [doi: 10.7717/peerj.1209] [Medline: 26380801]
19. O'Dea B, Achilles MR, Larsen ME, Batterham PJ, Calear AL, Christensen H. The rate of reply and nature of responses to suicide-related posts on Twitter. Internet Interv 2018 Sep;13:105-107 [FREE Full text] [doi: 10.1016/j.invent.2018.07.004] [Medline: 30206524]
20. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
21. Li A, Jiao D, Zhu T. Detecting depression stigma on social media: a linguistic analysis. J Affect Disord 2018 May;232:358-362. [doi: 10.1016/j.jad.2018.02.087] [Medline: 29510353]
22. Li A, Huang X, Jiao D, O'Dea B, Zhu T, Christensen H. An analysis of stigma and suicide literacy in responses to suicides broadcast on social media. Asia Pac Psychiatry 2018 Mar;10(1):e12314. [doi: 10.1111/appy.12314] [Medline: 29383880]
23. Li L, Li A, Hao B, Guan Z, Zhu T. Predicting active users' personality based on micro-blogging behaviors. PLoS One 2014 Jan;9(1):e84997 [FREE Full text] [doi: 10.1371/journal.pone.0084997] [Medline: 24465462]
24. Griffiths KM, Nakane Y, Christensen H, Yoshioka K, Jorm AF, Nakane H. Stigma in response to mental disorders: a comparison of Australia and Japan. BMC Psychiatry 2006 May;6:21 [FREE Full text] [doi: 10.1186/1471-244X-6-21] [Medline: 16716231]
25. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: randomised controlled trial. Br J Psychiatry 2004 Oct;185(4):342-349 [FREE Full text] [doi: 10.1192/bjp.185.4.342] [Medline: 15458995]
26. Wood L, Birtel M, Alsawy S, Pyle M, Morrison A. Public perceptions of stigma towards people with schizophrenia, depression, and anxiety. Psychiatry Res 2014 Dec;220(1-2):604-608. [doi: 10.1016/j.psychres.2014.07.012] [Medline: 25064387]
27. Chen M, Lawrie S. Newspaper depictions of mental and physical health. BJPsych Bull 2017 Dec;41(6):308-313 [FREE Full text] [doi: 10.1192/pb.bp.116.054775] [Medline: 29234506]
28. Zhao N, Jiao D, Bai S, Zhu T. Evaluating the validity of Simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services. PLoS One 2016 Jun;11(6):e0157947 [FREE Full text] [doi: 10.1371/journal.pone.0157947] [Medline: 27322382]
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
30. Passerello GL, Hazelwood JE, Lawrie S. Using Twitter to assess attitudes to schizophrenia and psychosis. BJPsych Bull 2019 Aug;43(4):158-166 [FREE Full text] [doi: 10.1192/bjb.2018.115] [Medline: 30784393]
31. Wong DFK, He X. Schizophrenia literacy among Chinese in Shanghai, China: a comparison with Chinese-speaking Australians in Melbourne and Chinese in Hong Kong. Aust N Z J Psychiatry 2011 Jul;45(7):524-531. [doi: 10.3109/00048674.2011.585604] [Medline: 21718122]
32. Gong AT, Furnham A. Mental health literacy: public knowledge and beliefs about mental disorders in mainland China. Psych J 2014 Jun;3(2):144-158. [doi: 10.1002/pchj.55] [Medline: 26271766]
33. Chen S, Wu Q, Qi C, Deng H, Wang X, He H, et al. Mental health literacy about schizophrenia and depression: a survey among Chinese caregivers of patients with mental disorder. BMC Psychiatry 2017 Mar;17:89 [FREE Full text] [doi: 10.1186/s12888-017-1245-y] [Medline: 28274209]
34. Angermeyer MC, Matschinger H. Public beliefs about schizophrenia and depression: similarities and differences. Soc Psychiatry Psychiatr Epidemiol 2003 Sep;38(9):526-534. [doi: 10.1007/s00127-003-0676-6] [Medline: 14504738]
35. Marie D, Miles B. Social distance and perceived dangerousness across four diagnostic categories of mental disorder. Aust N Z J Psychiatry 2008 Feb;42(2):126-133. [doi: 10.1080/00048670701787545] [Medline: 18197507]
36. Björkman T, Angelman T, Jönsson M. Attitudes towards people with mental illness: a cross-sectional study among nursing staff in psychiatric and somatic care. Scand J Caring Sci 2008 Jun;22(2):170-177. [doi: 10.1111/j.1471-6712.2007.00509.x] [Medline: 18489686]
37. Angermeyer MC, Millier A, Rémuzat C, Refaï T, Toumi M. Attitudes and beliefs of the French public about schizophrenia and major depression: results from a vignette-based population survey. BMC Psychiatry 2013 Nov;13:313 [FREE Full text] [doi: 10.1186/1471-244X-13-313] [Medline: 24252540]
38. Yilmaz E, Okanli A. The effect of internalized stigma on the adherence to treatment in patients with schizophrenia. Arch Psychiatr Nurs 2015 Oct;29(5):297-301. [doi: 10.1016/j.apnu.2015.05.006] [Medline: 26397432]
39. Link BG, Yang LH, Phelan JC, Collins PY. Measuring mental illness stigma. Schizophr Bull 2004 Jan;30(3):511-541. [doi: 10.1093/oxfordjournals.schbul.a007098] [Medline: 15631243]

40. China Internet Network Information Center. 2016 research report on behaviors of social media users in China URL: http://www.cnnic.cn/hlwfzyj/hlwxzbg/sqbg/201712/P020180103485975797840.pdf [accessed 2019-09-26]

Abbreviations

API: application programming interface
LIWC: Linguistic Inquiry and Word Count
SVM: support vector machine
NB: naïve Bayes
MPNN: multilayer perceptron neural network
LMT: logistic model trees

Edited by G Eysenbach; submitted 02.10.19; peer-reviewed by S Lawrie, Y Zhang; comments to author 09.11.19; revised version received 13.11.19; accepted 07.02.20; published 21.04.20.

Please cite as:
Li A, Jiao D, Liu X, Zhu T
A Comparison of the Psycholinguistic Styles of Schizophrenia-Related Stigma and Depression-Related Stigma on Social Media: Content Analysis
J Med Internet Res 2020;22(4):e16470
URL: http://www.jmir.org/2020/4/e16470/
doi:10.2196/16470
PMID:32314969

Original Paper

Mediation Effect of Suicide-Related Social Media Use Behaviors on the Association Between Suicidal Ideation and Suicide Attempt: Cross-Sectional Questionnaire Study

Xingyun Liu[1,2,3], MA; Jiasheng Huang[3], MA; Nancy Xiaonan Yu[3], PhD; 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, Beiijng, 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 15010965509
Email: tszhu@psych.ac.cn

Abstract

Background: A limited number of studies have examined the differences in suicide-related social media use behaviors between suicide ideators and suicide attempters or have sought to elucidate how these social media usage behaviors contributed to the transition from suicidal ideation to suicide attempt.

Objective: Suicide attempts can be acquired through suicide-related social media use behaviors. This study aimed to propose 3 suicide-related social media use behaviors (ie, attending to suicide information, commenting on or reposting suicide information, or talking about suicide) based on social cognitive theory, which proposes that successive processes governing behavior transition include attentional, retention, production, and motivational processes.

Methods: We aimed to examine the mediating role of suicide-related social media use behaviors in Chinese social media users with suicidal risks. A sample of 569 Chinese social media users with suicidal ideation completed measures on suicidal ideation, suicide attempt, and suicide-related social media use behaviors.

Results: The results demonstrated that suicide attempters showed a significantly higher level of suicidal ideation ($t_{563.64}$=5.04; P<.001; two-tailed) and more suicide-related social media use behaviors, which included attending to suicide information (t_{567}=1.94; P=.05; two-tailed), commenting on or reposting suicide information (t_{567}=2.12; P=.03; two-tailed), or talking about suicide ($t_{542.22}$=5.12; P<.001; two-tailed). Suicidal ideation also affected suicide attempts through the mediational chains.

Conclusions: Our findings thus support the social cognitive theory, and there are implications for population-based suicide prevention that can be achieved by identifying behavioral signals.

(J Med Internet Res 2020;22(4):e14940) doi:10.2196/14940

KEYWORDS

suicidal ideation; suicide; attempted; social media; suicide-related social media use behaviors

Introduction

Background

Suicide has been a critical public health problem, with approximately 1 million people committing suicide worldwide annually [1]. Although significant efforts have been made in the field of suicide prevention, suicide rates have seen an increase in recent decades [2]. In fact, it was reported that one-third of people with suicidal ideation would take actions to kill themselves [3]. Although a large amount of evidence has been reported in relation to risk factors (eg, depression, alcohol

use disorders, and hopelessness) that can be associated with suicidal ideation, little is known regarding how suicide ideators become suicide attempters [4]. As such, to facilitate precise suicide risk assessment and prevention, it is crucial to identify the mechanism that illustrates how suicide ideators progress into suicide attempters.

The ideation-to-action framework states that the development of suicidal ideation and the progression from ideation to suicide attempt are two distinct processes [5]. Pain and hopelessness account for suicidal ideation, whereas suicide capabilities transform suicidal ideation into suicide attempt. Suicide capabilities include acquired (eg, habituation to fear and pain involved in death), dispositional (eg, genetic fearlessness and pain tolerance), and practical (eg, access to means and knowledge of attempted lethality) capabilities that may serve in attempting suicide. Indeed, existing research about the progression from ideation to suicide attempt has mainly focused on disorders such as posttraumatic stress disorder, depression disorder, and anxiety disorder [4]. To elucidate how suicide capabilities can turn suicidal ideation into suicide attempt, studies investigating observable behaviors as warning signals that foreshadow how suicidal ideation turns into suicide attempt are warranted, and such information would enhance population-based suicide screening and improve the efficiency of suicide prevention [6].

Today, the internet can provide new data sources because people record their lives in varying degrees on the website, and it has become an active interaction platform for young people, where they sometimes exchange thoughts portraying a susceptibility to suicidality [7]. Live streams of suicides are found on social media [8], and people often use such media to share their distress [7]. Indeed, young people are reluctant to disclose their suicidal signs in medical visits, but they are somehow willing to share these on social media platforms such as Facebook and Twitter [9]. This aforementioned research indicates the possibility of using web-based behaviors in social media as helpful markers for suicide risk assessment. Although web suicidal behaviors may be a proxy for suicidal ideation, the web behavioral pattern of suicide attempters is not clear [10]. Researchers found that factors such as high levels of internet use, internet addiction, and exposure to websites with self-harm or suicidal content were particularly associated with suicidal behavior [11]. However, we are still unclear as to how web social media use behaviors contribute to the progression of suicidal ideation to suicide attempt [10,12].

People like to create and exchange user-generated content on social media [13]. In Sina Weibo (similar to Twitter), people can post microblogs (ie, tweets); can search, read, comment on, and repost others' microblogs; and would also be able to join in the groups they are interested in [14]. O'Connor and Kirtley found that exposure to suicide or suicidal behavior can turn suicide ideators into suicide attempters [15]. As such, social media platforms can function like a hatchery for suicidal behaviors in certain ways. Indeed, social cognitive theory suggests that people can acquire new behaviors through observing others. The theory proposes that successive processes governing observational learning include attentional, retention, production, and motivational processes [16]. Thus, suicidal behaviors can be seen as modeled behaviors, which are acquired through observing others' suicidal behaviors on social media. Consequently, the social cognitive theory has been used to examine the effect of mass media on suicide [16,17]. Fu et al [17] found a positive association between media influences and suicidal ideation based on social cognitive theory. However, they only took the internet as a traditional form of media, akin to television, and neglected the interactions that people made on the website. Therefore, we extended the existing study by using the social cognitive theory in this analysis to illustrate the effects of suicide-related social media use behaviors on the process from suicidal ideation to suicide attempt. On the basis of previous research [14], we mainly focused on 3 suicide-related social media use behaviors: (1) attending to suicide information (*attending to*), (2) commenting on or reposting suicide information (*commenting-reposting*), or (3) talking about suicide (*talking about*).

Attentional process is the first step of observational learning. The attentional process revolves around what people selectively pay attention to about the observed models and what message they procure from a wealth of information. Preconceptions, value preferences, and other factors determine this process [16]. People with suicidal ideation would have preconceptions of suicide, and they may inadvertently come across or actively seek suicidal posts on the website because social networking sites are important sources of suicide stories [18]. It has been confirmed that suicidal ideation was found to be significantly associated with the accessing of suicide or self-injury information on the website [19]. The reciprocal relationship between suicidal ideation and suicide-related information is thus an upward spiral feedback loop [17], particularly when the personalized recommender system works. Suicidal posts often include evocative words and images, which have been proven to be a risk factor for suicide behaviors [20]. Considerable evidence has indicated that exposure to stories of others' suicides can influence young people at suicide risk to attempt the same act [17,18]. Therefore, *attending to* is the first suicide-related social media use behavior we suggested that would be related to suicide attempts.

The processes for observational learning are retention and production. Retention involves the process of remembering the modeled activities. This process would involve not only simply copying but also proactively reconstructing the observed events. Production would be the deep processing of retention when learned behavior is generated through a "conception-matching process" (ie, through a production process, cognitive activities are developed into corresponding explicit behaviors) [16]. These two processes are closely linked. As we focus on warning behaviors in this study, we joined these two processes together. When compared against traditional media, social networking sites offer additional opportunities for transmission of suicide portrayals and knowledge of suicide among friends [18], which means that the internet offers a large number of opportunities to retain and produce suicide information the users can attend to through functions such as commenting or reposting. Through commenting or reposting, people with suicidal ideation may process suicide information more deeply than they would by just attending to suicide information. The suicide news would

become part of their memory, and they would put themselves into the suicide stories they saw, imagining how the stories could happen to them. This process may then accelerate the progression from suicidal ideation to suicide attempt. Adolescents use public websites to display comments about their attitude toward these suicide stories or their own suicidal ideation or suicide plan, actions that have been confirmed to have a close relation to suicide attempts [15]. Therefore, *commenting-reposting* is the second suicide-related social media use behavior we suggested that would be related to suicide attempts.

Finally, the motivational process is the final stage to make the decision on whether one would act out on the acquired modeled behaviors. Social cognitive theory differentiates internal acquisition from external behaviors because people would not show all the things that they have ever learned. They would have a higher possibility of performing the modeled behaviors that they felt were similar to their situation, and they value the outcomes [16]. By attending to suicide-related information on social media, suicidal social media users may adore what the models did to end a miserable life by committing suicide, particularly when this practice is shared by similar people. They may participate in online suicide groups to discuss with like-minded individuals about successful suicide cases, general issues associated with suicide, or their own suicide plan [21,22]. Moreover, they could directly talk about their own substantial suicide plan on the website publicly. Those explicit expressions of suicidal behaviors can not only send out signals that the subjects have a strong mind for suicide but also obtain feedback from others, who may in turn encourage them to commit suicide because the interaction can diminish the fear about death through further mental imagery or the normalizing, reinforcing, and even glorification of suicide [23]. In particular, the interactions may foster peer pressure to die by suicide or facilitate suicide pacts [22]. Such pacts are not rare globally in today's society because they improve the possibility of successful death by suicide [24-26]. It has been demonstrated that talking about suicidal ideation on Twitter can be significantly associated with having a suicide plan and attempting suicide [20,27]. Therefore, *talking about* is the final suicide-related social media use behavior we proposed that would be related to suicide attempt.

Objectives

In summary, the mechanism of the progression from suicide ideators to suicide attempters among social media users is unclear. On the basis of earlier research [14], we focused on 3 suicide-related social media use behaviors in sequence (ie, *attending to*, *commenting-reposting*, and *talking about*) according to social cognitive theory [16], as behavioral warning signals that may serve to illuminate the transmitting path.

There are 2 hypotheses for suicidal social media users:

H1: Compared with suicide ideators, suicide attempters would report a higher level of suicide-related social media use behaviors, including *attending to*, *commenting-reposting*, and *talking about.*

H2: Suicidal ideation would predict suicide attempt through the mediating chains of *attending to*, *commenting-reposting*, and *talking about*.

Methods

Participants

As the most popular social media platform in China, Sina Weibo has nearly 300 million users, and most of them are younger than 30 years [28,29], which makes it a perfect platform to prevent Chinese youth from committing suicide. We found a blogger on Weibo who committed suicide because of depression, and many people follow her suicide note even though she died 7 years ago. Until October 1, 2019, over 1 million comments were posted on her web suicide note. This unique post has become a "secret garden" to attract many people who suffer from depression or suicidal ideation. We manually annotated the comments for this post from March 2016 to June 2016, and 4616 social media users were identified as suicidal. The reason why we chose those comments is that it is nearly impossible to search suicidal posts randomly from Weibo, as, on average, 100 million new posts are generated every day on this platform [28]. We took this web suicide note as a breakthrough to trace suicidal social media users.

We sent a direct message that provided social support, referrals, and a link to questionnaires to these 4616 suicidal users. More details can be found in the study by Liu et al [30]. We reported the results according to the Checklist for Reporting Results of Internet E-Surveys [31]. A total of 725 individuals completed the questionnaires after signing informed consent, with no compensation being provided. Ethical approval was obtained from the Institute of Psychology, Chinese Academy of Sciences.

Measures

The demographic information of participants was collected, including their sex, age, education level, marital status, and living status (living alone, with family or partner, with friends, or with others).

Suicidal ideation was measured by the 4-item version of the Adult Suicidal Ideation Questionnaire (ASIQ) [32]. Each item is rated on a 7-point scale rating from 0 (I never have this thought) to 6 (almost everyday life) to measure the severity of suicidal ideation. A sample item was "I thought about killing myself." We used the total average score of the 4 items to represent suicidal ideation as a continuous variable. Both the original [32] and Chinese version [33] of the scale displayed good psychometric properties. The internal consistency, indicating whether items measure a unidimensional concept and whether a value higher than 0.70 is desirable [34], in this study was high (alpha=.91).

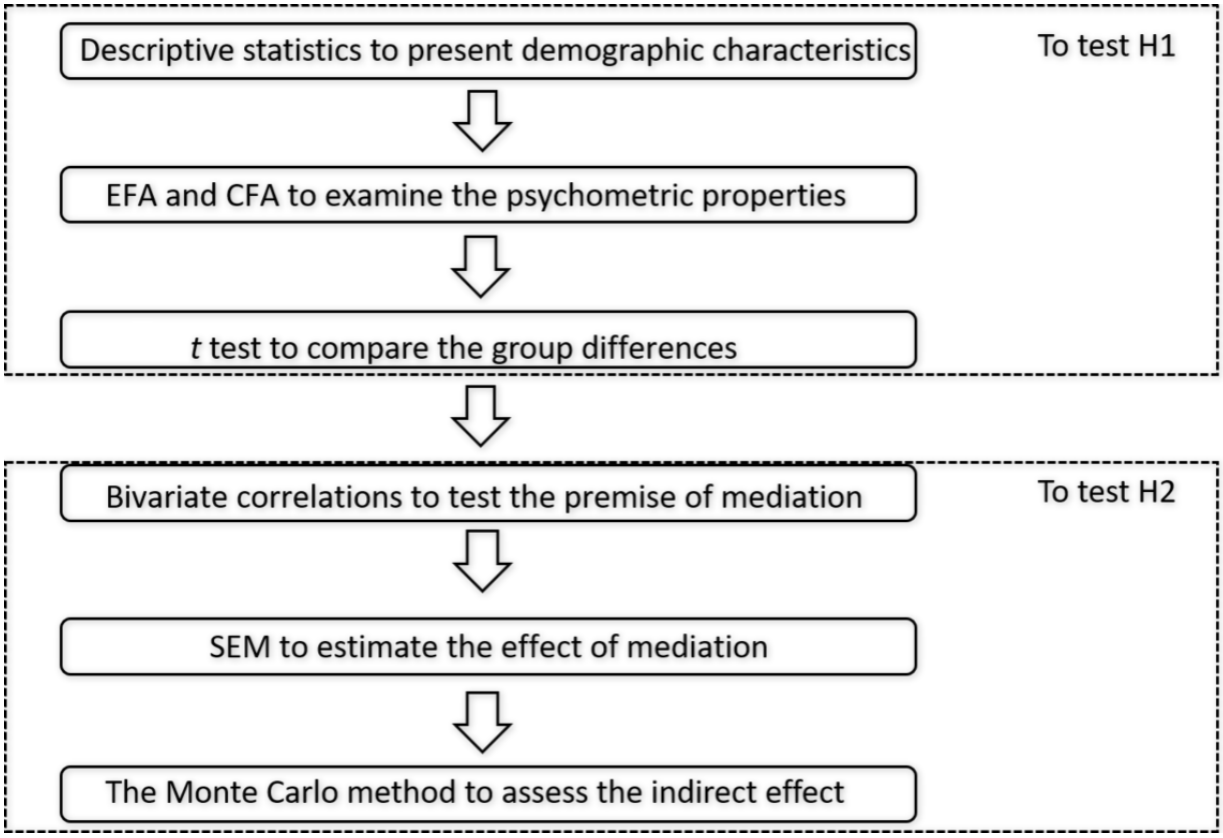
Suicide attempt was measured by one item "Have you ever tried to kill yourself?" Participants responded with binary choices (yes or no). As suicide attempt is a binary dependent variable, participants responded no (0) or yes (1) for suicide attempt. Participants who responded no to the suicide attempt item were grouped as suicide ideators, and those who responded yes to the suicide attempt item were categorized as suicide attempters [35].

Suicide-related social media use behaviors were developed by our team. This measure consists of *attending to* (2 items: "Attending to suicide news" and "Attending to friends who said they wanted to commit suicide"), *commenting-reposting* (2 items: "Commenting on or reposting suicide news" and "Commenting on or reposting other people's posts about killing themselves"), and *talking about* (2 items: "Talking about suicide in online suicide communities" and "Talking on the website about one's own concrete plan to commit suicide"). Participants indicated how often it had occurred in the last year based on a 5-point Likert scale that ranged from 1 (never) to 5 (mostly). The internal consistency values of the 3 subscales were 0.64, 0.74, and 0.72, respectively, in this study. The measurement items can be found in Multimedia Appendix 1.

Data Analysis

Mplus7 (Muthen and Muthen, Beijing, China) and SPSS21 (SPSS Inc, Beijing, China) were used for the statistical analysis. Descriptive statistics for demographic characteristics and study variables were also tabulated. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted to examine the psychometric properties of the self-developed measure of suicide-related social media use behaviors. *t* test was used to compare the differences between suicide ideators and suicide attempters. In addition, bivariate correlations were assessed. To evaluate the mediation model, structural equation modeling (SEM) was used to estimate latent variables because the latent variables can decrease measurement error and consider measurement and equation modeling at the same time [36]. Model fit was evaluated using several fit indices: (1) the chi-square (χ^2) test of model fit, (2) the root mean square error of approximation (RMSEA), (3) the comparative fit index (CFI), (4) the Tucker Lewis Index (TLI), and (5) the weighted root mean square (WRMR) because of the presence of binary data; that is, participants received 0 or 1 for suicide attempt. An RMSEA value of less than 0.06, a CFI at or above 0.90, a TLI at 0.90 or higher, and a WRMR value less than 1.00 indicate a relatively good fit [37,38]. Finally, the Monte Carlo method was used to further test the indirect effect in the mediation. Rather than testing new models with different constellations of the listed variables, the Monte Carlo method showed the estimated sizes of different indirect effects in the original model, so the model fit statistics were the same as those in the original model [39] (Figure 1).

Figure 1. The flowchart for data analyses. CFA: confirmatory factor analysis; EFA: exploratory factor analysis; SEM: structural equation modeling; H1: hypothesis 1; H2: hypothesis 2.



Results

Descriptive Statistics Results

After excluding respondents that reported conflicting or missing data (146/569, 20.1%) and no suicidal ideation (the ASIQ score=0; 10/569, 1.4%), there were 569 valid participants who reported a certain degree of suicidal ideation. This number exceeds the required sample size to test the study question, as the minimum sample size is 295 (power=0.80; alpha=.05; H0: RMSEA=0, H1: RMSEA=0.05; and df=36) [40]. The demographic characteristics of participants are displayed in Table 1. We grouped the participants into suicide ideators and suicide attempters according to their reports as to whether they had previously attempted suicide. The mean ages for all participants, suicide ideators, and suicide attempters were 21.94

(SD 3.31) years, 22.07 (SD 3.59) years, and 21.82 (SD 3.02) years, respectively. The age difference between the 2 groups was not significant ($t_{540.19}$=0.90, *P*=.37, two-tailed). As shown, there was no difference in demographic variables between the 2 groups. Most participants were unmarried young females with college degrees, and most of them were living with others (friends or families).

Table 1. Demographic characteristics of participants by suicidal status (N=569).

Characteristic	Total (N=569), n (%)	Suicide ideators (n=277), n (%)	Suicide attempters (n=292), n (%)	Chi-square value (*df*)	*P* value
Sex				**0.25 (1)**	**.62**
Male	78 (13.7)	40 (14.4)	38 (13.0)		
Female	491 (86.3)	237 (85.6)	254 (87.0)		
Educational level				**2.46 (2)**	**.29**
Primary	28 (4.9)	11 (4.0)	17 (5.8)		
Secondary	97 (17.1)	53 (19.1)	44 (15.1)		
Tertiary	444 (78.0)	213 (76.9)	231 (79.1)		
Marital status				**1.43 (2)**	**.49**
Single	531 (93.3)	256 (92.4)	275 (94.2)		
Married	24 (4.2)	12 (4.3)	12 (4.1)		
Divorced or others	14 (2.5)	9 (3.3)	5 (1.7)		
Living status				**4.95 (4)**	**.29**
With families	201 (35.3)	90 (32.5)	111 (38.0)		
With partner	33 (5.8)	16 (5.8)	17 (5.8)		
With friends	231 (40.6)	115 (41.5)	116 (39.7)		
Alone	75 (13.2)	44 (15.9)	31 (10.6)		
Others	29 (5.1)	12 (4.3)	17 (5.8)		

Results to Test Hypothesis 1

To examine its psychometric properties, we split the sample (n=569) into two random and equal halves. We randomly used half of the sample to conduct EFA and the other half to conduct CFA. See Table 2 for the results of the EFA; the factor number was set to 3. The 3 factors, in total, explained 77.4% of the total variance. The construct validity was supported by the EFA results.

The CFA model showed a good fit to the data: χ^2_6=2.1 (*P*=.09), RMSEA=0.06, CFI=0.99, TLI=0.97, and SRMR=0.02. Both the EFA and CFA results indicated that the scale had good psychometric properties.

On the basis of whether the participants attempted suicide or not, they were grouped as suicide ideators or suicide attempters as mentioned above. The group differences on study variables by suicide status are shown in Table 3. The attempters group showed a significantly higher level of suicidal ideation ($t_{563.64}$=5.04 *P*<.001; two-tailed) than that shown by the ideators group. Consistent with hypothesis 1, the attempters reported more *attending to* (t_{567}=1.94; *P*=.05; two-tailed), *commenting-reposting* (t_{567}=2.12; *P*=.03; two-tailed), and *talking about* ($t_{542.22}$=5.12; *P*<.001; two-tailed) behaviors than those reported by suicide ideators.

Descriptive statistics and correlations of study variables are shown in Table 4. As shown, suicidal ideation (*r*=0.21; *P*<.001) was significantly correlated with suicide attempt. The 3 suicide-related social media use behaviors (ie, *attending to:r*=0.08, *P*=.05; *commenting-reposting:r*=0.09, *P*=.03; and *talking about:r*=0.21, *P*<.001) were significantly correlated with suicide attempt. Therefore, we conducted subsequent mediation analysis with path analysis.

Table 2. Exploratory factor analysis of 6-item suicide-related social media use behaviors.

Items	Attending to[a]	Commenting-reposting[b]	Talking about[c]
Attending to suicide news	0.88	N/A[d]	N/A
Attending to friends who said they wanted to commit suicide	0.74	N/A	N/A
Commenting on or reposting suicide news	—[e]	0.71	N/A
Commenting on or reposting other people's posts about killing themselves	N/A	0.90	N/A
Talking about suicide in online suicide communities	N/A	N/A	0.85
Talking on the website about one's own concrete plan to commit suicide	N/A	N/A	0.82

[a]Attending to suicide information.

[b]Commenting on or reposting suicide information.

[c]Talking about suicide.

[d]N/A: not applicable.

[e]Loading <0.04 was not shown in the table.

Table 3. Group differences on study variables by suicidal status (N=569).

Variables	Suicide ideators (n=277), mean (SD)	Suicide attempters (n=292), mean (SD)	*t* test (*df*)	*P* value	Effect size
Suicidal ideation	2.34 (1.69)	3.10 (1.92)	−5.04 (563.64)	<.001	0.42
Attending to[a]	2.86 (0.77)	3.00 (0.91)	−1.94 (567)	.05	0.17
Commenting-reposting[b]	2.41 (0.95)	2.59 (1.04)	−2.12[e] (567)	.03	0.18
Talking about[c]	1.70 (0.88)	2.14 (1.15)	−5.12[a] (542.22)	<.001	0.43

[a]Attending to suicide information.

[b]Commenting on or reposting suicide information.

[c]Talking about suicide.

Table 4. Descriptive statistics and bivariate correlations of the study variables.

Variables	Correlation coefficient (*r*)				Value, mean (SD)
	Suicidal ideation	Suicide attempt	Attending to[a]	Commenting-reposting[b]	
Suicidal ideation					**2.73 (1.85)**
r	N/A[c]	N/A	N/A	N/A	
P value	N/A	N/A	N/A	N/A	
Suicide attempt					**292 (51.3)**
r	0.21	N/A	N/A	N/A	
P value	<.001				
Attending to					**2.93 (0.85)**
R	0.31	0.08	N/A	N/A	
P value	<.001	.05			
Commenting-reposting					**2.50 (1.00)**
R	0.30	0.09	0.52	N/A	
P value	<.001	.03	<.001		
Talking about[d]					**4.73 (1.42)**
R	0.46	0.21	0.46	0.56	
P value	<.001	<.001	<.001	<.001	

[a]Attending to suicide information.

[b]Commenting on or reposting suicide information.

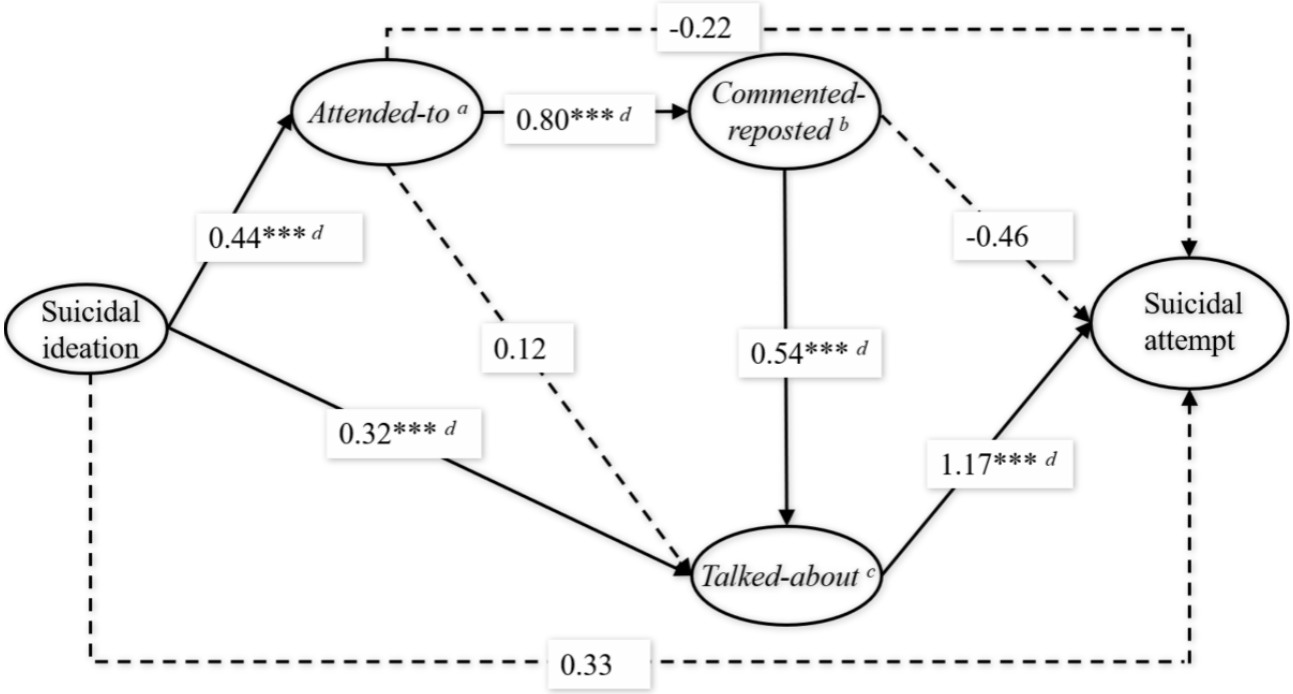
[c]N/A: not applicable.

[d]Talking about suicide.

Results to Test Hypothesis 2

Results of SEM are depicted in Figure 2. Model indices were acceptable, with χ^2_{36}=2.1 (*P*<.001), RMSEA=0.05, CFI=0.94, TLI=0.90, and WRMR=0.48. When considering the *attending to*, *commenting reposting*, and *talking about* variables, the association between suicidal ideation and suicide attempt was found to be no longer significant (β=.33; *P*=.14). Suicidal ideation was associated with a higher level of *attending to* (β=.44; *P*<.001) and *talking about* (β=.32; *P*<.001) behaviors. Only the *talking about* behavior (β=1.17; *P*<.001) was associated with suicide attempt. The relation between *attending to* and *talking about* behaviors was mediated by the *commenting-reposting* behavior. These associations supported hypothesis 2, which states that suicidal ideation predicted suicide attempts through the mediating chains of *attending to*, *commenting-reposting*, and *talking about* behaviors.

Figure 2. The mediating effect of social media use behaviors in the association between suicidal ideation and suicide attempt. a) Attended to suicide information; b) Commented on or reposted suicide information; c) Talked about suicide; d) *P*<.001.



Further mediation analysis with the Monte Carlo method found significant indirect effects of (1) *attending to*, *commenting-reposting*, and *talking about* behaviors as the mediators between suicidal ideation and suicide attempt and (2) *attending to* and *commenting-reposting* behaviors as the mediators between suicidal ideation and the *talking about* behavior. In addition, *attending to* and *commenting-reposting* behaviors did not show significant mediation effects in the association between suicidal ideation and suicide attempt. The estimated mediation effects are shown in Table 5, which provide support for hypothesis 2.

Table 5. Estimated mediation effects of attending to, commenting-reposting, and talking about behaviors in the association between suicidal ideation and suicide attempt.

Mediation effect	Indirect effect	95% CI	*P* value
Suicidal ideation→*Attending to*[a]→*Commenting-reposting*[b]→*Talking about*[c]	0.19	0.11 to 0.27	<.001
Suicidal ideation→*Attending to*[a]→Suicide attempt	−0.10	−0.50 to 0.31	.64
Suicidal ideation→*Attending to*[a]→*Commenting-reposting*[b]→Suicide attempt	−0.16	−0.48 to 0.16	.32
Suicidal ideation→*Attending to*[a]→*Commenting-reposting*[b]→*Talking about*[c]→Suicide attempt	0.22	0.09 to 0.36	.001

[a]Attending to suicide information.

[b]Commenting on or reposting suicide information.

[c]Talking about suicide.

Discussion

Principal Findings

On the basis of the social cognitive theory, this study aimed to identify web-based behavioral markers in social media that could distinguish suicide ideators from suicide attempters and illuminate the transition mechanisms that led from suicidal ideation to suicide attempt. Our findings demonstrated that when compared with Chinese social media users who were suicide ideators, suicide attempters reported more suicide-related social media use behaviors. More importantly, the effect of suicidal ideation on suicide attempt was mediated by suicide-related social media use behaviors. Our findings, therefore, make a theoretical contribution to the field by providing behavioral markers in the progression from suicidal ideation to suicide attempt with the Chinese social media user population that has suicidal ideation. To our knowledge, this is the first study to focus on the identification of suicide-related social media use behaviors and illustration of the progression from suicidal ideation to suicide attempt.

Consistent with previous studies [4], no difference in sex, marital status, or educational level between suicide ideators and suicide attempters among Chinese social media users was found. Our findings also indicated that demographic characteristics are not helpful in distinguishing suicide ideators from suicide attempters

in Chinese social media users, and there would be a need to use behavioral or other markers to differentiate between these 2 groups [6]. Interestingly, suicide attempters demonstrated more suicide-related social media active use behaviors than suicide ideators. Specifically, although they showed a marginal significance in the *attending to* behavior, they did report significantly more *commenting-reposting* and *talking about* behaviors as compared with suicide ideators. Our findings are also consistent with a Japanese study, which demonstrated that talking about one's own suicidal ideation on Twitter was positively associated with suicide attempt [27]. Moreover, previous evidence has indicated that social media may exacerbate suicidal behaviors because social media users could receive descriptions of how to kill oneself in pro-suicide groups [22]. Owing to the fact that the behavioral patterns among social media users that contribute to the transition from suicidal ideation to suicide attempt are unclear, more research is required to investigate how these 2 groups are different in terms of *attending to*, *commenting-reposting*, *talking about*, and other suicide-related social media use behaviors (eg, logging in to suicide-related blogs or clicking "like" for suicidal posts). Our findings indicate that behavioral risks of suicide attempters will pave the way for the identification of high-risk cases based on the web-based surveillance of social media use behaviors. For example, a few useful programs have been using a similar approach to prevent suicide online, such as "Psymap" in China [41] and Facebook's suicide-prevention tools [42]. These programs are helpful in detecting suicidal social media users through their suicide-related social media use behaviors, such as commenting on suicide notes. These existing programs have demonstrated feasibility and effectiveness in identifying suicidal cases with an attempt at monitoring their social media use behaviors. Our findings also provide empirical evidence as a rationale behind conducting these preventive measures. That being said, our study calls for more big-data studies on behavioral patterns of social media use among suicide attempters with interdisciplinary research involving psychology, psychiatry, computer science, and linguistics.

Interestingly, our findings demonstrated that suicidal ideation predicted suicide attempt through the mediating chains of social media use behaviors, which included the *attending to*, *commenting-reposting*, and *talking about* behaviors. Our results indicate that suicide-related social media use behaviors are a social learning process, which turns suicide ideators into suicide attempters through the process of attention (what people selectively notice, eg, attending to suicide information), retention, and production (what people constructively remember and what people externally acquire, eg, commenting on or reposting these posts) before finally moving to motivation (what people perform into action, eg, talking about their own suicide plan), as depicted in the social cognitive theory. Furthermore, our study demonstrated that the associations of *attending to* and *commenting-reposting* behaviors with suicide attempt were mediated by the *talking about* behavior. *Talking about* behavior made the biggest contribution in the mediating role in predicting suicide attempt. This result indicates that an individual may have a higher chance of committing suicide when he or she directly talks about suicide on the website compared with those who attend to suicide news and simply comment on or repost suicide information. This finding supports the social cognitive theory, which states that the attentional process, as well as the retention and production processes, contribute to behavioral acquisition, whereas the motivational process determines behavioral performance. From the attentional process to the motivational process, more processing and input are involved [16]. As the motivational process, talking about suicide, which requires an initiative with explicit expression, may involve more input and deeper processing than the *attending to* behavior, which is implicit, and the *commenting-reposting* behavior, which is responsive, do [17]. Our finding is consistent with previous results, which state that talking about suicide on the website is one of the most important warning signals for suicide [43]. Moreover, researchers have pointed out that suicide is only possible insofar as it is meaningful [44]. By talking about suicide on the website, suicidal people may complete the process of meaning-making, which leads to a higher possibility of suicide attempt. It is also worth noting that the association between suicidal ideation and the *talking about* behavior is significant. One possible explanation is that by being different from traditional media, such as newspaper and television [17], the internet provides a unique platform for direct interaction, thereby accelerating the social cognitive process to skip the beginning components (eg, attention, retention, and production) and reach the terminal component (ie, motivation).

Our mediation findings also provide insight into explaining how suicide capacity is acquired in social interaction online, potentially complementing the ideation-to-action framework [4]. This theory states that suicide capability is acquired through a habituation to fear and pain involved in death, access to means of committing suicide, knowledge of attempting deadliness, and so on [5]. Reading, commenting on, and reposting suicidal posts on the website, which often include evocative words and images [20], may also enhance one's suicide capability by eliminating the fear of death and increasing the knowledge of how to die. The "upward spiral" feedback loop between suicide-related information and suicidal ideation [17] would be boosted in the online interaction, and suicide capability is, therefore, enhanced in a web-based context. Moreover, talking about one's own suicidal ideation in social media may aggravate the situation because it may attract similar minds to attempt suicide together or normalize, reinforce, and even glorify suicide [23]. Hence, suicide capability may efficiently turn into action. Our study expands the ideation-to-action framework to a web-based situation by providing empirical evidence to confirm that *attending to*, *commenting-reposting*, and *talking about* behaviors are alarm signals that indicate the suicide ideators' intent to commit suicide.

Limitations

Although the study has made several contributions to the existing research, there are several limitations to this work. First, participants in this study were mainly suicidal, unmarried young females with college degrees. Our results are consistent with previous studies showing that single young females with a higher education were more inclined to talk about their suicidal ideation and seek help [30,45]. At the same time, because all those participants were recruited from one single social media blogging site, there is a possibility that they may differ from

other suicidal social media users who do not comment on this post. For example, they may have a higher level of suicidal ideation or use Weibo more frequently. In addition, they may be affected by other users who also left comments on this post while they interact through comments [46]. Caution should be taken in generalizing our findings to a larger sample size in other blogging sites, social media platforms, and social media users with different demographic characteristics and suicidal status. Moreover, although the suicide-related social media use behaviors examined in this study (ie, *attending to*, *commenting-reposting*, and *talking about*) seem to occur widely, these results need to be replicated in different cultures. Second, a previous study reported that there are potential risks of online suicide intervention programs because of the possibility of increasing contagion among the participants [47]. In addition, the frequency of social media use was positively associated with suicide risk [46]. We only examined 3 suicide-related social media use behaviors in this study. Future research should consider investigating other suicide-related social media use behaviors, such as logging in to suicide blogs, following suicide-provoking or suicide-preventing public accounts, and clicking "like" for suicidal posts. Third, even though Weibo is popular in China, there are still some drawbacks; for example, the accuracy of a web-based suicide claim needs to be further confirmed [48]. Chinese people even use various metaphors to describe death [49]. These phenomena make identification more difficult. In addition, not like Facebook or Twitter, in which the users are globally distributed and use different languages, Weibo bloggers are mainly Chinese and most use Chinese. Moreover, Sina Weibo's message limit to 140 characters may impede expression. Future study may pay attention to these language limitations to understand their impacts. We also suggest a combination of online logs and offline data such as school performance or health profile to identify suicide. Fourth, because of the cross-sectional study design, the associations presented in this study are correlational in nature. The effect of suicide-related social media use behaviors found in this study would need to be examined with longitudinal designs over various time frames (minutes, hours, days, weeks, and even months) [50]. Finally, this work relied on self-report data of suicidal issues, and this method is subject to potential biases. Future studies can consider adopting more objective indicators such as public log data because social media use behaviors can be recorded and observed from one's account history. Of course, ethics, real-world implementation, and legal liability are complex issues that cannot be overlooked [51].

Conclusions

This study aimed to investigate the behavioral markers available for distinguishing suicide ideators from suicide attempters and elucidate the behavioral process that transits from suicidal ideation to suicide attempt. Our findings demonstrated that suicide attempters showed a significantly higher level of suicidal ideation and more suicide-related social media use behaviors, including *attending to*, *commenting-reposting*, and *talking about* behaviors. Moreover, *attending to*, *commenting-reposting*, and *talking about* behaviors are chain mediators that convert suicidal ideation into suicide attempt. *Talking about* behavior mediates the relation between suicidal ideation and suicide attempt. These findings support the social cognitive theory and expand the ideation-to-action framework by clarifying the pathways of how suicide ideators turn into suicide attempters via the mediation role of suicide-related social media use behaviors. This study, therefore, paves the way for future research to focus on more behavioral markers that can serve in the identification of suicide attempters, subsequently contributing to the development of efficient and effective population-based suicide prevention.

Acknowledgments

The authors would like to gratefully acknowledge the generous support of the National Basic Research Program of China (2014CB744600), China Social Science Fund (17AZD041), National Social Science Fund of China (16AZD058), National Natural Science Foundation of China (31700984), and the Research Grants Council of the Hong Kong Special Administrative Region, China (Collaborative Research Fund, Project No. C1031-18G). The sponsors had no role in the study design; in 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 measurement items for the whole study.
[DOCX File , 16 KB - jmir_v22i4e14940_app1.docx]

References

1. World Health Organization. Preventing Suicide: A Global Imperative. Geneva: World Health Organization; 2014.
2. Bertolote JM, Fleischmann A. A global perspective in the epidemiology of suicide. Suicidologi 2002;7(2):6-8. [doi: 10.5617/suicidologi.2330]
3. Nock MK, Borges GJ, Bromet E, Alonso J, Angermeyer M, Beautrais A, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. Br J Psychiatry 2008 Feb;192(2):98-105 [FREE Full text] [doi: 10.1192/bjp.bp.107.040113] [Medline: 18245022]

4. May AM, Klonsky ED. What distinguishes suicide attempters from suicide ideators? A meta-analysis of potential factors. Clin Psychol Sci Pract 2016;23(1):5-20. [doi: 10.1111/cpsp.12136]
5. Klonsky ED, May AM. The three-step theory (3ST): a new theory of suicide rooted in the 'ideation-to-action' framework. Int J Cogn Ther 2015;8(2):114-129. [doi: 10.1521/ijct.2015.8.2.114]
6. Mars B, Heron J, Klonsky ED, Moran P, O'Connor RC, Tilling K, et al. What distinguishes adolescents with suicidal thoughts from those who have attempted suicide? A population-based birth cohort study. J Child Psychol Psychiatry 2019 Jan;60(1):91-99 [FREE Full text] [doi: 10.1111/jcpp.12878] [Medline: 29492978]
7. Belfort EL, Mezzacappa E, Ginnis K. Similarities and differences among adolescents who communicate suicidality to others via electronic versus other means: a pilot study. Adolesc Psychiatry 2012;2(3):258-262. [doi: 10.2174/2210676611202030258]
8. Li A, Huang X, Zhu T. A systematic analysis of online broadcasts of suicidality in China. Asia Pac Psychiatry 2018 Sep;10(3):e12302. [doi: 10.1111/appy.12302] [Medline: 28976066]
9. Pourmand A, Roberson J, Caggiula A, Monsalve N, Rahimi M, Torres-Llenza V. Social media and suicide: a review of technology-based epidemiology and risk assessment. Telemed J E Health 2019 Oct;25(10):880-888. [doi: 10.1089/tmj.2018.0203] [Medline: 30362903]
10. Padmanathan P, Biddle L, Carroll R, Derges J, Potokar J, Gunnell D. Suicide and self-harm related internet use. Crisis 2018 Nov;39(6):469-478 [FREE Full text] [doi: 10.1027/0227-5910/a000522] [Medline: 29848080]
11. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS One 2017;12(8):e0181722 [FREE Full text] [doi: 10.1371/journal.pone.0181722] [Medline: 28813437]
12. Bell J, Mok K, Gardiner E, Pirkis J. Suicide-related internet use among suicidal young people in the UK: Characteristics of users, effects of use, and barriers to offline help-seeking. Arch Suicide Res 2018;22(2):263-277. [doi: 10.1080/13811118.2017.1334609] [Medline: 28590842]
13. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
14. Dolan R, Conduit J, Fahy J, Goodman S. Social media engagement behaviour: a uses and gratifications perspective. J Strat Mark 2016;24(3-4):261-277. [doi: 10.1080/0965254x.2015.1095222]
15. O'Connor RC, Kirtley OJ. The integrated motivational-volitional model of suicidal behaviour. Philos Trans R Soc Lond B Biol Sci 2018 Sep 5;373(1754):pii: 20170268 [FREE Full text] [doi: 10.1098/rstb.2017.0268] [Medline: 30012735]
16. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001;3(3):265-299. [doi: 10.1207/s1532785xmep0303_03]
17. Fu KW, Chan YY, Yip PS. Testing a theoretical model based on social cognitive theory for media influences on suicidal ideation: Results from a panel study. Media Psychol 2009;12(1):26-49. [doi: 10.1080/15213260802669441]
18. Dunlop SM, More E, Romer D. Where do youth learn about suicides on the internet, and what influence does this have on suicidal ideation? J Child Psychol Psychiatry 2011 Oct;52(10):1073-1080. [doi: 10.1111/j.1469-7610.2011.02416.x] [Medline: 21658185]
19. Katsumata Y, Matsumoto T, Kitani M, Takeshima T. Electronic media use and suicidal ideation in Japanese adolescents. Psychiatry Clin Neurosci 2008 Dec;62(6):744-746 [FREE Full text] [doi: 10.1111/j.1440-1819.2008.01880.x] [Medline: 19068014]
20. Singaravelu V, Stewart A, Adams J, Simkin S, Hawton K. Information-seeking on the internet. Crisis 2015;36(3):211-219. [doi: 10.1027/0227-5910/a000307] [Medline: 26088826]
21. Masuda N, Kurahashi I, Onari H. Suicide ideation of individuals in online social networks. PLoS One 2013;8(4):e62262 [FREE Full text] [doi: 10.1371/journal.pone.0062262] [Medline: 23638019]
22. 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]
23. Gilat I, Shahar G. Suicide prevention by online support groups: an action theory-based model of emotional first aid. Arch Suicide Res 2009;13(1):52-63. [doi: 10.1080/13811110802572148] [Medline: 19123109]
24. Lee S, Kwon Y. Twitter as a place where people meet to make suicide pacts. Public Health 2018 Jun;159:21-26. [doi: 10.1016/j.puhe.2018.03.001] [Medline: 29689413]
25. Ozawa-de Silva C. Too lonely to die alone: internet suicide pacts and existential suffering in Japan. Cult Med Psychiatry 2008 Dec;32(4):516-551. [doi: 10.1007/s11013-008-9108-0] [Medline: 18800195]
26. Jiang F, Xu H, Liao H, Zhang T. Analysis of internet suicide pacts reported by the media in mainland China. Crisis 2017 Jan;38(1):36-43. [doi: 10.1027/0227-5910/a000402] [Medline: 27278566]
27. Sueki H. The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan. J Affect Disord 2015 Jan 1;170:155-160. [doi: 10.1016/j.jad.2014.08.047] [Medline: 25240843]
28. Microblog S. Sina microblog user development report. 2016 URL: http://blog.sina.com.cn/s/blog_e9475ade0102x1wv.html [accessed 2018-07-26] [WebCite Cache ID 71Bj4HljO]

29. Office of the Cyber Security and Information Committee of the CPC Central Committee. 2018 Jan 31. The 41th Statistical Report on Internet Development in China URL: http://www.cac.gov.cn/2018-01/31/c_1122347026.htm [accessed 2019-06-01] [WebCite Cache ID 71Bix5Mxc]
30. Liu X, Liu X, Sun J, Yu NX, Sun B, Li Q, et al. 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 May 8;21(5):e11705 [FREE Full text] [doi: 10.2196/11705] [Medline: 31344675]
31. 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]
32. Reynolds WM. Psychometric characteristics of the Adult Suicidal Ideation Questionnaire in college students. J Pers Assess 1991 Apr;56(2):289-307. [doi: 10.1207/s15327752jpa5602_9] [Medline: 2056423]
33. Fu K, Liu KY, Yip PS. Predictive validity of the Chinese version of the Adult Suicidal Ideation Questionnaire: psychometric properties and its short version. Psychol Assess 2007 Dec;19(4):422-429. [doi: 10.1037/1040-3590.19.4.422] [Medline: 18085934]
34. Bollen KA. Multiple indicators: internal consistency or no necessary relationship? Qual Quant 1984;18(4):377-385 [FREE Full text] [doi: 10.1007/bf00227593]
35. Asarnow J, Baraff L, Berk M, Grob C, Devich-Navarro M, Suddath R, et al. Pediatric emergency department suicidal patients: two-site evaluation of suicide ideators, single attempters, and repeat attempters. J Am Acad Child Adolesc Psychiatry 2008 Aug;47(8):958-966. [doi: 10.1097/CHI.0b013e3181799ee8] [Medline: 18596552]
36. Skrondal A, Rabe-Hesketh S. Semantic Scholar. Generalized linear latent and mixed models with composite links and exploded likelihoods URL: https://pdfs.semanticscholar.org/791f/ddc6383a2f4cd1ca88cb80ccbe5cd8a9b60f.pdf [accessed 2019-06-04] [WebCite Cache ID 78sOYIyW2]
37. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990 Mar;107(2):238-246. [doi: 10.1037/0033-2909.107.2.238] [Medline: 2320703]
38. -. The American Educational Research Association. In: Educational Researcher. 2016 Jul Presented at: Paper presented at the annual conference of the American Educational Research Association; ; New Orleans; 2002 April 4; Paper presented at the annual conference of the American Educational Research Association; ; New Orleans p. 13-16 URL: https://ci.nii.ac.jp/naid/10030318555/ [doi: 10.3102/0013189x012009000]
39. Preacher KJ, Selig JP. Advantages of Monte Carlo confidence intervals for indirect effects. Commun Method Meas 2012 Apr;6(2):77-98 [FREE Full text] [doi: 10.1080/19312458.2012.679848]
40. Preacher KJ, Coffman DL. Kristopher J Preacher. 2006 May. Computing Power and Minimum Sample Size for RMSEA URL: http://www.quantpsy.org/rmsea/rmsea.htm [accessed 2019-10-20]
41. China News Network. 2018 Feb 6. Psychological Map: How Artificial Intelligence Can Help Suicides URL: http://www.sohu.com/a/221203017_123753 [accessed 2019-06-04] [WebCite Cache ID 78sQ3Md1A]
42. de Andrade NN, Pawson D, Muriello D, Donahue L, Guadagno J. Ethics and artificial intelligence: suicide prevention on Facebook. Philos Technol 2018;31(4):669-684. [doi: 10.1007/s13347-018-0336-0]
43. Mandrusiak M, Rudd M, Joiner T, Berman A, van Orden KA, Witte T. Warning signs for suicide on the internet: a descriptive study. Suicide Life Threat Behav 2006 Jun;36(3):263-271. [doi: 10.1521/suli.2006.36.3.263] [Medline: 16805654]
44. Roen KJ, Scourfield J, McDermott E. Making sense of suicide: a discourse analysis of young people's talk about suicidal subjecthood. Soc Sci Med 2008 Dec;67(12):2089-2097. [doi: 10.1016/j.socscimed.2008.09.019] [Medline: 18950923]
45. 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]
46. Wang Z, Yu G, Tian X. Exploring behavior of people with suicidal ideation in a Chinese online suicidal community. Int J Environ Res Public Health 2018 Dec 26;16(1):pii: E54 [FREE Full text] [doi: 10.3390/ijerph16010054] [Medline: 30587805]
47. Bailey E, Rice S, Robinson J, Nedeljkovic M, Alvarez-Jimenez M. Theoretical and empirical foundations of a novel online social networking intervention for youth suicide prevention: A conceptual review. J Affect Disord 2018 Oct 1;238:499-505. [doi: 10.1016/j.jad.2018.06.028] [Medline: 29936387]
48. Horne J, Wiggins S. Doing being 'on the edge': managing the dilemma of being authentically suicidal in an online forum. Sociol Health Illn 2009 Mar;31(2):170-184 [FREE Full text] [doi: 10.1111/j.1467-9566.2008.01130.x] [Medline: 18983421]
49. Cheung WS, Ho SM. The use of death metaphors to understand personal meaning of death among Hong Kong Chinese undergraduates. Death Stud 2004 Jan;28(1):47-62. [doi: 10.1080/07481180490249265] [Medline: 14969278]
50. Klonsky ED, Saffer BY, Bryan CJ. Ideation-to-action theories of suicide: a conceptual and empirical update. Curr Opin Psychol 2018 Aug;22:38-43. [doi: 10.1016/j.copsyc.2017.07.020] [Medline: 30122276]
51. Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, et al. Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. Curr Psychiatry Rep 2018 Jun 28;20(7):51. [doi: 10.1007/s11920-018-0914-y] [Medline: 29956120]

Abbreviations

ASIQ: Adult Suicidal Ideation Questionnaire

CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
RMSEA: root mean square error of approximation
SEM: structural equation modeling
TLI: Tucker Lewis Index
WRMR: weighted root mean square

Edited by G Eysenbach; submitted 05.06.19; peer-reviewed by A Benis, F Lamy, R Guan; comments to author 29.09.19; revised version received 20.10.19; accepted 03.02.20; published 28.04.20.

Please cite as:
Liu X, Huang J, Yu NX, Li Q, Zhu T
Mediation Effect of Suicide-Related Social Media Use Behaviors on the Association Between Suicidal Ideation and Suicide Attempt: Cross-Sectional Questionnaire Study
J Med Internet Res 2020;22(4):e14940
URL: http://www.jmir.org/2020/4/e14940/
doi:10.2196/14940
PMID:32343249

Original Paper

Use of Rapid Online Surveys to Assess People's Perceptions During Infectious Disease Outbreaks: A Cross-sectional Survey on COVID-19

Pascal Geldsetzer[1], MD, MPH, PhD

Division of Primary Care and Population Health, Department of Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Pascal Geldsetzer, MD, MPH, PhD
Division of Primary Care and Population Health
Department of Medicine
Stanford University
1265 Welch Road
Stanford, CA, 94035
United States
Phone: 1 6507238596
Email: pgeldsetzer@stanford.edu

Abstract

Background: Given the extensive time needed to conduct a nationally representative household survey and the commonly low response rate of phone surveys, rapid online surveys may be a promising method to assess and track knowledge and perceptions among the general public during fast-moving infectious disease outbreaks.

Objective: This study aimed to apply rapid online surveying to determine knowledge and perceptions of coronavirus disease 2019 (COVID-19) among the general public in the United States and the United Kingdom.

Methods: An online questionnaire was administered to 3000 adults residing in the United States and 3000 adults residing in the United Kingdom who had registered with Prolific Academic to participate in online research. Prolific Academic established strata by age (18-27, 28-37, 38-47, 48-57, or ≥58 years), sex (male or female), and ethnicity (white, black or African American, Asian or Asian Indian, mixed, or "other"), as well as all permutations of these strata. The number of participants who could enroll in each of these strata was calculated to reflect the distribution in the US and UK general population. Enrollment into the survey within each stratum was on a first-come, first-served basis. Participants completed the questionnaire between February 23 and March 2, 2020.

Results: A total of 2986 and 2988 adults residing in the United States and the United Kingdom, respectively, completed the questionnaire. Of those, 64.4% (1924/2986) of US participants and 51.5% (1540/2988) of UK participants had a tertiary education degree, 67.5% (2015/2986) of US participants had a total household income between US $20,000 and US $99,999, and 74.4% (2223/2988) of UK participants had a total household income between £15,000 and £74,999. US and UK participants' median estimate for the probability of a fatal disease course among those infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was 5.0% (IQR 2.0%-15.0%) and 3.0% (IQR 2.0%-10.0%), respectively. Participants generally had good knowledge of the main mode of disease transmission and common symptoms of COVID-19. However, a substantial proportion of participants had misconceptions about how to prevent an infection and the recommended care-seeking behavior. For instance, 37.8% (95% CI 36.1%-39.6%) of US participants and 29.7% (95% CI 28.1%-31.4%) of UK participants thought that wearing a common surgical mask was "highly effective" in protecting them from acquiring COVID-19, and 25.6% (95% CI 24.1%-27.2%) of US participants and 29.6% (95% CI 28.0%-31.3%) of UK participants thought it was prudent to refrain from eating at Chinese restaurants. Around half (53.8%, 95% CI 52.1%-55.6%) of US participants and 39.1% (95% CI 37.4%-40.9%) of UK participants thought that children were at an especially high risk of death when infected with SARS-CoV-2.

Conclusions: The distribution of participants by total household income and education followed approximately that of the US and UK general population. The findings from this online survey could guide information campaigns by public health authorities, clinicians, and the media. More broadly, rapid online surveys could be an important tool in tracking the public's knowledge and misperceptions during rapidly moving infectious disease outbreaks.

(J Med Internet Res 2020;22(4):e18790) doi:10.2196/18790

KEYWORDS

rapid online surveys; perceptions; knowledge; coronavirus; SARS-CoV-2; pandemic; infectious disease; outbreak; survey; COVID-19; public health

Introduction

When faced with rapidly moving infectious disease outbreaks, such as the coronavirus disease 2019 (COVID-19), assessing knowledge and perceptions of relevant populations has to be accomplished in a short time frame if the findings are to be informative to the public health response. Population-representative household surveys generally take many months of preparation and data collection [1]. Phone surveys are faster to conduct but have increasingly suffered from low response rates (typically well below 10% [2]), which can be a major source of bias even when extensive weighting adjustments are made [3]. In addition, unless they use interactive voice response (which tends to further decrease the response rate [4]), phone surveys require substantial human resources to conduct the interviews. Given these limitations, rapid online surveys, which demand minimal human resources (beyond those needed to design the questionnaire) and could reach large numbers of respondents in a short time frame, may be a valuable tool to assess (and monitor over time) knowledge and perceptions of an infectious disease in the midst of an outbreak.

COVID-19 was first reported in Wuhan, China, in December 2019 [5]. On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic [6], and by March 17, 2020, there were more than 200,000 confirmed cases and over 8000 reported deaths from COVID-19 worldwide [7]. The course of the COVID-19 epidemic in the United States and the United Kingdom will likely be strongly impacted by how the population behaves, which is in turn influenced by what people know and believe about this disease [8]. A particular concern in this regard is the spread of dis- and misinformation about COVID-19 on social media sites, which has led the WHO to host a page with "myth busters" on their website and engage in discussions with social media companies [9]. Understanding what the general public knows about COVID-19 and which misperceptions they hold about the condition is important for US and UK public health authorities as well as the media to design effective information campaigns.

The speed with which COVID-19 is spreading across the world calls for rapid assessments of the population's knowledge and perceptions of this infection [7]. This study tests a rapid online survey methodology to determine knowledge and misperceptions of COVID-19 among the general adult population in both the United States and the United Kingdom.

Methods

Sampling Participants

This is a cross-sectional online survey conducted on the research platform created and managed by Prolific Academic Ltd. Prolific is an online platform that connects researchers with individuals around the world who are interested in participating in online research studies [10]. The platform's pool of participants consists of approximately 80,000 individuals, of whom about 43% reside in the United Kingdom and 33% reside in the United States [11]. Researchers are required to pay participants a minimum of US $6.50 per hour.

For this study, Prolific established strata by age group (18-27, 28-37, 38-47, 48-57, or ≥58 years), sex (male or female) and ethnicity (white, black or African American, Asian or Asian Indian, mixed, or "other") as well as all combinations of these strata. Using numbers from the latest census in the United States and the United Kingdom, a given number of places for taking the questionnaire were opened on the Prolific platform in each stratum to achieve the same distribution of participants by age, sex, and ethnicity as those in the general population. The targeted total sample size in each country was 1500 people. Participants' eligibility for the open places in a particular stratum was determined based on the information they had entered in their profile when registering with Prolific. Eligible participants enrolled in the study on a first-come, first-served basis. The study was implemented in two rounds of 1500 participants each in the United States and the United Kingdom, such that the total target sample size in each country was 3000. Participants had to have indicated that they are fluent in English when registering with Prolific to be eligible for this study.

Data Collection

Data were collected using the online questionnaire shown in Multimedia Appendix 1. Participants received US $1.50 (equivalent to £1.17 on March 1, 2020) for completing the questionnaire. Following an informed consent form, the questionnaire asked participants about the cause, current state, and future development of the COVID-19 epidemic; the risk of a fatal disease course; knowledge of symptoms and recommended health care–seeking behavior; measures to prevent an infection with SARS-CoV-2; and their perception of the risk posed by individuals of East-Asian ethnicity in their community. In order to investigate to what degree dis- and misinformation about COVID-19 has affected the general public's beliefs about the condition, participants were directly asked whether they believed several falsehoods listed on the WHO's "myth busters" website [12], which the WHO selected because these myths were circulating on social media [13]. Specifically, the questionnaire asked whether receiving a letter or package from China poses a risk of infection and whether using hand dryers, rinsing your nose with saline, eating garlic, applying sesame oil to the skin, taking antibiotics, and vaccinating against pneumonia are effective in preventing a SARS-CoV-2 infection. The questionnaire was built using Qualtrics software. Participants had to answer a question to reach the next question. Numerical entry questions did not allow for nonsensical inputs (eg, percentage questions were restricted to inputs between 0 and 100).

Data Analysis

For binary and categorical response options, the percentage of participants who selected each response was computed. For binomial proportions, two-sided 95% confidence intervals using the Wilson score interval were calculated [14]. No sampling weights were used, given that this was not a probabilistic sample of adults and that the survey was already self-weighting by the age, sex, and ethnicity groups used to establish the strata for sampling.

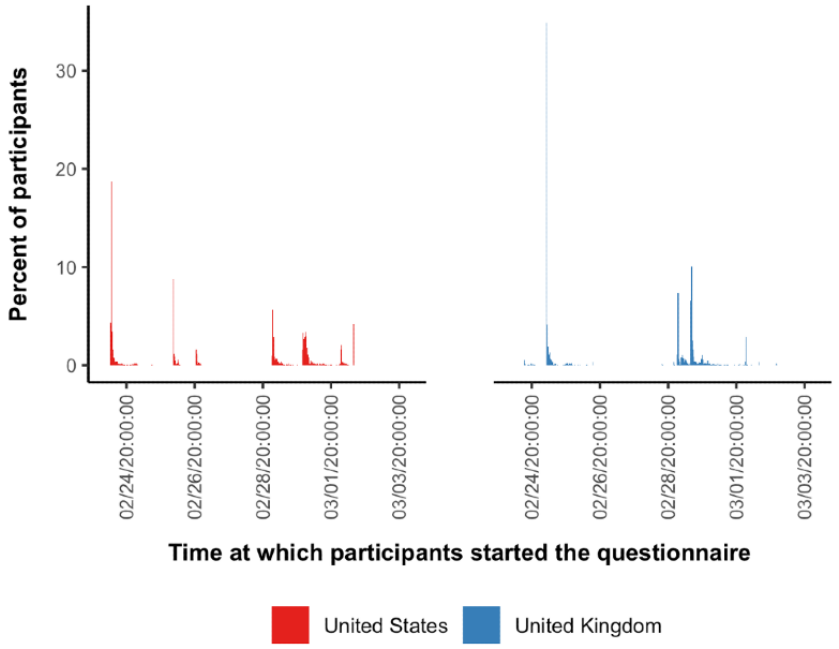
Three types of data quality checks were performed. First, participants who took less than 2 minutes to complete the questionnaire were excluded from the analysis because this indicated random clicking. This resulted in the exclusion of 2 participants. Second, if some respondents used random clicking to obtain the US $1.50 reward as fast as possible, a bimodal distribution in the time taken to complete the survey might be expected (with one group clicking as fast as possible and one reading the questions). I, therefore, plotted a histogram of the time taken to complete the survey. Third, participants were asked, at the end of the questionnaire, whether they looked up any answers online (*"It is natural to be tempted to look up the answer to a question, especially when it's only a click away. For approximately how many of the questions did you first look up the answer on Google or somewhere else before responding? The answer to this question will not affect your payment in any way."*) and if so, for which question. Those who self-reported looking up an answer online for a question were excluded from the analysis for that particular question. This was the case for 81 US participants and 63 UK participants who reported looking up the answer online for a median of 1 (IQR 1-5) and 1 (IQR 1-2.5) questions, respectively.

Results

Data Collection Time

Figure 1 shows that each of the two rounds of the survey took 2-3 days to conduct. There was no evidence of a bimodal distribution in the time taken to complete the survey (Multimedia Appendix 1).

Figure 1. Time at which participants started the questionnaire. Dates and times are given in Pacific Standard Time. Bins have a width equal to 30 minutes.



Participant Characteristics

Of 3000 adults residing in the United States and 3000 adults residing in the United Kingdom who could participate, 2986 and 2988, respectively, completed the questionnaire. Approximately two-thirds (1924/2986, 64.4%) of US participants and half (1540/2988, 51.5%) of UK participants had a tertiary education degree (Table 1). In addition, 67.5% (2015/2986) of US participants had a total household income between US $20,000 and US $99,999, and 74.4% (2223/2988) of UK participants had a total household income between £15,000 and £74,999. Furthermore, 17.3% (516/2986) of US participants and 13.7% (409/2988) of UK participants were currently students.

Table 1. Sample characteristics.

Characteristics	US	UK
Number of participants	2986	2988
Female, n (%)	1519 (50.9)	1531 (51.2)
Age group (years), n (%)		
18-27	655 (21.9)	550 (18.4)
28-37	687 (23.0)	557 (18.6)
38-47	531 (17.8)	563 (18.8)
48-57	493 (16.5)	480 (16.1)
≥58	620 (20.8)	838 (28.0)
Education, n (%)		
Less than a high school diploma/A-levels	24 (0.8)	396 (13.3)
High school degree/Completed A-levels	334 (11.2)	682 (22.8)
Some undergraduate education (no degree)	704 (23.6)	370 (12.4)
Associate degree	322 (10.8)	N/A[a]
Bachelor's degree	1068 (35.8)	1030 (34.5)
Master's degree	405 (13.6)	330 (11.0)
Professional degree	63 (2.1)	100 (3.3)
Doctorate	66 (2.2)	80 (2.7)
Total household income, n (%)		
<US $10,000/<£7500	165 (5.5)	172 (5.8)
US $10,000-US $19,000/£7500-£14,999	222 (7.4)	333 (11.1)
US $20,000-US $29,000/£15,000-£22,499	342 (11.5)	463 (15.5)
US $30,000-US $39,000/£22,500 - £29,999	325 (10.9)	473 (15.8)
US $40,000-US $49,000/£30,000-£37,499	280 (9.4)	358 (12.0)
US $50,000-US $59,000/£37,500-£44,999	304 (10.2)	312 (10.4)
US $60,000-US $69,000/£45,000-£52,499	230 (7.7)	242 (8.1)
US $70,000-US $79,000/£52,500-£59,999	242 (8.1)	156 (5.2)
US $80,000-US $89,000/£60,000-£67,499	138 (4.6)	121 (4.0)
US $90,000-US $99,000/£67,500-£74,999	154 (5.2)	98 (3.3)
US $100,000-US $149,000/£75,000-£99,999	401 (13.4)	168 (5.6)
≥US $150,000/≥£100,000	183 (6.1)	92 (3.1)
Race or ethnicity, n (%)		
White	2269 (76.0)	2540 (85.0)
Black or African American	392 (13.1)	110 (3.7)
Asian or Asian Indian	191 (6.4)	227 (7.6)
Mixed	74 (2.5)	62 (2.1)
Other	60 (2.0)	49 (1.6)
Current student, n (%)	516 (17.3)	409 (13.7)
Chinese descent, n (%)		
Born in China	11 (0.4)	15 (0.5)
Parents or grandparents born in China	57 (1.9)	27 (0.9)
Works as a health care provider, n (%)		
Nurse	33 (1.1)	44 (1.5)

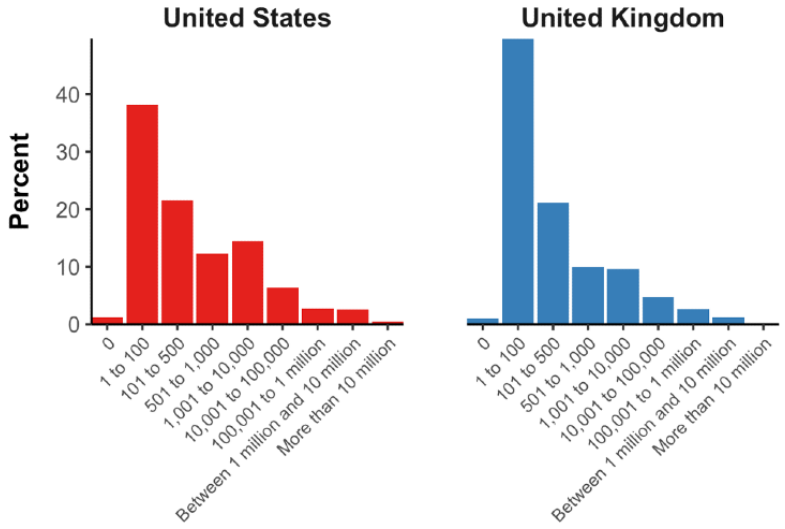
Characteristics	US	UK
Physician	5 (0.2)	15 (0.5)
Pharmacist	10 (0.3)	6 (0.2)
Other	102 (3.4)	118 (3.9)

[a]Associate degrees are not awarded in the UK.

Cause, Current State, and Future Development of the COVID-19 Epidemic

On a 7-point Likert scale ranging from "extremely unlikely" to "extremely likely," 23.9% (95% CI 22.4%-25.5%) of US participants and 18.4% (95% CI 17.1%-19.9%) of UK participants selected "slightly likely," "moderately likely," or "extremely likely" when asked whether SARS-CoV-2 is a bioweapon developed by a government or terrorist organization (Multimedia Appendix 1). The US and UK participants estimated that a median of 100 (IQR 20-500) and 40 (IQR 13-200) individuals in their respective country were currently infected with COVID-19. In addition, a mean of 61.0% (95% CI 59.3%-62.8%) of US and 71.7% (95% CI 70.1%-73.3%) of UK participants thought that the number of fatalities from COVID-19 in their country will be ≤500 people by the end of 2020 (Figure 2).

Figure 2. Proportion of participants who selected each category for their estimate of the number of COVID-19 deaths in their country by the end of 2020.

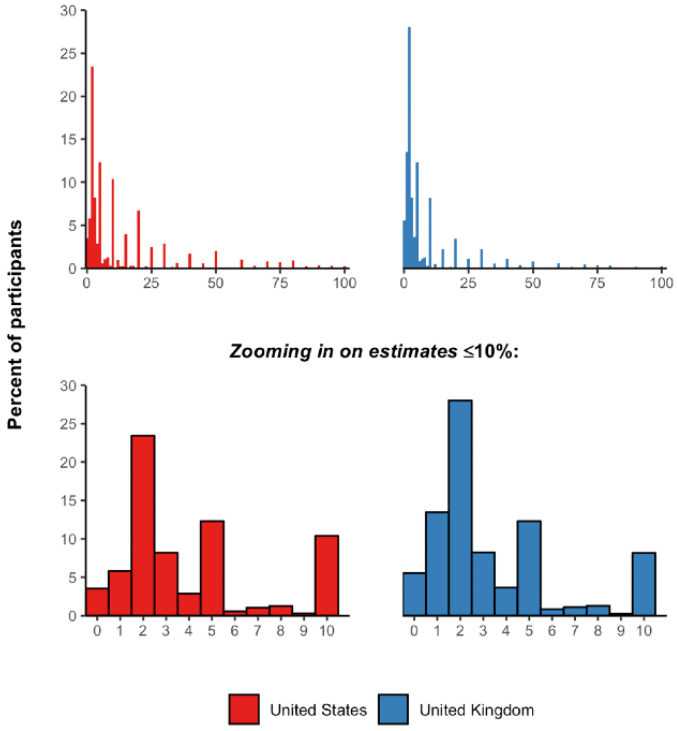


Case Fatality Rate

When asked what percent of individuals infected with COVID-19 experience a fatal disease course, the median estimate given by participants was 5.0% (IQR 2.0%-15.0%) among US participants and 3.0% (IQR 2.0%-10.0%) among UK participants. The full distribution of responses as well as a magnification of the distribution of responses among those who estimated a risk of death ≤10% is shown in Figure 3.

When asked "when they have been infected, what age groups are most likely to die from the illness caused by the new coronavirus" and presented with the option to select "children," "young adults," or "older adults" (selecting more than option was possible), 96.3% (95% CI 95.6%-96.9%) of participants in the United States and 98.3% (95% CI 97.7%-98.7%) of participants in the United Kingdom selected "older adults." However, 53.8% (95% CI 52.1%-55.6%) in the United States and 39.1% (95% CI 37.4%-40.9%) in the United Kingdom also thought that children were at a high risk of death when infected. Almost all participants in both countries (96.3%, 95% CI 95.6%-97.0% in the United States and 97.5%, 95% CI 96.9%-98.0% in the United Kingdom) responded that adults with other health problems were more likely to experience a fatal disease course than those without any other health problems.

Figure 3. Distribution of responses to the question, "What percent of people who get infected with the new coronavirus die from this infection?".



Symptoms of COVID-19 and Recommended Health Care–Seeking Behavior

Most participants in both the United States and the United Kingdom recognized fever, cough, and shortness of breath as three common symptoms and signs of COVID-19 (Figure 4).

When asked *"if you have a fever or cough and recently visited China, or spent time with someone who did, what would be the best course of action?"* 64.2% (95% CI 62.4%-65.9%) of US participants and 79.0% (95% CI 77.5%-80.5%) of UK participants responded with the recommended care-seeking option of staying home and contacting their health system. About a third of respondents stated that they would either delay care-seeking, attend the hospital emergency department unannounced, or take a taxi or public transport to their primary care provider (Figure 5).

Figure 4. Proportion of participants who replied with "yes" to whether each of seven symptoms or signs were common for COVID-19. The horizontal black bars represent the 95% CIs calculated using the Wilson method [14].

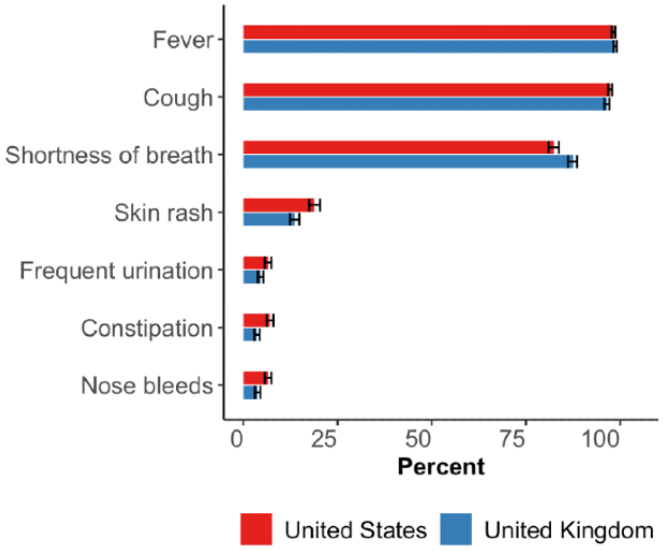
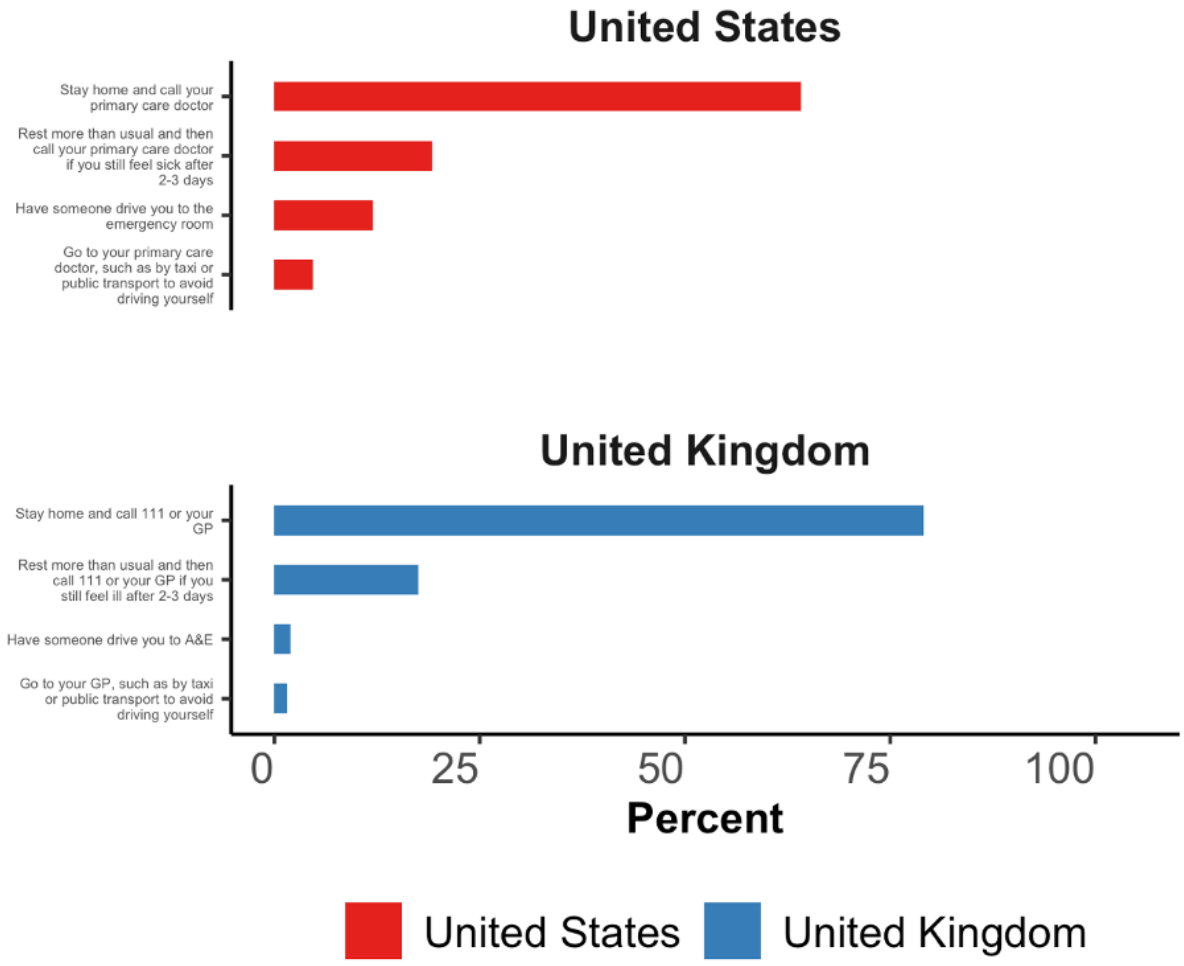


Figure 5. Responses to the question "If you have a fever or cough and recently visited China, or spent time with someone who did, what would be the best course of action?" GP: general practitioner; A&E: accident and emergency (department).



Preventing SARS-CoV-2 Infection

A total of 92.6% (95% CI 91.6%-93.4%) of US participants and 86.0% (95% CI 84.7%-87.2%) of UK participants selected each of the following three responses as effective measures for preventing infection with SARS-CoV-2: washing your hands; avoiding close contact with people who are sick; and avoiding touching your eyes, nose, and mouth with unwashed hands (Figure 6). However, a substantial proportion of participants also thought that using a hand dryer, rinsing your nose with saline, taking antibiotics, and gargling with mouthwash were effective prevention measures: 43.5% (95% CI 41.7%-45.2%) and 36.0% (95% CI 34.3%-37.8%) of US and UK participants, respectively, selected at least one of these options. Furthermore, 37.8% (95% CI 36.1%-39.6%) of US participants and 29.7% (95% CI 28.1%-31.4%) of UK participants agreed with the following statement: *"Consistently wearing a face mask is highly effective in protecting you from getting infected with the new coronavirus. For the purpose of this question, 'highly effective' is defined as reducing your risk of getting infected by >95% and a 'face mask' is a common medical mask."*

A total of 74.8% (95% CI 73.2%-76.4%) of US participants and 81.2% (95% CI 79.8%-82.6%) of UK participants correctly selected *"droplets of saliva that land in the mouths or noses of people who are nearby when an infected person sneezes or coughs"* as the main mode of SARS-CoV-2 transmission (Multimedia Appendix 1). Virtually all participants disagreed with the statement that *"only older adults can become infected with the new coronavirus"* (96.5%, 95% CI 95.8%-97.1% of US participants and 97.1%, 95% CI 96.5-97.7% of UK participants) and thought that there is currently no vaccine available that protects against COVID-19 (96.0%, 95% CI 95.3%-96.7% of US participants and 97.5%, 95% CI 96.9%-98.0% of UK participants). More than 20% of participants in both the United States and the United Kingdom thought that their government should quarantine everyone coming in from abroad for 14 days and suspend all air travel to their country (Figure 7).

Figure 6. Proportion of participants who replied with "yes" to whether each of 11 actions help prevent an infection with SARS-CoV-2. The horizontal black bars represent the 95% CIs calculated using the Wilson method [14].

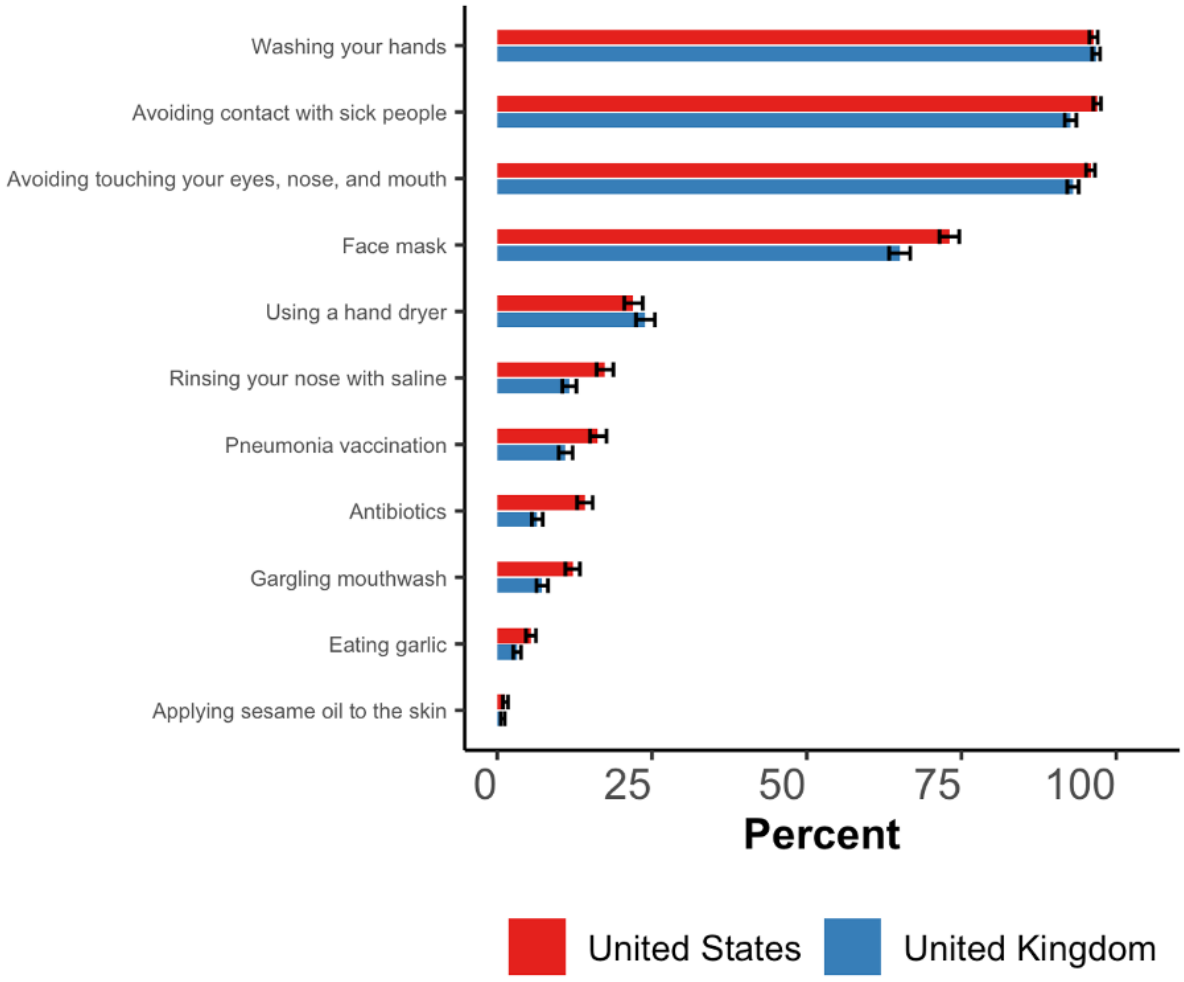
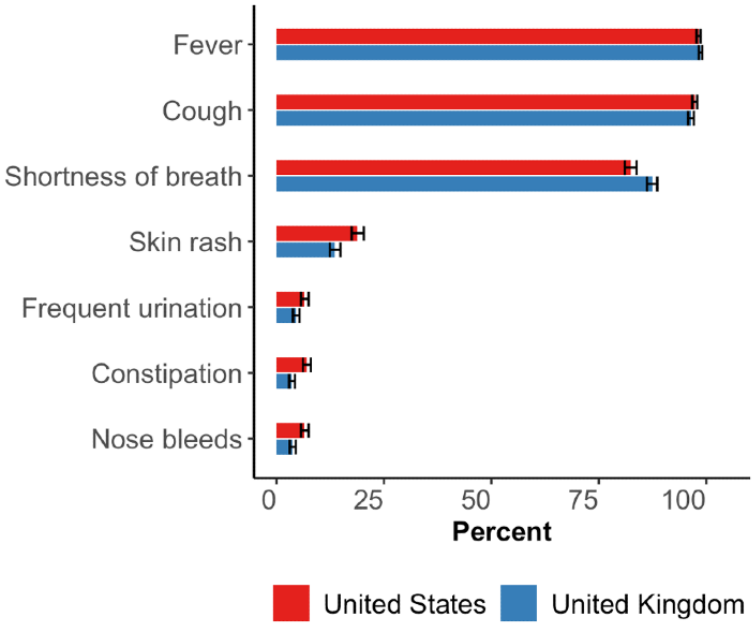


Figure 7. Proportion of participants who replied with "yes" to each government action in response to the question "At this point in the coronavirus epidemic, do you think your government should implement the following measures to prevent spreading of the virus?" The horizontal black bars represent the 95% CIs calculated using the Wilson method [14].



Perceptions of the Risk Posed by Community Members of East-Asian Ethnicity

When asked about the prevalence of an infection with SARS-CoV-2 among East-Asian individuals in their country, the median estimate among US and UK participants was 0.5% (IQR 0.0%-2.0%) and 0.5% (IQR 0.0%-1.0%), respectively (Figure 8). The median increased to 1.0% (IQR 0.0%-5.0%) among both US and UK participants when asking about the prevalence of COVID-19 among *"adults of East-Asian ethnicity in your neighborhood who wear a face mask."*

A total of 25.6% (95% CI 24.1%-27.2%) of US participants and 29.6% (95% CI 28.0%-31.3%) of UK participants responded with "yes" to the question, *"Do you think it would be prudent for you to not eat at Chinese restaurants for the next few weeks to reduce the risk of getting infected with the new coronavirus?"* Approximately a quarter of participants (29.0%, 95% CI 27.4%-30.7% of US participants and 24.4%, 95% CI 22.9%-26.0% of UK participants) thought one may become infected with SARS-CoV-2 by receiving a package from China. When asked *"If you were an Uber driver today, would you try to reject ride requests from people with East Asian-sounding names (or a profile photo of East-Asian ethnicity) to reduce your risk of getting infected with the new coronavirus?"* 29.7% (95% CI 28.1%-31.3%) of US participants and 40.8% (95% CI 39.0%-42.5%) of UK participants responded with "sometimes," "often," or "always" (Figure 9).

Figure 8. Distribution of the responses to questions on COVID-19 prevalence among individuals of East-Asian ethnicity. Of the total, 32 and 129 participants estimated a prevalence greater than 50% for the prevalence among East-Asian individuals in their country and East-Asian individuals wearing a face mask in their community, respectively. The responses from these individuals are not shown in the histogram below.

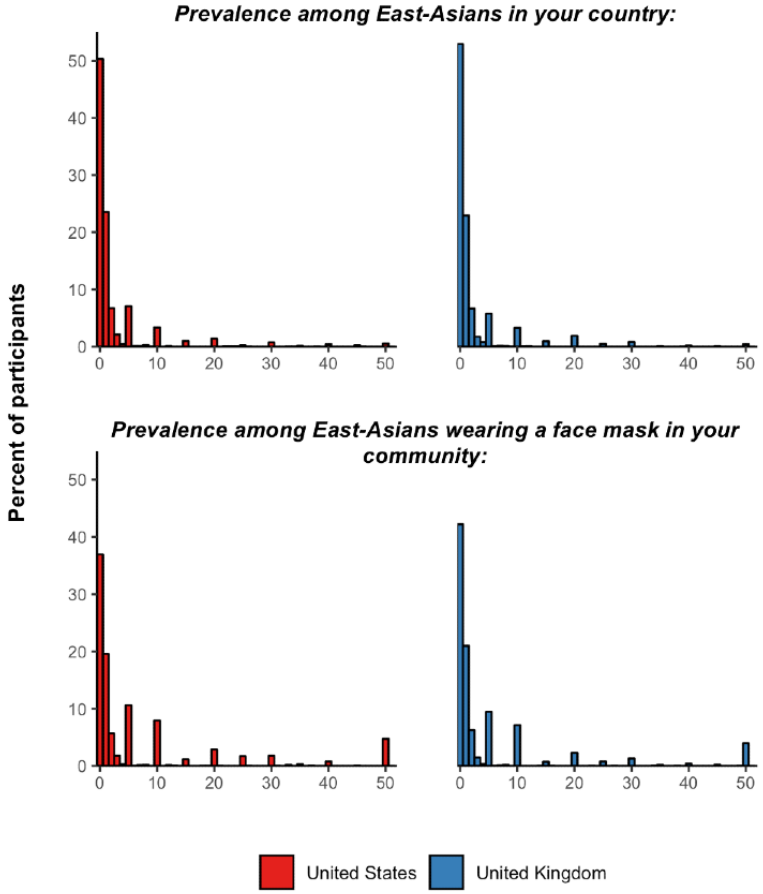
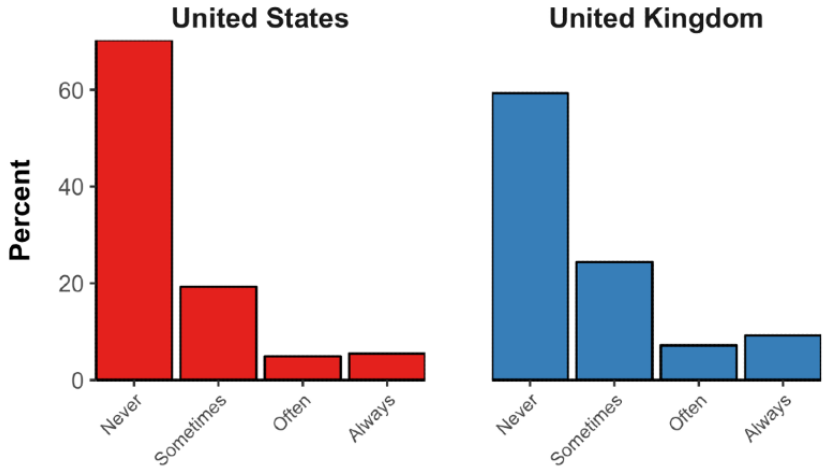


Figure 9. Distribution of responses to the question "If you were an Uber driver today, would you try to reject ride requests from people with East Asian-sounding names (or a profile photo of East-Asian ethnicity) to reduce your risk of getting infected with the new coronavirus?".



Discussion

It was possible to conduct an in-depth online assessment of knowledge and perceptions of COVID-19 among the general public in the United States and the United Kingdom in a short time frame. It took 2-3 days to obtain a completed questionnaire of 22 knowledge and perception questions from 1500 adults in each the United States and the United Kingdom, when allowing enrollment only in relatively granular strata by age, sex, and ethnicity (and each of these variables' combinations). Importantly, the distribution of participants by education and household income in this sample, although not part of the enrollment criteria, was similar to that of the general population in the United States and the United Kingdom [15,16]. In terms of data quality, there was no indication that participants randomly clicked on responses to earn the monetary reward as quickly as possible; only 2 participants took less than 2 minutes to complete the questionnaire and there was no evidence of a bimodal distribution in the time taken to complete the questionnaire. The direct cost to Prolific of administering the questionnaire was merely US $8961 (US $1.50 per completed questionnaire for a total of 5974 participants).

Regarding the survey findings, the general public in both the United States and the United Kingdom held several important misconceptions about COVID-19. Participants in both countries likely overestimated the probability of a fatal disease course among those infected with SARS-CoV-2 (while plagued by uncertainty, the case fatality rate is currently believed to lie below 1% [17]), thought that children were at an especially high risk of death from COVID-19 (which is currently not believed to be the case [5,17-19]), and believed that common surgical masks are highly effective in protecting them from SARS-CoV-2 infection. Participants also likely overestimated the prevalence of COVID-19 among East-Asian individuals in their communities. Likely as a result of this perception, a substantial proportion thought that they should refrain from frequenting Chinese restaurants, stated that they would refuse Uber rides to individuals of East-Asian ethnicity, and thought that receiving a package from China poses a risk of a SARS-CoV-2 infection. In general, differences in knowledge and misperceptions between US and UK participants were small.

This study's findings on the levels of knowledge and prevalence of misconceptions regarding COVID-19 could inform relevant information campaigns by public health authorities and the media as well as communication of health care workers with patients. For instance, such information provision may highlight the comparatively low case fatality rate, the low risk posed by individuals of East-Asian ethnicity living in the United States and the United Kingdom, and that children do not appear to be at a heightened risk of dying from COVID-19. In addition, a substantial proportion of participants appeared to believe that common surgical masks are highly effective in protecting the wearer from infection with SARS-CoV-2. Information campaigns may, therefore, want to emphasize the comparative effectiveness of common surgical masks versus other methods of prevention, particularly frequent and thorough handwashing and avoiding close contact with people who are sick. Lastly, it is important to note that while the general public appeared to be well informed about the common symptoms of COVID-19, over one-fourth of the participants selected a health care–seeking option that could lead to further transmission of SARS-CoV-2. Thus, clear messaging on the recommended care-seeking action when experiencing some of the core symptoms of COVID-19 will be crucial.

Public health information campaigns may also want to directly target some of the mis- and disinformation that has circulated on social media [9,12,20]. Such measures could include information that rinsing your nose with saline, using a hand dryer, taking antibiotics, and gargling with mouthwash are not effective prevention measures and that receiving a letter or package from China does not pose a great risk of SARS-CoV-2 infection. These are all falsehoods listed on the WHO's "myth busters" website, which a substantial proportion of participants in this study believed [12]. More broadly, this study underscores the need for the WHO and other public health bodies to continue working with social media campaigns to minimize the circulation of inaccurate information about COVID-19. In line with recent media reporting that this conspiracy theory has been actively spread on Twitter [20], about 1 in 5 participants believed it to be "slightly likely," "moderately likely," or "extremely likely" that SARS-CoV-2 is a bioweapon developed by a government or terrorist organization.

Participants did not expect a large number of individuals to die from COVID-19 in their country by the end of 2020. This finding may be surprising considering that fear-inducing headlines in the media may (at least up to a certain extent [21]) result in more attention by readers than more emotionally neutral ones, which could result in a catastrophizing of the epidemic. Moving forward, information campaigns on COVID-19 may need to balance the messaging of two important facts about the COVID-19 epidemic that could be interpreted by the general public to stand in direct conflict with each other: (1) The case fatality rate of COVID-19 appears to be lower than that of other recent infectious disease outbreaks such as Ebola infection [22], SARS [23], and the Middle East respiratory syndrome (MERS) [23,24]. On the other hand, however, the epidemic could cause a large number of fatalities, which implies that actions by governments and the general public to reduce transmission of SARS-CoV-2 could save many lives.

This study has several limitations. First, although the sample of participants is representative of the US and UK general population by age, sex, and ethnicity, and the distribution of participants by household income and education was similar to that in the US and UK general population, participants may still differ from the general population on a variety of other characteristics. These characteristics may be both correlated with their knowledge and perceptions of COVID-19 as well as with their decision to participate in the study or to create a profile with Prolific. Second, the estimates of discrimination against individuals of East-Asian ethnicity may be an underestimation because some participants may not have wanted to volunteer their discriminating tendencies to themselves or the researcher. However, I, as the researcher, had no access to any identifying information about the research participants, and participants were reminded of this fact prior to answering the question. In addition, such social desirability bias has been found to be lower in online surveys than in telephone or in-person surveys [25,26]. Third, it was possible for participants to randomly click responses in order to devote the least amount of time to earn the US $1.50 reward. In my view, this issue is unlikely to have caused major bias in this study because (1) there was no evidence of a bimodal distribution in the time taken to complete the survey (Multimedia Appendix 1); (2) while it was physically possible to complete the survey in under 90 seconds when randomly clicking on responses, only 2 participants completed the survey in under 2 minutes; and (3) the monetary reward (US $1.50) was relatively small, and thus, for most participants, unlikely to be the main motivation for participating in the study. Lastly, it is possible that individuals looked up the answers to some of the questions online prior to answering, which may have biased the results. Participants were reminded of the importance not to look up answers online prior to taking the survey and were asked at the end of the survey (while being reassured that their payment is not influenced by whether they volunteer information on having looked up an answer online) which, if any, questions they searched for an answer online prior to responding.

Rapid online surveys are a promising method to assess and track knowledge and perceptions in the midst of rapidly evolving infectious disease outbreaks. Such assessments are crucial because ensuring that the general public is well informed about a condition like COVID-19 could reduce unnecessary anxiety as well as reduce disease transmission and thus ultimately save lives.

Acknowledgments

PG conceived of the study, designed the questionnaire, conducted the data analysis, and wrote the manuscript. PG is the guarantor of the work. PG was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number KL2TR003143. The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. PG had full access to all the data in the study and had final responsibility for the decision to submit for publication.

I would like to thank all participants and the team at Prolific Academic Ltd. for their time and effort.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary appendix.
[DOCX File , 19214 KB - jmir_v22i4e18790_app1.docx]

References

1. Gong W, Taighoon Shah M, Firdous S, Jarrett BA, Moulton LH, Moss WJ, et al. Comparison of three rapid household survey sampling methods for vaccination coverage assessment in a peri-urban setting in Pakistan. Int J Epidemiol 2019 Apr 01;48(2):583-595. [doi: 10.1093/ije/dyy263] [Medline: 30508112]
2. Kennedy C, Hartig H. Pew Research Center. 2019. Response rates in telephone surveys have resumed their decline URL: https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/ [accessed 2020-03-30]

3. Abraham K, Presser S, Helms S. How social processes distort measurement: the impact of survey nonresponse on estimates of volunteer work in the United States. AJS 2009 Jan;114(4):1129-1165. [doi: 10.1086/595945] [Medline: 19824303]
4. Charles Q, Alexandra C, Leenisha M, Ashley A. In Search of the Optimal Mode for Mobile Phone Surveys in Developing Countries. A Comparison of IVR, SMS. and CATI in Nigeria. Survey Research Methods 12/10 2019;13(3):305-318. [doi: 10.18148/srm/2019.v13i3.7375]
5. World Health Organization. F1000Prime. 2019. Report of the WHO-China Joint Mission on Coronavirus Disease URL: https://f1000.com/prime/737509210#eval793572110 [accessed 2020-03-30]
6. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-03-30]
7. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 Mar 19. [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
8. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
9. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676 [FREE Full text] [doi: 10.1016/s0140-6736(20)30461-x]
10. Prolific. How it works URL: https://www.prolific.co/#researcher-content [accessed 2020-03-30]
11. Prolific. Explore our participant pool demographics URL: https://www.prolific.co/demographics [accessed 2020-03-30]
12. World Health Organization. Coronavirus disease (COVID-19) advice for the public: Myth busters URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2020-03-30]
13. Richtel M. W.H.O. Fights a Pandemic Besides Coronavirus: An 'Infodemic'. NYTimes URL: https://www.nytimes.com/2020/02/06/health/coronavirus-misinformation-social-media.html [accessed 2020-03-30]
14. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association 1927 Jun;22(158):209. [doi: 10.2307/2276774]
15. United States Census Bureau. URL: https://www.census.gov/data.html [accessed 2020-03-30]
16. Office for National Statistics. People, population and community URL: https://www.ons.gov.uk/peoplepopulationandcommunity [accessed 2020-03-30]
17. Fauci AS, Lane HC, Redfield RR. Covid-19 — Navigating the Uncharted. N Engl J Med 2020 Mar 26;382(13):1268-1269. [doi: 10.1056/nejme2002387]
18. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. Ahead of print. N Engl J Med 2020 Mar 28. [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
19. Centers for Disease Control and Prevention. Frequently Asked Questions: Coronavirus Disease-2019 (COVID-19) URL: https://www.cdc.gov/coronavirus/2019-ncov/specific-groups/children-faq.html [accessed 2020-03-30]
20. Romm T. The Washington Post. Millions of tweets peddled conspiracy theories about coronavirus in other countries, an unpublished U.S. report says URL: https://www.washingtonpost.com/technology/2020/02/29/twitter-coronavirus-misinformation-state-department/ [accessed 2020-03-30]
21. Kim HS, Forquer H, Rusko J, Hornik RC, Cappella JN. Selective Exposure to Health Information: The Role of Headline Features in the Choice of Health Newsletter Articles. Media Psychol 2016 Jan 12;19(4):614-637 [FREE Full text] [doi: 10.1080/15213269.2015.1090907] [Medline: 28255280]
22. Jacob ST, Crozier I, Fischer WA, Hewlett A, Kraft CS, Vega MDL, et al. Ebola virus disease. Nat Rev Dis Primers 2020 Mar 20;6(1):13. [doi: 10.1038/s41572-020-0147-3] [Medline: 32080199]
23. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 2016 Aug;14(8):523-534 [FREE Full text] [doi: 10.1038/nrmicro.2016.81] [Medline: 27344959]
24. Mahase E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ 2020 Mar 18;368:m641. [doi: 10.1136/bmj.m641] [Medline: 32071063]
25. Kreuter F, Presser S, Tourangeau R. Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. Public Opinion Quarterly 2009 Jan 26;72(5):847-865. [doi: 10.1093/poq/nfn063]
26. Keeter S. Pew Research Center: Methods. 2015. From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls URL: https://www.pewresearch.org/methods/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/ [accessed 2020-03-30]

Abbreviations

COVID-19: coronavirus disease 2019
MERS: Middle East respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 18.03.20; peer-reviewed by GE Iyawa, MA Bahrami; accepted 27.03.20; published 02.04.20.

Please cite as:
Geldsetzer P
Use of Rapid Online Surveys to Assess People's Perceptions During Infectious Disease Outbreaks: A Cross-sectional Survey on COVID-19
J Med Internet Res 2020;22(4):e18790
URL: http://www.jmir.org/2020/4/e18790/
doi:10.2196/18790
PMID:32240094

Original Paper

Measures of Patient Dissatisfaction With Health Care in Polycystic Ovary Syndrome: Retrospective Analysis

Luis R Hoyos[1], MD; Manesha Putra[2], MD; Abigail A Armstrong[1], MD; Connie Y Cheng[1], MD; Carrie K Riestenberg[1], MD; Tery A Schooler[1], BSc; Daniel A Dumesic[1], MD

[1]Department of Obstetrics and Gynecology, University of California, Los Angeles, CA, United States

[2]Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, MetroHealth Medical Center, University Hospital Cleveland Medical Center, Cleveland, OH, United States

Corresponding Author:
Daniel A Dumesic, MD
Department of Obstetrics and Gynecology
University of California
10833 Le Conte Avenue
Room 22-178 CHS
Los Angeles, CA
United States
Phone: 1 3107945542
Email: ddumesic@mednet.ucla.edu

Abstract

Background: Polycystic ovary syndrome (PCOS) is a common reproductive and metabolic disorder in women; however, many clinicians may not be well versed in scientific advances that aid understanding of the associated reproductive, metabolic, and psychological abnormalities. Women with PCOS are dissatisfied with health care providers, the diagnostic process, and the initial treatment of PCOS and seek information through alternative sources. This has affected the patient-physician relationship by allowing medical information acquired through the internet, whether correct or not, to become accessible to patients and reshape their health care perspective. Patient dissatisfaction with health care providers regarding PCOS raises questions about the responsibilities of academic institutions to adequately train and maintain the competence of clinicians and government agencies to sufficiently support scientific investigation in this field.

Objective: The primary aim was to examine internet searching behaviors of the public regarding PCOS vs another highly prevalent gynecologic disorder. The secondary aim was to explore satisfaction with health care among patients with PCOS and their internet use. The tertiary aim was to examine medical education in reproductive endocrinology and infertility (REI) during obstetrics and gynecology (Ob/Gyn) residency as a proxy for physician knowledge in this field.

Methods: Google search trends and StoryBase quantified monthly Google absolute search volumes for search terms related to PCOS and fibroids (January 2004 to December 2017; United States). The reproductive disorder, fibroids, was selected as a comparison group because of its high prevalence among women. Between female groups, monthly absolute search volumes and their trends were compared. A Web-based questionnaire (June 2015 to March 2018) explored health care experiences and the internet use of women with PCOS. REI rotation information during Ob/Gyn residency in the United States was obtained from the Association of Professors of Gynecology and Obstetrics website.

Results: For PCOS (R=0.89; P<.01), but not fibroids (R=0.09; P=.25), monthly absolute search volumes increased significantly. PCOS-related monthly absolute search volumes (mean 384,423 searches, SD 88,756) were significantly greater than fibroid-related monthly absolute search volumes (mean 348,502 searches, SD 37,317; P<.05). PCOS was diagnosed by an Ob/Gyn in 60.9% (462/759) of patients, and 57.3% (435/759) of patients were dissatisfied with overall care. Among patients with PCOS, 98.2% (716/729) searched for PCOS on the Web but only 18.8% (143/729) of patients joined an online PCOS support group or forum. On average, Ob/Gyn residencies dedicated only 4% (2/43) of total block time to REI, whereas 5.5% (11/200) of such residencies did not offer any REI rotations.

Conclusions: Over time, PCOS has been increasingly searched on the Web compared with another highly prevalent gynecologic disorder. Patients with PCOS are dissatisfied with their health care providers, who would benefit from an improved understanding of PCOS during Ob/Gyn residency training.

(J Med Internet Res 2020;22(4):e16541) doi:10.2196/16541

KEYWORDS

PCOS; fibroid; Google; healthcare quality; infoveillance; infodemiology; medical education; health care; internet; satisfaction

Introduction

Background

Polycystic ovary syndrome (PCOS) is a common reproductive and metabolic disorder in women, characterized by hyperandrogenism, menstrual irregularity, and polycystic ovarian morphology. Various PCOS phenotypes and different diagnostic criteria, as proposed by the National Institutes of Health (NIH), Rotterdam criteria, and the Androgen Excess and PCOS Society, often confuse both clinicians and the general public alike [1]. Moreover, many clinicians are not well versed in scientific advances that aid understanding of the reproductive, metabolic, and psychological abnormalities associated with PCOS. In a recent survey of gynecologists and reproductive endocrinologists, more than one-fourth of respondents did not know which PCOS diagnostic criteria they used and were unlikely to recognize associated comorbidities, complications, and benefits of lifestyle modification [2]. Worldwide, more than one-third of women see 3 different health care providers over 2 years before receiving a PCOS diagnosis [3]. Consequently, women with PCOS are dissatisfied with health care providers, the diagnostic process, and the initial treatment of PCOS [3-5] and seek information regarding weight loss, irregular menses, infertility, and excess hair growth through alternative sources [5-8]. This has led to the organization of educational events, such as the PCOS Awareness Symposium by PCOS Challenge, Inc, held at the University of California, Los Angeles (UCLA), in 2015, to provide the general public with up-to-date education on the diagnosis and management of PCOS as a crucial public health priority [9].

Public dissatisfaction with health care providers regarding PCOS care also raises questions about the responsibility of academic institutions to adequately train and maintain competence of clinicians in the care of women with PCOS. This responsibility also extends to government agencies to sufficiently support scientific investigation for research in PCOS.

Worldwide internet use now provides important health care information to everyone, with search engines, such as Google, improving the way one interacts with the world to ask questions and receive answers. Consequently, acquiring medical information through the internet has affected the patient-physician relationship by allowing information, whether correct or not, to become accessible to patients and reshape the perspective of their health care. In support of this, a study of online search behavior in the United States has shown that 72% of internet users use a search engine to obtain information about health care and clinical research [10-15]. Therefore, we hypothesized that increased public dissatisfaction with PCOS health care can be measured through exaggerated internet use.

Objectives

The primary aim of this study was to examine internet searching behaviors by the public regarding PCOS vs another highly prevalent gynecologic disorder (ie, fibroids). The secondary aim was to explore satisfaction with health care and use of the internet in patients with PCOS using a survey. The tertiary aim was to examine medical education in reproductive endocrinology and infertility (REI) during obstetrics and gynecology (Ob/Gyn) residency as a proxy for physician knowledge in this field.

Methods

Patient Survey

Institutional review board approval was obtained from UCLA to develop an open online survey oriented toward women with PCOS. This questionnaire assessed health care satisfaction and whether respondents sought Web-based information related to their health care. The survey was developed after the PCOS Awareness Symposium held at UCLA in 2015 and started recruitment during June of that year until March 2018, following the identification of critical education and health care gaps. Recruitment was voluntary and accomplished through in-person requests, campus flyers, and advertisement on the Ob/Gyn departmental PCOS website with the wording “Do you have PCOS and are 18 years of age or older? If so, we would like to invite you to participate in a survey” [16]. The website was secured, contained relevant information for those interested in PCOS, and provided our survey to any patient with PCOS.

The usability and technical functionality of the survey were tested beforehand, and patients were informed of the survey’s purpose, length, and principal investigator. The survey’s length was 5 pages with six to eight questions per page, which were not adaptive, randomized, or with a review step. No incentives were offered to complete the survey, and no personal identifying information was collected. Women had the option to email or mail the survey to the research team. The data were then checked for completeness and entered by the research team into the secured UCLA research electronic data capture system. Paper copies were stored in locked offices, and data spreadsheets were kept on password-protected computers. Each completed item in the survey was analyzed independently, and missing response rates were provided.

Google Trends

Google Trends is commonly used to monitor internet activities related to certain keywords by reporting an *index* of search activity. The fraction of queries that include the search term in a specific geography (ie, the United States) at a particular time relative to the total number of queries is measured by this index. The resulting numbers, defined as relative search volumes, are then scaled with the maximum value set at 100 [11,17]. Data regarding PCOS were compared with that of fibroids as another highly prevalent female reproductive disorder (NIH-defined PCOS, approximately 7% of reproductive-aged women; fibroids, approximately 70%-80% of women by age 50 years) [18,19]. Google Trends first examined the online search trend by generating monthly relative search volumes for the search terms

Polycystic Ovary Syndrome, *PCOS*, or *Polycystic Ovarian Syndrome* from January 2004 to December 2017 in the United States. This trend for PCOS was then compared with that for fibroids with the search terms *Fibroid*, *Fibroids*, *Uterine Fibroid*, *Uterine Fibroids*, *Leiomyoma*, and *Myoma* over the same time interval.

Unlike Google Trends, search engine optimization (SEO) tool StoryBase (SEO.dk, Lyngby, Denmark) shows absolute search volumes but only does so for the preceding 12 months. Therefore, using this tool, the monthly absolute search volumes throughout 2017 were generated to calculate the yearly absolute search volume. The 2017 relative search volume was then obtained by adding the corresponding monthly relative search volumes obtained by querying the search terms from 2004 to 2017 with Google Trends. The absolute search volume from 2017 was then divided by the 2017 relative search volume, and from this, the absolute search volume for each relative search volume unit was calculated. Using this value along with the monthly relative search volume for each individual month in the study period, the corresponding monthly absolute search volumes were calculated.

Data were analyzed using SPSS version 22 (IBM Corp, Armonk, NY). Linear regression was used to examine the search trend of PCOS- and fibroid-related terms over time. The mean cumulative monthly absolute search volume of PCOS-related terms was compared with that of fibroid-related terms with an unpaired Student *t* test.

SEO tool StoryBase (SEO.dk, Google, Mountain View, California) was also used to obtain the top 200 questions related to the PCOS search terms. These questions were then classified according to their content (definition, treatment, fertility, etc), and a top 10 list of related questions was created.

Obstetrics and Gynecology Residency Training

The percentage of block time devoted to REI during Ob/Gyn residency was examined, as information regarding PCOS-related questions correctly answered through the Council on Resident Education in Obstetrics and Gynecology (CREOG) examination was unavailable. The list of Ob/Gyn residency programs in the Association of Professors of Gynecology and Obstetrics (APGO) website was searched. Only programs in the United States were included, and all residency programs were classified as either university-based programs or non–university-based programs if they were either university-affiliated or community-based programs, respectively. For the latter, the Fellowship and Residency Electronic Interactive Database of the American Medical Association website [20] was reviewed. The characteristics of each program were searched, including numbers of graduating residents and REI rotations as well as percentages of rotation blocks spent in REI during residency. Rotation blocks instead of months were used as an outcome variable, as rotation duration was not consistently described. If listed, REI was assumed to be one block unless described otherwise. When one block was divided among multiple rotations, equal duration distribution was assumed. For example, if a block was labeled *REI, breast and US*, REI was recorded as one-third of a block.

Results

Patient Survey

Demographic characteristics of the 759 respondents to the UCLA-based PCOS survey are shown in Table 1.

Most respondents were diagnosed by an Ob/Gyn physician (462/759, 60.9%) and were often dissatisfied with explanations for the cause of their PCOS (319/759, 42.0%), treatment of their symptoms (389/759, 51.2%), and/or overall care of their symptoms (435/759, 57.3%; Table 2).

When asked about whether they had searched the internet for information about PCOS, 98.2% (716/729) of respondents replied yes, and 1.8% (13/729) replied no. Moreover, 3.9% (30/759) of women did not provide an answer. Only 18.8% (143/759) of respondents, however, had joined an online PCOS support group or forum.

Table 1. Demographic characteristics of survey respondents (n=759).

Demographic characteristics	Women, n (%)
Age at survey response (years)	
18-25	366 (48.2)
26-35	283 (37.3)
36-50	89 (11.7)
≥51	7 (0.9)
Missing answers	14 (1.8)
Age o f first symptoms (years)	
<10	23 (3.0)
10-15	226 (29.8)
16-25	385 (50.7)
26-35	96 (12.7)
≥36	16 (2.1)
Missing answers	13 (1.7)
Race	
White	363 (47.8)
African American	107 (14.1)
East Asian	25 (3.3)
South Asian	70 (9.2)
Native American or Alaskan native	14 (1.8)
Native Hawaiian or Pacific Islander	1 (0.1)
Mixed	52 (6.9)
Other	117 (15.4)
Missing answers	10 (1.3)
Ethnic background	
Hispanic or Latino	103 (13.6)
Non-Hispanic or Latino	602 (79.3)
Missing answers	54 (7.1)

Table 2. Survey of polycystic ovary syndrome patient satisfaction (n=759).

PCOS[a] survey questions	Women, n (%)
Have you received a formal diagnosis of PCOS by a health care professional?	
Yes	640 (84.3)
No	59 (7.8)
Not sure	50 (6.6)
Missing answers	10 (1.3)
What was the specialty of the doctor who diagnosed you with PCOS?	
Obstetrics and gynecology	462 (60.9)
Family medicine	106 (14.0)
Medical endocrinology	47 (6.2)
Reproductive endocrinology and infertility	34 (4.5)
Internal medicine	17 (2.2)
Pediatrics	15 (2.0)
Pediatric endocrinology	9 (1.2)
Other specialties	42 (5.5)
Missing answers	27 (3.5)
How many doctors did you see for your symptoms before you received a diagnosis of PCOS?	
1	275 (36.2)
2	213 (28.1)
3	126 (16.6)
4	58 (7.6)
5	23 (3.0)
≥6	35 (4.6)
Missing answers	29 (3.8)
At the time of your diagnosis, how satisfied were you with the explanation you received about the cause of PCOS?	
Completely satisfied	79 (10.4)
Mostly satisfied	158 (20.8)
Satisfied	179 (23.6)
Not satisfied	319 (42.0)
Missing answers	24 (3.2)
At the time of your diagnosis, how satisfied were you with the initial explanation of your treatment options for managing your PCOS symptoms?	
Completely satisfied	56 (7.4)
Mostly satisfied	119 (15.7)
Satisfied	167 (22.0)
Not satisfied	389 (51.2)
Missing answers	28 (3.7)
At the present time, how satisfied are you with the medical care you are receiving for your PCOS?	
Completely satisfied	38 (5.0)
Mostly satisfied	106 (14.0)
Satisfied	150 (19.8)
Not satisfied	435 (57.3)
Missing answers	30 (3.9)
Since your diagnosis, have you sought medical care for PCOS from a health care provider other than the one who diagnosed your PCOS?	

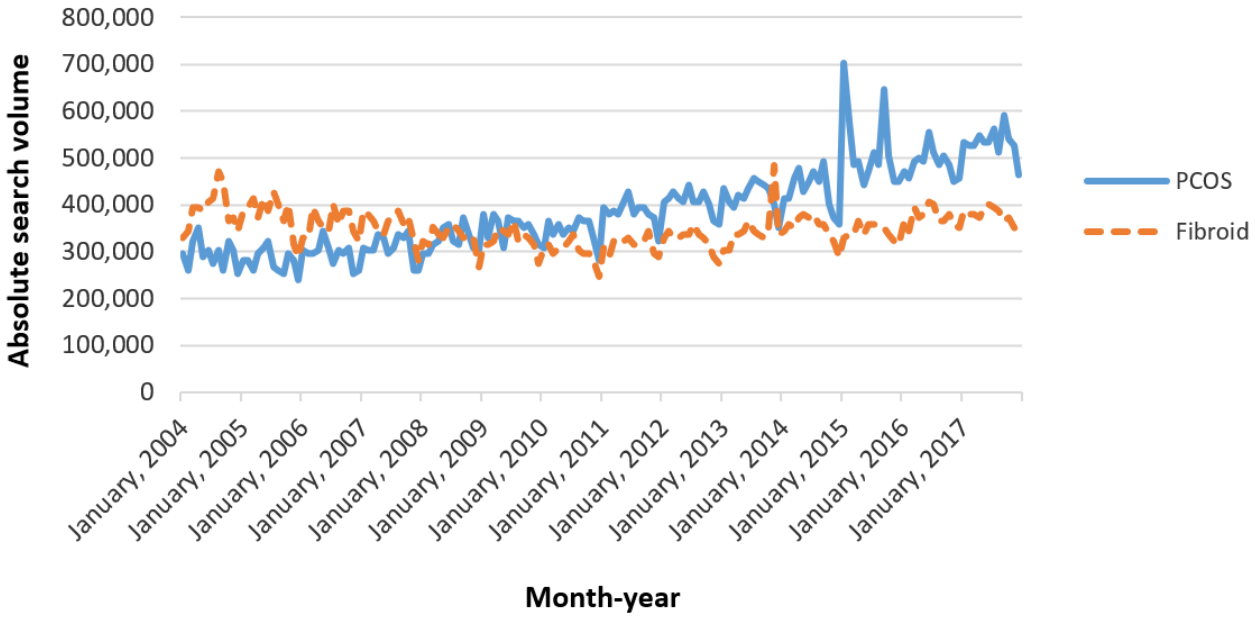
PCOS[a] survey questions	Women, n (%)
Yes	322 (42.4)
No	402 (53.0)
Missing answers	35 (4.6)

[a]PCOS: polycystic ovary syndrome.

Google Trends

Using Google Trends and SEO tool StoryBase (SEO.dk), we found that during the entire study period**,** there was a significant increase in the monthly absolute search volume of PCOS-related (R=0.89; P<.01), but not fibroid-related (R=0.09; P=.25), terms (Figure 1). Consequently, the mean monthly absolute search volume of PCOS-related terms between 2004 and 2017 was significantly greater than that of fibroid-related terms (PCOS: mean 384,423 searches, SD 88,756; fibroids: mean 348,502 searches, SD 37,317; P<.01) over the same time interval.

Figure 1. Monthly absolute search volumes for polycystic ovary syndrome and fibroids. PCOS: polycystic ovary syndrome.



The SEO tool StoryBase (SEO.dk) also examined the top 200 questions related to PCOS search terms by popularity, and these questions were then classified according to content with the top 10 PCOS-related questions listed in Textbox 1. Ten different topics related to PCOS were included within the first 22 questions. Among them, 23% (5/22) of topics were related to the PCOS definition, 14% (3/22) to availability of a cure, 14% (3/22) to achieving pregnancy, 14% (3/22) to diagnosis and testing, 9% (2/22) to weight control, 9% (2/22) to metformin, and 5% (1/22) related to other issues.

Textbox 1. Top 10 online questions related to polycystic ovary syndrome.

What is polycystic ovary syndrome (PCOS)?
How to get pregnant with PCOS?
How to lose weight with PCOS?
What causes PCOS?
How to treat PCOS?
How is PCOS diagnosed?
What does metformin do for PCOS?
Can PCOS be cured?
Can PCOS go away?
What is PCOS symptoms?

Obstetrics and Gynecology Residency Training

The APGO website provided the proportion of REI rotations among Ob/Gyn residencies in the United States. Of the 206 listed Ob/Gyn programs, six were excluded because of the lack of information about rotations. Of the remaining 200 Ob/Gyn programs, 57.5% (115/200) were university-based programs, and 42.5% (85/200) were non–university-based programs. University-based Ob/Gyn programs had 722 fourth-year residents listed, whereas non–university-based Ob/Gyn programs had 376 such residents. On average, university-based and non–university-based Ob/Gyn programs spent 4% (2/41) and 4% (2/46) of their residency blocks in REI, respectively. A formal REI rotation (ie, at least one complete block) was part of the Ob/Gyn residency curriculum in 92.0% (184/200) of programs, whereas 5.5% (11/200) did not list any REI rotation among their blocks.

Discussion

Principal Findings

This study demonstrates that the number of PCOS-related searches by Google has progressively increased over the past decade, with a heightened interest in PCOS shared by 98.2% (716/729) of respondents in our survey. The most common PCOS-related questions searched by Google pertained to its causes, definition, management, and natural history, with the most common PCOS-related question searched by Google being "What is PCOS?". Ironically, this question is not addressed by the American College of Obstetricians and Gynecologists patient education document about frequently asked questions for PCOS [21].

Specifically, 57.3% (435/759) of respondents in our survey were dissatisfied with their health care, with 42.0% (319/759) and 51.2% (389/759) of the same individuals being dissatisfied with explanations regarding the cause or treatment of PCOS, respectively. Only one in 5 of our respondents had joined an online PCOS support group or forum perhaps because of their preference for independent learning or anxiety to share personal experiences with others [22]. Furthermore, 42.4% (322/759) of survey respondents sought care from a health care provider other than the one who diagnosed PCOS.

A recent survey of gynecologists and reproductive endocrinologists by Dokras et al [2] has shown that 27.2% of these clinicians do not know the diagnostic criteria for PCOS they used. Although more than 85% of clinicians were aware of cardiometabolic comorbidities, fewer gynecologists recognized in patients with PCOS the possibility of concomitant mood-affective disorders, reduced quality of life, or the benefits of lifestyle modification [2].

These findings raise concerns regarding Ob/Gyn resident education on the diagnosis and management of PCOS-related reproductive and metabolic abnormalities, particularly as obstetrician/gynecologists were responsible for diagnosing PCOS in 60.9% (462/759) of our survey respondents. Our data further show that Ob/Gyn residency programs in the United States provide on average only 4% (2/43) of total block time to REI, with 5.5% (11/200) of such residencies not offering any REI rotation at all. Although Ob/Gyn residency programs often address the management of common PCOS-related symptoms, such as irregular menses and excess hair growth, such programs likely limit the clinical exposure of residents to complex PCOS-related metabolic and reproductive abnormalities that often exist within the context of assisted reproduction. Improved Ob/Gyn resident education on PCOS requires a revised curriculum that carefully integrates REI with primary health care aspects of Ob/Gyn. In this manner, REI specialists can interact with Ob/Gyn generalists to provide residents with a complete understanding of PCOS, along with its adverse reproductive and metabolic consequences and individualized clinical management.

Obstetrician/gynecologists also need to maintain their knowledge of recent advances in the field of PCOS. In July 2018, a new set of international guidelines was published for diagnosing and treating PCOS [23] with the goal of improving the clinical care of women with PCOS by physicians of various specialties. Continuing medical education courses that address up-to-date clinical guidelines could aid in this goal, particularly if they incorporate interactive cases or modules aimed at improving patient outcomes [24]. In support of this goal, professional societies should continue to promote interactive physician education, whereas government funding for PCOS research should be increased from its currently underfunded state (vs other chronic conditions) [25] to provide personalized, state-of-the-art health care for all women with PCOS based on multidisciplinary translational research.

The strengths of our study include the innovative use of Google to assess public internet searching behaviors regarding PCOS vs another highly prevalent gynecologic disorder. Information acquired over a decade regarding public interest in PCOS and questions asked online, supported by our survey, call for Ob/Gyn residency programs to improve clinical training in PCOS and for professional societies to maintain relevant educational materials.

Limitations

Limitations to our study include uncertainty as to whether individuals searching the internet for PCOS information were women affected by this syndrome or not. Sampling bias toward individuals dissatisfied with PCOS health care may also have affected our survey. In addition, information was unavailable regarding PCOS-related questions answered correctly by the CREOG examination as a measure of residents' competency in PCOS. In addition, the APGO and American Medical Association Fellowship and Residency Electronic Interactive Database websites may not have accurately represented the present Ob/Gyn residency curricula.

Comparison With Prior Work

Our findings agree with those of previous studies, in which women with PCOS seek health information on the internet [7] in a manner similar to that of individuals with other health issues [26]. That internet searches regarding PCOS exceeded those of fibroids, a highly prevalent gynecologic disorder, supports a previous report of significant patient dissatisfaction with health care regarding PCOS [4]. Unfortunately, sources such as

teenagers' and other women's digital magazines have social values and beliefs about women with PCOS embedded in the articles [27], further highlighting the need to provide patients with accurate and satisfactory information at the time of diagnosis.

In addition, 42.4% (322/759) of survey respondents sought care from a health care provider other than the one who diagnosed PCOS. In support of this, 60% of women in a previous study sought more than one health care provider before being diagnosed with PCOS [4,5].

Conclusions

Growing public use of internet PCOS search items above that of other highly prevalent gynecological disorder accompanies growing patient dissatisfaction with PCOS-related health care. Improved clinical care for women with PCOS, combined with continued scientific advances in this important area of women's health care, calls for academic medical institutions to improve education for clinicians to maintain current knowledge and for government agencies to increase research funding for PCOS as the most common endocrine disorder in women.

Conflicts of Interest

None declared.

References

1. ACOG Committee on Practice Bulletins--Gynecology. ACOG Practice Bulletin No. 108: Polycystic ovary syndrome. Obstet Gynecol 2009 Oct;114(4):936-949. [doi: 10.1097/AOG.0b013e3181bd12cb] [Medline: 19888063]
2. Dokras A, Saini S, Gibson-Helm M, Schulkin J, Cooney L, Teede H. Gaps in knowledge among physicians regarding diagnostic criteria and management of polycystic ovary syndrome. Fertil Steril 2017 Jun;107(6):1380-6.e1. [doi: 10.1016/j.fertnstert.2017.04.011] [Medline: 28483503]
3. Cree-Green M. Worldwide dissatisfaction with the diagnostic process and initial treatment of PCOS. J Clin Endocrinol Metab 2017 Feb 01;102(2):375-378 [FREE Full text] [doi: 10.1210/jc.2016-3808] [Medline: 28359108]
4. Gibson-Helm ME, Lucas IM, Boyle JA, Teede HJ. Women's experiences of polycystic ovary syndrome diagnosis. Fam Pract 2014 Oct;31(5):545-549. [doi: 10.1093/fampra/cmu028] [Medline: 24925927]
5. Gibson-Helm M, Teede H, Dunaif A, Dokras A. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. J Clin Endocrinol Metab 2017 Feb 01;102(2):604-612 [FREE Full text] [doi: 10.1210/jc.2016-2963] [Medline: 27906550]
6. Sills ES, Perloe M, Tucker MJ, Kaplan CR, Genton MG, Schattman GL. Diagnostic and treatment characteristics of polycystic ovary syndrome: descriptive measurements of patient perception and awareness from 657 confidential self-reports. BMC Womens Health 2001;1(1):3 [FREE Full text] [doi: 10.1186/1472-6874-1-3] [Medline: 11545683]
7. Ching HL, Burke V, Stuckey BG. Quality of life and psychological morbidity in women with polycystic ovary syndrome: body mass index, age and the provision of patient information are significant modifiers. Clin Endocrinol (Oxf) 2007 Mar;66(3):373-379. [doi: 10.1111/j.1365-2265.2007.02742.x] [Medline: 17302871]
8. Humphreys L, Costarelli V. Implementation of dietary and general lifestyle advice among women with polycystic ovarian syndrome. J R Soc Promot Health 2008 Jul;128(4):190-195. [doi: 10.1177/1466424008092230] [Medline: 18678115]
9. PCOS-Challenge. 2015. PCOS Awareness Symposium 2015 - Los Angeles Highlights URL: https://pcoschallenge.org/pcos-awareness-symposium-2015-los-angeles-highlights/ [accessed 2020-02-14]
10. 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]
11. Bragazzi NL, Bacigaluppi S, Robba C, Nardone R, Trinka E, Brigo F. Infodemiology of status epilepticus: a systematic validation of the Google Trends-based search queries. Epilepsy Behav 2016 Feb;55:120-123. [doi: 10.1016/j.yebeh.2015.12.017] [Medline: 26773681]
12. Brigo F, Trinka E. Google search behavior for status epilepticus. Epilepsy Behav 2015 Aug;49:146-149. [doi: 10.1016/j.yebeh.2015.02.029] [Medline: 25873438]
13. Lawson MA, Lawson MA, Kalff R, Walter J. Google search queries about neurosurgical topics: are they a suitable guide for neurosurgeons? World Neurosurg 2016 Jun;90:179-185. [doi: 10.1016/j.wneu.2016.02.045] [Medline: 26898496]
14. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
15. Telfer S, Woodburn J. Let me Google that for you: a time series analysis of seasonality in internet search trends for terms related to foot and ankle pain. J Foot Ankle Res 2015;8:27 [FREE Full text] [doi: 10.1186/s13047-015-0074-9] [Medline: 26146521]
16. UCLA Health. 2018. Polycystic Ovary Syndrome URL: http://obgyn.ucla.edu/pcos [accessed 2019-12-10]
17. Stephens-Davidowitz S, Varian H. UC Berkley. A Hands-on Guide to Google Data URL: http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf [accessed 2019-12-10]

18. ACOG Practice Bulletin No. 194: Polycystic Ovary Syndrome. Obstet Gynecol 2018 Jun;131(6):e157-e171. [doi: 10.1097/AOG.0000000000002656] [Medline: 29794677]
19. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol 2003 Jan;188(1):100-107. [doi: 10.1067/mob.2003.99] [Medline: 12548202]
20. American Medical Association. FREIDA Online Residency Fellowship Database URL: https://www.ama-assn.org/life-career/search-freida-online-ama-residency-fellowship-database [accessed 2018-08-17]
21. American College of Obstetricians and Gynecologists. Frequently Asked Questions Gynecologic Problems Polycystic Ovary Syndrome (PCOS) URL: https://www.acog.org/patient-resources/faqs/gynecologic-problems/polycystic-ovary-syndrome [accessed 2019-01-08]
22. Holbrey S, Coulson NS. A qualitative investigation of the impact of peer to peer online support for women living with polycystic ovary syndrome. BMC Womens Health 2013 Dec 17;13:51 [FREE Full text] [doi: 10.1186/1472-6874-13-51] [Medline: 24341398]
23. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Fertil Steril 2018 Aug;110(3):364-379 [FREE Full text] [doi: 10.1016/j.fertnstert.2018.05.004] [Medline: 30033227]
24. Zeiger RF. Toward continuous medical education. J Gen Intern Med 2005 Jan;20(1):91-94. [doi: 10.1111/j.1525-1497.2004.30049.x] [Medline: 15693934]
25. Brakta S, Lizneva D, Mykhalchenko K, Imam A, Walker W, Diamond MP, et al. Perspectives on polycystic ovary syndrome: is polycystic ovary syndrome research underfunded? J Clin Endocrinol Metab 2017 Dec 1;102(12):4421-4427. [doi: 10.1210/jc.2017-01415] [Medline: 29092064]
26. Lee K, Hoti K, Hughes JD, 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 02;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
27. Sanchez N, Jones H. 'Less Than A Wife': a study of polycystic ovary syndrome content in teen and women's digital magazines. J Med Internet Res 2016 Jun 02;18(6):e89 [FREE Full text] [doi: 10.2196/jmir.5417] [Medline: 27255809]

Abbreviations

APGO: Association of Professors of Gynecology and Obstetrics
CREOG: Council on Resident Education in Obstetrics and Gynecology
NIH: National Institutes of Health
Ob/Gyn: obstetrics and gynecology
PCOS: polycystic ovary syndrome
REI: reproductive endocrinology and infertility
SEO: search engine optimization
UCLA: University of California, Los Angeles

Edited by G Eysenbach; submitted 11.10.19; peer-reviewed by A Rambhatla, J Salisbury; comments to author 16.11.19; revised version received 18.12.19; accepted 27.01.20; published 21.04.20.

Please cite as:
Hoyos LR, Putra M, Armstrong AA, Cheng CY, Riestenberg CK, Schooler TA, Dumesic DA
Measures of Patient Dissatisfaction With Health Care in Polycystic Ovary Syndrome: Retrospective Analysis
J Med Internet Res 2020;22(4):e16541
URL: http://www.jmir.org/2020/4/e16541/
doi:10.2196/16541
PMID:32314967

Original Paper

Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study

Alaa Abd-Alrazaq[1], PhD; Dari Alhuwail[2,3], PhD; Mowafa Househ[1], PhD; Mounir Hamdi[1], PhD; Zubair Shah[1], PhD

[1]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[2]College of Life Sciences, Kuwait University, Kuwait, Kuwait

[3]Health Informatics Unit, Dasman Diabetes Institute, Kuwait, Kuwait

Corresponding Author:
Zubair Shah, PhD
College of Science and Engineering
Hamad Bin Khalifa University
Doha Al Luqta St, Ar- Rayyan
Doha
Qatar
Phone: 974 50744851
Email: ZShah@hbku.edu.qa

Abstract

Background: The recent coronavirus disease (COVID-19) pandemic is taking a toll on the world's health care infrastructure as well as the social, economic, and psychological well-being of humanity. Individuals, organizations, and governments are using social media to communicate with each other on a number of issues relating to the COVID-19 pandemic. Not much is known about the topics being shared on social media platforms relating to COVID-19. Analyzing such information can help policy makers and health care organizations assess the needs of their stakeholders and address them appropriately.

Objective: This study aims to identify the main topics posted by Twitter users related to the COVID-19 pandemic.

Methods: Leveraging a set of tools (Twitter's search application programming interface (API), Tweepy Python library, and PostgreSQL database) and using a set of predefined search terms ("corona," "2019-nCov," and "COVID-19"), we extracted the text and metadata (number of likes and retweets, and user profile information including the number of followers) of public English language tweets from February 2, 2020, to March 15, 2020. We analyzed the collected tweets using word frequencies of single (unigrams) and double words (bigrams). We leveraged latent Dirichlet allocation for topic modeling to identify topics discussed in the tweets. We also performed sentiment analysis and extracted the mean number of retweets, likes, and followers for each topic and calculated the interaction rate per topic.

Results: Out of approximately 2.8 million tweets included, 167,073 unique tweets from 160,829 unique users met the inclusion criteria. Our analysis identified 12 topics, which were grouped into four main themes: origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection. The mean sentiment was positive for 10 topics and negative for 2 topics (deaths caused by COVID-19 and increased racism). The mean for tweet topics of account followers ranged from 2722 (increased racism) to 13,413 (economic losses). The highest mean of likes for the tweets was 15.4 (economic loss), while the lowest was 3.94 (travel bans and warnings).

Conclusions: Public health crisis response activities on the ground and online are becoming increasingly simultaneous and intertwined. Social media provides an opportunity to directly communicate health information to the public. Health systems should work on building national and international disease detection and surveillance systems through monitoring social media. There is also a need for a more proactive and agile public health presence on social media to combat the spread of fake news.

(J Med Internet Res 2020;22(4):e19016) doi:10.2196/19016

KEYWORDS

coronavirus, COVID-19; SARS-CoV-2; 2019-nCov; social media; public health; Twitter; infoveillance; infodemiology; health informatics; disease surveillance

Introduction

Since the 1980s, human disease outbreaks have become increasingly frequent and diverse due to a plethora of ecological, environmental, and socioeconomic factors [1]. The family of coronaviruses was not considered to be highly pathogenic until 2003 and 2012 with the appearance of the severe acute respiratory syndrome in China followed by the Middle East respiratory syndrome in Saudi Arabia [2,3]. In December 2019, a series of patients with pneumonia of an unknown cause emerged in Wuhan, China [4]. Through contact tracing, these patients were linked back to a seafood and wet animal wholesale market in Wuhan [4]. To further investigate the symptoms, Chinese authorities conducted deep sequence analysis that provided ample evidence that the novel coronavirus was the causative agent of the disease [4], which is now known as the coronavirus disease (COVID-19). Since then, COVID-19 has quickly spread in China and other countries around the world. The disease is highly infectious, and, on average, each patient can spread the infection from 2 to 4 other individuals [5]. Worldwide, a total of 1,279,722 cases of COVID-19 and 72,614 deaths were confirmed in 212 countries by April 7, 2020 [6].

With the worldwide spread of the COVID-19 infection, individual activity on social media platforms such as Facebook, Twitter, and YouTube began to increase. A number of studies have shown that social media can play an important role as a source of data for detecting outbreaks but also in understanding public attitudes and behaviors during a crisis as a way to support crisis communication and health promotion messaging [7-11]. To assist public health professionals to make better decisions and aide their public health monitoring, advanced surveillance systems are developed to sort through large amounts of real time data from social media concerning public health information on a global scale [7]. Publicly accessible data posted on social media platforms by users around the world can be used to quickly identify the main thoughts, attitudes, feelings, and topics that are occupying the minds of individuals in relation to the COVID-19 pandemic. Such data can help policymakers, health care professionals, and the public identify primary issues that of concern and address them in a more appropriate manner.

A growing body of literature has been centered on examining the use of Twitter for public health research. A systematic review paper identified six main uses of Twitter for public health: analysis of shared content, surveillance of public health topics or diseases, public engagement, recruitment of research participants, Twitter-based public health interventions, and network analysis of Twitter users [9]. Other studies analyzed twitter data for sentiment analysis [12] and the use of Twitter to propagate credible vaccine-related web pages [8]. Building on previous work, this study aims to identify the main topics posted by Twitter users related to the COVID-19 pandemic. Analyzing such information can help policy makers and health care organizations assess the needs of their stakeholders and address them in an appropriate and relevant manner.

Methods

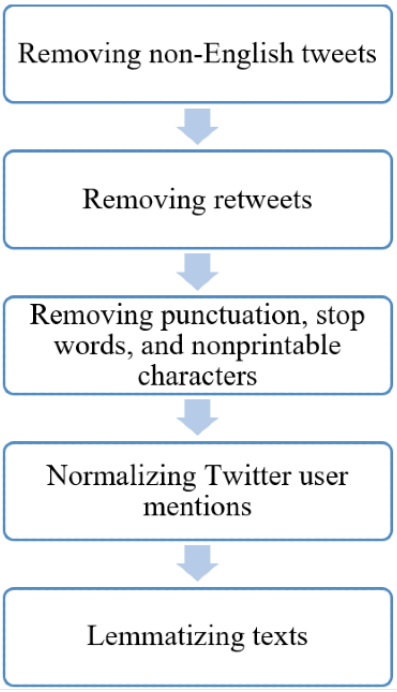
Data Collection

We collected coronavirus-related tweets between February 2, 2020, and March 15, 2020, using the Twitter standard search application programming interface (API) consisting of a set of predefined search terms ("corona," "2019-nCov," and "COVID-19"), which are the most widely used scientific and news media terms relating to the novel coronavirus. We extracted and stored the text and metadata of the tweets using the time stamp, number of likes and retweets, and user profile information including the number of followers. We stored the tweets in a database table, where the primary key of the table was tweet ID. As a result, the duplicates were not stored in our database. Only English language tweets were collected in the study. Since the metadata of tweets such as the number of likes and retweets might change over time, we recollected the updated metadata of the tweets at the end of the study period using the tweet IDs of the already collected tweets. Twitter standard search API allows the access of old tweets using tweet IDs. We used the Tweepy Python (Python Software Foundation) library for accessing the Twitter API and PostgreSQL (PostgreSQL Global Development Group) database for storing the collected tweets.

Data Preprocessing

We identified non-English tweets using the language field in the tweets metadata and removed them from the analysis. We identified and removed retweets from the analysis. We also removed punctuation, stop words such as *an* and *the*, and nonprintable characters such as emojis from the tweets. We normalized Twitter user mentions by converting, for example, "@Alaa" to "@username." Furthermore, various forms of the same word (eg, travels, traveling, and travel's) were lemmatized by converting them to the main word (eg, travel) using the WordNetLemmatizer module of the Natural Language Toolkit Python library. The data preprocessing is depicted in Figure 1. Following the terms and conditions, terms of use, and privacy policies of Twitter, all data were anonymized and were not reported verbatim to any third party.

Figure 1. Data preprocessing workflow.



Data Analysis

The processed tweets were analyzed using word frequencies of single words (unigram) and double-word (bigrams) combinations, and they were visualized through word clouds to identify the most common topics. In addition, we used the topic modeling technique [13] to identify the most common topics in the tweets. Topic modeling is an unsupervised machine learning technique that can find clusters in a collection of documents (tweets in this case). We used the latent Dirichlet allocation (LDA) algorithm from the Python sklearn package. LDA requires a fixed set of topics, where each topic is represented by a set of words. The objective of LDA is to map the given documents to the set of topics so that the words in each document are mostly captured by those topics. LDA is a widely used topic modeling algorithm. We used it to find natural clusters in the language of tweets. We applied topic modeling by specifying the number of topics required by the LDA to separate the set of tweets into various clusters. Based on our previous work, we selected 30 to be the number of topics for running the LDA [14].

We took the top representative words of each of the 30 topics produced by the LDA topic modelling algorithm (see LDA output in Multimedia Appendix 1) and the common words from the word cloud (see word cloud in Multimedia Appendix 2) and manually analyzed both sets of words. From this manual analysis, the authors reached a consensus on 12 topics and associated terms, unigram and bigram, for each topic (see associated terms for each topic in Multimedia Appendix 3). These terms were used to classify tweets, using a *rule-based classification script*, into different topics and compute the prevalence of each topic.

Next, we developed a rule-based classification script written in Python to check for the presence of any of the preidentified unigrams and bigrams in each tweet. The classification script used a simple string-matching technique to see if a given tweet contains the selected keywords of the topics. A tweet that contained a selected keyword related to a certain topic was classified as belonging to that topic.

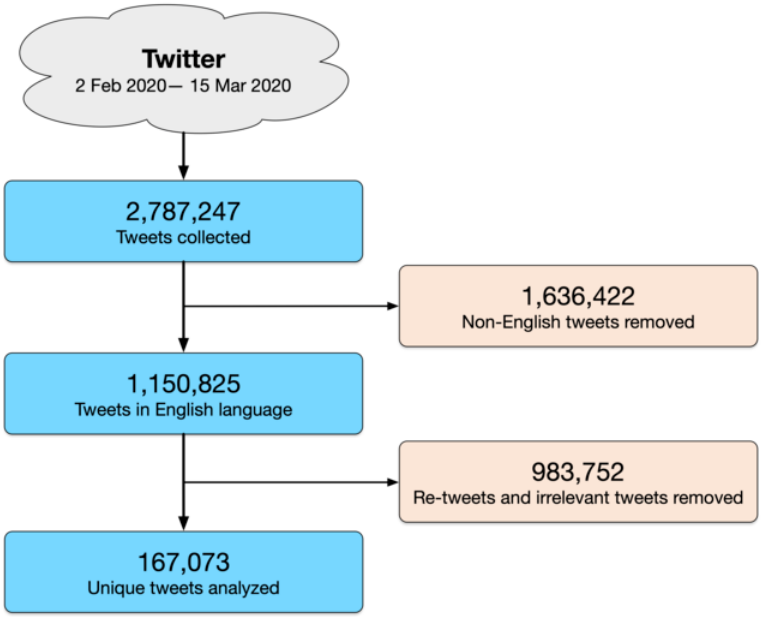
We also performed other analyses such as sentiment analysis, which extracts the mean number of retweets, likes, and followers for each topic and then calculates the interaction rate for each topic. The sentiment analysis was performed on the tweet text using the Python textblob library. The sentiment score varied between –1.0 to 1.0, with –1.0 as the most negative text and 1.0 as the most positive text. We calculated the mean sentiment and the mean number of likes, retweets, and followers for each topic. We also calculated the interaction rate for each topic by summing the total number of retweets and likes per topic divided by the sum of the total number of followers per topic. These measures provided additional insight into the topics and users who posted in these topics.

Results

Search Results

As shown in Figure 2, a total of 2,787,247 tweets were obtained between February 2, 2020, and March 15, 2020. Of these tweets, 1,636,422 (58.71%) non-English tweets were removed. Of the 1,150,825 remaining English tweets, 735,182 (63.88%) retweets were excluded. A further 248,570 (21.60%) tweets with no coronavirus-related terms in the text were also removed. These tweets were captured by Twitter API either because the name or the profile description of users matched the search terms. Accordingly, the study analyzed 167,073 unique tweets from 160,829 unique users.

Figure 2. Flowchart of selection of tweets.



Results of Tweet Analysis

Topics Emerged From Tweets

We identified 12 topics from the analyzed tweets. The 12 topics were grouped into four themes: the origin of COVID-19, the source of a novel coronavirus, the impact of COVID-19 on people and countries, and the methods for decreasing the spread of COVID-19. Table 1 summarizes the prevalence of the identified topics. Values on the diagonal of the table refer to numbers and percentages of tweets in a topic, and values in the off-diagonal of the table indicate numbers and percentages of tweets in the intersection of the two topics. For instance, a hypothetical tweet such as "while the death toll due to COVID-19 continues to rise, the travel ban imposed by countries to limit the spread of coronavirus infection started to affect the daily life of many people" could be classified under travel and death. The value at the intersection for these 2 topics in the table represents the number and percentage of tweets containing keywords related to both topics. More details about themes in these topics are elaborated in the following subsections.

Theme 1: Origin of COVID-19

This theme contains two topics that discuss the origin of COVID-19. The first topic was China, which was the most common topic of all identified topics. Tweeters talked about China as it was the country where the novel coronavirus originated from. The second topic was the outbreak. The tweets in this topic talked about the details of the outbreak, such as how, when, and where the outbreak emerged.

Theme 2: Source of the Novel Coronavirus

This theme included tweets about the causes leading to the transfer of COVID-19 to humans. Tweeters identified two sources of a novel coronavirus, which formed two topics in this study: eating meat and developing bioweapons. The former topic (eating meat) was identified in tweets mentioning the role of meat in the spread of COVID-19. Most of these tweets blamed nonvegetarians for the outbreak of COVID-19 and asked them to stop eating meat to stop the coronavirus spread. The latter topic (bioweapon) was formed by the tweets of individuals debating whether or not the COVID-19 virus originated from a Chinese biological military laboratory.

Theme 3: Impact of COVID-19 on People and Countries

The third theme was generated from tweets about the influence of COVID-19 on people, companies, and countries. The tweets in this theme identified six effects of COVID-19, which also formed six topics. The first topic related to the number of deaths caused by COVID-19. The tweets that belonged to this topic mainly showed statistics and numbers of deaths caused by a coronavirus in different cities and countries.

The second topic was the fear and stress caused by COVID-19. Twitter users in these tweets expressed their fear and stress about the coronavirus due to its quick spread and the lack of treatments or vaccines for the disease caused by the coronavirus.

The third topic was related to the effects of COVID-19 on travel from and to China and other countries. These tweets mostly discussed flight cancellations, postponements, travel bans, and restrictions as well as travel warnings imposed by many countries due to the coronavirus pandemic.

The impact of COVID-19 on the economy was the fourth topic. These tweets mostly showed actual or expected losses in the economy of many companies and countries due to, for example, closure of markets, a decrease of oil demands, delays in production, and canceling of important events, which came as a result of the COVID-19 outbreak.

Panic buying was the fifth topic identified. These tweets talked about how individuals in many countries became panic buyers in preparation for curfews, lockdowns, and stay-at-home orders due to the COVID-19 pandemic, and how supermarkets and shops controlled and prevented panic buying.

The last topic identified in this theme related to racism. Specifically, users in most of the tweets reported the spreading of racist, prejudiced, and xenophobic attacks (eg, rude comments or dirty looks) against East Asians given that COVID-19 originated from their countries.

Theme 4: Methods for Decreasing the Spread of COVID-19

The last theme brought together tweets that discussed methods for decreasing the spread of COVID-19. Two methods were identified from these tweets and formed the following two topics: wearing masks and the quarantine of people. Most of the tweets from the former topic talked about either the importance of face masks in decreasing the outbreak of the coronavirus or their shortage in several countries. Most of the tweets from the latter topic were about quarantining individuals who were infected with or suspected to have the coronavirus to reduce or prevent the spread of the disease.

As shown in the off-diagonal values in Table 1, the most common topic overlap was between China and deaths caused by COVID-19, followed by China and eating meat, China and the outbreak of COVID-19, deaths caused by COVID-19 and eating meat, and China and fear and stress about COVID-19.

Table 1. Numbers and percentages of tweets (N=167,073) related to each topic (diagonal values) and at the intersection of two topics (off-diagonal values).

Themes and subtopics	China, n (%)	Outbreak of COVID-19[a], n (%)	Eating meat, n (%)	Developing bioweapon, n (%)	Deaths caused by COVID-19, n (%)	Fear and stress about COVID-19, n (%)	Travel bans and warnings, n (%)	Economic losses, n (%)	Panic buying, n (%)	Increased racism, n (%)	Wearing masks, n (%)	Quarantining subjects, n (%)
Origin of COVID-19												
China	27,128 (16.24)	—[b]	—	—	—	—	—	—	—	—	—	—
Outbreak of COVID-19	2776 (1.66)	7468 (4.47)	—	—	—	—	—	—	—	—	—	—
Source of novel coronavirus												
Eating meat	4200 (2.51)	560 (0.34)	12,772 (7.65)	—	—	—	—	—	—	—	—	—
Developing bioweapon	808 (0.48)	151 (0.09)	220 (0.13)	2021 (1.21)	—	—	—	—	—	—	—	—
Impact of COVID-19 on people and countries												
Deaths caused by COVID-19	4332 (2.59)	905 (0.54)	2621 (1.57)	219 (0.13)	17,606 (10.54)	—	—	—	—	—	—	—
Fear and stress about COVID-19	1820 (1.09)	484 (0.29)	841 (0.50)	137 (0.08)	1421 (0.85)	8785 (5.26)	—	—	—	—	—	—
Travel bans and warnings	912 (0.55)	424 (0.25)	175 (0.10)	25 (0.01)	313 (0.19)	339 (0.20)	4358 (2.61)	—	—	—	—	—
Economic losses	1019 (0.61)	273 (0.16)	208 (0.12)	65 (0.04)	192 (0.11)	198 (0.12)	67 (0.04)	2565 (1.54)	—	—	—	—
Panic buying	598 (0.36)	175 (0.10)	115 (0.07)	39 (0.02)	183 (0.11)	161 (0.10)	83 (0.05)	826 (0.49)	2161 (1.29)	—	—	—
Increased racism	614 (0.37)	98 (0.06)	134 (0.08)	7 (0.01)	191 (0.11)	192 (0.11)	32 (0.02)	9 (0.01)	22 (0.01)	2136 (1.28)	—	—
Methods for decreasing COVID-19 spread												
Wearing masks	560 (0.34)	221 (0.13)	166 (0.10)	16 (0.01)	293 (0.18)	218 (0.13)	113 (0.07)	50 (0.03)	178 (0.10)	51 (0.03)	3397 (2.03)	—
Quarantining subjects	524 (0.31)	148 (0.09)	90 (0.05)	15 (0.01)	251 (0.15)	134 (0.08)	322 (0.19)	32 (0.02)	20 (0.01)	12 (0.01)	39 (0.02)	2014 (1.21)

[a]COVID-19: coronavirus disease.

[b]—: not available.

Results of Sentiment and Interaction Rate Analysis

As shown in Table 2, the mean of sentiment was positive in all topics except two: deaths caused by COVID-19 and increased racism. The highest mean of positive sentiments was for the eating meat topic, followed by the wearing masks topic. The highest mean of negative sentiments was for "deaths caused by COVID-19" topic.

Table 2. Results of sentiment and interaction analysis for tweets (N=167,073).

Topics	Sentiment, mean (SD)	Followers, mean (SD)	Likes, mean (SD)	Retweets, mean (SD)	Interaction rates	User mentions, n (%)	Link sharing, n (%)
China	0.028 (0.254)	5971.83 (182,938.26)	5.48 (128.42)	1.65 (51.08)	0.00120	10,323 (6.18)	11,041 (6.61)
Outbreak	0.037 (0.229)	20,498.22 (272,064.16)	6.48 (88.02)	2.69 (50.75)	0.00045	2038 (1.23)	3090 (1.85)
Eating meat	0.082 (0.282)	7177.12 (176,101.49)	12.34 (295.47)	7.09 (136.75)	0.00271	3815 (2.28)	7140 (4.27)
Developing bioweapon	0.016 (0.241)	3071.80 (22,697.08)	6.66 (114.81)	2.24 (37.53)	0.00290	1036 (0.62)	706 (0.42)
Deaths caused by COVID-19[a]	–0.057 (0.287)	9020.53 (204,289.34)	6.00 (86.42)	2.44 (39.75)	0.00094	6847 (4.10)	5924 (3.55)
Fear and stress about COVID-19	0.015 (0.247)	11,755.66 (310,842.61)	7.11 (129.05)	2.42 (48.22)	0.00081	3851 (2.30)	2693 (1.61)
Travel bans and warnings	0.032 (0.248)	9003.54 (154,933.20)	3.93 (33.27)	0.92 (8.07)	0.00054	2122 (1.27)	1210 (0.72)
Economic losses	0.035 (0.247)	13,361.82 (287,310.56)	15.33 (517.00)	3.58 (109.51)	0.00141	1225 (0.73)	846 (0.51)
Panic buying	0.031 (0.248)	12,121.17 (456,517.30)	4.07 (38.95)	0.89 (8.51)	0.00041	944 (0.56)	609 (0.36)
Increased racism	–0.033 (0.264)	2878.38 (64,604.27)	9.87 (80.57)	1.66 (14.89)	0.00400	685 (0.41)	427 (0.26)
Wearing masks	0.035 (0.262)	7557.34 (147,010.30)	8.08 (105.39)	1.88 (28.68)	0.00132	1200 (0.72)	1062 (0.64)
Quarantining subjects	0.012 (0.263)	6800.47 (87835.42)	5.64 (39.10)	1.90 (17.12)	0.00111	896 (0.54)	630 (0.38)

[a]COVID-19: coronavirus disease.

The mean of followers for tweeters who posted the collected tweets ranged from 2878 (in increased racism) to 13,361 followers (in economic losses). The economic loss topic had the highest mean of likes. On the other hand, travel ban and warning-related topics had the lowest mean of likes. The mean of retweets for the collected tweets varied between 0.89 (for panic buying) and 7.11 (for eating meat). The lowest interaction rate was for panic buying–related tweets, and the highest interaction rate was for racism-related tweets followed by bioweapon-related tweets and eating meat–related tweets (Table 2).

User mentions were the most common in China-related tweets, but they were the least common in racism-related tweets (Table 2). Similarly, link sharing was the most common in China-related tweets, whereas they were the least common in racism-related tweets (Table 2). Multimedia Appendix 4 shows more descriptive statistics (ie, medians, variances, standard deviations, maximums, and minimums) for all previously mentioned measures.

Discussion

Principal Findings

Users on Twitter discussed 12 main topics across four main themes related to COVID-19 between February 2, 2020, and March 15, 2020. User mentions and link sharing were the most common in the analyzed tweets. These findings might demonstrate that users on Twitter are interested in notifying or warning their friends and followers about COVID-19. These interpersonal communications indicate that people bond around the topic of COVID-19 on Twitter.

Users on Twitter also focused on the impact of coronavirus on people and countries. Specifically, numerous tweets were posted on the number of deaths linked to the coronavirus. Furthermore, the emotional and psychological impact of the coronavirus was mentioned in many tweets. Users on Twitter may show their fear and stress about COVID-19 and the lack of vaccine treatment options to prevent it or specific antiviral treatments [15]. However, the sensationalistic use of Twitter can be a great challenge for public health and outbreak response efforts because of the wild spread of misinformation and conspiracy theories [16]. The infectious outbreak of "fake news" and

"distorted evidence" in the digital world can create mass panic and cause damaging and devastating consequences in the real world, distorting evidence and impeding the response efforts and activities of health care workers and public health systems [17].

Additionally, the economic impact of COVID-19 on companies and countries were discussed in several tweets. Tweeters might talk about the economic impact of COVID-19 due to, for example, temporary closures of major fast-food chains and retailers (eg, McDonald's, KFC, Apple, and Adidas) [18], decreases in auto sales, drops in oil demand, production delays such as with the iPhone, the canceling or postponing of sporting events such as the Formula One World Championship, or decreases in airline revenues due to flight cancellations [18,19]. It has been estimated that the spread of COVID-19 could cost the worldwide economy a total of US $2.7 trillion [20]. The last impact of COVID-19 discussed by Twitter users was travel. This topic might have been common because most countries have banned travel from and to countries that confirmed the presence of COVID-19 inside their borders.

Tweets also focused on two possible sources of the coronavirus: the eating of meat and a Chinese biological military laboratory. Tweeters mentioned two main methods used to decrease the spread of COVID-19: masks and quarantine. The first method (masks) was discussed frequently on Twitter mainly due to the face mask shortage reported in several countries (eg, China, the United Kingdom, and the United States). The quarantine was a common topic in tweets because it was the first step that countries applied to control the outbreak of COVID-19.

Practical and Research Implications

Practical Implications

Research shows that crisis response activities in reality and online are becoming increasingly "simultaneous and intertwined" [21]. Social media provides a lucrative opportunity to spread and disseminate public health knowledge and information directly to the public [22]. However, social media can also be a powerful weapon and, if not used appropriately, can be destructive to public health efforts, especially during a public health crisis.

Therefore, more efforts are needed to build national and international detection and surveillance systems of diseases by examining online content published through the World Wide Web, including social media. There is a need for stronger and more proactive public health presence on social media. Governments and health systems should also "listen" or monitor the tweets from the public that relate to health, especially in a time of crisis, to help inform policies related to public health (eg, social distancing and quarantine) and supply chains among many others.

Research Implications

The global COVID-19 outbreak and its wild spread across countries demonstrates the need for more vigilant and timely responses aided by the research community. This was not the focus of this study, but future studies should investigate the spread of "fake news" in combination with infectious disease outbreaks [23]. Moreover, there is a need for providing access to a core corpus of social media posts available to the scientific and public health community while maintaining privacy. Additional work is necessary for multilingual sentiment analysis on social media platforms, as most research efforts have been devoted to English-language data [24], including this study. It could also be useful for future studies to consider longitudinal, multilingual sentiment analysis in addition to concurrent analysis of infectious disease outbreaks on different social media platforms, if feasible.

Strengths and Limitations

Several strengths and limitations can be attributed to this study analyzing tweets related to the recent COVID-19 outbreak. In this study, no geographical restrictions were applied on the tweets analyzed considering the worldwide spread of the disease. However, the study only analyzed tweets in the English language, which may limit the generalizability of the findings about this worldwide outbreak. In addition, given that the Twitter standard search API does not allow researchers to obtain tweets posted more than 1 week ago [25], we could not get COVID-19-related tweets posted before February 2, 2020. Thus, the findings may not be generalizable to that period. Moreover, this study could not collect tweets from accounts marked as private. Therefore, findings may not represent all the topics discussed by users on Twitter related to COVID-19. Only posts on Twitter were analyzed in this study, thereby, our findings may not be generalizable to other social media platforms. Furthermore, the findings reported in this study are limited to only those that have access to and use Twitter. Therefore, caution is advised before assuming the generalizability of the results, as Twitter is not used by everyone in the population.

Conclusion

The COVID-19 pandemic has been affecting many health care systems and nations, claiming the lives of many people. As a vibrant social media platform, Twitter projected this heavy toll through the interactions and posts people made related to COVID-19. It is clear that coordinating public health crisis response activities in the real world and online is paramount, and should be a top priority for all health care systems. We need to build more national and international detection and surveillance systems to detect the spread of infectious diseases and combat the fake news that is usually accompanied by these diseases.

Acknowledgments

The publication of this article was funded by the Qatar National Library.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Latent Dirichlet allocation output.
[PDF File (Adobe PDF File), 241 KB - jmir_v22i4e19016_app1.pdf]

Multimedia Appendix 2
Word cloud.
[PDF File (Adobe PDF File), 255 KB - jmir_v22i4e19016_app2.pdf]

Multimedia Appendix 3
Associated terms for each topic.
[PDF File (Adobe PDF File), 182 KB - jmir_v22i4e19016_app3.pdf]

Multimedia Appendix 4
Descriptive statistics for sentiment and interaction analysis.
[XLSX File (Microsoft Excel File), 35 KB - jmir_v22i4e19016_app4.xlsx]

References

1. Smith KF, Goldberg M, Rosenthal S, Carlson L, Chen J, Chen C, et al. Global rise in human infectious disease outbreaks. J R Soc Interface 2014 Dec 06;11(101):20140950 [FREE Full text] [doi: 10.1098/rsif.2014.0950] [Medline: 25401184]
2. Cui J, Li F, Shi Z. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2019 Mar;17(3):181-192 [FREE Full text] [doi: 10.1038/s41579-018-0118-9] [Medline: 30531947]
3. Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen K. Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov 2016 May;15(5):327-347 [FREE Full text] [doi: 10.1038/nrd.2015.37] [Medline: 26868298]
4. World Health Organization. 2020 Jan 21. Novel coronavirus (2019-nCoV) situation report - 1 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4
5. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280. [doi: 10.1111/tmi.13383] [Medline: 32052514]
6. World Health Organization. 2020 Apr 07. Coronavirus disease 2019 (COVID-19) situation report – 78 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200407-sitrep-78-covid-19.pdf?sfvrsn=bc43e1b_2
7. Jordan S, Hovet S, Fung I, Liang H, Fu K, Tse Z. Using Twitter for public health surveillance from monitoring and prediction to public response. Data 2018 Dec 29;4(1):6. [doi: 10.3390/data4010006]
8. Shah Z, Surian D, Dyda A, Coiera E, Mandl KD, Dunn AG. Automatically appraising the credibility of vaccine-related web pages shared on social media: a Twitter surveillance study. J Med Internet Res 2019 Nov 04;21(11):e14007. [doi: 10.2196/14007] [Medline: 31682571]
9. 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 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
10. Shah Z, Dunn AG. Event detection on Twitter by mapping unexpected changes in streaming data into a spatiotemporal lattice. IEEE Trans Big Data 2019:1-1. [doi: 10.1109/tbdata.2019.2948594]
11. Steffens MS, Dunn AG, Wiley KE, Leask J. How organisations promoting vaccination respond to misinformation on social media: a qualitative investigation. BMC Public Health 2019 Oct 23;19(1):1348. [doi: 10.1186/s12889-019-7659-3] [Medline: 31640660]
12. 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 May 08;21(5):e12881. [doi: 10.2196/12881] [Medline: 31344669]
13. Blei DM, Lafferty JD. Topic models. In: Text Mining. London, UK: Chapman and Hall/CRC; 2009:101-124.
14. Dyda A, Shah Z, Surian D, Martin P, Coiera E, Dey A, et al. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. Hum Vaccin Immunother 2019;15(7-8):1488-1495 [FREE Full text] [doi: 10.1080/21645515.2019.1596712] [Medline: 30978147]
15. Centers for Disease Control and Prevention. 2020. Prevention & treatment URL: https://www.cdc.gov/coronavirus/2019-ncov/about/prevention-treatment.html [accessed 2020-02-23]
16. The Lancet Infectious Diseases. Challenges of coronavirus disease 2019. Lancet Infect Dis 2020 Mar;20(3):261 [FREE Full text] [doi: 10.1016/S1473-3099(20)30072-4] [Medline: 32078810]
17. Vogel L. Viral misinformation threatens public health. CMAJ 2017 Dec 18;189(50):E1567. [doi: 10.1503/cmaj.109-5536] [Medline: 29255104]

18. Biron B, Zhu Y. Business Insider. 2020 Feb 07. Major fast-food chains and retailers in China are shutting their doors as the deadly coronavirus continues to spread. Here's a list of closures URL: https://www.businessinsider.com/coronavirus-fears-mcdonalds-starbucks-close-2020-1 [accessed 2020-02-23]
19. Hutt R. World Economic Forum. 2020 Feb 17. The economic effects of COVID-19 around the world URL: https://www.weforum.org/agenda/2020/02/coronavirus-economic-effects-global-economy-trade-travel/ [accessed 2020-02-23]
20. Orlik T, Rush J, Cousin M, Hong J. Bloomberg. 2020 Mar 06. Coronavirus Could Cost the Global Economy $2.7 Trillion. Here's How URL: https://www.bloomberg.com/graphics/2020-coronavirus-pandemic-global-economic-risk/
21. Veil SR, Buehner T, Palenchar MJ. A work‐in‐process literature review: incorporating social media in risk and crisis communication. Journal of Contingencies and Crisis Management 2011 Jun 01;19(2):110-122. [doi: 10.1111/j.1468-5973.2011.00639.x]
22. Breland JY, Quintiliani LM, Schneider KL, May CN, Pagoto S. Social media as a tool to increase the impact of public health research. Am J Public Health 2017 Dec;107(12):1890-1891. [doi: 10.2105/AJPH.2017.304098] [Medline: 29116846]
23. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
24. Dashtipour K, Poria S, Hussain A, Cambria E, Hawalah AYA, Gelbukh A, et al. Multilingual sentiment analysis: state of the art and independent comparison of techniques. Cognit Comput 2016;8:757-771 [FREE Full text] [doi: 10.1007/s12559-016-9415-7] [Medline: 27563360]
25. Twitter. Search tweets URL: https://developer.twitter.com/en/docs/tweets/search/overview [accessed 2020-04-06]

Abbreviations

API: application program interface
COVID-19: coronavirus disease
LDA: latent Dirichlet allocation

Edited by G Eysenbach; submitted 31.03.20; peer-reviewed by R Zowalla, E Da Silva; comments to author 02.04.20; revised version received 09.04.20; accepted 09.04.20; published 21.04.20.

Please cite as:
Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z
Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study
J Med Internet Res 2020;22(4):e19016
URL: http://www.jmir.org/2020/4/e19016/
doi:10.2196/19016
PMID:32287039

Original Paper

Public Reactions to the Cigarette Control Regulation on a Chinese Microblogging Platform: Empirical Analysis

Wanting Wen[1,2*], MSc; Zhu Zhang[1,3*], PhD; Ziqiang Li[1], MSc; Jiaqi Liang[1], PhD; Yongcheng Zhan[4], PhD; Daniel D Zeng[1,2,3], PhD; Scott J Leischow[5], PhD

[1]State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3]Shenzhen Artificial Intelligence and Data Science Institute (Longhua), Shenzhen, China
[4]Orfalea College of Business, California Polytechnic State University, San Luis Obispo, CA, United States
[5]College of Health Solutions, Arizona State University, Phoenix, AZ, United States
[*]these authors contributed equally

Corresponding Author:
Wanting Wen, MSc
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation
Chinese Academy of Sciences
95 Zhongguancun East Road
Beijing, 100190
China
Phone: 86 13488793849
Email: wanting.wen@ia.ac.cn

Abstract

Background: On January 1, 2019, a new regulation on the control of smoking in public places was officially implemented in Hangzhou, China. On the day of the implementation, a large number of Chinese media reported the contents of the regulation on the microblog platform Weibo, causing a strong response from and heated discussion among netizens.

Objective: This study aimed to conduct a content and network analysis to examine topics and patterns in the social media response to the new regulation.

Methods: We analyzed all microblogs on Weibo that mentioned and explained the regulation in the first 8 days following the implementation. We conducted a content analysis on these microblogs and used social network visualization and descriptive statistics to identify key users and key microblogs.

Results: Of 7924 microblogs, 12.85% (1018/7924) were in support of the smoking control regulation, 84.12% (6666/7924) were neutral, and 1.31% (104/7924) were opposed to the smoking regulation control. For the negative posts, the public had doubts about the intentions of the policy, its implementation, and the regulations on electronic cigarettes. In addition, 1.72% (136/7924) were irrelevant to the smoking regulation control. Among the 1043 users who explicitly expressed their positive or negative attitude toward the policy, a large proportion of users showed supportive attitudes (956/1043, 91.66%). A total of 5 topics and 11 subtopics were identified.

Conclusions: This study used a content and network analysis to examine topics and patterns in the social media response to the new smoking regulation. We found that the number of posts with a positive attitude toward the regulation was considerably higher than that of the posts with a negative attitude toward the regulation. Our findings may assist public health policy makers to better understand the policy's intentions, scope, and potential effects on public interest and support evidence-based public health regulations in the future.

(J Med Internet Res 2020;22(4):e14660) doi:10.2196/14660

KEYWORDS

cigarette smoking; regulations; social media; information networks

Introduction

Background

China has been the largest consumer market of cigarettes in the world. In China, approximately 27.7% of adults aged between 15 and 69 years and 1% of adolescents aged between 13 and 15 years are smokers [1]. According to a World Health Organization (WHO) report, more than 1 million people die from smoking-related diseases every year in China, whereas 740 million people including 180 million children suffer from secondhand smoking [2]. In addition to the high rates of cigarette use, electronic cigarettes (e-cigarettes) have become increasingly popular in China, especially among adolescents [3]. E-cigarette sellers claim that e-cigarettes are less harmful than tobacco and help to quit smoking [4]. Some research shows that e-cigarettes appear to be effective when used by smokers as an aid to quit smoking, and the hazard to health arising from long-term vapor inhalation from the e-cigarettes available is likely to be much less than the harm from smoking tobacco [5,6]. However, some research also shows the risks associated with e-cigarette use. The inhaled aerosols of e-cigarettes contain numerous potential toxins, some of which could be dangerous for health with long-term use [7], and inhalation of the non-nicotine–derived chemicals present in e-cigarette aerosols is actually harmful to small airway epithelium cells and alveolar macrophages in e-cigarette smokers. As the respiratory system of minors is not fully developed, inhalation of such aerosols can have adverse effects on lung function [8-10]. Improper use may also lead to various safety risks such as nicotine poisoning, and it is possible that youth using a new generation of e-cigarettes will become addicted [9]. Widespread smoking and tobacco use and the uncertainty of e-cigarette effects, including the potential of youth addiction to both cigarettes and e-cigarettes, raise severe public health concerns.

To alleviate the above challenges, the government agencies in China have taken actions to increase tobacco control regulations. In 2005, China ratified the WHO's Framework Convention on Tobacco Control to reduce the harms from smoking [11]. On November 24, 2014, the Health and Family Planning Commission of China drafted the "Regulations on the Control of Smoking in Public Places (Draft for Review)" [12] to solicit opinions from the public. This is the first time that China has proposed to formulate administrative regulations to comprehensively control tobacco throughout the country [12]. According to the statistics of the Health and Family Planning Commission of China, until the end of 2016, 18 cities in China had formulated local smoke-free environmental laws and regulations, covering one-tenth of the total population of China [13]. In recent years, tobacco control efforts in China have been strengthened and tobacco taxes have increased. As a result, smokers began to choose e-cigarettes as a nicotine delivery tool [14]. With the prominent usage of e-cigarettes among teenagers and more studies being conducted on the harm of e-cigarettes on minors, China issued the "Notice on Prohibiting the Sale of Electronic Cigarettes to Minors" in 2018 [10], which was the first of several regulations on e-cigarette control in China, to protect minors from e-cigarettes.

Social media platforms play an important role in spreading news and shaping the public's attitudes toward novel health issues [15,16]. Different groups of users can use different social media platforms to discuss smoking topics [17]. Public health community and policy makers can use these platforms for public health surveillance [18-20]. For example, for a newly issued policy, public health officials can use these platforms to judge the public's understanding of the policy to determine whether the policy needs to be further interpreted and publicized. By exploring the user-generated content on social media, researchers can study public opinions on cigarette use [21] and understand the influence of tobacco control policies on the smoking cessation intentions [22,23]. There have been some studies about public health on social media in the United States [17-20]. However, related research is sparse in China. Only a few research studies have focused on social media studies regarding the public health strategies about cigarette and e-cigarette [24-26]. For example, Cui et al [25] explored the nature and extent of discussions around the electronic nicotine delivery systems in Chinese social media, which have the power to influence a massive audience. Jin et al [26] evaluated the effectiveness of microblogging in tobacco control communication and provided new media era strategies for tobacco control. Thus, there is an urgent need to conduct social media–driven public opinion studies on smoking control policies.

Objectives

On January 1, 2019, the newly revised Regulations on the Control of Smoking in Public Places officially took effect in Hangzhou. The regulations expand the scope of application, strictly control the scope of smoking places, adopt a multisectoral supervision model, increase smoking control measures, and improve legal responsibilities [27]. In addition, e-cigarette control has been added to the new regulations [27], which fills the legislation gap of e-cigarette control in the local tobacco regulation of China. Owing to the regulations' significance in tobacco control of China, this study aimed to understand the public's responses to the revision and enforcement of tobacco control regulations in China by analyzing the messages about the regulations on Sina Weibo (ie, the largest microblogging platform in China). This study combines content analysis and network analysis to understand the public's perception and attitude toward the implementation of the regulations as well as the influence of the regulations on social media. To the best of our knowledge, this is the first study to leverage social media to analyze the impact of tobacco control regulations including e-cigarettes in China. This study will facilitate timely understanding of the impact of the implementation of the tobacco control regulations and the public's responses to the first batch of regulation of e-cigarettes in China for public health community and policy makers.

Methods

Data Collection

We collected microblogs about the smoking control regulations in Hangzhou on Sina Weibo by developing web crawlers with a programming language called Python (Python Software

Foundation) and a Python package called selenium [28]. The data were collected from January 1, 2019, to January 8, 2019. According to our observation, the data reached the peak on the second day and then decreased rapidly. During the week after January 8, 2019, we collected only 35 microblogs, which is only 0.44% (35/7924) of the data collected from January 1, 2019, to January 8, 2019, so the data of about 1 week were sufficient for this study. The reason for choosing Sina Weibo was that it has a large user base, with 376 million monthly active users and 165 million daily active users [29]. We first collected postings by 4 keywords, including "杭州 控烟令 (Hangzhou Smoking Control Order)," "杭州 禁烟令 (Hangzhou No Smoking Order)," "杭州 禁烟 (Hangzhou No Smoking)," and "杭州 控烟 (Hangzhou Smoking Control)." The keywords were very commonly used by Sina Weibo users in their discussions within the week after the regulation was announced. Sina Weibo has two approaches to forwarding postings. In the first approach, forwarded postings are visible to everyone; however, in the other approach, postings are only visible to some users allowed. For each relevant microblog, we also collected their reposted microblogs visible to all netizens.

We collected 7776 original microblogs by using our designed web crawlers. As to each microblog, we collected its metadata including its posted time, author name, originality, the number of times being forwarded, the number of times being liked, the number of comments, the number of times being praised, and the hyperlink of the original posting if it was not original. For some microblogs that we did not collect by web crawlers, we extracted contents and usernames based on the specific structure of their forwarding microblogs. For example, for a microblog, the user of the microblog is "Xiao Xia171," and the content of the microblog is "Support, I hope that Anji will also upgrade the smoking ban. //@Mu Jian: Support～." After the extraction process, we got two microblogs: "Xiao Xia171: Support, I hope that Anji will also upgrade the smoking ban" and "Mu Jian: Support～." The second microblog "Mu Jian: Support～," which we did not get by web crawlers, was added to our dataset. In this way, we got another 148 uncollected microblogs. In the end, we retrieved 7924 postings of 7255 users. Then, we collected the users' profile information, including the number of followers, the number of followees, the number of posts, and the verification label, which indicates a well-known celebrity or organization.

Content Analysis

To understand the public's perception of the tobacco control regulations in Hangzhou, we followed the coding practice as mentioned in Harris et al [30] to code our collected 7924 postings with different stances and themes. First, the two authors read all microblogs and proposed a list of themes independently, and then they came up with the final categories of themes by discussion. Second, the two authors labeled each post with stances and themes independently and reached a consensus for all posts by discussion. Finally, we developed five themes to code the posts, including regulation-related news sharing, policy development, policy implementation, related benefits, and health and science. Each theme contained several subthemes. Each microblog was assigned only one theme and one of its subthemes. The meanings of these themes and subthemes are depicted in Table 1. In total, we proposed a list of 5 themes and 11 subthemes. The Cohen kappa statistic was used to test the reliability of coding the themes of the posts [31]. The kappa statistic was 0.88, which indicates high reliability.

To gain further understanding of the public's attitude toward the new regulations, we classified each microblog into three types of stances: Pro-Regulation (Pro-R), Anti-Regulation (Anti-R), and Discussion (Discuss). Pro-Regulation was used to describe the messages that promote the new regulations. Anti-regulation meant that the messages of this type expressed negative attitudes toward the new regulation. Any messages about the regulations that cannot be classified as either Pro-Regulation or Anti-Regulation were coded as "Discussion." Similarly, the kappa coefficient for coding the stances was 0.90, which indicates high reliability.

Table 1. The description of themes and subthemes used for coding.

Themes	Description of themes	Subthemes	Description of subthemes
Regulation-related news sharing	Posting or forwarding regulation content	Regulations; discussions about the regulations	Direct posted or forwarded policy content without comments; comments on the details of the regulations
Policy development	Discussing the rationality of policy development	Scope of tobacco control; sources of production and sale	Discussion on the tobacco control scope; discussion on whether to control the sources of tobacco production and sale
Policy implementation	Effect of policy implementation and punitive measures	Effect of the policy; punishment	Discussion on the effect of policy implementation; discussion on what kind of punishment measures to use
Related benefits	Reasons for the introduction of new regulations, the impact of taxes or e-cigarettes on traditional cigarettes	E-cigarette–related benefits; tax-related benefits	E-cigarettes may affect the market of traditional cigarettes; the impact of the escalation of smoking control on taxes
Health and science	Health and scientific content related to e-cigarettes and traditional cigarettes	Addiction and quitting smoking; harms of e-cigarettes; harms of traditional cigarettes and secondhand smoke	Discussion on addiction and quitting smoking; impact of e-cigarettes on health; impact of traditional cigarettes and secondhand smoke on health

Network Analysis

In public health, network analysis has been used to study disease transmission, information transmission, the influence of social networks on health behavior [32-36], and so on. In this study, we used network analysis and visualization methods to analyze the diffusion pattern of microblogs related to smoking control. Specifically, we constructed a transmission network in which nodes represent microblog users and edges represent reposting relationships between users. The transmission network is directed, and its direction indicates the flow of information. For example, if user B reposts a microblog of user A, the directed edge from A to B is established. In such a directed network, the outdegree centrality indicates the number of users who repost the posts of the target user.

Highly central network members aid in the dissemination of information [37]. Therefore, we examined central network members in the transmission network to understand their role in information dissemination on smoking control. The transmission network was constructed based on the dataset of reposts we collected. The number of forwarded microblogs collected from a third-party perspective will be less than the actual number of forwarded microblogs because Sina Weibo allows users to specify some users who are allowed to read when forwarding microblogs and the unspecified users cannot see the forwarded microblogs. When we evaluated the key members in the transmission network, we considered both outdegree centrality of the transmission network and network members' microblog forwarding number. We defined a key member as a network node whose outdegree centrality is more than 35, and the forwarding number of every microblog that this member has posted is more than 100. Most microblogs (80%) were posted or reposted by these key members.

In addition to outdegree centrality, we also measured the key original posts with the following metrics: the number of forwarded microblogs, the number of collected forwarded microblogs, effective transmission rate, effective support rate, and effective opposition rate.

The aim of the transmission network was to analyze the transmission of stance from microblog forwarding. If the collected forwarded microblogs did not contain the comment information, it could not reflect the netizens' stance on the regulations, so it was not what we should focus on. Effective forwarded microblogs meant the forwarded microblogs we collected contained users' support (stance "Pro-R") or opposition (stance "Anti-R") information. In contrast, ineffective forwarded microblogs meant the microblogs that were just forwarded without any comments or with unclear attitudes (stance "Discuss"). Effective transmission rate was defined as the number of effective forwarded microblogs divided by the number of collected forwarded microblogs. Effective supportive (opposition) rate was defined as the number of effective forwarded microblogs with stance "Pro-R" (stance "Anti-R) divided by the number of effective forwarded microblogs.

Results

Microblog Content

Among the 7924 microblogs, 12.85% (1018/7924) were in support of the smoking control regulations, 1.31% (104/7924) were opposed to the regulations, 84.12% (6666/7924) were neutral, and 1.72% (136/7924) were irrelevant to the regulations. Of the five themes identified, the most discussed theme was regulation-related news sharing with 83.22% (6594/7924) posts, and the least discussed theme was related benefits with only 0.98% (78/7924).

Multimedia Appendix 1 shows the summary of the collected microblogs by theme and stance. Here, we mention the most discussed subtopics for each theme. For the theme of regulation-related news sharing, there were 72.38% (5735/7924) of microblogs discussing subtopic 1, of which all posts were neutral for the Hangzhou regulations; 10.84% (859/7924) of posts were about subtopic 2, of which 59.3% (509/859) posts supported the regulations. For the theme of policy development, there were 3.41% (270/7924) of microblogs discussing subtopic 1, of which 88.9% (240/270) posts supported the regulations; 0.92% (73/7924) of posts were about subtopic 2, of which 58% (42/73) posts supported the regulations. For the theme of health and science, 2.51% (199/7924) of microblogs were about subtopic 1, of which 85.4% (170/199) were neutral to the regulations; 3.69% (292/7924) of posts were about subtopic 2, of which 71.2% (208/292) were neutral; 1.64% (130/7924) of posts were about subtopic 3, of which 50.0% (65/130) supported the regulations in Hangzhou. As to the theme policy implementation, there were 1.73% (137/7924) of microblogs discussing subtopic 1, of which 55.5% (76/137) were neutral to the regulations; 0.37% (29/7924) of posts were about subtopic 2, of which 52% (15/29) were neutral. For the theme of related benefits, 0.38% (30/7924) of microblogs discussed subtopic 1, of which 77% (23/30) were neutral; 0.61% (48/7924) of posts were about subtopic 2, of which 88% (42/48) were neutral.

In these 7924 microblogs, there were 14.75% (1169/7924) original microblogs and 85.25% (6755 /7924) forwarded microblogs. For the original microblogs, the most discussed theme was regulation-related news sharing (879/1169), and of these, 92.7% (815/879) were neutral to the regulations. Health and science (169/1169) was the second most popular topic, and of these, 89.3% (151/169) supported the regulations. As to the forwarded microblogs, the most discussed theme was also regulation-related news sharing (6594/6755), and of these, 78.89% (5202/6594) were neutral. The average number of posts for one Weibo user was 1.1, and the user with the most posts was Lawyer Zhiming Zhuang. Lawyer Zhuang's views on the Hangzhou regulations were mixed, but the theme he participated in was mainly policy development.

Transmission Network

The distribution of antipolicy and propolicy microblogs and repost microblogs is shown in Figure 1. Most microblogs were sent between January 1, 2019, and January 2, 2019. The propolicy microblogs and antipolicy microblogs peaked on January 2, but the former were far greater than the latter.

Throughout the week, there were fewer antipolicy microblogs and repost microblogs.

According to the aforementioned network analysis method, 12 key accounts involving 13 key original microblogs were extracted from the data, named RP-1 to RP-13. Among them, account "@CCTV News" had two related original microblogs (RP-2 and RP-9). Table 2 shows the 13 key original microblogs, and detailed content is presented in Table 3. The new regulation was implemented, starting January 1, 2019, and the authoritative internet media took the lead in reporting (RP-1 and RP-3). On the next day, several media agencies followed up the new regulation report (RP-2, RP-4, and RP-5), and some authoritative health public accounts began to popularize cigarette hazards (RP-11, RP-12, and RP-13), which led to a lot of discussion. After a few days, local media started to promote new regulations (RP-8), and authoritative central media popularized the dangers of e-cigarettes (RP-9 and RP-10) on January 5, 2019, which brought the second peak of discussion. Then, local media continued to promote new regulations (RP-7 and RP-8) until January 8, 2019. During the week after January 8, 2019, we collected only 35 microblogs, which is only 0.44% (35/7924) of the data collected from January 1, 2019, to January 8, 2019, so we did not analyze the data after January 8, 2019.

Figure 1. Timeline showing the distribution of antipolicy, propolicy microblogs, and repost microblogs.

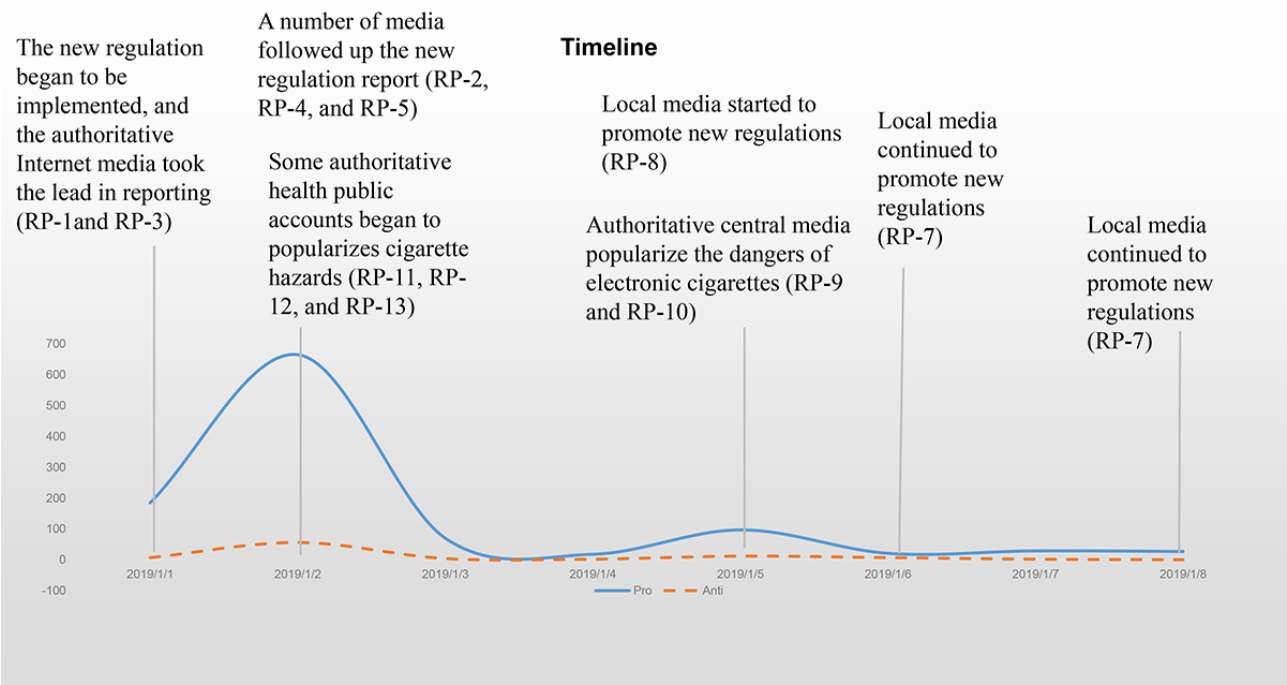


Table 2. Summary of the key original microblogs.

Categories, type of account, and account[a]	Content	Collected forwarded rate, n/N (%)[b]	Effective transmission rate, n (%)[c]	Effective support rate, n (%)[d]	Effective opposition rate, n (%)[d]
Regulation-related news sharing					
News media					
The Paper	RP-1	1651/4584 (36.02)	190 (11.51)	178 (93.7)	12 (6.3)
CCTV News	RP-2	414/469 (88.3)	151 (36.5)	148 (98.0)	3 (2.0)
NewsHead	RP-3	237/262 (90.5)	94 (39.7)	87 (93)	7 (7)
People's network	RP-4	157/165 (95.2)	32 (20.4)	29 (91)	3 (9)
China News Network	RP-5	87/111 (78.4)	21 (24)	18 (86)	3 (14)
Total	N/A[e]	2546/5591 (45.54)	488 (19.17)	460 (94.3)	28 (5.7)
Local news media					
Hangzhou' s Big Popular Life	RP-6	155/1221 (12.69)	21 (13.5)	21 (100)	0 (0)
Hot Events in Hangzhou	RP-7	97/1270 (7.64)	0 (0)	0 (0)	0 (0)
Hangzhou Information Headlines	RP-8	45/952 (4.72)	1 (2)	1 (100)	0 (0)
Total	N/A	297/3443 (8.63)	22 (7.3)	22 (100)	0 (0)
Health promotion					
News media					
CCTV NEWS	RP-9	1120/1315 (85)	104 (9.29)	86 (82.7)	18 (17.3)
Local news media					
Hangzhou Top Information List	RP-10	122/1034 (12)	1 (0.8)	0 (0)	1 (100)
Self-media Weibo accounts					
Clove doctor	RP-11	470/593 (79.3)	62 (13.2)	55 (89)	7 (11)
Health preservation - Lao Yang	RP-12	37/133 (27.8)	1 (3)	1 (100)	0 (0)
Rice cake mother	RP-13	73/112 (65.2)	3 (4)	3 (100)	0 (0)
Total	N/A	580/838 (69.2)	66 (11.4)	59 (89)	7 (11)

[a]See Multimedia Appendix 2 for the Weibo usernames in Chinese and English.

[b]The N value for each microblog reflects the number of retweets.

[c]The denominator of the n values is the corresponding n value of the Collected forwarded rate column.

[d]The denominator of the n values is the corresponding n values of the Effective transmission rate column.

[e]N/A: not applicable.

Table 3. The detailed content.

Number of post	Content[a]
RP-1	• Hangzhou is Upgrading Smoking Control Ordinance: Electronic cigarettes are included in the scope of smoking ban; the maximum penalty for illegal smoking is 20,000 RMB. • The newly revised Hangzhou Regulations on Smoking Control in Public Places was formally implemented on January 1. Hangzhou is one of the earliest cities in China to carry out tobacco control legislation, and tobacco control work has been steadily and firmly promoted. the latest "smoking control order" has expanded the scope of regulation, strictly controlled the scope of smoking places, and put forward the requirements of smoking ban in outdoor areas of some public places. the implementation of a multi-sectoral supervision model has strengthened law enforcement. electronic cigarettes are also included in the smoking ban. If smokers are found to be smoking in non-smoking areas, they may be asked to stop smoking or leave the place immediately. If not dissuaded, citizens can call "12345" to report complaints. after registering and accepting complaints, the cases will be handed over to the corresponding tobacco control regulatory authorities for disposal in accordance with the division of duties of tobacco control supervision.
RP-2	• Hangzhou Smoking Control Ordinance: Electronic cigarettes are included in the smoking ban with a maximum penalty of 20,000 RMB. • The newly revised Hangzhou Regulations on Smoking Control in Public Places were formally implemented on January 1. it is clearly stipulated that smoking is prohibited in indoor public places, indoor workplaces and public transport. therefore, smoking is prohibited even in one's own office. smoking places are prohibited not only from lighting tobacco products and smoking traditional cigarettes, but also from smoking electronic cigarettes.
RP-3	• Hangzhou Upgrading Smoking Control Ordinance: Electronic cigarettes are included in the scope of smoking ban, the maximum penalty for illegal smoking is 20,000 RMB. • The newly revised Hangzhou Regulations on Smoking Control in Public Places was formally implemented on January 1. the latest "smoking control order" has expanded the scope of application, strictly controlled the scope of smoking places, and put forward the requirements of smoking ban in outdoor areas of some public places. the implementation of a multi-sectoral supervision model has strengthened law enforcement. electronic cigarettes are also included in the smoking ban.
RP-4	• Hangzhou Upgrading Smoking Control Ordinance: Electronic cigarettes were included in the smoking ban with a maximum penalty of 20,000 RMB. • The newly revised Hangzhou Regulations on Smoking Control in Public Places were formally implemented on January 1. the latest "smoking control order" has set a buffer period for smoking places. electronic cigarettes are included in the scope of smoking ban. if operators and managers fail to fulfill their duties of smoking control, they can be fined up to 20,000 yuan.
RP-5	• Hangzhou Upgraded Smoking Control Ordinance: Electronic cigarettes were included in the smoking ban, with a maximum penalty of 20,000 RMB for illegal smoking. • The newly revised Hangzhou Regulations on Smoking Control in Public Places was formally implemented on January 1. the latest "smoking control order" has expanded the scope of application, strictly controlled the scope of smoking places, and put forward the requirements of smoking ban in outdoor areas of some public places. the implementation of a multi-sectoral supervision model has strengthened law enforcement. electronic cigarettes are also included in the smoking ban.
RP-6	• Hangzhou Upgrading Smoking Control Ordinance: Electronic cigarettes were included in the smoking ban with a maximum penalty of 20,000 RMB. • The newly revised Hangzhou Regulations on Smoking Control in Public Places were formally implemented on January 1. The latest "smoking control order" has set a buffer period for smoking places. electronic cigarettes are included in the scope of smoking ban. if operators and managers fail to fulfill their duties of smoking control, they can be fined up to 20,000 yuan.
RP-7	• # Hangzhou Hot Events # The newly revised Regulations on Smoking Control in Public Places of Hangzhou have been formally implemented. the latest "smoking control order" has expanded the scope of application, strictly controlled the scope of smoking places, and put forward the requirements of smoking ban in outdoor areas of some public places. the implementation of a multi-sectoral supervision model has strengthened law enforcement. electronic cigarettes are also included in the smoking ban.
RP-8	• Hangzhou Upgrading Smoking Control Ordinance: Electronic cigarettes are included in the scope of prohibition, with a maximum penalty of 20,000 RMB for illegal smoking. • The newly revised Hangzhou Regulations on Smoking Control in Public Places have been formally implemented. the latest "smoking control order" has expanded the scope of application, strictly controlled the scope of smoking places, and put forward the requirements of smoking ban in outdoor areas of some public places. the implementation of a multi-sectoral supervision model has strengthened law enforcement. electronic cigarettes are also included in the smoking ban.
RP-9	• This kind of smoke is harmful! • The newly revised Hangzhou Regulations on Smoking Control in Public Places came into effect on January 1. no smoking places not only prohibit lighting tobacco products and smoking traditional cigarettes, but also prohibit smoking electronic cigarettes. why are electronic cigarettes banned?

Number of post	Content[a]
RP-10	• # Hangzhou Headline. This kind of smoke, harmful! • The newly revised Hangzhou Regulations on Smoking Control in Public Places came into effect on January 1. no smoking places not only prohibit lighting tobacco products and smoking traditional cigarettes, but also prohibit smoking electronic cigarettes. why are electronic cigarettes banned?
RP-11	• Smoking is harmful to health. what about electronic cigarettes? • Let's do a science popularization. • # Hangzhou electronic cigarettes are included in the smoking ban#
RP-12	• # Hangzhou electronic cigarettes are included in the smoking ban# A video tells you how harmful smoking is. after watching it, everyone decides to quit smoking. [Health preservation—Lao Yang's Weibo Video Link] @ Health preservation—LaoYang
RP-13	• # Hangzhou electronic cigarettes are included in the smoking ban# Comprehensive tobacco control from family to society is the right way for parents to protect their children from second-hand smoke. smoke-free and healthy environment is the best gift we give our children. [Rice cake mother's Second Shot Video Link] @ Rice cake mother

[a]See Multimedia Appendix 3 for the detailed content in Chinese and English.

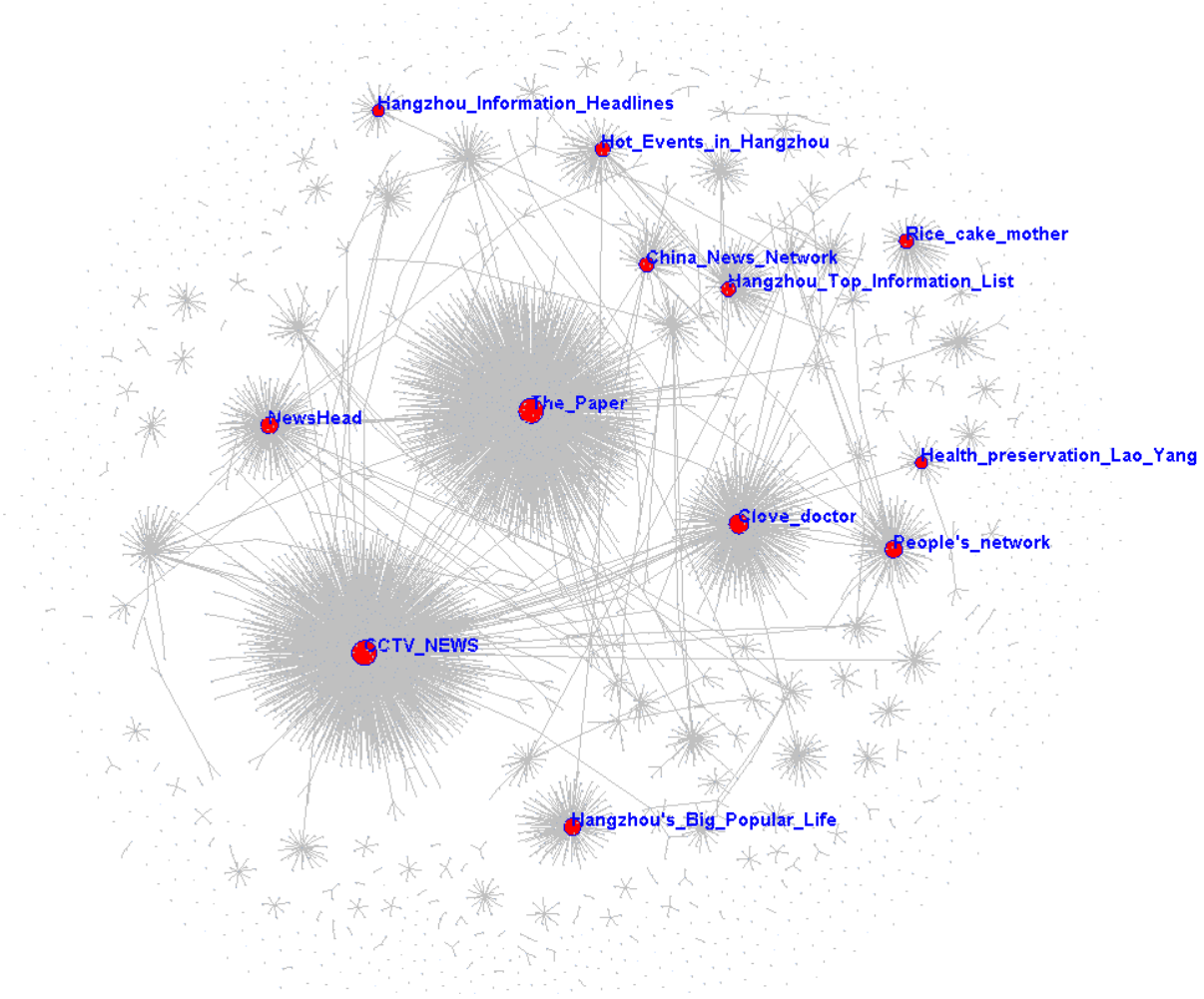
The microblogs in Table 2 can be divided into two themes: regulation-related news sharing and health. There were 8 microblogs about "regulation-related news sharing," and the news media and local news media accounts were the main communication accounts. For the microblogs about "regulation-related news sharing," the effective transmission rate of news media was high, and the effective support rate was high. The effective transmission rate of local news media was very low, but the effective support rate was high. This may be because the audiences of the local news media were mostly local users, and they were more inclined to use the comment function as a means of chatting with friends. The content of the comment was mostly the content of chatting between friends who did not want the content to be seen by others, so the effective transmission rate was low. There were 5 microblogs about health, and the corresponding accounts were news media, local news media, and self-media Weibo accounts. These microblogs were mainly about the dangers of e-cigarettes (RP-9, RP-10, and RP-11) and the harms of traditional cigarettes (RP-12) and secondhand smoke to young people (RP-13). The total effective transmission rate of these microblogs was medium, but the total effective opposition rate was high. Especially with regard to the health hazards of e-cigarettes (RP-9, RP-10, and RP-11), their opposition position was significantly higher than the other two microblogs (RP-12 and RP-13) in the same theme, resulting in high total effective opposition rate of the theme "health," which reflects the great differences among netizens in the impact of e-cigarettes on health.

The transmission network consisted of 6436 Weibo users who reposted at least one post from other users, and 819 isolated users who did not repost posts. There were 6330 asymmetric ties and 22 mutual ties in the network. A user usually has 1 to 3 other users who repost his or her posts, except for user "Lawyer Zhiming Zhuang" with 27 users reposting his posts.

Figure 2 shows the transmission network where node size represents outdegree: the larger a node, the higher the frequency of its posts being reposted. In the transmission network, there were 12 key Weibo users with 13 key original microblogs.

Microblogs RP-1 to RP-5 were related to a newspaper article detailing the differences between the latest antismoking policy and previous policies. It elaborated the buffer period to construct suitable smoking rooms for customers, multisector supervision mode, prohibition of e-cigarettes, the scope of smoking places (including indoor public places, indoor workplaces, public transportation, and some outdoor areas), and so on. Owing to the relatively earlier time of publication and rich content, the number of reposts reached 5591. The main categories of the reposted microblogs included discussion of the regulations, scope of tobacco control, effect of the policy, and sources of production and sale.

Figure 2. Transmission network with nodes sized by outdegree and the 12 key Weibo users shown in red.



Microblogs RP-6 and RP-8 outlined the significant differences between the latest antismoking policies and previous policies. Their accounts were the local information accounts. The total number of forwarding posts was 3473, and the number of collected forwarding posts was 300, but the effective rate was only 7%.

Microblogs RP-9 to RP-10 focused on the e-cigarettes mentioned in the regulations by using long images. Referring to a large amount of survey data, the microblogs illustrated the definition of e-cigarettes, the differences between e-cigarettes and combustible cigarettes, and whether e-cigarette use was really “harmless.” They also discussed that e-cigarettes should not be ignored and that e-cigarettes should be used to assist in quitting smoking according to doctors’ advice. The main categories of repost microblogs included harms of e-cigarettes, discussion of the regulations, taxes, harms of traditional cigarettes and secondhand smoke, addiction and quitting smoking, places and scopes of tobacco control, and interest issues. Microblog RP-11 contained a long image that referred to a journal article related to e-cigarette policy published by the American Heart Association. This microblog outlined the regulatory status of e-cigarettes around the world and analyzed in detail the composition of e-cigarettes and their health harms, which was reposted 593 times. The main categories of these repost microblogs were as follows: harms of e-cigarettes, discussion of the regulations, addiction and quitting smoking, and tobacco ban and quitting smoking.

Microblog RP-12 contained a video that simulated the health hazards of direct and secondhand smoke through an experiment. The microblog was reposted 133 times. The main category was harms of traditional cigarettes. Microblog RP-13 contained a Web hyperlink to a video about interviewing innocent children with regard to their views on secondhand smoking. It was reposted 112 times.

Discussion

Principal Findings

When the tobacco control regulations in Hangzhou were promoted on social media, the content of the regulations was highlighted by different types of social media accounts from several angles: the expansion of the regulation scope, the expansion of the scope of smoking control, the definition of multisectoral supervision, and the inclusion of e-cigarettes. For the temporal distribution of the collected data, the revised smoking control legislation took effect in Hangzhou on January 1, and the discussion peaked on January 2, when people began to rationally analyze the reasons and effects for smoking control.

Within a few days after the publication of the regulations, several discussions about the regulations took place around

several major accounts. Most users were just involved in the forwarding and discussion of the regulations (6119/7255, 84.34%), without explicit support or opposition. Among the 1043 users who explicitly expressed their support or opposition to the policy, a large proportion of users showed supportive attitudes (956/1043, 91.66%). Among the 965 users who showed supportive attitude, 273 users expressed their views on policy development. Some thought that the scope of the regulations should be extended to other cities, and some people thought that tobacco sales should be included in the regulations. A total of 480 users participated in the discussion of the regulations. They mainly focused on and supported three aspects of the regulations. First, the regulations strictly control the scope of tobacco control and clearly stipulate that smoking is prohibited in indoor public places, indoor workplaces, and public transportation. They are supported by most Weibo users. For example, one user posted, “No Smoking in indoor workplaces.” Second, the regulations also impose a smoking ban on outdoor areas in some public places. They also save people from the pain of passive smoking in some places outside. Another user posted, “The smoking ban in Hangzhou is worthy of praise! The regulations also impose smoking bans on outdoor areas in some public places.” Third, the regulations have incorporated e-cigarettes into the smoking ban and have received support from many users who hate e-cigarettes. For instance, a user posted, “In public, smoking e-cigarettes is also inappropriate. Support the ban on all cigarettes in public.”

In terms of topics of concern to netizens, as shown in the microblog content analysis results and Multimedia Appendix 1, Weibo users were generally concerned about (including support, opposition and discuss) policy implementation (n=165) and the impact of smoking and secondhand smoke on health (n=130). In particular, because of the introduction of e-cigarettes into the scope of supervision for the first time, netizens were also concerned about the hazards of e-cigarettes (n=286). Some Weibo users, especially users in Beijing and Guangzhou, expressed their hope that the policy could be extended to their cities (n=270) when forwarding microblogs.

The public needs to raise awareness about the motivations behind the regulations. Some netizens were worried that the tobacco control regulations would accelerate the increase in the tobacco tax rate, which would lead to an increase in the price of tobacco products, even though China's tobacco tax is far below the average in the world [38]. Therefore, to solve these concerns, more effort is needed to promote and explain the tobacco control regulations through social media.

The public health community and the government need to put more effort into popularizing the knowledge about e-cigarettes to enhance the public’s acceptance of e-cigarette control. Although many netizens believe that smoking leads to a health hazard, some people oppose the control of e-cigarettes and believe that e-cigarettes are not harmful at all or less harmful than cigarettes (n=13). Some netizens (n=73) believe that the control of e-cigarettes was added to the regulations because of the interest divergence between e-cigarettes and traditional cigarettes. They also believe that the government intends to impose high taxes on e-cigarettes, so they first introduced e-cigarette control in the regulatory policies.

The public needs to increase their trust in the enforcement of the regulations. Some users believe that the regulation is well formulated, but the most important thing is the implementation effect of the regulation. Some netizens (n=76) criticized that the previous smoking control regulations were ineffective, and the regulatory measures did not take effort. This reflects the fact that although Chinese netizens generally support the smoking control regulations in public places, they distrust the effectiveness of the corresponding policy implementation. Therefore, the government should pay attention not only to the rationality of policy formulation but also to the actual effect of policy implementation. In addition, policy makers can use social media to gain feedback on policy implementation to improve the quality of tobacco control.

Limitations

Some limitations should be considered when interpreting the results. First, more than 80% of Sina Weibo users are under the age of 30 years. The number of male users (56.3%) is slightly higher than that of female users (43.7%). Users from third- and fourth-tier cities account for more than 50% of Weibo’s monthly active users [29]. Although our data were collected from the most popular social media platforms, further research is needed on other social media platforms in China. Second, our data only cover the first week after the regulations were issued in Hangzhou. On the basis of the temporal distribution of the collected data, the discussions about the regulations decreased gradually after the first peak, so the data were still reasonable for this study although the time interval was not very long. Third, the tobacco control regulations mentioned in this paper were issued by the Hangzhou government, so the regulations’ influence is still limited because they are not a national policy.

Conclusions

Given the very high prevalence of smoking in China, reducing smoking and tobacco use is difficult because it requires both policy changes and changes in social norms. The recent tobacco control policy in Hangzhou led to a heated discussion on social media, and we conducted this study to understand how the public responded to the impact of the new policy on consumers, tobacco industry, and the public interest. We found that the number of posts supporting the policy were significantly higher than those that were against the policy. The topics of these posts were diverse, including regulation scope, health effects of cigarettes and e-cigarettes, taxes, policy enforcement, and so on. We also found that most of the key social media accounts in spreading the policy were professional new media accounts rather than individual accounts. Our findings demonstrate the possibility of using social media to evaluate the influence of public health policy and understand public perceptions. As the use of social media grows, more research is needed to better understand how to develop and implement sound tobacco control policies and relevant campaigns to engage the public to improve their health.

Acknowledgments

WW and ZZ conceived the idea for this study. WW designed the study. WW, LZ, and LJ conducted data analysis and drafted the manuscript. ZZ, ZY, SL, and DZ provided critical feedback, helped interpret the analysis of results, and revised the manuscript. All authors read and approved the final manuscript. This work is also supported in part by the Ministry of Science and Technology of China Major Grants No. 2017YFC0820105 and No. 2016QY02D0305, and the Ministry of Health of China under Grant No. 2017ZX10303401-002. This work is also supported by National Natural Science Foundation of China (NSFC) Grants No. 71621002 and No. 71974187. This work is also supported by CAS Key Grant No. ZDRW-XH-2017-3 and the Research Foundation of SKL-MCCS for Young Scientists under Grant No. 20190212. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Ministry of Science and Technology of China Major Grant, NSFC, and CAS Key Grant.

Conflicts of Interest

SL is the editor-in-chief of the journal *Tobacco Regulatory Science*.

Multimedia Appendix 1
The original text of Microblogs.
[DOCX File , 30 KB - jmir_v22i4e14660_app1.docx]

Multimedia Appendix 2
Weibo usernames in Chinese and English.
[DOCX File , 12 KB - jmir_v22i4e14660_app2.docx]

Multimedia Appendix 3
Chinese-English contrast table of microblog contents.
[DOCX File , 16 KB - jmir_v22i4e14660_app3.docx]

References

1. World Health Organization. 2017. WHO Report on the Global Tobacco Epidemic 2017 URL: https://www.who.int/tobacco/global_report/2017/en/ [accessed 2019-03-28] [WebCite Cache ID 77CrdTgLq]
2. Tobacco Control Resource Center. 2014. 2014 China Youth Tobacco Survey Report URL: http://www.tcrc.org.cn/html/zy/cbw/jc/2827.html [accessed 2020-02-10] [WebCite Cache ID 78DU94R3z]
3. Wang X, Zhang X, Xu X, Gao Y. Electronic cigarette use and smoking cessation behavior among adolescents in China. Addict Behav 2018 Jul;82:129-134 [FREE Full text] [doi: 10.1016/j.addbeh.2018.02.029] [Medline: 29522934]
4. Trtchounian A, Talbot P. Electronic nicotine delivery systems: is there a need for regulation? Tob Control 2011 Jan;20(1):47-52. [doi: 10.1136/tc.2010.037259] [Medline: 21139013]
5. Royal College of Physicians. 2016. Nicotine Without Smoke: Tobacco Harm Reduction URL: https://www.rcplondon.ac.uk/projects/outputs/nicotine-without-smoke-tobacco-harm-reduction-0 [accessed 2019-05-08] [WebCite Cache ID 78DJA8xL6]
6. Hajek P, Etter J, Benowitz N, Eissenberg T, McRobbie H. Electronic cigarettes: review of use, content, safety, effects on smokers and potential for harm and benefit. Addiction 2014 Nov;109(11):1801-1810 [FREE Full text] [doi: 10.1111/add.12659] [Medline: 25078252]
7. Tegin G, Mekala HM, Sarai SK, Lippmann S. E-cigarette toxicity? South Med J 2018 Jan;111(1):35-38. [doi: 10.14423/SMJ.0000000000000749] [Medline: 29298367]
8. Staudt MR, Salit J, Kaner RJ, Hollmann C, Crystal RG. Altered lung biology of healthy never smokers following acute inhalation of E-cigarettes. Respir Res 2018 May 14;19(1):78 [FREE Full text] [doi: 10.1186/s12931-018-0778-z] [Medline: 29754582]
9. China National Market Supervision Administration, China State Tobacco Monopoly Bureau. China Tobacco. Notice of the State Administration of Market Supervision and Administration of the State Administration of Tobacco Monopoly on Banning the Sale of Electronic Cigarettes to Minors URL: http://www.tobacco.gov.cn/html/49/86870396_n.html [accessed 2019-03-28] [WebCite Cache ID 77CrwqP2e]
10. Korfei M. The underestimated danger of E-cigarettes - also in the absence of nicotine. Respir Res 2018 Aug 29;19(1):159. [doi: 10.1186/s12931-018-0870-4] [Medline: 30157845]
11. Li Z, Elton-Marshall T, Fong GT, Quah AC, Feng G, Jiang Y, et al. Erratum to: noticing cigarette health warnings and support for new health warnings among non-smokers in China: findings from the International Tobacco Control project (ITC) China survey. BMC Public Health 2017 Jun 19;17(1):581 [FREE Full text] [doi: 10.1186/s12889-017-4495-1] [Medline: 28629441]
12. National Health and Family Planning Commission of the People's Republic of China. The regulations on smoking control in public places (draft). Chin Pract J Rural Doc 2015(10):3-5 [FREE Full text] [doi: 10.3969/j.issn.1672-7185.2015.10.003]

13. Danzi Z. China city news. 18 cities across the country have established smoke-free environmental regulations URL: http://www.tobaccochina.com/kongyan/20176/201765175120_752613.shtml [accessed 2019-03-28]
14. China Baogao. 2018-2023 China Electronic Cigarette Market Development Survey and Development Trend Evaluation Report URL: http://baogao.chinabaogao.com/dianzishebei/296669296669.html [accessed 2019-03-28] [WebCite Cache ID 77CswxPH6]
15. Liang Y, Zheng X, Zeng DD, Zhou X, Leischow SJ, Chung W. Characterizing social interaction in tobacco-oriented social networks: an empirical analysis. Sci Rep 2015 Jun 19;5:10060 [FREE Full text] [doi: 10.1038/srep10060] [Medline: 26091553]
16. Chu KH, Unger JB, Allem JP, Pattarroyo M, Soto D, Cruz TB, et al. Diffusion of messages from an electronic cigarette brand to potential users through Twitter. PLoS One 2015;10(12):e0145387 [FREE Full text] [doi: 10.1371/journal.pone.0145387] [Medline: 26684746]
17. Merianos AL, Gittens OE, Mahabee-Gittens EM. Depiction of health effects of electronic cigarettes on YouTube. J Subst Use 2016;21(6):614-619 [FREE Full text] [doi: 10.3109/14659891.2015.1118565] [Medline: 28217030]
18. Lazard AJ, Wilcox GB, Tuttle HM, Glowacki EM, Pikowski J. Public reactions to e-cigarette regulations on Twitter: a text mining analysis. Tob Control 2017 Dec;26(e2):e112-e116. [doi: 10.1136/tobaccocontrol-2016-053295] [Medline: 28341768]
19. Allem JP, Escobedo P, Chu KH, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. Tob Control 2017 Mar;26(2):226-229 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052757] [Medline: 26956467]
20. Chung JE. A smoking cessation campaign on Twitter: understanding the use of Twitter and identifying major players in a health campaign. J Health Commun 2016 May;21(5):517-526. [doi: 10.1080/10810730.2015.1103332] [Medline: 27115179]
21. Kavuluru R, Sabbir AK. Toward automated e-cigarette surveillance: spotting e-cigarette proponents on Twitter. J Biomed Inform 2016 Jun;61:19-26 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.006] [Medline: 26975599]
22. Troelstra SA, Bosdriesz JR, Boer MRD, Kunst AE. Effect of Tobacco Control Policies on Information Seeking for Smoking Cessation in the Netherlands: A Google Trends Study. PLoS One 2016;11(2):e0148489 [FREE Full text] [doi: 10.1371/journal.pone.0148489] [Medline: 26849567]
23. Thrul J, Ramo DE. Cessation strategies young adult smokers use after participating in a Facebook intervention. Subst Use Misuse 2017 Jan 28;52(2):259-264 [FREE Full text] [doi: 10.1080/10826084.2016.1223690] [Medline: 27759475]
24. Wang F, Zheng P, Yang D, Freeman B, Fu H, Chapman S. Chinese tobacco industry promotional activity on the microblog Weibo. PLoS One 2014;9(6):e99336 [FREE Full text] [doi: 10.1371/journal.pone.0099336] [Medline: 24914739]
25. Cui K, Zheng X, Zeng D, Leischow S. An Empirical Analysis on Communications about Electronic Nicotine Delivery Systems (ENDS) in Chinese Social Media. In: Proceedings of the International Conference on Smart Health.: Springer, Cham; 2014 Presented at: ICSH'14; July 10-11, 2014; Beijing, China p. 37-43. [doi: 10.1007/978-3-319-08416-9_4]
26. Jin XZ, Jun S, Bai-Jing HU, Shi Q, Xia J, Xiao L, et al. Effectiveness evaluation of tobacco control communications in the new media era. Chinese Journal of Health Education 2013;29(4):319-321 [FREE Full text]
27. Hangzhou People's Congress-Home. Regulations on Smoking Control in Public Places in Hangzhou URL: http://www.hzrd.gov.cn/zxzx/wsgg/201811/t20181106_737284.html [accessed 2020-02-10] [WebCite Cache ID 77CtADg72]
28. Python Software Foundation. selenium 3.6.0 URL: https://pypi.org/project/selenium/3.6.0/ [accessed 2020-02-10] [WebCite Cache ID 77CtMJTlG]
29. Sina Weibo Data Center. Weibo report publishing platform-micro report. 2017. 2017 Weibo User Development Report URL: http://data.weibo.com/report/reportdetail?id=404 [accessed 2019-03-28] [WebCite Cache ID 77CtVRn1n]
30. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014 Oct 16;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
31. Carletta J. Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics 1996;22(2):249-254 [FREE Full text]
32. Luke DA, Harris JK. Network analysis in public health: history, methods, and applications. Annu Rev Public Health 2007;28:69-93. [doi: 10.1146/annurev.publhealth.28.021406.144132] [Medline: 17222078]
33. Harris JK, Clements B. Using social network analysis to understand Missouri's system of public health emergency planners. Public Health Rep 2007;122(4):488-498 [FREE Full text] [doi: 10.1177/003335490712200410] [Medline: 17639652]
34. Merrill J, Bakken S, Rockoff M, Gebbie K, Carley KM. Description of a method to support public health information management: organizational network analysis. J Biomed Inform 2007 Aug;40(4):422-428 [FREE Full text] [doi: 10.1016/j.jbi.2006.09.004] [Medline: 17098480]
35. De Brún A, McAuliffe E. Social network analysis as a methodological approach to explore health systems: a case study exploring support among senior managers/executives in a hospital network. Int J Environ Res Public Health 2018 Mar 13;15(3):pii: E511 [FREE Full text] [doi: 10.3390/ijerph15030511] [Medline: 29534038]
36. de Beurs DD. Network analysis: a novel approach to understand suicidal behaviour. Int J Environ Res Pub Health 2017 Feb 23;14(3):219. [doi: 10.3390/ijerph14030219]
37. Luke DA. Viewing dissemination through a network lens. In: Brownson RC, Colditz GA, Proctor EK, editors. Dissemination and Implementation Research in Health: Translating Science to Practice. USA: Oxford University Press; 2012:154-174.

38. Li J, White JS, Hu T, Fong GT, Yuan J. The heterogeneous effects of cigarette prices on brand choice in China: implications for tobacco control policy. Tob Control 2015 Jul;24(Suppl 3):iii25-iii32 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051887] [Medline: 25855642]

Abbreviations

e-cigarette: electronic cigarette
NSFC: National Natural Science Foundation of China
WHO: World Health Organization

Edited by G Eysenbach; submitted 09.05.19; peer-reviewed by D Carvalho, DR Colombo Dias, MY Song; comments to author 03.10.19; revised version received 27.11.19; accepted 24.01.20; published 27.04.20.

Please cite as:
Wen W, Zhang Z, Li Z, Liang J, Zhan Y, Zeng DD, Leischow SJ
Public Reactions to the Cigarette Control Regulation on a Chinese Microblogging Platform: Empirical Analysis
J Med Internet Res 2020;22(4):e14660
URL: http://www.jmir.org/2020/4/e14660/
doi:10.2196/14660
PMID:32338615

Original Paper

Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach

Qian Liu[1,2], PhD, MSc; Zequan Zheng[3,4*], MBBS; Jiabin Zheng[3,4*], MBBS; Qiuyi Chen[1], MA; Guan Liu[5], PhD; Sihan Chen[3], BA; Bojia Chu[3], BA; Hongyu Zhu[3], BA; Babatunde Akinwunmi[6,7], MD, MPH, MMSc; Jian Huang[8], PhD; Casper J P Zhang[9], PhD; Wai-Kit Ming[3], MD, PhD, MPH, MMSc, EMBA

[1]School of Journalism and Communication, National Media Experimental Teaching Demonstration Center, Jinan University, Guangzhou, Guangdong Province, China

[2]Department of Communication, University at Albany, State University of New York, Albany, New York State, NY, United States

[3]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, Guangdong Province, China

[4]International School, Jinan University, Guangzhou, Guangdong Province, China

[5]Computer Centre, Jinan University, Guangzhou, Guangdong Province, China

[6]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

[7]Pulmonary and Critical Care Medicine Unit, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[8]Multidisciplinary, Collaborative Research Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, St Mary's Campus, Imperial College London, London, United Kingdom

[9]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

[*]these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, PhD, MPH, MMSc, EMBA
Department of Public Health and Preventive Medicine
School of Medicine
Jinan University
601 Huangpu W Ave, Tianhe District
Guangzhou, Guangdong Province, 510632
China
Phone: 86 85228852
Email: wkming@alumni.harvard.edu

Abstract

Background: In December 2019, a few coronavirus disease (COVID-19) cases were first reported in Wuhan, Hubei, China. Soon after, increasing numbers of cases were detected in other parts of China, eventually leading to a disease outbreak in China. As this dreadful disease spreads rapidly, the mass media has been active in community education on COVID-19 by delivering health information about this novel coronavirus, such as its pathogenesis, spread, prevention, and containment.

Objective: The aim of this study was to collect media reports on COVID-19 and investigate the patterns of media-directed health communications as well as the role of the media in this ongoing COVID-19 crisis in China.

Methods: We adopted the WiseSearch database to extract related news articles about the coronavirus from major press media between January 1, 2020, and February 20, 2020. We then sorted and analyzed the data using Python software and Python package Jieba. We sought a suitable topic number with evidence of the coherence number. We operated latent Dirichlet allocation topic modeling with a suitable topic number and generated corresponding keywords and topic names. We then divided these topics into different themes by plotting them into a 2D plane via multidimensional scaling.

Results: After removing duplications and irrelevant reports, our search identified 7791 relevant news reports. We listed the number of articles published per day. According to the coherence value, we chose 20 as the number of topics and generated the topics' themes and keywords. These topics were categorized into nine main primary themes based on the topic visualization figure. The top three most popular themes were prevention and control procedures, medical treatment and research, and global

or local social and economic influences, accounting for 32.57% (n=2538), 16.08% (n=1258), and 11.79% (n=919) of the collected reports, respectively.

Conclusions: Topic modeling of news articles can produce useful information about the significance of mass media for early health communication. Comparing the number of articles for each day and the outbreak development, we noted that mass media news reports in China lagged behind the development of COVID-19. The major themes accounted for around half the content and tended to focus on the larger society rather than on individuals. The COVID-19 crisis has become a worldwide issue, and society has become concerned about donations and support as well as mental health among others. We recommend that future work addresses the mass media's actual impact on readers during the COVID-19 crisis through sentiment analysis of news data.

(J Med Internet Res 2020;22(4):e19118) doi:10.2196/19118

KEYWORDS

coronavirus; COVID-19; outbreak; health communication; mass media; public crisis; topic modeling

Introduction

In December 2019, some pneumonia cases caused by an unknown pathogen were reported in Wuhan, Hubei, China, and similar cases were soon reported in other provinces of China. After multiple sample collections and laboratory analyses, the pathogen was identified as a novel coronavirus named severe acute respiratory syndrome coronavirus 2 by the International Committee of Taxonomy of Viruses [1], and the disease was named coronavirus disease (COVID-19) by the World Health Organization on February 11, 2020 [2]. According to the National Health Commission (NHC) of the People's Republic of China, until February 2020, there had been approximately 80,000 confirmed cases and more than 2000 deaths in China [3]. Other countries such as Japan, South Korea, Thailand, Singapore, and the United States also reported COVID-19 cases in their countries [4]. Although the cases at the early stage in these countries were identified as imported cases from Wuhan or other cities in Hubei Province, some domestic cases and local transmission were also reported.

The rapid spread of COVID-19 has already caused great public attention and many heated discussions, and the Chinese mass media have been reporting relevant information about the virus and the outbreak. As effective public health measures are required to be implemented in time to avoid the breakdown of the health system [5], the media can certainly play a crucial role in conveying updated policies and regulations from authorities to the citizens.

Since no COVID-19 vaccine is yet available, each citizen should be aware of the harm caused by this novel coronavirus, the prevention methods, and the designated hospital in their local area to access at any time. If misleading or incorrect information was transmitted to the public, the people may get anxious and react to the information in many ways, including making a panic purchase and trying unnecessary or even detrimental medicine regimens. Therefore, it requires mass media information dissemination activities in conjunction with the health stakeholders to help individuals, authorities, the government, and others to understand the precarious worldwide and public health conditions posed by COVID-19 and identify health-related knowledge and training required in facing this menace.

Given the desire to know whether the media works efficiently in delivering the latest COVID-19 information to the public audience, major media reports were collected and analyzed. Multimodal data modeling can combine multiple information reports from various resources. To cope with multimodal data, topic modeling was used. Topic modeling is a type of statistical model that arranges unstructured data structurally in accordance with latent themes. With this model, we could investigate the patterns of health communication through the media and the role the media has played so far during the COVID-19 crisis in China.

Methods

Data Collection

We collected Chinese news and articles related to COVID-19 from January 1, 2020, to February 20, 2020. We then applied the latent Dirichlet allocation (LDA) modeling method to derive useful information from these news reports.

Data from Chinese news and related articles were collected from the WiseSearch database [6]. The WiseSearch database is one of the most reputable, ever-growing Chinese media content databases, containing the news and article data from more than 1500 print media sources and over 10,000 internet media sources. It is famous for its reproducibility, timeliness, great coverage, and high data integrity compared with the other database [7]. The news and article data in the WiseSearch database are updated in a timely manner [6].

To gain insights into the early period of health information communication related to the coronavirus, we conducted a search with the keyword "coronavirus" in the WiseSearch database.

LDA is a generative probabilistic topic modeling method that is widely applied in text mining [8], medicine [9,10], and social network analysis [11] due to its excellent capability of converting visual words, a small part of an image that conveys a certain message about the image or alternation of the pixels, into images and visual word documents [12-14]. It is a generative statistical model with a three-level hierarchical Bayesian model. The basic assumption of this model is a combination of words belonging to different topics [15]. LDA indicates that there may be various topics in an article and that the wording in that article is attributable to one of its topics. We

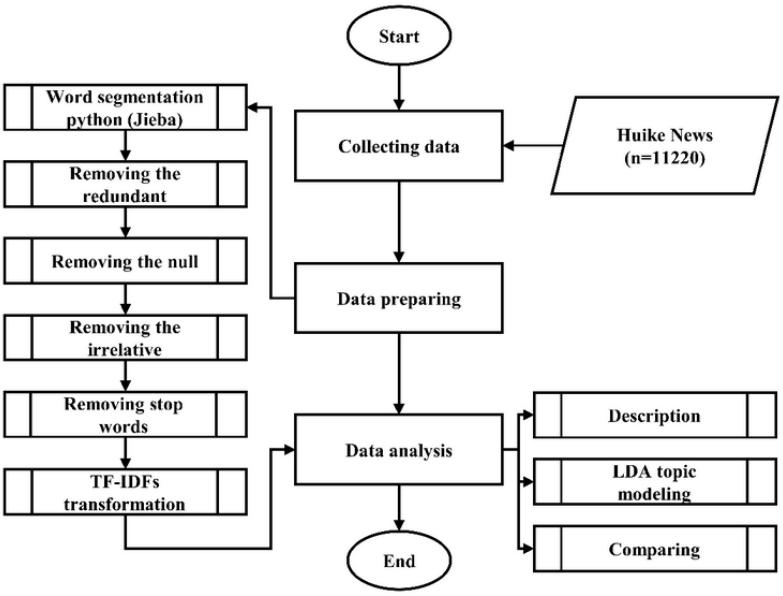
can discover the topics among the data pool by using Gibbs sampling techniques [16].

Processing

A total of 11,220 articles were found with the keyword search "coronavirus," dated between January 1, 2020, and February 20, 2020. After cleaning the data, 7791 articles remained.

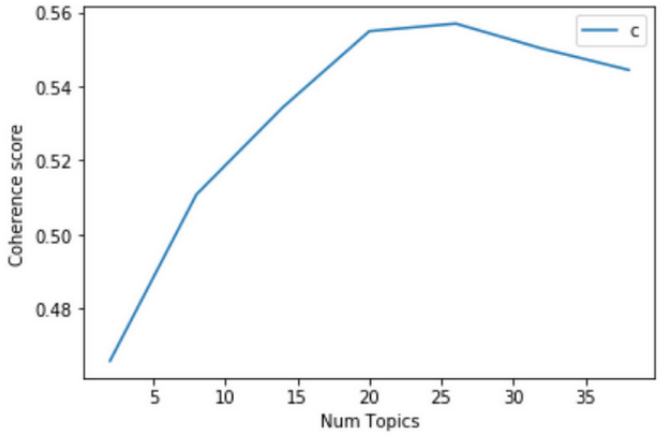
Before applying the LDA modeling, we used Python (Python Software Foundation) to perform data cleaning and used the Python package Jieba for data processing [17,18]. The detailed data process is illustrated in Figure 1. We next removed common Chinese stop characters such as "ten," "a," "of," and "it." We removed duplicate news reports. We then excluded news reports about other coronaviruses like severe acute respiratory syndrome-related coronavirus or Middle East respiratory syndrome-related coronavirus manually. We also built a document-term matrix and used term frequency–inverse document frequency (TF–IDF) to process the data. TF–IDF is a numerical statistic that is used to reflect the importance of a word to an article in a corpus [19].

Figure 1. Data processing flowchart. LDA: latent Dirichlet allocation; TF–IDF: term frequency–inverse document frequency.



To seek a suitable LDA topic number and the explanations to investigate the relationship between the COVID-19 crisis and news reports, we conducted multiple studies. We used a coherence score to evaluate the selection of a suitable number of topics [20]. Topic coherence measures the consistency of a single topic by measuring the semantic similarity between words with high scores in a topic, which contributes to improving the semantic understanding of the topic. That is, words are represented as vectors by the word's co-occurrence relation, and semantic similarity is the cosine similarity between word vectors. The coherence is the arithmetic mean of these similarities [21]. We used the Coherence Model from Gensim (RARE Technologies Ltd), the Python package for natural language processing, to calculate the coherence value [22]. According to Figure 2, the coherence score increased and reached a stable score as the number of topics increased to 20, then declined after the number of topics reached 25. However, we found that the results would be uninterpretable for humans if only statistical measures were applied [23]. As a result, we combined statistical measures and manual interpretation and chose 20 topics to analyze with the help of Python version 3.6.1 and the LDAvis tool [15]. We set λ=1 and set 20 topics and their keywords. Topics' names were generated according to their corresponding keywords to expatiate the topics.

Figure 2. Coherence score for the topic numbers.



We also divided these topics into different themes to study them better. In the visualization, that is the 2D plane (Figures 3 and 4), 20 topics were represented as cycles. These circles overlapped, and their centers are determined by computed topic distance [15]. By this approach, these 20 topics were classified into nine main primary themes and are shown in Table 1. Textbox 1 shows illustrative quotes for each theme.

Figure 3. Intertopic distance map. PC: principal component.

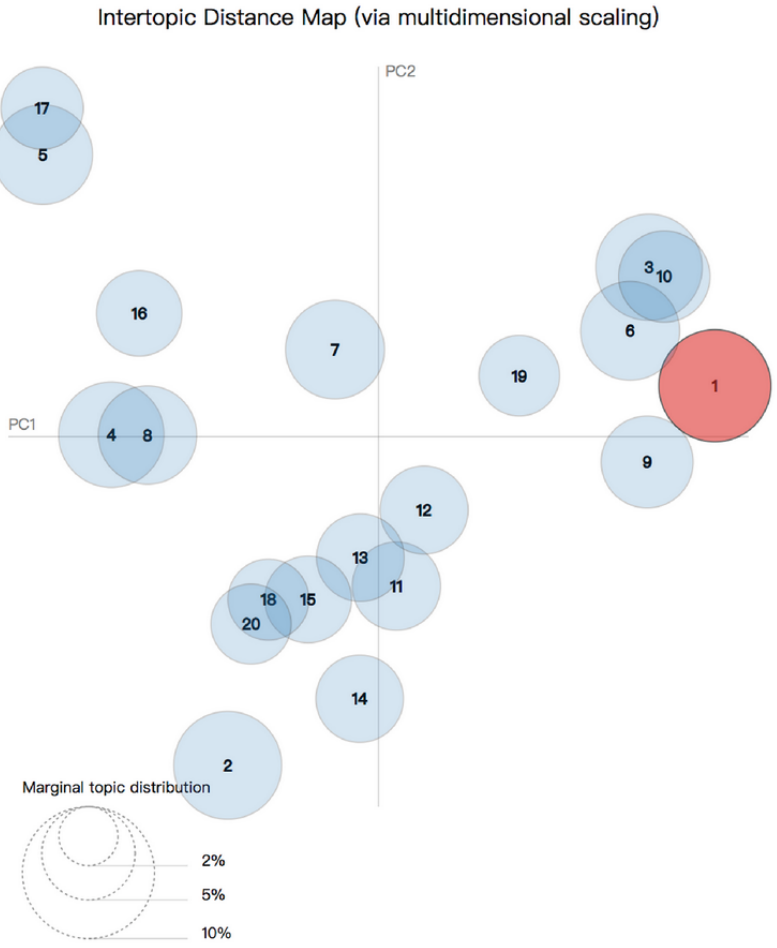


Figure 4. Top 30 most relevant terms for Topic 1 (7.18% of tokens).

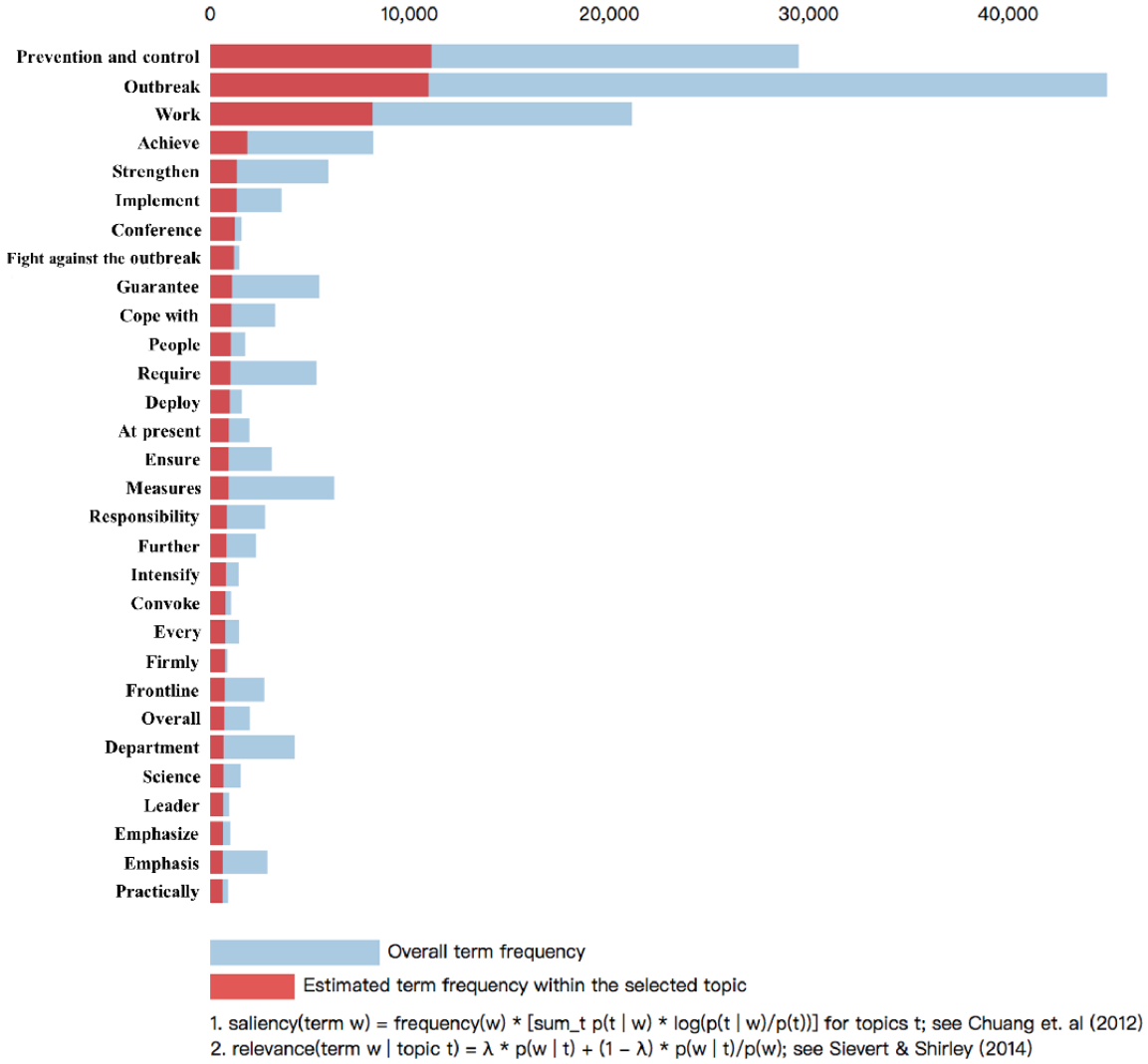


Table 1. Topic classification and keywords.

Theme, topics, and keywords	News reports (N=7791), n (%)[a]
Theme 1: Confirmed cases	747 (9.58)
• Topic 5 • Keywords: cases, confirmed, patient, pneumonia, novel, coronavirus, infection	444 (5.69)
• Topic 17 • Keywords: new type, novel coronavirus, pneumonia, infection	303 (3.88)
Theme 2: Medical supplies	436 (5.59)
• Topic 7: Medical supplies • Keywords: mask, disinfection, protection, contact, symptom	436 (5.59)
Theme 3: Medical treatment and research	1253 (16.08)
• Topic 16: Virus investigation and drug research • Keywords: detection, research, laboratory, treatment, coronavirus, drug	327 (4.19)
• Topic 4: Epidemiologic study • Keywords: virus, infection, spread	498 (6.39)
• Topic 8: Medical affiliation and staff • Keywords: hospital, patient, medical staff, Wuhan, medical team	428 (5.49)
Theme 4: Prevention and control procedures	2538 (32.57)
• Topic 1: The progress in prevention and control • Keywords: prevention and control, work, meeting, outbreak, fight against the outbreak, conference	560 (7.18)
• Topic 6: Community prevention and control work • Keywords: personnel, prevention and control, community, outbreak, quarantine	436 (5.59)
• Topic 10: Prevention and control policy • Keywords: prevention and control, work regulation, department, outbreak, measure in accordance with the law, quarantine	374 (4.80)
• Topic 3: Prevention and control measures • Keywords: prevention and control, measures, outbreak	506 (6.49)
• Topic 19: Company fight against the outbreak • Keywords: outbreak, company, prevention and control, coronavirus, pneumonia, impact, fight against, employee	288 (3.69)
• Topic 9: Prevention and control methods in industries and sectors • Keywords: enterprise, outbreak, service, prevention and control, guarantee, support, manufacture	374 (4.80)
Theme 5: Wuhan's story	522 (6.70)
• Topic 2: Wuhan's story • Keywords: Wuhan, work, Spring Festival, frontline, family member, together	
Theme 6: Mental health	342 (4.38)
• Topic 14: Mental health • Keywords: outbreak, information, mental, society, outbreak, platform, people, nationwide, epidemic control	
Theme 7: Global/local social/economic influences	919 (11.79)
• Topic 20: Impact on Mainland China and Special Administrative Region of the People's Republic of China • Keywords: Hong Kong, mainland, Taiwan, Macao, pneumonia, outbreak, government, impact	288 (3.69)
• Topic 18: Influence during the Spring Festival • Keywords: cancel, event, hotel, visitor, Spring Festival, tourism, announce, journalist, Wuhan	296 (3.79)

Theme, topics, and keywords	News reports (N=7791), n (%)[a]
• Topic 15: National and international response • Keywords: China, international, response, take measures, outbreak	335 (4.29)
Theme 8: Materials supplies and society support	692 (8.88)
• Topic 13: Material supplies and donations • Keywords: materials, donation, mask, Wuhan, antiattack, medical, prevention and control, Hubei	342 (4.38)
• Topic 11: Mask supply • Keywords: mask, production, enterprise, supply, price, manufacture, market	350 (4.49)
Theme 9: Detection at public transportation	342 (4.38)
• Topic 12: Detection at public transportation • Keywords: passenger, Wuhan, body temperature, detection, airport, vehicle	

[a]The total percentage is not 100% due to automatic rounding while exporting the results.

Textbox 1. Further description of each theme.

Theme 1

Confirmed cases of coronavirus disease

Theme 2

The medical supply situation such as the shortage of surgical masks, protection suits, and safety goggles in the initial stage of the outbreak

Theme 3

The latest medical treatment and research about the disease, such as the designated hospital, medical staff, route of transmission, and drugs

Theme 4

Different aspects of the prevention and control procedures

Theme 5

Stories from individuals in Wuhan, such as the frontline workers combating the outbreak and lives of individuals during the crisis

Theme 6

The mental health of the medical staff and national citizens

Theme 7

Influence of coronavirus disease in China and other regions and countries, and the influence on the economy and society

Theme 8

The Chinese society's cooperation to provide material support

Theme 9

The policy and application of detection at public transportation

Results

Figure 3 shows the design of the topic model, in which 20 different topics are plotted as circles. The areas of the circles indicate the overall prevalence, and the center of the circles was determined by computing the distance between topics. Intertopic distances are shown on a 2D plane [24] via multidimensional scaling. The principal component (PC)1 represents the transverse axis, and the PC2 represents the longitudinal axis.

In Figure 4, we show the top 30 most relevant terms for topic 1, which had the highest proportion of all topics, as an example. We selected topic 1 and the system visualized the word frequency distribution relative to the full corpus. Each bar shows the given term's overall frequency and the estimated frequency within topic 1. In topic 1, the news reports mainly talked about the prevention and control work deployment, and they mentioned prevention and control, outbreak work, and outbreak most frequently. In this way, we could study the content of this topic and give the topic's name. This approach is illustrated in the literature [25].

In Figure 5, 5b, 5c, and 5d are partial magnifications of 5a. The data of daily confirmed cases and deaths between January 1, 2020, and January 16, 2020, was extracted from the figure in a transmission dynamics study published on March 26, 2020 [26]. Figure 5 shows that the amount of relevant news slightly increased after a new death was reported on January 9, 2020.

There was also a slight decrease between January 24 and 25, as these 2 days were Chinese New Year's Eve and Chinese New Year, and because the Chinese government had decided to lockdown 13 cities in Hubei Province, which was accompanied by the shutdown of the transportation system on January 31, 2020. Between January 20 and 23, 2020, we observed a sharp increase in relevant news as hundreds of daily new cases occurred. We also found that there was a transient sharp decrease between January 1 and 2, 2020, after the NHC released the Protection Guideline for Population at Different Risk Levels and the Prevention Guideline of Facemask Usage [27]. As the daily new cases decreased on January 4, 2020, the number of daily news reports began to drop. The increase in the number of cases on February 12 and 13, 2020, was due to the updated diagnosis criteria in the COVID-19 protocol (fifth version) [28].

Figure 5. Timeseries of news streams with daily confirmed cases and deaths.

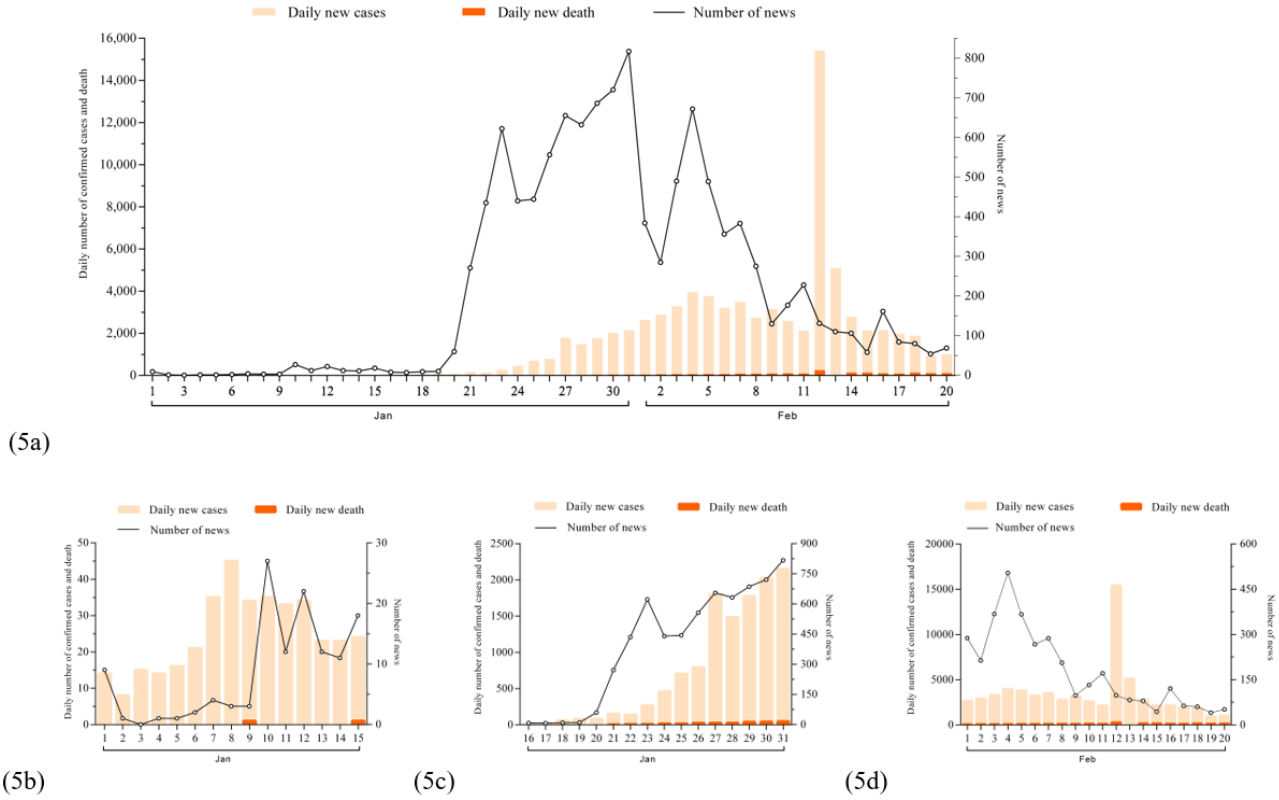


Table 1 shows the theme percentage allocation of our collected news reports. Given our analysis, theme 4 (prevention and control procedures) was the most popular theme. Theme 3 (medical treatment and research) was involved in less than one-sixth of the related news. Theme 7 (global/local social/economic influences) was included in less than one-eighth of all news reports about the coronavirus. The other 6 themes each accounted for less than 10% of the news stories.

As shown in Table 2, China News Service was the most productive media source, followed by the Securities Times and China Securities Journal. Local and national newspapers all participated in reporting recent updates.

Table 2. The most represented media sources for the collected news reports (N=7791).

Media sources	News reports, n (%)
China News Service	1155 (14.82)
Securities Times	176 (2.26)
China Securities Journal	159 (2.04)
Gansu Daily	121 (1.55)
Changsha Evening News	102 (1.31)
Qinghai Daily (digital newspaper)	100 (1.28)
Shenzhen Special Zone Daily	97 (1.25)
Dalian Daily (digital newspaper)	95 (1.22)
Youjiang Daily	87 (1.12)
Inner Mongolia Daily (Chinese version)	82 (1.05)

Our collected news reports mentioned various organizations and companies as shown in Table 3. Wuhan University and Huangzhong University of Science and Technology, as the top two universities in Wuhan, were the most mentioned, followed by Zhejiang University. University affinitive hospitals and university alumni associations participated actively in the fight against COVID-19.

Table 3. Organizations and companies mentioned in news reports (N=7791).

Organization or company	News articles, n (%)
Wuhan University	102 (1.31)
Huazhong University of Science and Technology	66 (0.85)
Zhejiang University	65 (0.83)
Pension and compensation benefits	35 (0.45)
Peking University	28 (0.36)
Wuhan Tianhe International Airport	28 (0.36)
Lanzhou University	22 (0.28)
China Construction Bank	21 (0.27)
Nanchang University	17 (0.22)
Industrial and Commercial Bank of China	17 (0.22)

Discussion

Principal Findings

The COVID-19 crisis has aroused great public concern in China and around the world. Topic modeling provides an alternative perspective to investigate the relationship between media reports and the COVID-19 outbreak. We collected media reports, listed the reports number each day (see Multimedia Appendix 1) and used topic modeling to analyze them. Although several COVID-19 cases were found in December 2019, we observed few news reports about them, showing that the press media did not focus on this disease at that time. As the outbreak became severe and more pneumonia cases were confirmed, the number of news reports began to steadily increase and then rapidly increased on January 19, 2020. In general, news trends peak and wane, according to the confirmed cases during other infectious disease outbreak periods; however, in some cases, the mass media cannot capture the outbreak in time and, therefore, fails to become the leading indicator [29]. This is because it takes time and rigorous effort for journalists to choose a topic, investigate the situation, collect data, and verify the authenticity of the material before they can finally present the news; as a result, a delay ensues. Mass media news reports lag behind the real time coronavirus developments, indicating that the media does not play an adequately forewarning function in public health communication and sensitization.

There was a rapid increase of related news after January 19, 2020, showing that the mass media started to pay more attention to this outbreak. However, since the virus is novel and there are not enough studies on it, the mass media might have conveyed misinformation, which may have induced negative psychological effects in the public like fear, anger, or sadness [30]. In addition, being overfed with reports will result in mass communication fatigue that will dampen the media's effect [31]. Therefore, the government and the mass media should figure out the suitable news themes and daily news numbers to enable the public to keep alert about the outbreak with less harmful mental pressure. The media should also be obliged to ensure the reports' accuracy.

The topics focused on by the mass media can be divided into nine classifications. Theme 4 (prevention and control procedure) and theme 3 (medical treatment and research) were two major themes that together accounted for around half of the content. It is important for the government to communicate with the citizens using mass media during the disease outbreak [32]; therefore, in these reports, the management of important government departments, medical institutions, and community control methods are emphasized. Positive and enthusiastic forecasts backed with active public health interventions are disseminated, which can eliminate unnecessary public worry and extreme panic, aimed at asserting the nation's confidence in virus containment and victory within a short period.

Control of the sources of infection, interruptions of transmission routes, and the protection of susceptible people are three major principles to prevent and control infectious diseases. To cope with the COVID-19 crisis, the Chinese government took measures based on these three principles. The detection of viral infections within public transportation networks aroused great public concern, given that the outbreak coincided with the Spring Festival when many were traveling. Few news reports about this are included in theme 4 (prevention and control procedure), suggesting that the mass media might not have been providing sufficient health information about detection within the transportation network.

The scale of medical treatments and research was the second most popular topic. Our results showed that the mass media conveyed this kind of health information by paying attention to the detection of suspicious cases, drugs that might cure patients, and the transmission routes of the virus. However, reports within theme 4 (prevention and control procedure) and theme 3 (medical treatment and research) mainly focused on

the whole society, while instructions on personal prevention and clinic and medicine choices were less mentioned.

Influences on activities (home and abroad) were also reported together with economic influences, which was included in theme 7 (global/local social/economic influences). These data indicate that the impact of the COVID-19 crisis is not limited to the medical field but also extends to social and economic fields. It is also a worldwide health issue that requires people around the world to work closely together.

The term "confirmed cases" appeared in 9.58% (n=747/7791) of the articles. This indicates that the mass media has served a public health function, as case numbers and their changing rates in news reports can directly give the public intuitive feelings about the speed of viral spread, the momentum, and the hazard of this coronavirus. It can also help citizens remain alert about virus transmission and, therefore, change their daily habits accordingly.

Theme 2 (medical supplies) and theme 8 (material supplies and society support) connect the material supplies with the COVID-19 crisis. Since the outbreak was so sudden and the transmission is so rapid, people in affected areas require medical material and other necessities, especially after the Chinese government shut the major entrance to Hubei to control the outbreak. The mass media can communicate with other parts of China to call for donations and support.

There was an emphasis on Wuhan stories, where news stories focused on the lives of individuals instead of the whole city. We also observed that theme 6 (mental health) accounts for 4.38% (342/7791) of all news articles. Previous studies have shown that there was an increase in mental health problems in both the medical staff [33] and the residents under quarantine [34] during other previous disease outbreaks. Therefore, these kinds of news reports can help the readers refocus on this easily neglected area, and therefore, early interventions can be made. These two themes indicate that the mass media adopts a people-oriented principle when reporting on the COVID-19 crisis, contributing to the warm society phenomenon.

Limitations

This study is the first step to understanding the Chinese mass media's role during the COVID-19 crisis. However, there are still several limitations in our study. First, we included a large number of Chinese news articles about COVID-19 from the WiseSearch mass media database, which only covers text news articles. However, the mass media has recently used new media platforms such as TikTok (video social media) and WeChat (the largest Chinese instant messaging app) to deliver health information through images, snapshots, and short videos. Therefore, we may have omitted news content and the impact of mass media in these media platforms. Second, we only selected a certain period of the outbreak. The pandemic is still ongoing, and the topics and themes are changing; therefore, we may have missed some novel topics and themes. Third, the LDA model has its own limitations such as a lack of nuances for qualitative thematic analysis and poor performance on short articles. Some relative studies introduced sentiment analysis to investigate the emotional differences in the message content [35]; it would be valuable if we could also apply sentiment analysis to supervise the news and investigate the public's reaction to news related to COVID-19.

Conclusion

Collecting and analyzing reports on the novel coronavirus shed light on how the Chinese media have delivered health information during the COVID-19 crisis. Our study provides evidence that the Chinese mass media news lags behind when reporting the major developments of the viral spread. Prevention and control procedures, medical treatment, and research are major themes of the press but mainly focus on the whole society, while instructions on personal and individual prevention, clinic and medicine choices, and detection need to be further enhanced. Global and local influences were reported as the COVID-19 crisis started to impose pressure on public health worldwide and urged cooperation among all humankind. Further research should be considered to explore the impacts of mass media on the readers through sentiment analysis of news data and the influences of misinformation about COVID-19 delivered through the mass media.

Acknowledgments

This paper was funded by the National Social Science Foundation of China (18CXW021).

Authors' Contributions

QL and W-KM conceived the original idea and designed the whole research process. QL, GL, and QC collected and cleaned the data. QL and W-KM did the data analysis and data interpretation, and wrote the first version of the manuscript. QL, JZ, and ZZ made the figures. SC, BC, HZ, JH, CZ, and BA contributed to the administration of the project, data analysis, and data interpretation. Both ZZ and JZ contributed to the final version of the manuscript. BA and W-KM reviewed the manuscript. All authors contributed to the interpretation of the results and the final manuscript. All authors discussed and agreed on the implications of the study findings and approved the final version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Timeseries news streams.
[DOCX File , 17 KB - jmir_v22i4e19118_app1.docx]

References

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr;5(4):536-544. [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]
2. World Health Organization. Coronavirus disease (COVID-19) outbreak URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019
3. National Health Commission of the People's Republic of China. 2020. Latest developments in epidemic control on Feb 29 URL: http://www.nhc.gov.cn/xcs/yqtb/202003/9d462194284840ad96ce75eb8e4c8039.shtml
4. World Health Organization. 2020 Feb 13. Coronavirus disease 2019 (COVID-19) situation report – 24 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200213-sitrep-24-covid-19.pdf?sfvrsn=9a7406a4_4
5. Ming WK, Huang J, Zhang CJP. Breaking down of healthcare system: mathematical modelling for controlling the novel coronavirus (2019-nCoV) outbreak in Wuhan, China. bioRxiv 2020 Jan 30:e. [doi: 10.1101/2020.01.27.922443]
6. WiseSearch. URL: http://wisenews.wisers.net.cn
7. Luo J. WiseSearch database and its practical application. Journal of Library and Informational Science in Agriculture 2016;28(7):19-23.
8. Hassanpour S, Langlotz CP. Unsupervised topic modeling in a large free text radiology report repository. J Digit Imaging 2016 Feb;29(1):59-62 [FREE Full text] [doi: 10.1007/s10278-015-9823-3] [Medline: 26353748]
9. Goyal N, Gomeni R. A latent variable approach in simultaneous modeling of longitudinal and dropout data in schizophrenia trials. Eur Neuropsychopharmacol 2013 Nov;23(11):1570-1576. [doi: 10.1016/j.euroneuro.2013.03.004] [Medline: 23602612]
10. Kandula S, Curtis D, Hill B, Zeng-Treitler Q. Use of topic modeling for recommending relevant education material to diabetic patients. AMIA Annu Symp Proc 2011;2011:674-682 [FREE Full text] [Medline: 22195123]
11. Li A, Huang X, Hao B, O'Dea B, Christensen H, Zhu T. Attitudes towards suicide attempts broadcast on social media: an exploratory study of Chinese microblogs. PeerJ 2015;3:e1209. [doi: 10.7717/peerj.1209] [Medline: 26380801]
12. McLaurin E, McDonald AD, Lee JD, Aksan N, Dawson J, Tippin J, et al. Variations on a theme: topic modeling of naturalistic driving data. Proc Hum Factors Ergon Soc Annu Meet 2014 Sep;58(1):2107-2111 [FREE Full text] [doi: 10.1177/1541931214581443] [Medline: 26190948]
13. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 2015 Dec 1;16(S13):e. [doi: 10.1186/1471-2105-16-s13-s8]
14. Zheng Y, Zhang Y, Larochelle H. A Deep and Autoregressive Approach for Topic Modeling of Multimodal Data. IEEE Trans Pattern Anal Mach Intell 2016 Jun 1;38(6):1056-1069. [doi: 10.1109/tpami.2015.2476802]
15. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research 2003 Jan;3:993-1022.
16. He BD, De Sa CM, Mitliagkas I, Ré C. Scan order in Gibbs sampling: models in which it matters and bounds on how much. In: Advances in neural information processing systems. San Diego, CA: NIPS Proceedings; 2016.
17. Day MY, Lee CC. Deep learning for financial sentiment analysis on finance news providers. IEEE 2016. [doi: 10.1109/ASONAM.2016.7752381]
18. Zhao W, Luo X, Qui T. Recent Developments in Smart Healthcare. Basel, Switzerland: MDPI; 2018.
19. Rajaraman A, Ullman J. Mining Of Massive Datasets. Cambridge, United Kingdom: Cambridge University Press; 2011.
20. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring topic coherence over many models and many topics. 2012 Presented at: 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; 2012; Jeju Island, Korea.
21. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. 2015 Feb Presented at: Eighth ACM International Conference on Web Search and Data Mining; 2015; Shanghai, China. [doi: 10.1145/2684822.2685324]
22. Gensim. models.coherencemodel – Topic coherence pipeline URL: https://radimrehurek.com/gensim/models/coherencemodel.html
23. Grimmer J, Stewart BM. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Polit anal 2017 Jan 04;21(3):267-297. [doi: 10.1093/pan/mps028]
24. Chuang J, Ramage D, Manning C, Heer J. Interpretation and trust: designing model-driven visualizations for text analysis. : Association for Computing Machinery; 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2012; Austin, Texas. [doi: 10.1145/2207676.2207738]
25. Chuang J, Manning C, Heer J. Termite: visualization techniques for assessing textual topic models. 2012 Presented at: International Working Conference on Advanced Visual Interfaces; 2012; Capri Island, Italy. [doi: 10.1145/2254556.2254572]
26. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207. [doi: 10.1056/nejmoa2001316]
27. National Health Commission of the People's Republic of China. 2020. The notification of protection guideline for population at different risk levels and prevention guideline of facemask usage URL: http://www.nhc.gov.cn/xcs/zhengcwj/202001/a3a261dabfcf4c3fa365d4eb07ddab34.shtml

28. National Health Commission of the People's Republic of China. The protocol for the novel coronavirus pneumonia (the fifth version) (in Chinese) URL: http://www.nhc.gov.cn/yzygj/s7653p/202002/d4b895337e19445f8d728fcaf1e3e13a.shtml [accessed 2020-03-11]
29. Ghosh S, Chakraborty P, Nsoesie EO, Cohn E, Mekaru SR, Brownstein JS, et al. Temporal topic modeling to assess associations between news trends and infectious disease outbreaks. Sci Rep 2017 Jan 19;7(1):40841. [doi: 10.1038/srep40841] [Medline: 28102319]
30. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552. [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
31. Collinson S, Khan K, Heffernan JM. The effects of media reports on disease spread and important public health measurements. PLoS One 2015;10(11):e0141423. [doi: 10.1371/journal.pone.0141423] [Medline: 26528909]
32. Wang S, Wang B, Peng C, Song C, Zhang H, Sun D, et al. [Awareness on SARS and public health emergencies among general publics]. Zhonghua Liu Xing Bing Xue Za Zhi 2006 Jun;27(6):503-507. [Medline: 17152511]
33. Chong M, Wang W, Hsieh W, Lee C, Chiu N, Yeh W, et al. Psychological impact of severe acute respiratory syndrome on health workers in a tertiary hospital. Br J Psychiatry 2004 Aug;185:127-133. [doi: 10.1192/bjp.185.2.127] [Medline: 15286063]
34. Jeong H, Yim HW, Song Y, Ki M, Min J, Cho J, et al. Mental health status of people isolated due to Middle East Respiratory Syndrome. Epidemiol Health 2016;38:e2016048. [doi: 10.4178/epih.e2016048] [Medline: 28196409]
35. Kim EH, Jeong YK, Kim Y, Kang KY, Song M. Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. Journal of Information Science 2016 Jul 11;42(6):763-781. [doi: 10.1177/0165551515608733]

Abbreviations

COVID-19: coronavirus disease
LDA: latent Dirichlet allocation
NHC: National Health Commission
PC: principal component
TF–IDF: term frequency–inverse document frequency

Edited by G Eysenbach; submitted 04.04.20; peer-reviewed by MA Bahrami, V Osadchiy, M Lamba, X Shi; comments to author 09.04.20; revised version received 15.04.20; accepted 16.04.20; published 28.04.20.

Please cite as:
Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, Chu B, Zhu H, Akinwunmi B, Huang J, Zhang CJP, Ming WK
Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach
J Med Internet Res 2020;22(4):e19118
URL: http://www.jmir.org/2020/4/e19118/
doi:10.2196/19118
PMID:32302966

Original Paper

Web-Based Medical Information Searching by Chinese Patients With Breast Cancer and its Influence on Survival: Observational Study

Yan Li[1*], MD; Shan Ye[1*], RN; Yidong Zhou[1*], MD; Feng Mao[1*], MD; Hailing Guo[1*], RN; Yan Lin[1], MD; Xiaohui Zhang[1], MD; Songjie Shen[1], MD; Na Shi[1], RN; Xiaojie Wang[1], RN; Qiang Sun[1], MD

Department of Breast Surgery, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Qiang Sun, MD
Department of Breast Surgery
Peking Union Medical College Hospital
Peking Union Medical College, Chinese Academy of Medical Sciences
1, Shuaifuyuan, Dongcheng District
Beijing, 100730
China
Phone: 1 69158721
Fax: 1 69158720
Email: sunqiangpumch@126.com

Abstract

Background: The internet allows patients to easily look for health information. However, how Chinese patients with breast cancer use the internet has rarely been investigated, and there is a scarcity of information about the influence of internet use on survival.

Objective: This observational study aimed to investigate the details of online medical information searching by Chinese patients with breast cancer and to determine whether internet use has any survival benefits.

Methods: Patients who were diagnosed with invasive breast cancer at Peking Union Medical College Hospital between January 2014 and December 2015 were enrolled. We obtained information on their internet-searching behavior and gathered data from the patients' medical and follow-up records. The associations between internet use and other clinic-pathological factors were analyzed. A Cox proportional-hazards model and the Kaplan-Meier method were used for disease-free survival (DFS) analyses.

Results: A total of 973 patients with invasive breast cancer who underwent definitive surgery took part in the study. Among them, 477 cases (49.0%) performed web-based breast cancer information searching before the initial treatment. A multivariate logistic regression analysis suggested that web-based breast cancer information searching was significantly associated with younger age (odds ratio [OR] 0.95, 95% CI 0.94-0.97, *P*<.001), higher education level (OR 1.37, 95% CI 1.01–1.86, *P*=.04), and breast conserving surgery (OR 1.35, 95% CI 1.04-1.77, *P*=.03). Baidu (73.4%, 350/477) and WeChat (66.7%, 318/477) were the two most popular online information sources for breast cancer; however, only 44.9% (214/477) felt satisfied with the online information. In contrast to the nonweb searching group, the web-using patients who were satisfied with online information showed significantly improved DFS (hazard ratio 0.26; 95% CI 0.08-0.88, *P*=.03).

Conclusions: The patients who were most likely to search the internet for breast cancer information were younger and well-educated, and they were more likely to have breast conserving therapy. Web-using patients who were satisfied with the internet information showed significantly improved DFS. Patients should browse credible websites offering accurate and updated information, and website developers should provide high-quality and easy-to-understand information to better meet the needs of patients with breast cancer.

(J Med Internet Res 2020;22(4):e16768) doi:10.2196/16768

KEYWORDS

breast cancer; internet; disease-free survival; breast conserving therapy; online information; satisfaction level

Introduction

The general population has been increasing their use of the internet since the early 1990s. By the end of 2017, China had 772 million internet users, which is the largest population of "netizens" in the world. The internet penetration rate reached 55.8%, with increases of 2.6% per year, which was higher than the world average. The number of mobile internet users in China increased to 753 million by the end of 2017, and there was a yearly increase of 57.34 million [1].

The internet enables patients to conveniently search for health information and it now plays a larger part in their treatment decision making [2]. Using health information from the internet is associated with stronger participation in decision making, better decisions, and more frequent changes in health behavior, and it may enable patients to communicate with doctors more effectively [2,3]. Some surveys have found considerably high rates of online health information-searching behavior between patients who are more educated, younger, and of a higher socioeconomic status [4,5]. However, other studies showed that educational level and residence had no effect on the rates of patients searching the internet for health information [4,6]. Although searching for internet medical information may benefit patients' understanding of the disease and medical processes, some studies showed there is a significant lack of high-quality websites providing accurate and complete information [7-9]. Many websites provide inaccurate, conflicting, or misleading information on guidelines and recommendations established by professional medical groups, and it may lead to confusion, unnecessary anxiety, or treatment delay [10,11].

Many studies have revealed a significant proportion of patients with cancer use the internet to find information about their diseases [12,13]. To date, most of the studies about online cancer information research are carried out in Europe and North America [14,15]. Castleton reported that about 63% of North American patients with cancer searched the internet for health-related information [16]. Breast cancer is the most frequent cancer that occurs in Chinese women, and there is an age-standardized rate of 21.6 cases per 100,000 women [17]. For patients with breast cancer in western countries, the rate of internet use for health-related information was reported to be between 29% and 57% [12,18,19]. Patients with breast cancer search the internet for information related to treatment, effect on daily life, prognosis, and cause of disease [18,19]. Internet information is also important in patients' decision making regarding the type of surgery for breast cancer [20]. However, the use of the internet has rarely been investigated in Chinese patients with breast cancer, and there is a scarcity of information about the influence of internet use on the recurrence and survival of breast cancer.

In this observational study, we examined how patients with breast cancer use the internet to obtain information pertaining to their disease. The aims of our study were to identify the factors that relate to the use of the internet as a source of health information, investigate the details of online information searching, and determine whether internet use would have any survival benefits for patients with breast cancer. We hope that by recognizing the online information searching details of patients, we will be able to suggest strategies to facilitate internet searches on breast cancer information and help patients better cope with their disease.

Methods

Ethics Statement

This study was endorsed by the ethics committee of Peking Union Medical College Hospital (PUMCH), and all the participants provided informed consent.

Patients

Patients who were diagnosed with invasive breast cancers in PUMCH between January 2014 and December 2015 were enrolled in this study. To be included, the participants must have had definitive surgery of both the breast and axilla. The exclusion criteria were as follows: neoadjuvant therapy received prior to definitive surgery or distant metastasis of breast cancer (Stage IV). All eligible participants had pertinent medical records, and they had regular follow-ups. The last follow-up date was January 8, 2019.

Data Collection

On the day of discharge from the hospital, the patients, who had just received a definitive surgery of breast cancer, were approached by one member of the research team. The researcher explained the process and significance of the study. After the patients signed the statement of informed consent, the researcher had a face-to-face interview with them. The interview aimed to investigate the web-based breast cancer information searching from the detection of a breast tumor to the initial treatment. The survey included 7 questions relating to sociodemographics (age, education, and residence), web-based information-seeking information (devices, channels and websites, and satisfaction level with web-based information), and self-rated anxiety level.

We gathered data on tumor characteristics from the patients' medical records, such as molecular subtype, cancer stage, and type of surgery. Follow-up data were retrieved from the database. The patients treated at PUMCH were instructed to return for follow-up visits every 6 months after the surgery. A follow-up via telephone was used if a patient failed to attend their appointment. The follow-up procedures included medical history, physical examination, laboratory tests, and imaging examinations (eg, ultrasound, mammography, computed tomography scan, or bone scan).

The endpoint of this study was disease-free survival (DFS). DFS was calculated from the initial treatment to a second primary cancer, recurrence, or death without evidence of recurrence or second primary cancer. Breast cancer recurrence included distant metastases and locoregional recurrence.

Statistical Analysis

Quantitative data were compared with a Student's *t* test. Categorical data were compared with a 2-sided chi-square test and multivariate logistic regression model. DFS was studied using the Kaplan-Meier curve. Time-to-event endpoints were analyzed with 2-sided log-rank tests. A multivariate Cox

proportional-hazards regression model survival analysis was performed after adjusting potential factors that might affect the patient's survival, including age, molecular subtype, stage, and type of surgery. Differences were considered significant at $P<.05$. Statistical analyses were performed using the STATA statistical software package (version 14.0).

Data Availability Statement

All the data that were generated or analyzed in this study are included in this article.

Results

Patient

A total of 1004 patients with invasive breast cancer who underwent definitive surgery were eligible for this study and 31 of them refused. All 973 final participants were female and of Chinese descent. The median age was 47.2 years (range 21-79); 548 (56.3%) of the patients had breast conserving surgery, and 425 (43.7%) of them had a mastectomy. Postoperative adjuvant treatment was determined according to the National Comprehensive Cancer Network guidelines and the patient's actual situation.

Web Use

Among the 973 participants, 477 cases (49.02%) performed web-based breast cancer information searching before initial treatment (web group), while 496 cases (50.98%) did not seek medical information via the internet (nonweb group). By *t* test, age was significantly younger in the web group. Analyses with a chi-square test showed that tumor stage, molecular subtype, and self-rated anxiety levels were not significantly different between the two groups. However, higher education level, urban residence, and breast conserving therapy (BCT) were significantly associated with the web group (Table 1).

Table 1. Patient characteristics within subgroups.

Characteristic	Web group (n=477)	Nonweb group (n=496)	*P* value
Age (years), mean (SD)	43.81 (11.73)	50.49 (14.00)	<.001
Education, n (%)			.01
< high school	376 (78.8)	356 (71.8)	
< high school	101 (21.2)	140 (28.2)	
Residence, n (%)			<.001
Urban	278 (58.3)	211 (42.5)	
Rural	199 (41.7)	285 (57.5)	
Stage, n (%)			.14
I	244 (51.2)	224 (45.2)	
II	141 (29.6)	157 (31.7)	
III	92 (19.3)	115 (23.2)	
Molecular subtype, n (%)			.79
Luminal A	223 (46.8)	246 (49.6)	
Luminal B	112 (23.5)	113 (22.8)	
Triple-negative breast cancer	95 (19.9)	95 (19.2)	
HER2[a]-positive	47 (9.9)	42 (8.5)	
Self-rated anxiety level, n (%)			.52
Low	225 (47.2)	251 (50.6)	
Medium	125 (26.2)	126 (25.4)	
High	127 (26.6)	119 (24.0)	
Type of surgery, n (%)			.047
Breast conserving therapy	284 (59.5)	264 (53.2)	
Mastectomy	193 (40.5)	232 (46.8)	

[a]HER2: human epidermal growth factor receptor 2.

A multivariate logistic regression was also carried out. The outcomes suggested that web-based breast cancer information searching was significantly associated with younger age, higher education, and breast conserving surgery. However, for residence, tumor stage, molecular subtype, and self-rated anxiety level, correlations with web-based information searching were not significant (Table 2).

Table 2. Multivariate logistic analysis of the association between web-based information seeking and patient characteristics.

Characteristic	Web use (n=477), n (%)	Odds ratio (95% CI)	*P* value
Age	N/A[a]	0.95 (0.94-0.97)	<.001
Education			
< high school (reference)	101 (41.9)	N/A	N/A
< high school	376 (51.4)	1.37 (1.01-1.86)	.04
Residence			
Rural (reference)	199 (41.1)	N/A	N/A
Urban	278 (56.9)	0.78 (0.53-1.14)	.197
Stage			
I (reference)	244 (52.1)	N/A	N/A
II	141 (47.3)	0.81 (0.59-1.09)	.16
III	92 (44.4)	0.73 (0.53-1.03)	.08
Molecular subtype			
Luminal A (reference)	223 (47.6)	N/A	N/A
Luminal B	112 (49.8)	1.01 (0.71-1.41)	.97
Triple-negative breast cancer	95 (50.0)	1.04 (0.73-1.48)	.83
HER2[b]-positive	47 (52.8)	1.20 (0.75-1.93)	.45
Self-rated anxiety level			
Low (reference)	225 (47.3)	N/A	N/A
Medium	125 (49.8)	1.03 (0.75-1.42)	.86
High	127 (51.6)	1.13 (0.82-1.56)	.45
Type of surgery			
Mastectomy (reference)	193 (45.4)	N/A	N/A
Breast conserving therapy	284 (51.8)	1.35 (1.04-1.77)	.03

[a]N/A: not applicable.

[b]HER2: human epidermal growth factor receptor 2.

We further investigated the details of web-based breast cancer information-seeking behavior, which are shown in Table 3. Regarding web channels, Baidu (Chinese “Google”) and WeChat (Chinese “Facebook”) were the top two most popular online information sources for breast cancer. Less than half of participants browsed professional websites.

Table 3. Details of patients' web-based breast cancer information-searching behavior.

Variables	Web-using participants (n=477), n (%)
Devices	
Computer (desktop, laptop, tablets)	178 (37.3)
Mobile phones	203 (42.6)
Both	96 (20.1)
Web channels	
Baidu	350 (73.4)
WeChat	318 (66.7)
Professional websites	213 (44.7)
Discussion forums	129 (27.0)
Others	147 (30.8)
Satisfaction level	
Satisfied	214 (44.9)
Unsatisfied	263 (55.1)

Survival Analysis Between the Web Group and Nonweb Group

The median follow-up time was 35 months (range 12-60). In the web group, there were 19 DFS events: 10 cases of bone metastases, 3 cases of locoregional recurrence, 3 cases of lung metastases, 1 case of distant lymph node metastasis, 1 case of liver metastasis, and 1 case of multiple metastases. In the nonweb group, there were 26 DFS events: 11 cases of bone metastases, 6 cases of locoregional recurrence, 2 cases of lung metastases, 2 cases of distant lymph node metastases, 2 cases of liver metastases, 1 case of brain metastasis, 1 case of multiple metastases, and 1 case of death. The 3-year Kaplan-Meier estimates for DFS were 94.32% in the web group, and 93.94% in the nonweb group (Figure 1). The log-rank comparison showed no significant difference in DFS (P=.29). If we further divided the web group into a "Satisfied" subgroup and an "Unsatisfied" subgroup, the log-rank comparison showed a significant difference among all three groups (web-satisfied, web-unsatisfied, and nonweb) (P=.03; Figure 2).

Figure 1. Kaplan-Meier survival curves of disease-free survival by web usage.

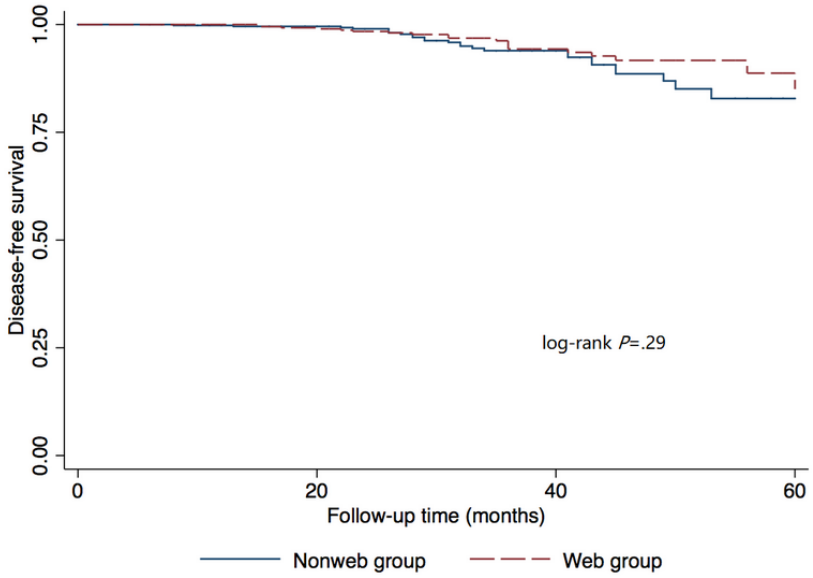
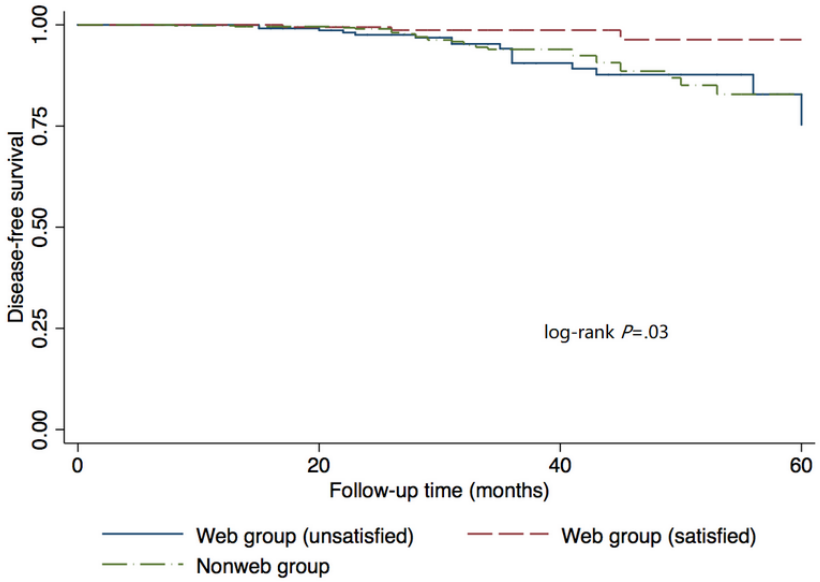


Figure 2. Kaplan-Meier survival curves of disease-free survival (DFS) according to web usage and satisfaction level of internet information.



Multivariate Analysis for DFS

In the multivariate analysis, the included prognostic factors for DFS were age, tumor stage, molecular subtype, type of surgery, and web-seeking condition (Table 4). Compared to the nonweb group, the web-using patients who were unsatisfied with the internet information showed no significant difference for DFS; however, the web-using patients who were satisfied with the internet information showed significantly improved DFS, after adjustment for all other factors. The web-satisfied subgroup also showed significantly improved DFS than the web-unsatisfied subgroup (hazard ratio 0.26; 95% CI, 0.08-0.90, *P*=.03).

Table 4. Cox proportional hazards regression model analysis of disease-free survival.

Factor	Hazard ratio (95% CI)	*P* value
Age (continuous)	1.00 (0.98-1.03)	.54
Stage		
I (reference)	N/A[a]	N/A
II	1.20 (0.57-2.51)	.63
III	1.86 (0.90-3.82)	.09
Molecular subtype		
Luminal A (reference)	N/A	N/A
Luminal B	2.12 (0.86-5.23)	.10
Triple-negative breast cancer	3.79 (1.63-8.77)	.002
HER2[b]-positive	7.11 (2.89-17.48)	<.001
Type of surgery		
Mastectomy (reference)	N/A	N/A
Breast conserving therapy	1.85 (0.98-3.48)	.06
Web-based information seeking condition		
No web (reference)	N/A	N/A
Web use-unsatisfied	0.99 (0.51-1.96)	.99
Web use-satisfied	0.26 (0.08-0.88)	.03

[a]N/A: not applicable.

[b]HER2: human epidermal growth factor receptor 2.

Discussion

The use of the internet for obtaining health information has expanded significantly in the past several years. Because of its frequently updated content, wide availability, and multimedia forms of presenting data, many patients use the internet to acquire health information [21-23]. The reported rates of internet use by patients with cancer have varied widely, with estimates ranging from 4.8% to 77% [19,24-28]. The rate of 49.0% (477/973) in this study is within this range, and it is consistent with the data available specifically for breast cancer: 21% to 51% [12,19,24].

The results from this study showed a definite association between education level, age, location of residence, BCT, and the likelihood that a patient uses the internet for breast cancer information. However, tumor stage, molecular subtype, and anxiety were not as statistically significant as factors affecting internet use among patients with breast cancer. Multivariate analysis suggested that age, education level, and type of surgery were associated with use of the internet but not location of residence. The current literature has reported similar findings [19,24,25]. Younger patients have more opportunities to access the internet, and they are more open to online information, so the link between age and internet use is easily understood [1]. For older adults, there are usually more obstacles to access the internet. They are less familiar with the necessary skills to browse the webpages (eg, typing or setting up an account) and the basic logic of doing an online search. Assistance from family, community, and government is needed to help them overcome these difficulties. Some older adults may also have physical problems that limit their use of the internet, such as low vision and trembling hands. Special software such as interactive voice response systems need to be developed and provided to them [15,22]. As for education level, a possible explanation might be that people with lower education are more likely to have lower socioeconomic status and a lower income level. This has led to concerns about a “digital divide”, that is, the existence of a barrier for patients in lower educational and socioeconomic groups to benefit from online health information [29]. To eliminate or reduce this “digital divide”, training programs of internet use should be offered to these less educated people. In addition, the government should build high-speed and more affordable broadband infrastructure and offer subsidies to those who cannot afford the cost of basic internet devices [29-31]. However, in our multivariate analysis, the likelihood of internet use was not different between urban and rural residences, which is not consistent with previous studies [30,31]. In China, because of the lack of an effective appointment and referral system, receiving medical service in a large hospital in a metropolis such as Beijing and Shanghai is an expensive, complicated, and time-consuming process [32]. Founded in 1921 by the Rockefeller Foundation, PUMCH is a national center guiding the diagnosis and treatment of difficult and serious diseases appointed by the National Health Commission, and it ranks first in a series of different ranking systems [33]. As a result, the rural residents who can visit PUMCH are more likely to have a relatively higher economic status compared to those living in remote and poor areas. For these people, the internet using condition is likely to be similar to that of urban residents. Further studies should pay more attention to those people with extreme poverty who are in more need of various sources of medical information.

Our study implied that patients with breast cancer who were searching for online medical information were more likely to receive BCT. Jordan et al [34] reported that high-quality and easily accessible web-based information could help to significantly improve patients’ participation in decision making for breast cancer surgery. Currently in China, the rates of BCT are relatively low compared with the international average level [35]. One of the reasons is that the patients lack knowledge about BCT, which is misunderstood to be “unsafe” or “incomplete”, and so they may opt to avoid this treatment even when it is suggested by their clinician [35]. The internet can play an important role in educating patients with breast cancer to accept less invasive treatments as the standard procedure if they are indicated as being the most suitable treatment by their clinicians [36].

We also reported that the self-rated anxiety level was not significantly different whether the patients sought online medical information or not. On the one hand, high-quality online information can facilitate patients’ understanding of the disease, and make them more prepared to face it, both physically and psychologically [37]. On the other hand, the online information is usually too complex and misleading for the general public to comprehend, and it sometimes even contains obvious mistakes. This will lead to confusion and anxiety, and offset the possible improvement of psychological well-being brought by internet searching [34,38,39]. However, some studies reported that online communities, forums, and peer-support groups, which offer the patients more opportunities to express their feelings and receive peer education, have the potential to improve the emotional well-being of patients with cancer [40-42]. Therefore, online interactive and individualized social media tools about breast cancer should be developed to enable patients to establish one-to-one, one-to-many, or many-to-many communications with professionals and their peers [43].

This study demonstrated that, altogether, more than 62% of the web group used mobile smartphones to search for online information on breast cancer. By the end of 2017, the percentage of Chinese netizens using mobile phones to access the internet was 97.5%, hitting a new high record [1]. This was 2.4 points more than at the end of 2016. WeChat is an innovative application that is installed on over 90% of smartphones and 60% of computers, and it is currently performing as an important tool that is transforming the daily life of users in various ways [1]. Many different types of medical information were continuously created and transmitted among millions of users through WeChat, which provided WeChat with the enormous potential to affect the general public’s health knowledge and possible health status. In our study, 66.7% (318/477) of the web group searched for breast cancer information via WeChat. However, Baidu is still on the top of the web channel list, with 73.4% (350/477) of the web group using Baidu as the information source. The largest concern connected with the internet as a health resource is that it is unregulated; high-quality information sits side-by-side with poor or misleading

information. In this study, the general satisfaction level with internet information was not high, with a satisfaction rate of only 44.9% (214/477). This rate is consistent with previous articles [18]. Accuracy and readability are the most concerning issues about online medical information. Patients should be recommended web channels with professional sources, updated information, interactive communication, and patient-friendly language [8,13,14].

Regarding survival analysis, univariate and multivariate regression analyses indicated that there was no DFS difference between the web and nonweb groups; however, web-using patients who were satisfied with internet information showed significantly improved DFS. This result can be explained as follows: accurate online information is beneficial to patients' recurrence and survival condition, while inaccurate or "harmful" information can offset this positive effect. Misled by harmful information, patients may refer to less proven treatments, which can lead to the delay of standard therapy [11]. In China, the online medical information is often shown side-by-side with commercial promotions, which undermines its independency and accuracy. The government should have strict laws against false advertising of online medical information and recommend credible information sources to the public. Physicians should help patients by referring them to reliable and patient-friendly websites and communicating with the patients online.

This study has some limitations. As a study undertaken in a single surgical unit, this population may not be typical of the Chinese population as a whole; in particular, we did not include patients with late-stage cancer or those who were receiving chemotherapy. In addition, we did not analyze some factors that may influence the DFS results, such as time duration from breast tumor detection to initial treatment, adjuvant therapy (radiation therapy, chemotherapy, and targeted therapy), complications of surgery and adjuvant therapy, compliance to therapy, and web-searching behavior after surgery. These factors should be included in future studies. In addition, a multicenter study with larger patient numbers would increase the sample size and add more weight to these results.

Findings from this study have significant implications for clinical practice. First, physicians should ask patients with breast cancer about their use of the internet as an information source and help them to identify inaccurate and misleading information. Second, physicians should discuss the online information with their patients and help them to better understand the disease and treatment options. Third, health care providers should direct patients to accurate and credible websites and provide high-quality online medical information. When discussions of online medical information between physicians and patients are incorporated into the consultations, both shared clinical decision making and the physician-patient relationship will be improved [18].

In conclusion, patients with breast cancer frequently use the internet to obtain health-related information. This study has shown that patients who are most likely to use the internet for breast cancer information are younger and well-educated, and they are more likely to have BCT. Web-using patients who were concurrently satisfied with internet information showed significantly improved DFS compared to those who did not search online for medical information. Patients should browse credible websites offering accurate and updated information, and website developers should provide high-quality and easy-to-understand information to better meet the needs of patients with breast cancer.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grant Number 3332019023) and by Beijing Talents foundation from the Organization Department of Beijing Municipal Committee.

Authors' Contributions

Conception and design of the study were done by authors YL, SY, YZ, FM, HG, and QS. YL and QS provided administrative support. Data collection and interpretation were done by authors YL, YZ, FM, SY, YLin, XZ, SS, NS, and XW. All authors contributed to the writing of the manuscript and final approval of the manuscript.

Conflicts of Interest

None declared.

References

1. China Internet Network Information Center. URL: http://cnnic.com.cn/IDR/ReportDownloads/201911/P020191112539794960687.pdf [accessed 2019-10-01]
2. Bratucu R, Gheorghe IR, Purcarea RM, Gheorghe CM, Popa Velea O, Purcarea VL. Cause and effect: the linkage between the health information seeking behavior and the online environment--a review. J Med Life 2014 Sep 15;7(3):310-316 [FREE Full text] [Medline: 25408746]
3. Cao W, Zhang X, Xu K, Wang Y. Modeling online health information-seeking behavior in China: the roles of source characteristics, reward assessment, and internet self-efficacy. Health Commun 2016 Sep;31(9):1105-1114. [doi: 10.1080/10410236.2015.1045236] [Medline: 26861963]
4. 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]

5. Or CKL, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
6. Galloway S, Graydon J, Harrison D, Evans-Boyden B, Palmer-Wickham S, Burlein-Hall S, et al. Informational needs of women with a recent diagnosis of breast cancer: development and initial testing of a tool. J Adv Nurs 1997 Jun;25(6):1175-1183. [doi: 10.1046/j.1365-2648.1997.19970251175.x] [Medline: 9181414]
7. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
8. Peters P, Smith A, Funk Y, Boyages J. Language, terminology and the readability of online cancer information. Med Humanit 2016 Mar;42(1):36-41. [doi: 10.1136/medhum-2015-010766] [Medline: 26530827]
9. Warren E, Footman K, Tinelli M, McKee M, Knai C. Do cancer-specific websites meet patient's information needs? Patient Educ Couns 2014 Apr;95(1):126-136. [doi: 10.1016/j.pec.2013.12.013] [Medline: 24447523]
10. Arif N, Ghezzi P. Quality of online information on breast cancer treatment options. Breast 2018 Feb;37:6-12. [doi: 10.1016/j.breast.2017.10.004] [Medline: 29040893]
11. Li Y, Zhou Y, Mao F, Guan J, Lin Y, Wang X, et al. The influence on survival of delay in the treatment initiation of screening detected non-symptomatic breast cancer. Sci Rep 2019 Jul 15;9(1):10158 [FREE Full text] [doi: 10.1038/s41598-019-46736-1] [Medline: 31308467]
12. Littlechild SA, Barr L. Using the Internet for information about breast cancer: a questionnaire-based study. Patient Educ Couns 2013 Sep;92(3):413-417. [doi: 10.1016/j.pec.2013.06.018] [Medline: 23891419]
13. Xie B, Su Z, Liu Y, Wang M, Zhang M. Health information sources for different types of information used by Chinese patients with cancer and their family caregivers. Health Expect 2017 Aug;20(4):665-674 [FREE Full text] [doi: 10.1111/hex.12498] [Medline: 27604843]
14. Foroughi F, Lam AK, Lim MSC, Saremi N, Ahmadvand A. "Googling" for cancer: an infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. JMIR Cancer 2016 May 04;2(1):e5 [FREE Full text] [doi: 10.2196/cancer.5212] [Medline: 28410185]
15. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, et al. Web conversations about complementary and alternative medicines and cancer: content and sentiment analysis. J Med Internet Res 2016 Jun 16;18(6):e120 [FREE Full text] [doi: 10.2196/jmir.5521] [Medline: 27311444]
16. Castleton K, Fong T, Wang-Gillam A, Waqar MA, Jeffe DB, Kehlenbrink L, et al. A survey of Internet utilization among patients with cancer. Support Care Cancer 2011 Aug;19(8):1183-1190. [doi: 10.1007/s00520-010-0935-5] [Medline: 20556435]
17. Fan L, Strasser-Weippl K, Li J, St Louis J, Finkelstein DM, Yu K, et al. Breast cancer in China. Lancet Oncol 2014 Jun;15(7):e279-e289. [doi: 10.1016/S1470-2045(13)70567-9] [Medline: 24872111]
18. Nguyen SKA, Ingledew P. Tangled in the breast cancer web: an evaluation of the usage of web-based information resources by breast cancer patients. J Cancer Educ 2013 Dec;28(4):662-668. [doi: 10.1007/s13187-013-0509-6] [Medline: 23832398]
19. Valero-Aguilera B, Bermúdez-Tamayo C, García-Gutiérrez JF, Jiménez-Pernett J, Cózar-Olmo JM, Guerrero-Tejada R, et al. Information needs and Internet use in urological and breast cancer patients. Support Care Cancer 2014 Feb;22(2):545-552. [doi: 10.1007/s00520-013-2009-y] [Medline: 24122406]
20. Hawley ST, Li Y, An LC, Resnicow K, Janz NK, Sabel MS, et al. Improving breast cancer surgical treatment decision making: the iCanDecide randomized clinical trial. J Clin Oncol 2018 Mar 01;36(7):659-666 [FREE Full text] [doi: 10.1200/JCO.2017.74.8442] [Medline: 29364772]
21. Ilic D. The role of the internet on patient knowledge management, education, and decision-making. Telemed J E Health 2010;16(6):664-669. [doi: 10.1089/tmj.2010.0003] [Medline: 20575610]
22. Hartzband P, Groopman J. Untangling the Web--patients, doctors, and the Internet. N Engl J Med 2010 Mar 25;362(12):1063-1066. [doi: 10.1056/NEJMp0911938] [Medline: 20335581]
23. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
24. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Internet use and social support in women with breast cancer. Health Psychol 2002 Jul;21(4):398-404. [Medline: 12090683]
25. Xu X, Hong T, Li B, Liu W. Effect of internet on Chinese patients undergoing elective laparoscopic cholecystectomy. Patient 2015 Apr;8(2):179-185. [doi: 10.1007/s40271-014-0072-1] [Medline: 24958467]
26. Laurent MR, Cremers S, Verhoef G, Dierickx D. Internet use for health information among haematology outpatients: a cross-sectional survey. Inform Health Soc Care 2012 Mar;37(2):62-73. [doi: 10.3109/17538157.2011.606481] [Medline: 22017473]
27. Metz JM, Devine P, DeNittis A, Jones H, Hampshire M, Goldwein J, et al. A multi-institutional study of Internet utilization by radiation oncology patients. Int J Radiat Oncol Biol Phys 2003 Jul 15;56(4):1201-1205. [doi: 10.1016/s0360-3016(03)00407-3] [Medline: 12829160]
28. Buntrock S, Hopfgarten T, Adolfsson J, Onelöv E, Steineck G. The Internet and prostate cancer patients: searching for and finding information. Scand J Urol Nephrol 2007;41(5):367-374. [doi: 10.1080/00365590701303827] [Medline: 17853004]

29. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the Internet, and the digital divide. Health Aff (Millwood) 2000;19(6):255-265. [doi: 10.1377/hlthaff.19.6.255] [Medline: 11192412]
30. 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 May 31;21(5):e12693 [FREE Full text] [doi: 10.2196/12693] [Medline: 31152526]
31. Qiu Y, Ren W, Liu Y, Yin P, Ren J. Online health information in a rural residential population in Zhejiang Province, China: a cross-sectional study. BMJ Open 2019 May 05;9(5):e026202 [FREE Full text] [doi: 10.1136/bmjopen-2018-026202] [Medline: 31061032]
32. Yu W, Yu X, Hu H, Duan G, Liu Z, Wang Y. Use of hospital appointment registration systems in China: a survey study. Glob J Health Sci 2013 Jul 22;5(5):193-201 [FREE Full text] [doi: 10.5539/gjhs.v5n5p193] [Medline: 23985121]
33. Peking Union Medical College Hospital. 2019. URL: https://www.pumch.cn/en/introduction.html [accessed 2020-01-19]
34. Bruce JG, Tucholka JL, Steffens NM, Neuman HB. Quality of online information to support patient decision-making in breast cancer surgery. J Surg Oncol 2015 Nov;112(6):575-580 [FREE Full text] [doi: 10.1002/jso.24046] [Medline: 26417898]
35. Zhang B, Sivasubramaniam PG, Zhang Q, Wang J, Zhang B, Gao J, et al. Trends in radical surgical treatment methods for breast malignancies in China: a multicenter 10-Year retrospective study. Oncologist 2015 Sep;20(9):1036-1043 [FREE Full text] [doi: 10.1634/theoncologist.2014-0281] [Medline: 26253559]
36. Vromans R, Tenfelde K, Pauws S, van Eenbergen M, Mares-Engelberts I, Velikova G, et al. Assessing the quality and communicative aspects of patient decision aids for early-stage breast cancer treatment: a systematic review. Breast Cancer Res Treat 2019 Nov;178(1):1-15 [FREE Full text] [doi: 10.1007/s10549-019-05351-4] [Medline: 31342311]
37. Shim M, Cappella JN, Han JY. How does insightful and emotional disclosure bring potential health benefits?: study based on online support groups for women with breast cancer. J Commun 2011 Jun;61(3):432-464 [FREE Full text] [doi: 10.1111/j.1460-2466.2011.01555.x] [Medline: 25568496]
38. Basch CH, MacLean SA, Garcia P, Basch CE. Readability of online breast cancer information. Breast J 2019 May;25(3):562-563. [doi: 10.1111/tbj.13276] [Medline: 30972837]
39. Miles RC, Baird GL, Choi P, Falomo E, Dibble EH, Garg M. Readability of online patient educational materials related to breast lesions requiring surgery. Radiology 2019 Apr;291(1):112-118. [doi: 10.1148/radiol.2019182082] [Medline: 30694156]
40. Bender JL, Katz J, Ferris LE, Jadad AR. What is the role of online support from the perspective of facilitators of face-to-face support groups? A multi-method study of the use of breast cancer online communities. Patient Educ Couns 2013 Dec;93(3):472-479. [doi: 10.1016/j.pec.2013.07.009] [Medline: 23928354]
41. Stanton AL, Thompson EH, Crespi CM, Link JS, Waisman JR. Project connect online: randomized trial of an internet-based program to chronicle the cancer experience and facilitate communication. J Clin Oncol 2013 Sep 20;31(27):3411-3417 [FREE Full text] [doi: 10.1200/JCO.2012.46.9015] [Medline: 23940228]
42. Zhu J, Ebert L, Liu X, Chan SW. A mobile application of breast cancer e-support program versus routine Care in the treatment of Chinese women with breast cancer undergoing chemotherapy: study protocol for a randomized controlled trial. BMC Cancer 2017 Apr 26;17(1):291 [FREE Full text] [doi: 10.1186/s12885-017-3276-7] [Medline: 28446141]
43. Briceño AC, Gospodarowicz M, Jadad AR. Fighting cancer with the internet and social networking. Lancet Oncol 2008 Nov;9(11):1037-1038. [doi: 10.1016/s1470-2045(08)70275-4] [Medline: 19018618]

Abbreviations

BCT: breast conserving therapy
DFS: disease-free survival
OR: odds ratio
PUMCH: Peking Union Medical College Hospital.

Edited by G Eysenbach; submitted 23.10.19; peer-reviewed by Y Qiu, J Zhu; comments to author 30.12.19; revised version received 26.01.20; accepted 29.02.20; published 17.04.20.

Please cite as:
Li Y, Ye S, Zhou Y, Mao F, Guo H, Lin Y, Zhang X, Shen S, Shi N, Wang X, Sun Q
Web-Based Medical Information Searching by Chinese Patients With Breast Cancer and its Influence on Survival: Observational Study
J Med Internet Res 2020;22(4):e16768
URL: http://www.jmir.org/2020/4/e16768/
doi:10.2196/16768
PMID:32301740

Review

Generative Participatory Design Methodology to Develop Electronic Health Interventions: Systematic Literature Review

Pieter Vandekerckhove[1], MSc; Marleen de Mul[1], PhD; Wichor M Bramer[2], PhD; Antoinette A de Bont[1], PhD

[1]Erasmus School of Health Policy and Management, Rotterdam, Netherlands

[2]Medical Library Erasmus MC, Erasmus Medical Center Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Pieter Vandekerckhove, MSc
Erasmus School of Health Policy and Management
Bayle (J) Building, 6th Floor
Burgemeester Oudlaan 50
Rotterdam, 3062 PA
Netherlands
Phone: 31 10 408 8555
Email: vandekerckhove@eshpm.eur.nl

Abstract

Background: Generative participatory design (PD) may help in developing electronic health (eHealth) interventions. PD is characterized by the involvement of all stakeholders in creative activities. This is different from the traditional user-centered design, where users are less involved. When looking at PD from a *research through design* perspective, it is important to summarize the reasons for choosing a certain form of generative PD to further develop its methodology. However, the scientific literature is currently unclear about which forms of PD are used to develop eHealth and which arguments are used to substantiate the decision to use a certain form of generative PD.

Objective: This study aimed to explore the reporting and substantiation of generative PD methodologies in empirical eHealth studies published in scientific journals to further develop PD methodology in the field of eHealth.

Methods: A systematic literature review following the Cochrane guidelines was conducted in several databases (EMBASE, MEDLINE Ovid, Web of Science, and CINAHL EBSCOhost). Data were extracted on the recruitment and management of stakeholders, the use of tools, and the use of outcome measures.

Results: Of the 3131 studies initially identified, 69 were selected for qualitative synthesis. The reporting was very variable, depending to a large extent on whether the study stated that reporting on the PD process was a major aim. The different levels of reporting and substantiation of the choices of a recruitment strategy, stakeholder management, and tools and outcome measures are presented. Only a few authors explicitly used arguments directly related to PD guiding principles such as democratic, mutual learning, tacit and latent knowledge, and collective creativity. Even though PD principles were not always explicitly discussed in the method descriptions of the studies, they were implicitly present, mostly in the descriptions of the use of PD tools. The arguments used to substantiate the choices made in stakeholder management, PD tools, and the type of outcome measures adopted point to the involvement of PD principles.

Conclusions: Studies that have used a PD research methodology to develop eHealth primarily substantiate the choice of tools made and much less the use of stakeholders and outcome measures.

(J Med Internet Res 2020;22(4):e13780) doi:10.2196/13780

KEYWORDS

cocreation; co-design; participatory design; telemedicine; eHealth; medical informatics; method; methodology; review

Introduction

Participatory Design Methodology

Stakeholder participation is considered to play an important role in developing electronic health interventions (eHealth) [1-4]. However, during the development of eHealth, challenges remain in gaining the trust of stakeholders, managing multiple stakeholders, and involving end users [1,5]. In contrast to more traditional forms of user-centered design, where stakeholders are less involved, generative participatory design (PD) focuses

on including stakeholders in creative activities [3,4]. Therefore, PD is promising in that it could overcome the challenges seen in the development of eHealth [6-8].

PD is becoming increasingly intertwined with research and is therefore also considered to be a research methodology. Looking at PD from a research perspective, the methodological choices to be made are of particular interest. Methodological elements that play a key role in PD research are the recruitment and management of stakeholders [4], the use of outcome measures [4], and the use of tools [9,10]. The literature indicates that the application of each of these elements varies when PD is employed.

Looking at the literature on participatory methods to develop eHealth, a recent systematic literature review showed that 24 frameworks have been used [11]. However, as many studies do not refer to a framework, more attention is needed on the methodologies employed [11].

Methodological Elements

Turning to stakeholders, the varying involvement of patients as end users has been widely discussed in the literature [12,13]. Warnings have been given regarding the ability of users to express their needs and about the prejudices of PD practitioners regarding the participants [14], and the involvement of end users remains debated. When it comes to outcome measures, there is a wide variety that can be used to evaluate PD outputs related to the PD process itself and to the eHealth technology output [15,16]. Tools describe the actions that take place between participants [17], and PD scholars have categorized these tools into make, tell, and enact tools [3,10,17]. Make tools are material components such as a prototype to facilitate the embodiment of thoughts in physical artifacts [10]. Tell tools facilitate the telling of stories to capture implicit information about the use of a technology and how people may wish to use it in the future [10]. Enacting refers to the activities where one or more people act out possible futures by physically trying things out in settings that resemble the possible futures [10]. Finally, PD toolkits can involve make, tell, and enact tools and are used to push people to start thinking about their experiences so that using the tools in the PD process can yield better results.

How stakeholders, tools, and outcome measures are employed in the PD process depends on which PD methodology is followed. Furthermore, there is a lack of a strong methodological explanation that could help develop a more rigorous science of PD [2,4]. Using methodological arguments to make each methodological decision applied in studies employing PD more explicit could improve the scientific rigor of PD as a research methodology [18].

Guiding Principles

The PD literature encompasses various theories that form the foundations for methodologies [2-4,9,10]. Value-laden concepts such as democracy, participation, empowerment, and empathy [4,9] contain values such as inclusion and equality [9] and play a fundamental role in PD. On the basis of the work by Van der Velden and Mörtberg [9] and of Sanders and Stappers [3], four key guiding PD principles can be discerned:

- Democracy: In contrast to traditional design practices, the aim is to involve all stakeholders including nondesigners and future users who will be affected by new technologies. Users can become part of the design team as *experts of their experiences* given appropriate tools to express themselves [13]. The aim is to increase diversity of experience, values, and knowledge. This is believed to foster trust among those involved and to facilitate a learning process and a commitment to taking responsibility for each other and the design result.
- Mutual learning: Participants (both designers and nondesigners) learn from each other, but they also learn from themselves when reflecting on their own work.
- Tacit or latent knowledge: To assess the needs of people beyond the observable or easily detectable, that is, in the form of tacit needs. This deeper knowledge includes explicit and implicit day-to-day technological expertise from the present, future, and past [19]. Sanders has defined tacit needs as being conscious but not expressed and latent needs as subconscious needs that cannot be expressed in words [3,20].
- Collective creativity: PD is considered to be essentially a process of collective creativity [3]. Sanders and Stappers [3] refer to social creativity in which people follow a process referred to as *the path of expression*. Creativity facilitates a design process from which values emerge and become inscribed in the product or service [9]. Everyone is assumed to possess some creative ability, although a design role requires a certain level of creativity [13].

Given the developing nature of the PD methodology, the theoretical and empirical literature does not always incorporate these insights and the four guiding principles. In the theoretical design literature, the relationship between PD principles and the use of stakeholders, tools, and outcome measures is only implicitly suggested [2,4]. For instance, PD principles seem to be implicit in the description of make tools. The democratic principle is implicitly present as make tools include both designers and nondesigners in *making things* [10]. As such, make tools can be used to enhance the democratic involvement of stakeholders. In addition, the collective creativity principle is also implicitly present. Tools, depending on the aim, can be used within a PD project to (1) probe participants, (2) prime participants—to immerse them in a domain, (3) to gain a better understanding of their experiences, or (4) to generate new ideas [17]. Depending on the aim, make tools can be used as part of a probing approach (to inspire ideas), a participatory prototyping approach (stakeholders provide feedback on an existing prototype), or a generative approach (stakeholders give ideas a physical form) [10,19]. It has been suggested that the probing and generative approaches are better suited to early design, or the so-called fuzzy front end, and that prototyping is more useful in later, less fuzzy, design stages [19]. Therefore, the democratic principle and the creativity principles can be used to argue in favor of adopting make tools at different times in the design process.

Little has been reported on the specific arguments used to explain the choice of specific stakeholders, tools, and outcome measures. Although stakeholders can be involved in various

ways in the development of gerontology [8], mobile health (mHealth) [7], and serious games [6,21], a discussion on the methodological considerations is missing. Second, various tools are described for developing health information technology [22], gerontology [8], and mHealth [7], but without methodological substantiation.

In addition, given the very limited presence of evaluations in the empirical literature, it is difficult to establish the outcome measures that are used, let alone the principles upon which they are selected. Eyles et al [7] failed to find any mHealth studies that reported outcome measures. Merkel and Kucharski [8] found a few studies that evaluated some eHealth results, for example, by testing a prototype. However, they did not report the results of the evaluations [8]. Merkel and Kucharski [8] also stated that there were no studies that had evaluated the process of PD itself. Exceptionally, DeSmet et al [21] did evaluate the effectiveness of PD in serious games. They expected that the use of PD in the development of serious games was less effective than when users were involved merely as testers in the game (albeit without taking sample size and strength of effect into account) [21].

Aim

Given these uncertainties, the aim of this study was to explore the substantiation behind the methodological choice to use a certain form of PD in developing eHealth. This paper was intended to be a start in looking at the state of reporting of PD research methodology and, therefore, used a systematic literature review to summarize the current status of reporting in peer-reviewed scientific journals. This research has the potential to guide researchers and practitioners to areas where greater substantiation is needed when using or reporting PD. By considering the current methodological choices, some recommendations are also provided that may also help researchers and practitioners select a method that helps them better achieve their aims.

Methods

Systematic Literature Review

A systematic literature review with qualitative synthesis was conducted to summarize existing knowledge on PD methodology in the development of eHealth technology. In the medical field, the Cochrane review process is considered the gold standard. Given that this review is focused on eHealth, this systematic review follows the Cochrane guidelines [23]. To ensure completeness and transparency, a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting statement is included [24].

Given that PD methodology is developing, the focus was on the reported use and justification of the choices made in using PD tools, stakeholder management, stakeholder recruitment, and the outcome measures selected. The first research question focuses on the use of PD: “How is the use of PD, in particular the involvement of stakeholders, the use of tools, and the use of outcome measures described in the empirical literature about eHealth development?” The second research question focuses on the justification for a type of PD: “What reasons, related to the guiding principles of PD, are offered to substantiate the preference for a given use of stakeholders, tools, and outcome measures?”

Selection of Studies

Search queries were developed by an experienced medical information specialist (WB) and the searches used terms such as participatory design, co-design, cocreation, and collaborative design in the field of telehealth. In addition to these terms, we used a more descriptive approach where we combined human centeredness, patient involvement, etc, with shared decision making or doctor-patient relations in the field of telehealth. The term user involvement was also added to the search. The term participatory research was not used as the terms “co-creation,” “co-design,” and “participatory” were assumed to cover this field.

The search strategies for all the databases that were used can be found in Multimedia Appendix 1. The following databases have been searched from their inception until November 12, 2019 (date last searched): EMBASE (1974-), MEDLINE ALL (Ovid, 1946-), Web of Science Core Collection (Web of Knowledge, 1900-), and CINAHL (EBSCOhost, 1937-). All the references from searches on electronic databases were exported and duplicates removed in Endnote X9 (Thompson Reuters Inc) software. The identified titles and abstracts were then screened for eligibility by two independent researchers.

The following working definition for PD was used: PD refers to the collective creative design process of designers and nondesigners, whereby users are considered partners during the design process. PD activities can generally be described as cocreation workshops or cocreation exercises, or they can be more specifically described by referring to make (ie, collage), tell (ie, cards), and act (ie, acting out) tools. Studies that used other terms were also included if they were described by the authors as co-design or PD-related activities [10,17]. Studies that used other popular terms such as cocreation were only included if, as part of the methodology, PD tools were described.

The selection criteria for inclusion and exclusion are shown in Textboxes 1 and 2. Studies that had as their main objective developing eHealth technology were included. Articles in conference proceedings were also included. Study protocols and conference abstracts were excluded as these included insufficient information about the execution of the PD study and its results. Non-English language publications were excluded.

All types of empirical study designs were included, and no restrictions were placed on the types of participants. For instance, studies involving only patients or only care professionals in PD were included. The presence of PD activities was chosen as the inclusion criterion rather than other features of PD because this area has the most clearly defined consensus in the literature. Other aspects of PD, such as stakeholder recruitment, stakeholder management, and PD outcome measures, were not used as the inclusion or exclusion criteria as these terms can be used in somewhat arbitrary ways.

Textbox 1. Inclusion criteria for screening.

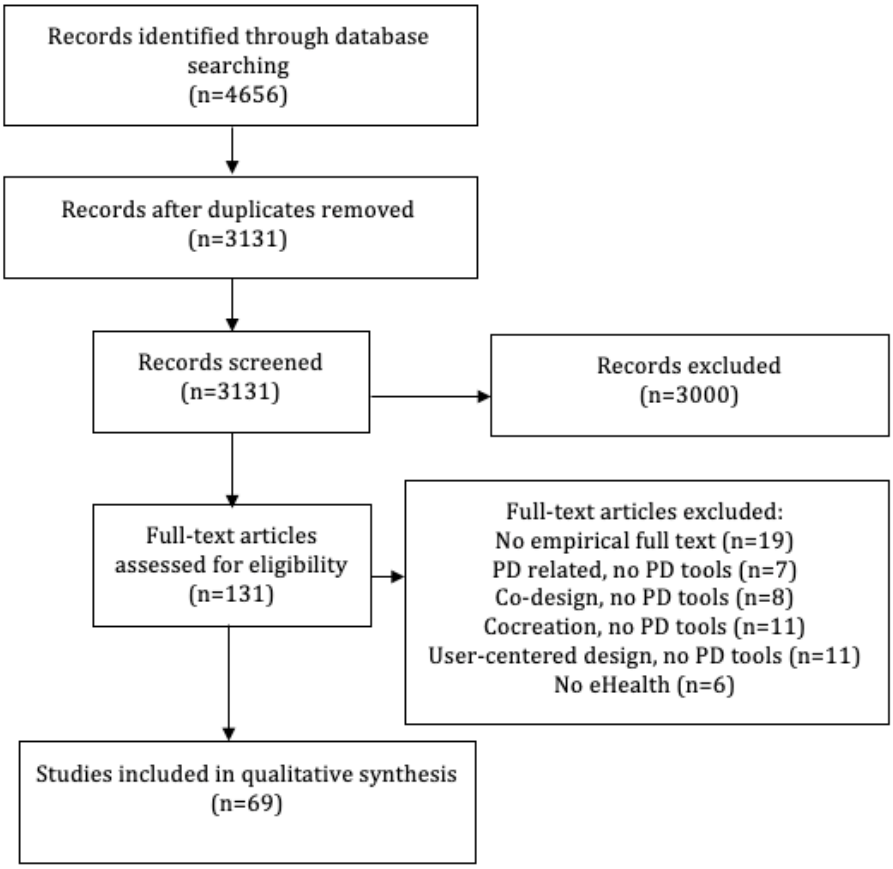
- Language: English language
- Format: Full text available (including full conference papers)
- Study design: Empirical study describing the direct or indirect observation or experience of using participatory design (PD) to develop electronic health (eHealth) published in a peer-reviewed journal or conference proceedings. The aim of the paper was to report on the use of PD to develop eHealth.
- Product or service developed: eHealth related
- Method of development: PD as a collective creative design process of designers and nondesigners whereby users are considered to be partners in the process and the use of PD activities is described with this mindset (including participatory prototyping)
- Design development phases: All innovation phases included (predesign, early design [discover], and design and make)
- Setting: at least one of the PD tools used must be in a group setting (ie, more than one individual involved)

Textbox 2. Exclusion criteria for screening.

- Language: Other than English
- Format: Only abstract or full text unavailable
- Study design: Nonempirical studies (ie, reviews, editorials, discussion papers, methodological papers, papers reflecting on eHealth developed with PD), studies not peer reviewed (eg, dissertations)
- Product or service developed: Other than electronic health (eHealth
- Method of development: Nonparticipatory design, participatory design (PD) where users are considered as subjects in the design process (user-centered design), the use of PD is not described (ie, only qualitative research tools such as focus groups or interviews)
- Setting: All PD tools used only by individuals
- Design development phases: Value cocreation excluded (market phase and later marketing phases)

The identification and selection of studies is summarized in Figure 1 according to the PRISMA guidelines [24]. Following the removal of duplicates, 3131 articles were identified through the search strategy, of which 3000 articles were then excluded based on the title and contents of the abstract. This left 131 unique full-text studies for review, of which 69 met the inclusion criteria (see Multimedia Appendix 2 for full-text studies excluded). The main reasons for full-text exclusion were (1) not considered to be empirical studies or full-text peer-reviewed documents (eg, conference abstracts, protocols, and a PhD thesis; n=19), (2) mentioned PD-related activities, but no PD tools (n=7), (3) mentioned co-design but no PD tools (n=8), (4) mentioned cocreation but no PD tools (n=11), (5) mentioned user-centered design, but no PD tools (n=11), and (6) did not mention eHealth (n=6).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. eHealth: electronic health; PD: participatory design.



Data Extractions and Synthesis

To provide an overview of the general characteristics of the studies, the eHealth domain, the health domain, and the theoretical references used to refer to PD were summarized. In addition, the use of stakeholders, tools, and outcome measures was assessed as follows.

First, regarding the use of stakeholders, different strategies could be used depending on the interpretation of PD principles. Therefore, data were extracted related to the number and type of stakeholders, stakeholder recruitment, and stakeholder management. Second, regarding the use of tools, different tools can be used at different times depending on the PD principles. Therefore, the type of tool and the purpose in using the tool were extracted. Finally, the study was placed in a design phase depending on the stage in which the study started: predesign, early design, or post first prototype (Table 1).

Table 1. Description of design phases.

Design phase	Description
Predesign (including fuzzy front end)	Phase of understanding and defining the problem, often these studies would focus on the unmet need of a certain population.
Early design	In this phase, there is already some understanding of the problem and the aim is to develop a first concrete idea, often these studies would aim to develop or enhance a first idea or prototype.
Post first prototype	In this phase, there is already a first idea for a solution, which will be iterated or enhanced.

Third, as the use of outcome measures is influenced by the general methodological aim and the principles that are emphasized, the type of outcome measure that was used to evaluate eHealth and the PD itself was extracted.

Assessment of Sufficiency of Reporting

Owing to the variety of study designs, a quality assessment was not appropriate. Instead, an assessment of sufficiency of reporting was conducted, as used in a previous systematic review [7]. This was done with an 8-item checklist:

1. Setting: Is it clear where the PD development of the intervention took place?
2. Stakeholders: Is it clear who was involved in the PD, and does one know all that one needs to know about the participants?
3. Facilitators: Is it clear who facilitated the PD process?
4. Procedure: Is it clear what PD methods were used?
5. Materials: Are any physical materials used in the PD process adequately described?
6. Intensity: Is the length of the PD phase and individual sessions clear?

7. Schedule: Is the interval and frequency of the PD sessions clear?
8. Clarity: Is the description of the overall PD process clear?

Results

Overall Findings

The general health and eHealth technology characteristics and the theoretical references used in the studies when referring to PD are described below. The year of publication ranged from 2006 to 2019. The 69 studies cover 65 unique eHealth technology products and services. The majority of these were either Web-based tools such as *online self-management tools* [25], *person-centered Web support* [26], or a *Web-based plan for integrated care* [27], or mHealth apps.

There is a large diversity in the health domains considered. The mental health domain was most often addressed by the eHealth technology. The most frequent aims of the eHealth were disease-specific interventions (weight loss, psychosocial care, and rehabilitation) and self-management. The prevalence of self-management aims could be expected because the PD democratic principle emphasizes the involvement of users, and this may help the later uptake by these users of eHealth focused on self-management.

In addition to the health and eHealth technology characteristics, the theoretical references of PD are presented here. Nearly all studies, 65, mentioned a theory of PD. Clemensen et al's description of the PD methodology [28,29] was referenced in 10 of the reviewed studies [30-39], and that by Sanders and Stappers [13] was referenced in 9 [27,30,31,40-45]. A handbook on PD by Simonsen et al [10] was referenced 7 times [27,31,36,46-49]. PD principles and practices [50,51] were also referenced on several occasions [38,47,52-54]. In addition, the methodology by Spinuzzi [2] was referenced in 4 papers [25,26,31,32]. References to other design theories were also used, such as experience-based design [55] in studies by Wherton et al [39] and Crosby [56], design thinking [57,58] in various studies [37,56,59-61], human-centered design [62] in the study by Das and Svanæs [63], and prototyping [64] in the study by Hetrick et al [65].

Reporting on Stakeholders, Tools, and Outcome Measures

The reporting on stakeholder recruitment, stakeholder management, PD tools (make, tell, or enact), and outcomes measures to evaluate eHealth and the PD process is presented in Multimedia Appendix 3. The amount of reporting varied widely between 8 and 36 on a reporting scale of 40. All studies naturally reported on some kind of PD tools being used as this was an inclusion criterion.

Overall, 25 of the studies stated that an aim of the study was to describe the PD process or provide details of the PD process or of a design process similar to it (see gray-shaded rows in Multimedia Appendix 3). These studies scored highest on the reporting scale, with 13 of the 17 studies scoring above 30 stating that describing the PD process was an aim.

Overall, 38 studies reported on stakeholder recruitment and 30 studies reported on stakeholder management. In addition, 23 studies reported outcome measures to evaluate the eHealth technology under development, and 3 studies reported outcomes to evaluate a PD process that was already employed.

Stakeholders

Types of Stakeholders

Overall, the number of participants taking part in the PD activities varied across studies. The number depended on the different types of stakeholders and the timing of the PD activities.

A total of 63 studies reported on the stakeholders involved. All of these studies involved the main intended user of the eHealth technology in the design process: the patient, the care professional, or both (see Multimedia Appendix 4 and Table 2). Among the patient, or content expert, stakeholder group, young adults and children were involved in 17 studies. Many other stakeholder types were also involved in some studies. For instance, dieticians, psychologists, a social worker, and a journalist were all involved in 1 study [46], 1 study involved a business analyst [38], 1 study a pharmacist [66], and another involved government representatives [35]. In all, 3 studies also involved, alongside a core group of stakeholders, advisory groups to provide feedback at different times [25,67,68].

Table 2. Types of stakeholders included in the participatory design process (n=69).

Stakeholder	Studies
Patient or content expert	[25-27,30-35,37-40,42-46,49,52-54,56,59-61,63,65,68-93]
Care professional	[26,27,31,32,34,35,37,39,41,45,46,48,56,60,61,63,65,67,68,71,73,74,76,84,87,94-99]
Informal caregiver (ie, parent)	[32,35,46,60,65,68,78,87,97]
Designer	[25,26,42,46,52,65,71,76,98]
Software developer	[25-27,38,39,42,46,48,61,63,68,74,97]
Researcher	[25-27,32,37,41,42,52,54,61,63,65,68,98]

Stakeholder Recruitment

The reporting on recruitment was mostly about the patient or content experts and not the other stakeholders. For instance, no study clearly explained how they recruited designers or software developers. This may be because these stakeholders were not recruited but already part of the project team. The most common recruitment strategy was purposive or convenience sampling

[30,31,33,37,42,46,52,63,70,71,80,83] followed by snowball or in-person recruitment [40,65,71,79]. One study used representative sampling to include all potential users [67]. In all, 5 studies aimed for diversity in recruitment [25,45,48,81,85].

Most studies that report recruitment criteria focused on age and health care exposure. A total of 7 studies also mentioned access to internet and basic knowledge in using phones or a computer and the internet. Overall, 4 studies also reported criteria related to personal traits such as social or communicative skills, creativity, motivation, and capabilities to engage actively [31,48,85,90]. Financial incentives were also often used in the recruitment process.

In general, there is a lack of methodological arguments provided for the recruitment choices. It is unclear why designers are involved in so few studies. The PD projects may have worked with researchers who were trained in design, or they may have consulted designers before or after the PD project. Furthermore, methodological argumentation is missing on how the recruitment criteria serve the PD process and PD design aims. For instance, arguments referring to PD principles could be used to substantiate the criteria chosen. As an example, the decision to use personal trait criteria could be substantiated by stating that people who are more communicative and motivated may share more relevant knowledge than others and help others to learn from each other. These arguments could refer to the PD principle of mutual learning. Optimizing mutual learning may be particularly relevant in a health care context, given health care professionals' limited available time.

Stakeholder Management

In terms of stakeholder management, creating a safe environment is important. Many approaches were reported, for example, a safe environment was sometimes fostered by creating small groups [37,63]. Sessions were deliberately shortened to reduce the burden on chronically ill patients and to give them time to reflect between sessions [49,91]. On other occasions, reassurance was provided by a researcher that no judgement was involved to avoid intimidation [40], or an explanation was provided that there was a flat communication structure [27,63].

Others mentioned the use of an icebreaker [80]. Introductions were given and sometimes also refreshments [85]. Games were used to establish the aims and rules of a workshop [71]. Others used a quick design exercise as an icebreaker, especially to get the participants used to participating in design activities [32].

Moderation was also used to reduce doubt and to seek consensus [65]. Field kits [41] or graphics [31] were used to clarify and explain concepts to clinicians and developers. Some reported that training sessions had been provided [32,47,85,92]. Information was provided using popular metaphors on key data points that were important in the design of the product or service [32]. Some studies helped children by explaining the interface and what was technically feasible during the exercises [75,89,90]. The expectations regarding a creative exploration component were clearly explained to nondesigners in one study. Elsewhere, it was made clear to the participants that the focus was on creativity and that they should not reflect on implementation at that stage [91,98]. One study [91] explicitly chose not to explain the existing technologies in order to not influence the participants and constrain their ideas.

Various approaches were taken toward the mixing of groups. Some studies chose to address the power imbalance between health professionals and patients by separating stakeholders [63,65]. Others wanted to mix stakeholders to cross-fertilize perspectives in some instances but keep subgroups by type to highlight the perceptions of a stakeholder group such as caregivers [67].

Some measures were also taken to stimulate creativity when tools were being used. To stimulate intuitive representations [32], participants were given blank cards and were invited to write on them directly [98]. Some facilitators also took an active role in helping participants suggest creative ideas but without trying to be dominant [80]. Another measure that was taken at the end of a PD session was to invite participants to walk around and look at the creations of other teams (world cafés) to increase the diversity of perspectives [32,93]. Consensus over a range of created ideas was moderated by inviting teams to evaluate the differences between ideas.

The reported facilitation varied between involving researchers and designers [42], a team of clinicians and designers [71], or a clinician and researchers [44]. Facilitation was intended to support creativity and hands-on exercises [37,48]. A mental health professional was also present during a workshop with participants who were at risk of psychological distress [73].

On some occasions, arguments related to PD principles are provided to substantiate the stakeholder management. For instance, when justifying exercises that are meant to stimulate creativity. However, further argumentation could have been provided about the relationship between creativity and the design goals.

Tools

A variety of PD tools are used in the studies that report the development of eHealth in the predesign, early design, and post first prototype phases (see Table 3). Looking at all three phases, most combinations of tools are used in the predesign phase [31,37,39,44,97]. In this phase, 4 studies used combinations of three different types of PD tools (make, tell, and enact) [46,49,89,90]. The predesign phase is also characterized by mainly make tools that adopt a generative approach. Some studies also used a toolkit or field kit [41,47], which indicates the emphasis on helping people generate new ideas. This is different from the early design, and post first prototype phase, where fewer tools and fewer combinations of tools are used.

In all, 8 studies referred to specific techniques for a participatory prototyping approach such as *thinking aloud* [42,46,52,65,70,98], and 1 study referred to a card sorting technique for tell tools (Collaborative Analysis of Requirements and Design; CARD) [63]. Furthermore, methodological references were made to Design studio [65], Scaffold [41], the *good enough* model [71], and future workshops [80,91].

Table 3. Tools (n=69).

Phase and tools	Studies
Predesign	
2D mapping, brainstorm, post-it, mind map, Chinese portrait [26]	[30,34,37,41,44-46,49,61,66,77,80,83-85,88,91,92,97-100]
Prototyping, 2D mockup, 2D design, sketch	[30,31,34,37,40,42-44,46,49,61,67,77,78,89-92,98,100,101]
Personas	[37,49,71,88]
Cards	[31,37,39,47,49,67,84,100]
Artifact for discussion	[85]
Storyboarding	[31,37,39,46]
Scenarios, customer journey	[44,66,89,90]
Service blueprint	[66]
Role-play	[46,49,82,89,97]
Design journal notebook	[91]
Early design	
2D mapping	[63,69]
2D mockups, sketch	[25,32,65,68-70,72-74]
Cards	[32,63]
Storyboarding	[26]
Scenarios	[56]
Post first prototype	
2D mapping, brainstorm, post-its	[36,52,59,76,79,81,93,96,102]
Prototyping, 2D or 3D mockup, sketch	[27,33,35,36,38,48,52-54,59,71,75,76,81,86,96]
Persona	[35,79,93]
Cards	[79,86]
Storyboarding	[53,76]
Scenarios, user journey	[35,54,81,93,102]
Role-play	[79]

When looking at the substantiation offered for the PD tools used, different types of methodological arguments can be identified. Most studies argued that their main goal was to gather information or to develop, organize, or test new ideas to improve the product or service design (type 1). In many studies, an argument based on analogy is used to explain why they chose certain tools by referring to other PD literature where similar tools were used with similar design process aims (type 2).

Some authors specifically argued why they used certain generative tools by explaining the type of knowledge that they seek to capture (Type 3). Phillips et al [88] explained why they used empathy maps with people living with HIV was precisely because it is a good tool for exploring topics people feel shameful about. Ahmed et al [32] specifically highlighted their aim of using PD to visualize information in an actionable way. Some visualization tools, such as a timeline, were specifically used to capture hopes and beliefs about the future [59]. How et al stressed that their aim with PD was to merge different domains of knowledge brought together in the co-design process in their project [29]. In doing so, “the ‘Technology Domain’ comprises of selected emergent technologies that could inspire new design ideas, and the ‘Health-care Domain’ comprises of health areas that are of interest for developing new technological applications.” The authors explained that the co-design tools were specifically chosen to bring these knowledge domains together and develop a solution in this knowledge-sharing process. One study also referred to the use of certain tools including storyboards to help stakeholders express their deeper tacit knowledge [31]. In all, 4 studies [30,41,69,91] used specific generative tools such as field kits, workbooks, and design journals without explicitly reporting why these specific tools were chosen. As implied by Peters et al [30], one might assume that they were used to sensitize in the sense that they can help stakeholders express their deeper or tacit knowledge.

Some studies also related the knowledge advantage of using tools to the stakeholders involved in the PD project (type 4). This type of study justifies identifying knowledge domains related to stakeholders and then choosing outcome measures to capture that knowledge. One study explicitly stated the value of having a design expert in the teams to help select appropriate tools [37]. Another study referred to PD principles in involving clinicians as nondesigners in the design decision-making process

to enhance their views and facilitate insights of others in the design [75]. This suggests that the authors related their recruitment strategy and stakeholder management to the use of PD activities and tools.

Outcome Measures

Some of the studies evaluated the eHealth product or service output after the PD activities were concluded. The eHealth output varies depending on whether the development is in the predesign, early design, or post first prototype stage. Overall, 50 studies considered that the outputs of the PD process were in agreement with findings from similar studies or, in the case of an eHealth product, that after testing, they were effective. For instance, in an early design study, it was reported that "our design considerations show agreement with previous work related to human-factors for telerehabilitation technologies" [41]. A study where eHealth technology had been developed to a later stage reported that "we constructed an EHR-tethered PHR module named MyHealthKeeper and implemented this software in an EHR-friendly hospital" [74], which can be seen as indicating that the technology output was considered effective. Only 1 study [102] reported a negative experience: an app that had been developed for nurses did not improve the workflow, although important lessons were drawn.

Of these 50 studies that considered the outputs to be positive or effective, 22 studies reported outcome measures. These outcome measures concerned the development of the eHealth (ie, ideas developed), the quality of the eHealth (ie, usability), and the outcomes for the user (eg, body weight, managing medication, or education on health topics; see Table 4). Most of the reported outcome measures were related to usability and user feedback. As an outcome of the idea generation process, 2 studies measured the number of ideas [90,96]. Another measured the quality of new ideas: they were grouped under labels and then rated by clinicians [41]. 2 studies reported outcome measures based on clinical parameters and participation in activities for care transitioning, managing medication and education on topics such as health insurance [59,74]. There was another study reporting clinical outcome measures (not reported in Table 4); however, the authors did not make it clear whether they considered the eHealth to be effective [100].

In terms of substantiating the choices for certain outcome measures for evaluating eHealth, methodological arguments were generally missing. However, the outcome measures that How et al [41] used, such as idea grouping and the use of labels, suggest that their intention was to evaluate the knowledge development process. This could have been further substantiated by referring to PD principles related to the principles of mutual learning or creativity, for instance, to measure the impact of tools on ideas developed or shared.

Next to evaluating eHealth technology, some studies also evaluated the development process itself. Overall, 55 studies, based on the experience of the authors, considered the PD method to have successfully contributed to the eHealth development. For instance [41]:

> *Through a mediated exploration with clinicians and technology co-designers, we could broadly explore opportunity areas for new technologies within a healthcare domain and unravel initial design considerations related to this intersection.*

Of these 55 studies that considered the method to have effectively contributed to the eHealth development, 3 studies reported outcome measures [41,45,93] (see Table 4). Outcome measures were reported regarding the quality of the knowledge development process (ie, unique ideas) and stakeholder management (ie, voices heard [45]).

When it came to substantiating the outcome measures chosen for method evaluation, methodological argumentation was again generally missing. However, the outcome measures that How et al [41] used do suggest that the intention was to evaluate the knowledge development process. The authors measured how stakeholders rated the extent to which they had an understanding of the new technology and the extent to which the use of clinical knowledge was enabled in the co-design process. Similar arguments related to knowledge expression may have driven the choice of stakeholder management outcome measures made by Revenas et al [45].

Table 4. Outcome measures used when electronic health technology and the participatory design method were positively evaluated (n=69).

Outcomes measures	Studies
eHealth[a] evaluation	
eHealth development (number of ideas for development)	[41,90,96]
eHealth quality (usability, feasibility)	[30,35,46,52,53,56,63,66,68,69,71,72,75,84,90,92,96]
User outcome (effectiveness)	[59,74]
Participatory design method evaluation	
Quality of ideas (ie, unique ideas)	[41]
Understanding of new technology through co-design process	[41]
Enablement of clinical knowledge through co-design process	[41]
Overall experience	[45,93]
Workshop content in line with the aim	[45]
Voices heard (perception)	[45]
Balance between voiced patients and care professionals	[45]

[a]eHealth: electronic health.

Discussion

Principal Findings

Overall, reporting on PD methods varied significantly in studies where PD is used to develop eHealth. The extent of the reporting depended on whether or not the aim of the study was to report on the PD process itself. When it came to substantiating the methodological choices made, the justification for the tools used tended to be given the most attention.

Only a few authors explicitly used arguments directly related to PD guiding principles such as democratic, mutual learning, tacit and latent knowledge, and collective creativity. Even though the PD principles were not explicitly discussed in the method of many studies, they were implicitly identified in some. The arguments used to substantiate the choices made in stakeholder management, PD tools, and the type of outcome measures point to these principles being considered. In this discussion, the results regarding the stakeholders, tools, and outcomes are discussed separately and considered alongside other literature.

A few studies had a clear recruitment strategy, and two studies aimed for diversity in recruitment. Purposive and convenience sampling were most often used. Some studies, when reporting on recruitment, gave the recruitment strategy or the recruitment criteria. However, it was often unclear why certain stakeholders were included or excluded or why certain recruitment criteria were used. For instance, in line with the mutual learning and creativity principle, it could be expected that the recruitment strategy would aim to include designers, and this was rarely the case.

The recruitment criteria that were mentioned included age, health care exposure, access to internet, knowledge of using phones and internet, communicative skills, motivation, and capabilities to engage actively. Few studies included criteria related to personal characteristics such as communication, motivation, and engagement. This is perhaps surprising given the importance of knowledge transmission in relation to the principles of mutual learning and collective creativity. Furthermore, some studies used financial incentives to recruit individuals.

In the PD literature, the levels of expertise, passion, and creativity are suggested to play important roles in the PD process [13]. Expertise has also been suggested by others as an important condition in enhancing the creative process [103]. A meta-analysis of the PD of serious games also mentioned expertise being included as a factor of interest, but it was not found in the included studies [21]. Diversity has also been stated to play an important role in the creativity process [3,104]. Considering these personal characteristics as a whole, diversity was only identified in the recruitment strategy of a few studies in this review. This is surprising, and we would have expected the assessment of personal traits to be more prominent in the recruitment strategies in the studies included in this review.

In terms of stakeholder management, the results of this study show that various actions were taken. Moderation was aimed at providing a safe environment for equal participation, and facilitation was adopted to enhance knowledge sharing between stakeholders and to enhance creativity. This shows that some studies did consider the democratic and creativity principles of PD. Consideration was given to managing the PD process by providing a presentation about its content. In line with the principles of mutual learning and collective creativity, it may be important to manage explanations, given the different levels of expertise of care professionals, software developers, and patients involved. Overall, we had the sense that there was an implicit emphasis on creativity and understanding in some studies, but it remained unclear why a certain form of stakeholder management was chosen.

As noted above, one study may have considered the cognitive abilities of the users involved. This was also suggested in a recent meta-analysis of PD used to develop serious games where it was stressed that one should facilitate the PD tools according to the users' cognitive abilities to increase the quality of idea

generation [21]. In addition, others have also stressed that creativity can be managed on an individual level or on a group level [3,105]. Overall, it seems that adequate attention is being given to facilitating the creative process. On a personal level, creativity is correlated with a mental state of flow, and therefore, facilitating this state may play an important role in developing high-quality ideas in the PD process [16,103].

Various combinations of tools are used across the various design phases of eHealth. Some studies also described the use of toolkits, the scaffold method, the CARD technique, and the think aloud technique. The greater use of combinations of tools with a generative approach in the predesign phase may indicate that authors used these combinations to generate more new ideas. This is in line with the principles of collective creativity and tacit and latent knowledge. When looking at the arguments used to select tools, the argumentation could be categorized into four types of arguments related to knowledge development: (1) tools are used to harvest ideas for the product or service development, (2) arguments in favor of the tools based on other literature, (3) arguments explaining the aim of the tools to retrieve specific type of knowledge, and (4) arguments explaining the aim of the tools in relation to the stakeholders involved.

This focus on knowledge arguments was expected as this is implied in other publications. However, it has not yet been explicitly summarized in terms of levels of argumentation. Others have stated the importance of recognizing the fundamental role of knowledge development in PD. Given the nature of PD, this implies gaining an understanding and a generative creativity that leads in itself to different ways of knowing [4]. In terms of epistemology, the field of knowledge development is closely related to creative processes. Sanders and Stappers [3] have hinted at using social creativity theory and a *path of expression*. Even though the knowledge development theory could be a building block in a methodological framework, it is remote from practical methodological guidelines on selecting between PD tools.

In terms of outcome measures, only a limited number of studies reported outcome measures to evaluate eHealth development and the use of the PD process itself. One study in this review described the outcome measures in considerable depth for the evaluation of both eHealth and the method [41]. Compared with the other studies reviewed, this study had a more rigorous methodological framework, which also substantiated the chosen tools. This study explicitly explained that the focus was on the development of ideas and the use of different fields of expertise and knowledge. It also hinted at considerations related to knowledge developments related to the chosen tools. Nevertheless, it remains challenging to propose appropriate outcome measures to capture the output of creativity given our current understanding of it. These methodological challenges may prevent reporting the use of certain epistemological argumentations.

The identified lack of outcome measures is in line with findings elsewhere. Previous systematic reviews have also highlighted the lack of transparency about the evaluations of PD [7,8]. However, depending on the methodology and design phase, different outcome measures are suggested to evaluate the method [15,106]. Three output domains have been suggested related to the stakeholders (ie, empowerment), to knowledge (ie, tacit, pragmatic, and technical), and to implementation (ie, ownership) [16].

Limitations

The results of this study are limited for several reasons. First, the search strategy for relevant research is limited by the focus on papers published in scientific journals. Given that many reports on PD in developing eHealth are not in scientific journals, the review only provides a partial view of the state of reporting PD methodology, namely only that in the empirical scientific literature.

The screening process is limited by the definitions applied for the terms used in the inclusion and exclusion criteria. As there is no universally agreed definition of PD, a working definition was chosen that focuses on one strand of PD research, namely where stakeholders are a partner in the process. Consequently, studies describing PD in a more user-centered way were excluded, and their inclusion may have led to different results.

Turning to the analysis and conclusions, the following limitations were identified. First, it is challenging to draw conclusions based on the reporting of the PD methods as described in the papers selected in the systematic literature review. The actual methodological intentions and considerations made during the PD project may differ from what is reported in the studies. The limited number of studies reporting outputs and outcome measures may be related to the recognized publication bias toward reporting positive results and eHealth products and services that are already fairly developed. In addition, the evaluation of the eHealth technology may have been reported in a separate publication; for example, in the paper by Waller et al [98] included in this review, it is noted that the results of the randomized controlled trial of the eHealth technology are reported elsewhere [107], and the latter paper did not meet the inclusion criteria for this review. This was because studies that focused on the outcome measures of the eHealth technology were excluded from this review.

Implications

The PD methodology is still under development [2,4,108]. Providing methodological reasoning in a transparent way about the choices of stakeholders, tools, and outcome measures employed is important for methodological progress. A clear PD methodology could well enhance the development of eHealth in practice as practitioners would then be able to argue more rigorously for a certain form of PD. A clear methodology may also improve the rigor and accountability of the science of PD. For instance, given a methodology, evaluation criteria could be used to evaluate the method, which can then inform other researchers about how it can be further improved. A clear methodology may also help to select an appropriate form of PD for a specific research design.

Reflection

The fact that the methodological reasonings behind the use of PD are not widely reported could be because of several reasons.

From a scientific perspective, PD has mixed origins, ranging from social science through action research to the design sciences [3]. This may result in different scientific reporting styles appearing across the scientific literature; for example, the theoretical underpinnings of a methodology tend to be much less described in empirical literature in the health sciences than elsewhere.

The academic design culture is still developing alongside other different cultures such as engineering, the arts, and the social sciences [109]. Although classical research methods and design methods are closely related, they are different. In the PD science field, one sees many different crossovers; for example, one can involve research for a single aspect during a design project but also fully incorporate research methods at every design step. Depending on how research is used in a PD project, the reporting will differ. When the emphasis is on scientific reporting, the methodological steps tend to be explained, but when the emphasis is on design reporting, the design products will be more heavily emphasized. Looking at the results of our study from this perspective, one could argue that the majority of the authors have put the emphasis on design reporting and less on scientific reporting.

This observation can be further explained using the observations by Spinuzzi [2], who claimed that there is no strong methodological justification for PD in the first place. Although there are some principles, stated in this study, on how PD should be conducted, a methodological framework for PD is scarcely discussed [2,4]. This may leave researchers confused as to how to employ and report on PD methodology.

PD reporting could be improved if PD researchers were to adopt a more *scientific* attitude toward carrying out a PD project. Improving documentation on the choices of certain PD recruitment strategies, the use of certain tools, or the use of outcome measures could provide more information that could then be reported in scientific journals. Improving education about the scientific documentation of PD projects for designers and eHealth developers could help to improve future reporting. One key challenge here is to translate design terminology to scientific terminology and vice versa; for example, prototype testing in design might be translated as hypothesis testing in science.

Further Research

Further research can help improve the methodological framework for PD in eHealth. A particular focus on the knowledge development process, as a core aspect of PD, would greatly help in substantiating methodological choices and in measuring the outputs of a PD process, especially in eHealth given the various areas of technical knowledge involved. There is a growing interest in the methodology of design known as *Research through Design* [109], which could help foster the development of a methodological framework for PD that would help develop better eHealth.

Conclusions

Studies that use a PD research methodology to develop eHealth primarily substantiate the choice of tools and much less the selection of stakeholders and outcome measures.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Systematic review: search strategy.
[DOCX File , 17 KB - jmir_v22i4e13780_app1.docx]

Multimedia Appendix 2
Excluded studies.
[DOCX File , 33 KB - jmir_v22i4e13780_app2.docx]

Multimedia Appendix 3
Reporting on stakeholders, tools and outcomes.
[DOCX File , 31 KB - jmir_v22i4e13780_app3.docx]

Multimedia Appendix 4
Stakeholder recruitment and management.
[DOCX File , 37 KB - jmir_v22i4e13780_app4.docx]

References

1. 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 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
2. Spinuzzi C. The methodology of participatory design. Tech Commun 2005;52(2):163-174 [FREE Full text]
3. Sanders L, Stappers P. Convivial Toolbox: Generative Research for the Front End of Design. Amsterdam: BIS; 2012.

4. Frauenberger C, Good J, Fitzpatrick G, Iversen OS. In pursuit of rigour and accountability in participatory design. Int J Hum Comput Stud 2015 Feb;74:93-106 [FREE Full text] [doi: 10.1016/j.ijhcs.2014.09.004] [Medline: 26109833]
5. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. Interact J Med Res 2013 Apr 16;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
6. Dekker MR, Williams AD. The use of user-centered participatory design in serious games for anxiety and depression. Games Health J 2017 Dec;6(6):327-333. [doi: 10.1089/g4h.2017.0058] [Medline: 28956617]
7. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
8. Merkel S, Kucharski A. Participatory design in gerontechnology: a systematic literature review. Gerontologist 2019 Jan 9;59(1):e16-e25. [doi: 10.1093/geront/gny034] [Medline: 29788319]
9. van der Velden M, Mörtberg C. Participatory design and design for values. In: van den Hoven J, Vermaas PE, van de Poel I, editors. Handbook of Ethics, Values, and Technological Design. Dordrecht: Springer; 2014:1-22.
10. Brandt E, Binder T, Sanders EB. Tools and techniques: Ways to engage telling, making and enacting. In: Simonsen J, editor. Routledge International Handbook of Participatory Design. Abingdon, United Kingdom: Routledge; 2013:145-181.
11. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Participat Med 2019;11(1):e11474. [doi: 10.2196/11474]
12. Palumbo R. The Bright Side and the Dark Side of Patient Empowerment: Co-creation and Co-destruction of Value in the Healthcare Environment. New York, United States: Springer; 2017.
13. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
14. Steen M. Tensions in human-centred design. CoDesign 2011 Mar;7(1):45-60. [doi: 10.1080/15710882.2011.563314]
15. Steen M, Manschot M, de Koning N. Benefits of co-design in service design projects. Int J Des 2011;5(2):53-60 [FREE Full text]
16. 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]
17. Sanders E, Brandt E, Binder T. A Framework for Organizing the Tools and Techniques of Participatory Design. In: Proceedings of the 11th Biennial Participatory Design Conference. 2010 Presented at: PDC'10; November 29 - December 3, 2010; Sydney, Australia p. 195-198. [doi: 10.1145/1900441.1900476]
18. Wahyuni D. The research design maze: understanding paradigms, cases, methods and methodologies. J Appl Manag Account Res 2012 Jun;10(1):69-80.
19. Sanders E, Stappers P. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
20. Sanders E. Converging perspectives: product development research for the 1990s. Des Manag J 1992;3(4):49-54. [doi: 10.1111/j.1948-7169.1992.tb00604.x]
21. DeSmet A, Thompson D, Baranowski T, Palmeira A, Verloigne M, de Bourdeaudhuij I. Is participatory design associated with the effectiveness of serious digital games for healthy lifestyle promotion? A meta-analysis. J Med Internet Res 2016 Apr 29;18(4):e94 [FREE Full text] [doi: 10.2196/jmir.4444] [Medline: 27129447]
22. Kanstrup A, Madsen J, Nøhr C, Bygholm A, Bertelsen P. Developments in participatory design of health information technology - a review of PDC publications from 1990-2016. Stud Health Technol Inform 2017;233:1-13. [Medline: 28125408]
23. Cochrane Training. Cochrane Handbook for Systematic Reviews of Interventions URL: https://training.cochrane.org/handbook [accessed 2018-10-12]
24. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
25. Allin S, Shepherd J, Tomasone J, Munce S, Linassi G, Hossain SN, et al. Participatory design of an online self-management tool for users with spinal cord injury: qualitative study. JMIR Rehabil Assist Technol 2018 Mar 21;5(1):e6 [FREE Full text] [doi: 10.2196/rehab.8158] [Medline: 29563075]
26. Berg M, Adolfsson A, Ranerup A, Sparud-Lundin C, University of Gothenburg Centre for Person-Centred Care. Person-centered web support to women with type 1 diabetes in pregnancy and early motherhood--the development process. Diabetes Technol Ther 2013 Jan;15(1):20-25. [doi: 10.1089/dia.2012.0217] [Medline: 23297670]
27. Bjerkan J, Hedlund M, Hellesø R. Patients' contribution to the development of a web-based plan for integrated care - a participatory design study. Inform Health Soc Care 2015 Mar;40(2):167-184. [doi: 10.3109/17538157.2014.907803] [Medline: 24786524]
28. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]

29. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
30. Peters D, Davis S, Calvo RA, Sawyer SM, Smith L, Foster JM. Young people's preferences for an asthma self-management app highlight psychological needs: a participatory study. J Med Internet Res 2017 Apr 11;19(4):e113 [FREE Full text] [doi: 10.2196/jmir.6994] [Medline: 28400353]
31. Terp M, Laursen BS, Jørgensen R, Mainz J, Bjørnes CD. A room for design: through participatory design young adults with schizophrenia become strong collaborators. Int J Ment Health Nurs 2016 Dec;25(6):496-506. [doi: 10.1111/inm.12231] [Medline: 27293176]
32. Ahmed R, Toscos T, Ghahari RR, Holden RJ, Martin E, Wagner S, et al. Visualization of cardiac implantable electronic device data for older adults using participatory design. Appl Clin Inform 2019 Aug;10(4):707-718. [doi: 10.1055/s-0039-1695794] [Medline: 31533172]
33. Danbjørg DB, Villadsen A, Gill E, Rothmann MJ, Clemensen J. Usage of an exercise app in the care for people with osteoarthritis: user-driven exploratory study. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e11 [FREE Full text] [doi: 10.2196/mhealth.7734] [Medline: 29326092]
34. Jakobsen PR, Hermann A, Søndergaard J, Wiil U, Clemensen J. Development of an mHealth application for women newly diagnosed with osteoporosis without preceding fractures: a participatory design approach. Int J Environ Res Public Health 2018 Feb 13;15(2):pii: E330 [FREE Full text] [doi: 10.3390/ijerph15020330] [Medline: 29438343]
35. LaMonica HM, Davenport TA, Burns J, Cross S, Hodson S, Veitch J, et al. Technology-enabled mental health service reform for open arms - veterans and families counselling: participatory design study. JMIR Form Res 2019 Sep 19;3(3):e13662 [FREE Full text] [doi: 10.2196/13662] [Medline: 31538937]
36. Nielsen C, Agerskov H, Bistrup C, Clemensen J. User involvement in the development of a telehealth solution to improve the kidney transplantation process: A participatory design study. Health Informatics J 2019 Sep 30:1460458219876188. [doi: 10.1177/1460458219876188] [Medline: 31566460]
37. O'Brien N, Heaven B, Teal G, Evans EH, Cleland C, Moffatt S, et al. Integrating evidence from systematic reviews, qualitative research, and expert knowledge using co-design techniques to develop a web-based intervention for people in the retirement transition. J Med Internet Res 2016 Aug 3;18(8):e210 [FREE Full text] [doi: 10.2196/jmir.5790] [Medline: 27489143]
38. Tang T, Lim ME, Mansfield E, McLachlan A, Quan SD. Clinician user involvement in the real world: designing an electronic tool to improve interprofessional communication and collaboration in a hospital setting. Int J Med Inform 2018 Feb;110:90-97 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.11.011] [Medline: 29331258]
39. Wherton J, Sugarhood P, Procter R, Hinder S, Greenhalgh T. Co-production in practice: how people with assisted living needs can help design and evolve technologies and services. Implement Sci 2015 May 26;10:75 [FREE Full text] [doi: 10.1186/s13012-015-0271-8] [Medline: 26004047]
40. AlJaberi H. Social interaction needs and entertainment approaches to pregnancy well-being in mHealth technology design for low-income transmigrant women: qualitative codesign study. JMIR Mhealth Uhealth 2018 Apr 13;6(4):e61 [FREE Full text] [doi: 10.2196/mhealth.7708] [Medline: 29653919]
41. How TV, Hwang AS, Green RE, Mihailidis A. Envisioning future cognitive telerehabilitation technologies: a co-design process with clinicians. Disabil Rehabil Assist Technol 2017 Apr;12(3):244-261. [doi: 10.3109/17483107.2015.1129457] [Medline: 26746683]
42. Kristiansen AM, Svanholm JR, Schjødt I, Jensen KM, Silén C, Karlgren K. Patients with heart failure as co-designers of an educational website: implications for medical education. Int J Med Educ 2017 Feb 25;8:47-58 [FREE Full text] [doi: 10.5116/ijme.5898.309e] [Medline: 28237976]
43. Rochat J, Nap HH, Ricci A, Cornelisse L, Lukkien D, Lovis C, et al. Designing an online social support platform through co-creation with seniors. Stud Health Technol Inform 2018;247:760-764. [Medline: 29678063]
44. Switsers L, Dauwe A, Vanhoudt A, van Dyck H, Lombaerts K, Oldenburg J. Users' perspectives on mHealth self-management of bipolar disorder: qualitative focus group study. JMIR Mhealth Uhealth 2018 May 2;6(5):e108 [FREE Full text] [doi: 10.2196/mhealth.9529] [Medline: 29720363]
45. Revenäs Å, Forsberg HH, Granström E, Wannheden C. Co-designing an eHealth service for the co-care of Parkinson disease: explorative study of values and challenges. JMIR Res Protoc 2018 Oct 30;7(10):e11278 [FREE Full text] [doi: 10.2196/11278] [Medline: 30377143]
46. Castensøe-Seidenfaden P, Husted GR, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a self-management app for young people with type 1 diabetes: methodological challenges, experiences, and recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]
47. Rassmus-Gröhn K, Magnusson C, Rydeman B, Randall G, Belson S. Conjuring up new technology - using magic objects in co-ideation with stroke survivors. Stud Health Technol Inform 2017;242:429-436. [Medline: 28873835]
48. Noergaard B, Sandvei M, Rottmann N, Johannessen H, Wiil U, Schmidt T, et al. Development of a web-based health care intervention for patients with heart disease: lessons learned from a participatory design study. JMIR Res Protoc 2017 May 17;6(5):e75 [FREE Full text] [doi: 10.2196/resprot.7084] [Medline: 28526674]

49. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
50. Schuler D, Namioka A. Participatory Design: Principles and Practices. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993.
51. Muller MJ, Kuhn S. Participatory design. Commun ACM 1993;36(6):24-28. [doi: 10.1145/153571.255960]
52. Cheng VW, Davenport TA, Johnson D, Vella K, Mitchell J, Hickie IB. An app that incorporates gamification, mini-games, and social connection to improve men's mental health and well-being (MindMax): participatory design process. JMIR Ment Health 2018 Nov 19;5(4):e11068 [FREE Full text] [doi: 10.2196/11068] [Medline: 30455165]
53. Scandurra I, Hägglund M, Koch S. Visualisation and interaction design solutions to address specific demands in shared home care. Stud Health Technol Inform 2006;124:71-76. [Medline: 17108506]
54. Wärnestål P, Svedberg P, Lindberg S, Nygren JM. Effects of using child personas in the development of a digital peer support service for childhood cancer survivors. J Med Internet Res 2017 May 18;19(5):e161 [FREE Full text] [doi: 10.2196/jmir.7175] [Medline: 28526663]
55. Bate P, Robert G. Bringing User Experience to Healthcare Improvement: The Concepts, Methods and Practices of Experience-Based Design. Abingdon, UK: Radcliffe Publishing; 2007.
56. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: a self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan;64(1):139-145. [doi: 10.1002/pbc.26177] [Medline: 27574031]
57. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]
58. Stanford d.school. A Virtual Crash Course in Design Thinking URL: https://dschool.stanford.edu/resources/a-virtual-crash-course-in-design-thinking [accessed 2019-12-03]
59. Beaudry J, Consigli A, Clark C, Robinson KJ. Getting ready for adult healthcare: Designing a chatbot to coach adolescents with special health needs through the transitions of care. J Pediatr Nurs 2019;49:85-91. [doi: 10.1016/j.pedn.2019.09.004] [Medline: 31644960]
60. Woods L, Duff J, Roehrer E, Walker K, Cummings E. Design of a consumer mobile health app for heart failure: findings from the nurse-led co-design of Care4myHeart. JMIR Nursing 2019;2(1):e14633 [FREE Full text] [doi: 10.2196/14633]
61. Woods L, Cummings E, Duff J, Walker K. Design thinking for mHealth application co-design to support heart failure self-management. Stud Health Technol Inform 2017;241:97-102. [Medline: 28809190]
62. The British Standards Institution. British Standards Document BS EN ISO 9241-210 Ergonomics of Human-System Interaction. Human-Centred Design for Interactive Systems URL: https://landingpage.bsigroup.com/LandingPage/Undated?UPI=000000000030388991 [accessed 2020-02-26]
63. Das A, Svanæs D. Human-centred methods in the design of an e-health solution for patients undergoing weight loss treatment. Int J Med Inform 2013 Nov;82(11):1075-1091. [doi: 10.1016/j.ijmedinf.2013.06.008] [Medline: 23886482]
64. Warfel TZ. Prototyping: A Practitioner's Guide. New York, NY: Rosenfeld Media; 2009.
65. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth Codesign of a Mobile Phone App to Facilitate Self-Monitoring and Management of Mood Symptoms in Young People With Major Depression, Suicidal Ideation, and Self-Harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
66. Teixeira JG, Pinho NF, Patrício L. Bringing service design to the development of health information systems: the case of the Portuguese national electronic health record. Int J Med Inform 2019 Dec;132:103942. [doi: 10.1016/j.ijmedinf.2019.08.002] [Medline: 31627031]
67. Giroux D, Tremblay M, Latulippe K, Provencher V, Poulin V, Giguere A, et al. Promoting identification and use of aid resources by caregivers of seniors: co-design of an electronic health tool. JMIR Aging 2019 Aug 22;2(2):e12314 [FREE Full text] [doi: 10.2196/12314] [Medline: 31518284]
68. Coyne I, Prizeman G, Sheehan A, Malone H, While AE. An e-health intervention to support the transition of young people with long-term illnesses to adult healthcare services: Design and early use. Patient Educ Couns 2016 Sep;99(9):1496-1504. [doi: 10.1016/j.pec.2016.06.005] [Medline: 27372524]
69. Davis SR, Peters D, Calvo RA, Sawyer SM, Foster JM, Smith L. 'Kiss myAsthma': using a participatory design approach to develop a self-management app with young people with asthma. J Asthma 2018 Sep;55(9):1018-1027. [doi: 10.1080/02770903.2017.1388391] [Medline: 29064746]
70. Grim K, Rosenberg D, Svedberg P, Schön UK. Development and usability testing of a web-based decision support for users and health professionals in psychiatric services. Psychiatr Rehabil J 2017 Sep;40(3):293-302. [doi: 10.1037/prj0000278] [Medline: 28737415]
71. Hobson EV, Baird WO, Partridge R, Cooper CL, Mawson S, Quinn A, et al. The TiM system: developing a novel telehealth service to improve access to specialist care in motor neurone disease using user-centered design. Amyotroph Lateral Scler Frontotemporal Degener 2018 Aug;19(5-6):351-361. [doi: 10.1080/21678421.2018.1440408] [Medline: 29451026]
72. March S, Day J, Zieschank K, Ireland M. The interactive child distress screener: development and preliminary feasibility testing. JMIR Mhealth Uhealth 2018 Apr 19;6(4):e90 [FREE Full text] [doi: 10.2196/mhealth.9456] [Medline: 29674310]

73. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eClinic to improve access to and quality of mental health care for young people: using participatory design as research methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
74. Ryu B, Kim N, Heo E, Yoo S, Lee K, Hwang H, et al. Impact of an electronic health record-integrated personal health record on patient participation in health care: development and randomized controlled trial of MyHealthKeeper. J Med Internet Res 2017 Dec 7;19(12):e401 [FREE Full text] [doi: 10.2196/jmir.8867] [Medline: 29217503]
75. Stålberg A, Sandberg A, Söderbäck M, Larsson T. The child's perspective as a guiding principle: young children as co-designers in the design of an interactive application meant to facilitate participation in healthcare situations. J Biomed Inform 2016 Jun;61:149-158 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.024] [Medline: 27050824]
76. Whitehouse SR, Lam P, Balka E, McLellan S, Deevska M, Penn D, et al. Co-Creation With TickiT: Designing and evaluating a clinical eHealth platform for youth. JMIR Res Protoc 2013 Oct 18;2(2):e42 [FREE Full text] [doi: 10.2196/resprot.2865] [Medline: 24140595]
77. Wiljer D, Johnson A, McDiarmid E, Abi-Jaoude A, Ferguson G, Hollenberg E, et al. Thought spot: co-creating mental health solutions with post-secondary students. Stud Health Technol Inform 2017;234:370-375. [Medline: 28186070]
78. Chomutare T, Johansen S, Hartvigsen G, Årsand E. Serious game co-design for children with type 1 diabetes. In: Mantas J, Hasman A, Gallos P, Kolokathi A, Househ MS, editors. Unifying the Applications and Foundations of Biomedical and Health Informatics. Amsterdam, Netherlands: IOS Press; 2016:83-86.
79. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. J Med Internet Res 2019 May 30;21(5):e12996 [FREE Full text] [doi: 10.2196/12996] [Medline: 31148545]
80. Gardsten C, Mörtberg C, Blomqvist K. Designing an ICT self-management service: suggestions from persons with type 2 diabetes. Health Technol (Berl) 2017;7(2):197-206 [FREE Full text] [doi: 10.1007/s12553-016-0176-9] [Medline: 29201589]
81. Gonsalves PP, Hodgson ES, Kumar A, Aurora T, Chandak Y, Sharma R, et al. Design and development of the 'POD Adventures' smartphone game: a blended problem-solving intervention for adolescent mental health in India. Front Public Health 2019;7:238 [FREE Full text] [doi: 10.3389/fpubh.2019.00238] [Medline: 31508404]
82. Gordon M, Henderson R, Holmes JH, Wolters MK, Bennett IM, SPIRIT (Stress in Pregnancy: Improving Results with Interactive Technology) Group. Participatory design of ehealth solutions for women from vulnerable populations with perinatal depression. J Am Med Inform Assoc 2016 Jan;23(1):105-109. [doi: 10.1093/jamia/ocv109] [Medline: 26342219]
83. Iribarren S, Stonbraker S, Suero-Tejeda N, Granja M, Luchsinger JA, Mittelman M, et al. Information, communication, and online tool needs of Hispanic family caregivers of individuals with Alzheimer's disease and related dementias. Inform Health Soc Care 2019;44(2):115-134 [FREE Full text] [doi: 10.1080/17538157.2018.1433674] [Medline: 29504837]
84. Joensson K, Melholt C, Hansen J, Leth S, Spindler H, Olsen MV, et al. Listening to the patients: using participatory design in the development of a cardiac telerehabilitation web portal. Mhealth 2019;5:33 [FREE Full text] [doi: 10.21037/mhealth.2019.08.06] [Medline: 31620460]
85. Konnerup U. Engaging people with aphasia in design of rehabilitation through participatory design: a way to learn what they really want. Stud Health Technol Inform 2017;233:148-157. [Medline: 28125421]
86. Lucero R, Sheehan B, Yen P, Velez O, Nobile-Hernandez D, Tiase V. Identifying consumer's needs of health information technology through an innovative participatory design approach among English- and Spanish-speaking urban older adults. Appl Clin Inform 2014;5(4):943-957 [FREE Full text] [doi: 10.4338/ACI-2014-07-RA-0058] [Medline: 25589909]
87. Ospina-Pinillos L, Davenport T, Diaz AM, Navarro-Mancilla A, Scott EM, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the mental health eClinic: qualitative study. J Med Internet Res 2019 Aug 2;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
88. Phillips JC, Rowsell DJ, Boomer J, Kwon J, Currie LM. Personas to guide understanding traditions of gay men living with HIV who smoke. Qual Health Res 2016 Jan;26(1):41-54. [doi: 10.1177/1049732315581614] [Medline: 25881965]
89. Ruland CM, Slaughter L, Starren J, Vatne TM. Children as design partners in the development of a support system for children with cancer. Stud Health Technol Inform 2006;122:80-85. [Medline: 17102222]
90. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. J Biomed Inform 2008 Aug;41(4):624-635 [FREE Full text] [doi: 10.1016/j.jbi.2007.10.004] [Medline: 18082468]
91. Skeels MM, Unruh KT, Powell C, Pratt W. Catalyzing social support for breast cancer patients. Proc SIGCHI Conf Hum Factor Comput Syst 2010 Apr;2010:173-182 [FREE Full text] [doi: 10.1145/1753326.1753353] [Medline: 21654894]
92. Swallow D, Petrie H, Power C, Lewis A, Edwards A. Involving older adults in the technology design process: a case study on mobility and wellbeing in the built environment. Stud Health Technol Inform 2016;229:615-623. [Medline: 27534357]
93. VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, et al. The optimization of an eHealth solution (Thought Spot) with transition-aged youth in postsecondary settings: participatory design research. J Med Internet Res 2018 Mar 6;20(3):e79 [FREE Full text] [doi: 10.2196/jmir.8102] [Medline: 29510970]
94. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth Codesign of a Mobile Phone App to Facilitate Self-Monitoring and Management of Mood Symptoms in Young People With Major Depression, Suicidal Ideation, and Self-Harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]

95. Hobson EV, Baird WO, Partridge R, Cooper CL, Mawson S, Quinn A, et al. The TiM system: developing a novel telehealth service to improve access to specialist care in motor neurone disease using user-centered design. Amyotroph Lateral Scler Frontotemporal Degener 2018 Aug;19(5-6):351-361. [doi: 10.1080/21678421.2018.1440408] [Medline: 29451026]
96. Ruland CM, Borosund E, Varsi C. User requirements for a practice-integrated nurse-administered online communication service for cancer patients. Stud Health Technol Inform 2009;146:221-225. [Medline: 19592838]
97. Slomian J, Emonts P, Vigneron L, Acconcia A, Reginster J, Oumourgh M, et al. Meeting the needs of mothers during the postpartum period: using co-creation workshops to find technological solutions. JMIR Res Protoc 2017 May 3;6(5):e76 [FREE Full text] [doi: 10.2196/resprot.6831] [Medline: 28468746]
98. Waller A, Franklin V, Pagliari C, Greene S. Participatory design of a text message scheduling system to support young people with diabetes. Health Informatics J 2006 Dec;12(4):304-318. [doi: 10.1177/1460458206070023] [Medline: 17093001]
99. Woods L, Cummings E, Duff J, Walker K. Conceptual design and iterative development of a mHealth app by clinicians, patients and their families. Stud Health Technol Inform 2018;252:170-175. [Medline: 30040701]
100. Garzo A, Silva PA, Garay-Vitoria N, Hernandez E, Cullen S, de Cock VC, et al. Design and development of a gait training system for Parkinson's disease. PLoS One 2018;13(11):e0207136 [FREE Full text] [doi: 10.1371/journal.pone.0207136] [Medline: 30418993]
101. Verbiest ME, Corrigan C, Dalhousie S, Firestone R, Funaki T, Goodwin D, et al. Using codesign to develop a culturally tailored, behavior change mHealth intervention for indigenous and other priority communities: A case study in New Zealand. Transl Behav Med 2019 Jul 16;9(4):720-736. [doi: 10.1093/tbm/iby093] [Medline: 30388262]
102. Hasvold PE, Scholl J. Disrupted rhythms and mobile ICT in a surgical department. Int J Med Inform 2011 Aug;80(8):e72-e84. [doi: 10.1016/j.ijmedinf.2011.01.006] [Medline: 21317028]
103. Csikszentmihalyi M, Sawyer K. Creative insight: the social dimension of a solitary moment. In: The Systems Model of Creativity. Netherlands: Springer; 2014:73-98.
104. Pentland A. Social Physics: How Good Ideas Spread-the Lessons from a New Science. London, United Kingdom: Penguin Press; 2014.
105. Sawyer RK. Explaining Creativity: The Science Of Human Innovation. Oxford, UK: Oxford University Press; 2012.
106. Roser T, Samson A. aicep Portugal Global. 2009. Co-creation: New Pathways to Value. An Overview URL: http://www.portugalglobal.pt/PT/RoadShow/Documents/2016/GuimaraesCo_creationNewPathways_to_value_An_overview.pdf [accessed 2020-02-26]
107. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med 2006 Dec;23(12):1332-1338. [doi: 10.1111/j.1464-5491.2006.01989.x] [Medline: 17116184]
108. Muller M. Participatory design. In: Sears A, Jacko JA, editors. Human-Computer Interaction: Development Process. Boca Raton, Florida: CRC Press; 2009.
109. Stappers P, Giaccardi E. Interaction Design Foundation. The Encyclopedia of Human-Computer Interaction, 2nd Ed: 43. Research through Design URL: http://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design [accessed 2020-02-21]

Abbreviations

CARD: Collaborative Analysis of Requirements and Design
eHealth: electronic health
mHealth: mobile health
PD: participatory design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 15.03.19; peer-reviewed by C Frauenberger, Å Revenäs; comments to author 03.10.19; revised version received 10.12.19; accepted 16.12.19; published 27.04.20.

Please cite as:
Vandekerckhove P, de Mul M, Bramer WM, de Bont AA
Generative Participatory Design Methodology to Develop Electronic Health Interventions: Systematic Literature Review
J Med Internet Res 2020;22(4):e13780
URL: http://www.jmir.org/2020/4/e13780/
doi:10.2196/13780
PMID:32338617

Original Paper

Interactive Digital Health Tools to Engage Patients and Caregivers in Discharge Preparation: Implementation Study

Theresa E Fuller[1*], BSc; Denise D Pong[1*], MD, MPH; Nicholas Piniella[1], BSc; Michael Pardo[1], MD; Nathaniel Bessa[1,2], BSc; Catherine Yoon[1], MSc; Robert B Boxer[1,2], MD, PhD; Jeffrey Lawrence Schnipper[1,2], MD, MPH; Anuj K Dalal[1,2], MD

[1]Brigham and Women's Hospital, Boston, MA, United States

[2]Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Anuj K Dalal, MD
Brigham and Women's Hospital
Brigham Circle, 1620 Tremont Street
Suite BC-3-002HH
Boston, MA, 02120-1613
United States
Phone: 1 (617) 525 8891
Fax: 1 (617) 732 7072
Email: adalal1@bwh.harvard.edu

Abstract

Background: Poor discharge preparation during hospitalization may lead to adverse events after discharge. Checklists and videos that systematically engage patients in preparing for discharge have the potential to improve safety, especially when integrated into clinician workflow via the electronic health record (EHR).

Objective: This study aims to evaluate the implementation of a suite of digital health tools integrated with the EHR to engage hospitalized patients, caregivers, and their care team in preparing for discharge.

Methods: We used the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework to identify pertinent research questions related to implementation. We iteratively refined patient and clinician-facing intervention components using a participatory process involving end users and institutional stakeholders. The intervention was implemented at a large academic medical center from December 2017 to July 2018. Patients who agreed to participate were coached to watch a discharge video, complete a checklist assessing discharge readiness, and request postdischarge text messaging with a physician 24 to 48 hours before their expected discharge date, which was displayed via a patient portal and bedside display. Clinicians could view concerns reported by patients based on their checklist responses in real time via a safety dashboard integrated with the EHR and choose to open a secure messaging thread with the patient for up to 7 days after discharge. We used mixed methods to evaluate our implementation experience.

Results: Of 752 patient admissions, 510 (67.8%) patients or caregivers participated: 416 (55.3%) watched the video and completed the checklist, and 94 (12.5%) completed the checklist alone. On average, 4.24 concerns were reported per each of the 510 checklist submissions, most commonly about medications (664/2164, 30.7%) and follow-up (656/2164, 30.3%). Of the 510 completed checklists, a member of the care team accessed the safety dashboard to view 210 (41.2%) patient-reported concerns. For 422 patient admissions where postdischarge messaging was available, 141 (33.4%) patients requested this service; of these, a physician initiated secure messaging for 3 (2.1%) discharges. Most patient survey participants perceived that the intervention promoted self-management and communication with their care team. Patient interview participants endorsed gaps in communication with their care team and thought that the video and checklist would be useful closer toward discharge. Clinicians participating in focus groups perceived the value for patients but suggested that low awareness and variable workflow regarding the intervention, lack of technical optimization, and inconsistent clinician leadership limited the use of clinician-facing components.

Conclusions: A suite of EHR-integrated digital health tools to engage patients, caregivers, and clinicians in discharge preparation during hospitalization was feasible, acceptable, and valuable; however, important challenges were identified during implementation. We offer strategies to address implementation barriers and promote adoption of these tools.

Trial Registration: ClinicalTrials.gov NCT03116074; https://clinicaltrials.gov/ct2/show/NCT03116074.

(J Med Internet Res 2020;22(4):e15573) doi:10.2196/15573

KEYWORDS

patient engagement; care transitions; health information technology; implementation science

Introduction

The transition from hospital to ambulatory care is a vulnerable time for patients and stressful for their caregivers: new treatments have been initiated, conditions require monitoring, and the plan is in flux. Approximately 19% to 28% of patients experience preventable adverse events after discharge, many due to suboptimal monitoring of conditions, medication errors or nonadherence, and failure to execute the recovery plan [1-6]. During hospitalization, discharge planning is often initiated late, and input from patients regarding their preparedness is frequently lacking, which may lead to delays and dissatisfaction [7]. After discharge, patients report problems related to follow-up, medications, and self-care; have unanswered questions that could have easily been addressed before discharge [8]; and often feel *more relieved than burdened* when readmitted [9]. Lack of patient engagement during the process of discharge preparation may contribute to avoidable adverse events and costly readmissions [10], particularly those that occur early after hospitalization [11,12].

To date, efforts to enhance and standardize discharge practices have typically targeted clinicians [13,14]; interventions directed at patients provide an opportunity to improve patient understanding, self-management, and postdischarge outcomes [15]. National agencies (eg, Agency for Healthcare Research and Quality [AHRQ], Centers for Medicare and Medicaid Services) are attempting to engage patients and caregivers more broadly by offering access to discharge preparation materials that include checklists for patients [16,17]. Few institutions have determined how best to operationalize these tools for patients. Digital health technology could be leveraged to more proactively engage patients, caregivers, and clinicians during the process of discharge preparation [18-21]; however, currently available patient-facing digital health tools such as patient portals have gaps in functionality with regard to assessing discharge readiness, are not well integrated with the electronic health record (EHR), and present challenges when used during hospitalization [18,21-23]. Although it is technically feasible to administer a discharge checklist through a patient portal or mobile device [24,25], hospitals lack knowledge about the potential for adoption and perceived utility of these tools for patients and clinicians in a real-world clinical setting, as well as potential barriers for sustaining the intervention from an organizational perspective.

To address this knowledge gap, we designed and developed an interactive patient-centered discharge toolkit (PDTK), a suite of EHR-integrated digital health tools that enabled patients to self-assess and communicate discharge preparedness to their care team and request secure text messaging with a hospital physician after discharge as part of a project funded by the AHRQ. Guided by the Reach, Effectiveness, Adoption, Implementation and Maintenance (RE-AIM) framework, we conducted a mixed methods study to describe the use, adoption, and perceived utility of the PDTK, as well as the key challenges encountered during implementation [26].

Methods

Overview

We used RE-AIM (Table 1), a framework designed to address issues related to the implementation of health services and clinical informatics research [26], to identify research questions to evaluate the feasibility and acceptability of the PDTK (Table 2) in a real-world clinical setting. Specifically, we employed a variety of quantitative and qualitative methods to assess the Reach and potential for Adoption, while identifying barriers to Implementation and strategies to Maintain the PDTK for hospitalized general medicine patients. Effectiveness of the intervention on outcomes will be evaluated in future studies.

Table 1. Reach, Effectiveness, Adoption, Implementation, and Maintenance framework: research questions and methods of analysis by dimension.

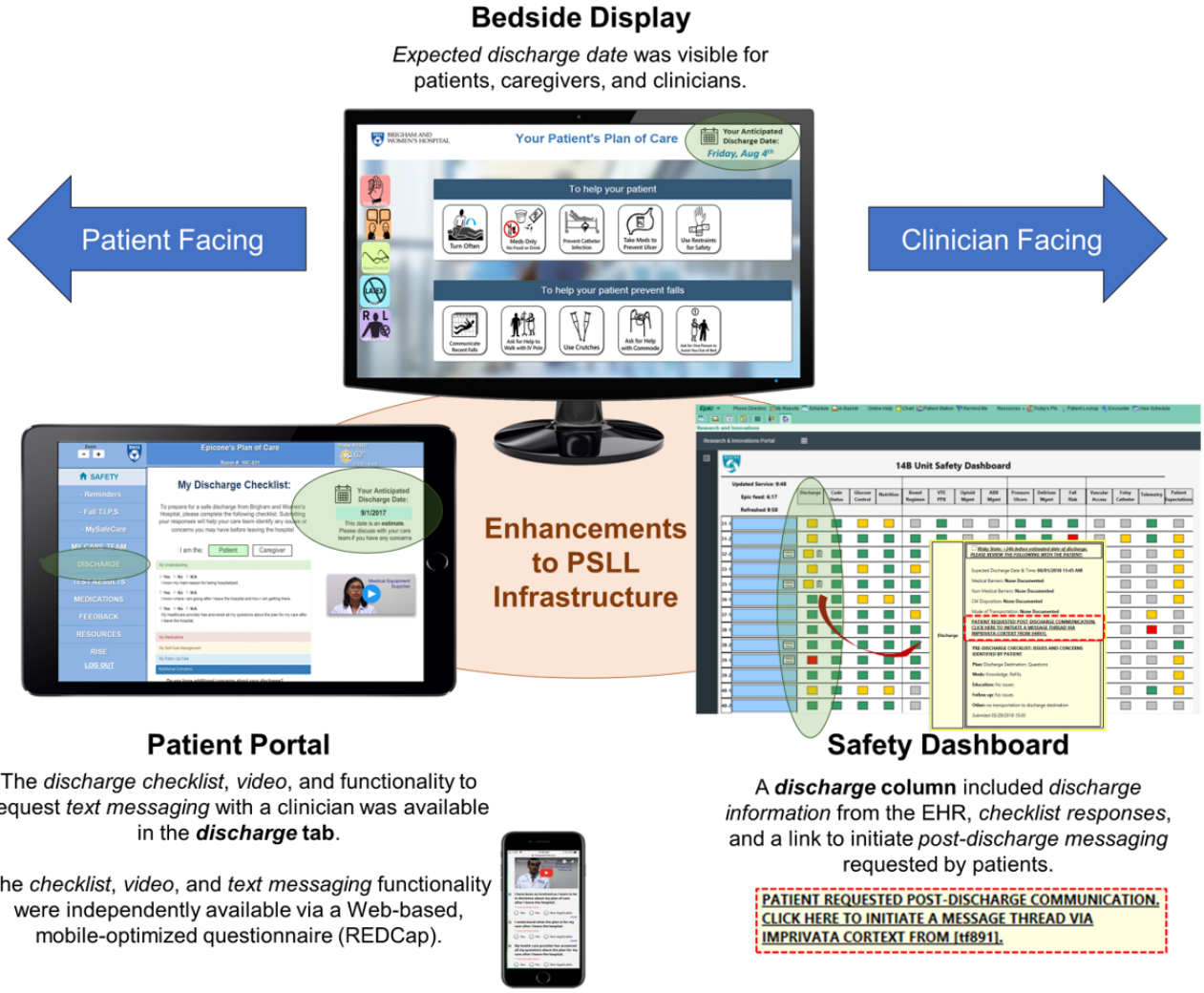
Dimension		Methods	Results
Reach	• How many patients participate and why do they choose to decline?	• Descriptive analysis of patients approached, and enrolled, including reasons for declining	• Main results • Table 3
	• What types of patients use the patient-facing PDTK[a] components?	• Descriptive analysis of patient characteristics and hospitalization metrics from administrative databases, and whether they did or did not submit a checklist, or watch the video	• Main results • Table 3
Effectiveness	• Does the PDTK activate patients at discharge?	• Interviews at discharge to assess proportion of patients with Patient Activation Measure scores >55 (level 3 or 4)	• Future study
	• Will the PDTK favorably impact health care resource utilization after discharge?	• Medical record review and phone interviews (30 days after discharge) to determine the proportion of patients with ≥1 unscheduled emergency department visit or readmission	• Future study
	• Can a checklist identify patients' discharge concerns?	• Descriptive analysis of patients' responses to checklist items	• Main results • Table 4
Adoption	• How many clinicians participate, and what types of clinicians use the clinician-facing PDTK components?	• Total number and percentage of clinicians of different types accessing the dashboard column and initiating postdischarge messaging	• Table 4
Implementation	• How frequently is each PDTK component utilized by patient and clinician participants?	• Percentage of approached patients watching the video, completing checklist, and requesting postdischarge messaging • Percentage of clinicians accessing dashboard column and initiating postdischarge messaging	• Table 4
	• Is the PDTK perceived to be valuable for patients and clinicians?	• Descriptive analysis of survey results administered to patient participants • Thematic analysis of content from semistructured interviews of patients and focus groups of clinicians	• Patient survey results • Table 5
Maintenance	• What barriers, unintended consequences, and workflow challenges are encountered? • What strategies are required to incorporate the PDTK into operations?	• Thematic analysis of content from semistructured interviews of patients and focus groups of clinicians	• Tables 5 and 6

[a]PDTK: patient-centered discharge toolkit.

Setting and Participants

The PDTK study (Clinicaltrials.gov NCT03116074) was approved by the Partners' institutional review board and was conducted at Brigham and Women's Hospital in Boston, MA. The study was conducted on three 30-bed general medicine units from December 2017 through July 2018 in parallel with our AHRQ-funded Patient Safety Learning Laboratory (PSLL; Clinicaltrials.gov NCT02969343) reported elsewhere [27-30]. As part of the PSLL, we integrated a *bedside display* for patients and clinicians, a *patient portal* for patients and caregivers, and a *safety dashboard* for clinicians into our EHR environment (Epic Systems, Inc, Verona, WI) [29,30]. These applications used enterprise data services to obtain clinical data from the EHR in real time [19,30-33]. This EHR-integrated digital health infrastructure (Figure 1) served as a platform on which to incorporate enhancements for the independently funded PDTK study. The key focus of the PSLL was to use systems engineering and human factors methods to design, develop, and implement tools to prevent harm *in the hospital* (eg, falls, catheter-associated urinary tract infections) [29]. The goal of the PDTK study was to improve safety *during transitions out of the hospital* by designing, developing, and implementing enhancements to the PSLL infrastructure (Table 2) based on clinician end-user requirements and organizational priorities.

Figure 1. Patient-centered discharge toolkit: Enhancements to the EHR-integrated digital health infrastructure. PSLL: Patient Safety Learning Laboratory; EHR: electronic health record.



The PDTK comprised enhancements to each of the 3 components of the PSLL technical infrastructure: bedside display, patient portal, and safety dashboard (see Multimedia Appendix 1).

Each of the 3 study units was codirected by a physician and nurse pair and staffed by its own group of nurses. Patients admitted to each unit were cared for by 1 of 2 geographically localized general medicine teams comprising residents or physician assistants and a supervising hospitalist attending physician [34]. These medical teams rotated approximately every 2 weeks. A few off-service patients (ie, admitted to a service other than general medicine) were admitted to these units under the care of a different attending physician.

Any English- and Spanish-speaking patient admitted to these units under the general medicine service was eligible to participate by using any of the patient-facing PDTK components (checklist, video, and secure messaging). For patients who did not have the capacity to consent (as determined by a member of the care team), a caregiver (a designated health care proxy) could participate on their behalf. Patient or caregiver participants were offered access to an acute care patient portal on either personal devices or study-issued mobile devices (iPad Air, Apple, Inc, Cupertino, CA) as part of the concurrent PSLL study [19]. Any clinician (nurse, resident, physician assistant, and attending) caring for a general medicine patient admitted to these units was eligible to participate: all clinicians had access to EHR-integrated digital health infrastructure as part of the concurrent PSLL study and could therefore access the clinician-facing PDTK components (safety dashboard and secure messaging).

Table 2. Description of core components of the patient-centered discharge toolkit. Patient-facing Patient Safety Learning Laboratory (PSLL) tools: patient portal and bedside display; clinician-facing PSLL tools: bedside display and safety dashboard.

Component	Description
EDD[a] display	• Current EDD from the EHR[b] was visible to patients on the patient portal and bedside display, and to clinicians on the bedside display and safety dashboard (Figure 1, green circles)
Discharge video	• Patients could choose to watch a Web-based video of a clinician talking through each checklist item at an appropriate health literacy level • Embedded via a hyperlink into the patient portal and REDCap (Research Electronic Data Capture, Nashville, TN) survey • Available in English (clinician) and Spanish (medical interpreter)
Discharge checklist	• A 16-item checklist that was available in English or Spanish could be completed by patient or caregiver via the patient portal or REDCap survey on a mobile device approximately 24 to 48 hours before EDD • Dichotomous responses were sent to EHR-integrated safety dashboard in real time via API[c]
Clinician dashboard discharge column	• Green flags identified patients with an EDD more than 1 day from the current date • Yellow flags identified patients with an EDD less than 1 day from or equal to the current date • Red flags identified patients with an EDD that was either not entered or past the current date; for patient portal enrollees, indicated that a checklist had not been completed when the current date was within 1 day of the EDD • Checklist icon identified patients who had completed checklist and were awaiting clinician review • *No* or *unsure* responses to checklist items were displayed by domain; free-text entries were displayed as additional patient-reported concerns; clinicians could address any unsatisfied item as needed (eg, unable to pay for medication and patient unaware of follow-up) • Displayed key data from the EHR (medical and nonmedical barriers to discharge, discharge destination, and transportation) • A link to initiate secure messaging was displayed for patients who requested postdischarge messaging
Secure messaging postdischarge	• A secure messaging thread was opened by a clinician (opt-in process) via a link in safety dashboard (Figure 1, red dashed box) • Patients were invited by their discharging clinician (attending, senior resident) to communicate up to 7 days on receiving an SMS text with a hyperlink to a mobile-optimized messaging portal • Clinicians messaged with the patient via a HIPAA[d]-compliant app (Imprivata Cortext) on their mobile phone (without giving the patient access to their mobile phone number)

[a]EDD: expected discharge date.

[b]EHR: electronic health record.

[c]API: application programming interface.

[d]HIPAA: Health Insurance Portability and Accountability Act.

Iterative Refinement of Intervention Components

In previous work, we engaged patient advisors, clinical stakeholders, information system professionals, and quality and safety leaders to identify gaps in discharge processes [19,35]. For this study, during the design and development phase, we conducted informal workflow observations on study units and interviews with stakeholders to identify end-user requirements for addressing these gaps by engaging patients and clinicians in discharge preparation while aligning with key organizational priorities: engaging patients to improve patient satisfaction, improving expected discharge date (EDD) documentation in the EHR, and reducing 30-day hospital readmissions [31]. For example, improving EDD accuracy—defined by our institution as the percent of final EDD entries equal to the actual discharge date (ADD)—was an organizational priority for improving operational throughput. Thus, to ensure timeliness of checklist completion and review of checklist responses by the care team, we enhanced the EHR-integrated *patient portal*, *bedside display*, and *safety dashboard* to improve the visibility of the EDD for both patients and clinicians (Figure 1). We presumed that the likelihood of checklist submission by patients and review by clinicians would be dependent on where patients were in their hospital course as well as their currently documented EDD.

As in the concurrent PSLL study [27,29,36], we applied user-centered design principles to refine patient- and clinician-facing intervention components (Figure 1) to ensure that we addressed end-user needs [31]. For the discharge checklist (Figure 2, left), our goal was to improve structure and organization, validate content, and clarify wording and utility. Key refinements were identified through multiple iterations of the original checklist within our research team (in part based on our experience with a transitions study funded by the Patient-Centered Outcomes Research Institute [37]), 2 sessions with our hospital's patient and family advisory council, and a short pilot in which we administered a paper-based prototype to a convenience sample of 10 hospitalized patients and requested feedback. On the basis of the feedback from unit nurses and patient advisors, we also created a video to help hospitalized patients understand the purpose of completing the checklist to prepare for discharge. To develop the discharge

video (Figure 2, right), we adapted a method previously demonstrated to improve patients' understanding of their medical condition and care plan [38]. Finally, we determined that patients would need to watch the video and complete the checklist via one of several workflows: using the patient portal on a hospital-issued mobile device, using their own mobile device, or having research staff coach patients or caregivers to complete the checklist and then submit responses on their behalf.

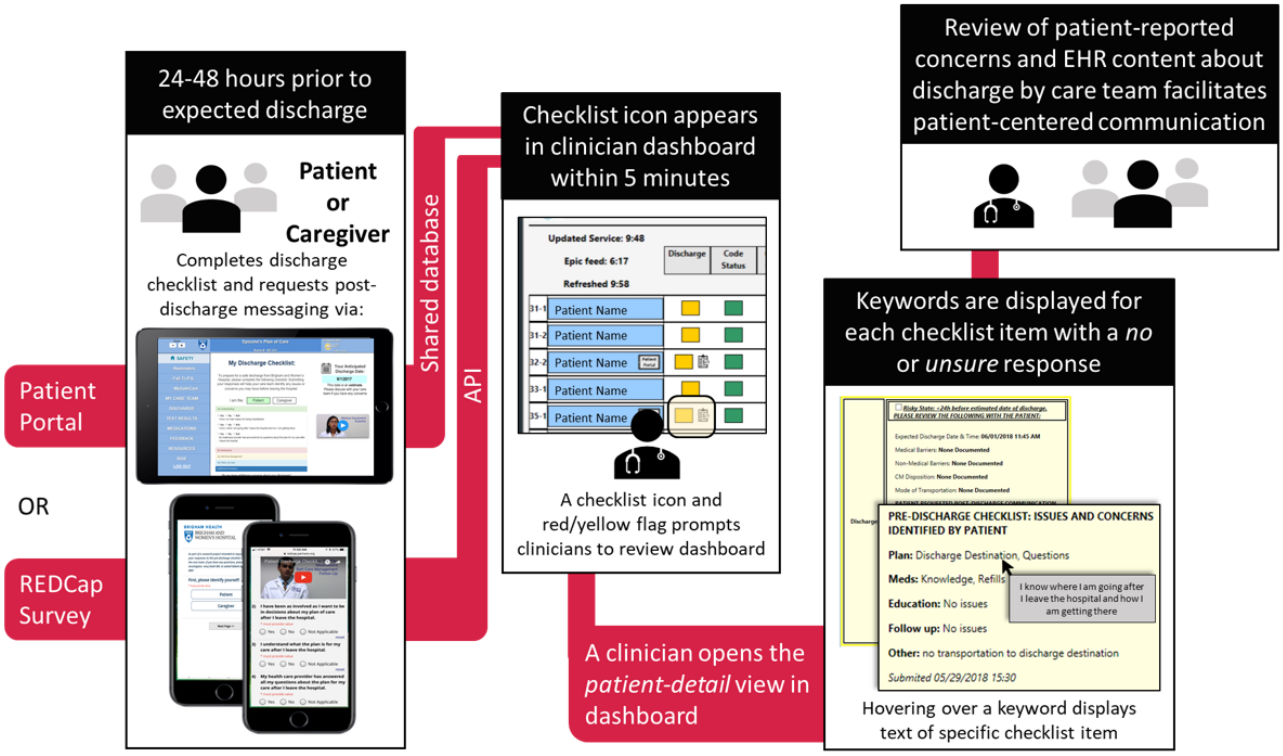
Figure 2. Discharge checklist and video.

Next, we made discharge process–based enhancements to the PSLL technical infrastructure integrated with our EHR to ensure that patient-reported information from the checklist would be communicated to the care team as the EDD approached. On the basis of the feedback from patients and clinical unit leadership, we confirmed that patients would want their care team to have access to the checklist responses sufficiently before actual discharge to allow time for the care team to review and address any issues (see *Enrollment* below). Thus, we developed a checklist submission and review process (Figure 3) to ensure that checklist responses submitted by patients would be visible for clinicians to review in the EHR in real time.

The discharge checklist (Figure 2, left), originally created by Coleman [24], was adapted for our institution [4], and further refined into a four-domain, 16-item discharge checklist. Questions were simplified, reordered, and separated into four domains: *My Understanding*, *My Medications*, *My Self-Care Management*, and *My Follow-up*. Dichotomous checklist responses (*Yes*/*No*) were determined to be too strict (ie, patients were uncertain of their response and did not want to check yes or no); thus, a third option (*Unsure*) and a box for a free-text response was added. A caregiver version to be completed by health care proxies, as well as a Spanish version (approved by our hospital's interpreter services) was created to make the checklist more inclusive of patients who lacked capacity or who did not speak English as their primary language (Multimedia Appendix 2).

The discharge video (Figure 2, right) was developed by creating an English version of a script in which a clinician guided the viewer through each checklist domain. This script was translated into Spanish and approved by our hospital's interpreter services. A mobile device (iPad Air, Apple, Inc) was used to film English and Spanish versions. Video editing software (iMovie, Apple, Inc) was used to produce the videos. The videos were uploaded onto a video-hosting site (YouTube, LLC); hyperlinks to these videos were incorporated into the patient portal and REDCap (Research Electronic Data Capture, Nashville, TN; see Checklist Submission section).

Figure 3. Checklist submission and review. EHR: electronic health record.



Checklist Submission

The two options available for patients to complete and submit the checklist were as follows: the patient portal or a Web-based REDCap survey [19,39]. REDCap submissions could be completed by patients on a mobile device (via a hyperlink emailed to the patient), or by study staff who would submit the responses on their behalf. Checklist responses submitted via the patient portal were made visible in the safety dashboard via a shared database. Checklist responses submitted via REDCap were routed to the safety dashboard using the REDCap application programming interface and matched to the corresponding patient using key identifiers (medical record number and admission date). In either case, checklist responses were displayed on the safety dashboard for clinicians within 5 min of submission.

Checklist Review

A new discharge column in the safety dashboard displayed key data elements (medical and nonmedical barriers, EDD) from the EHR (Epic Systems, Inc, Verona, WI) using enterprise Web services [27]. EHR data were transformed into clinical decision support using color-coded *flags*: red=*action needed*; yellow=*risky state*; green=*guideline compliant*; gray=*not applicable*. For example, a red flag appeared if the current date was past the EDD, no EDD was entered, or the current date was within 24 hours of the EDD, but no checklist had been submitted. A checklist icon appeared on the safety dashboard on successful submission. Checklist items with a *no* or *unsure* response were briefly summarized as a keyword on the *patient-detail* view of the safety dashboard (eg, *Meds: Access/Adherence*) with a hover-over displaying the specific item answered by the patient (eg, *I understand how to get my medications and will take them as prescribed after I leave the hospital*). Patients were permitted to enter their mobile phone number during checklist submission to request secure messaging (Imprivata, Lexington, MA) with a hospital physician (attending or senior resident) for up to 7 days after discharge [19,27,40]. A link appeared in the safety dashboard for the clinician to initiate a message thread with the patient.

Patient and Caregiver Enrollment

The intervention went live in December 2017, starting with a 1-month wash-in period in which we debugged various technical components. During business days, recruitment lists were created in which the care unit and patient approach order were randomized to minimize confounding (eg, always starting enrollment on the top floor of the hospital and ending at the bottom). All patients admitted to intervention units for at least 24 hours with EDDs within 24-48 hours were eligible. Research assistants asked the nurses caring for these patients if they were appropriate to approach, temporarily excluding any patient who was nonverbal, incapacitated, or behaviorally not safe. When an eligible patient was confirmed as not capable of participating, the research assistant attempted to identify a caregiver (a designated health care proxy). Patients and caregivers who were not available for recruitment (eg, off unit getting a test) were reapproached later that day or on subsequent days if they still met eligibility criteria. Research assistants then asked eligible patients (or caregivers) to watch the video and complete the checklist (including entering their mobile phone number to request postdischarge messaging) via the patient portal or REDCap workflow. A patient was considered enrolled on successful submission of the checklist.

Measurements and Data Collection

We used mixed methods to evaluate our implementation experience per each RE-AIM dimension (Table 1). Specifically,

we used quantitative methods to measure usage and perceived utility of intervention components (Reach, Adoption, and Implementation). We used qualitative methods to assess barriers to and facilitators of Implementation and Maintenance from the patient and clinician perspective.

Usage of Intervention Components (Quantitative)

We captured the number of times the video was watched, the checklist was submitted, and postdischarge messaging was requested by patients. We captured the number and type of patient-reported concerns for each checklist submitted, as well as the number of times clinicians accessed the discharge column on the safety dashboard to view patient-reported concerns and click on the link to initiate postdischarge messaging.

Patient Surveys (Quantitative)

We previously reported *usability* of the patient portal, which included a discharge module with an earlier, noninteractive version of the checklist [19]. To better understand *perceived utility* of the PDTK, we asked a convenience sample of enrolled patients to participate in a survey guiding them through each intervention component. The 2-part survey, based on prior work [32], asked participants to rate their perceived readiness for discharge and willingness to self-assess discharge preparedness on a 5-point Likert scale. After watching the video, completing the checklist, and viewing how clinicians could visualize checklist responses in the safety dashboard, participants rated statements about their confidence that the intervention would facilitate self-identification and communication of discharge concerns to their care team.

Patient Interviews (Qualitative)

We verbally consented and conducted semistructured interviews with a convenience sample of English-speaking patients within 24-48 hours of anticipated discharge. Care team members (nurse, physician assistant, and resident or attending) were asked to identify patients within this time frame who were present in their rooms and would be amenable to participating in a brief interview. To reduce sampling bias, potential participants were selected to create a diverse sample based on age, gender, and reason for hospitalization. To minimize selection bias, patients were assured that their participation would not influence their care team's medical decisions, and care teams were not told which patients agreed to participate. All participants completed the checklist and watched the video either before or at the time of the interview. Study staff (DP and NP) trained in qualitative research methods conducted the interviews using a semistructured interview guide that explored (1) patient experiences completing the checklist and viewing the video, (2) addressal of discharge concerns by the care team, and (3) pros and cons of using these tools. Interviews were digitally recorded, transcribed, and reviewed for accuracy, and conducted until thematic saturation was achieved [41].

Clinician Focus Groups (Qualitative)

After the study was completed, we conducted focus groups with physicians, physician assistants, and nurses to assess implementation barriers and facilitators until thematic saturation. Using a structured guide, we asked about EDD entry via the EHR, perception of the checklist workflow, usage of the discharge column of the safety dashboard to review checklist responses and initiate postdischarge messaging, and awareness of patient-facing components (bedside display and patient portal). Focus group discussions were digitally recorded, transcribed, and reviewed for accuracy.

Statistical and Qualitative Analyses

Descriptive statistics were used to report patient demographic and administrative data, quantify patient-reported concerns, calculate the frequency of tool use by patients and clinicians, and quantify survey data. We calculated the proportion of patients (or caregivers) completing the checklist via the patient portal or REDCap. All qualitative data collected from patient and clinician participants were transcribed, openly coded, and analyzed using the constant comparative method [42]. Two researchers (DP and NP) independently coded all transcripts line by line using Word (Microsoft, Inc, Redmond, WA). The code structure was revised as needed to capture novel concepts, adapt, and merge existing concepts; transcripts were coded with each iteration of the codebook, with any discrepancies resolved during consensus meetings. This process was repeated until no novel concepts were identified, at which point, the 2 researchers again independently applied the final code structure to all transcripts. Key themes were identified in a final group consensus meeting with the study staff. Key implementation barriers were identified by study staff via a group consensus approach based on quantitative and qualitative data.

Results

Of 752 patient-admissions, the patient (or caregiver) watched the video and completed the checklist in 416 (55.3%), and the patient/caregiver completed the checklist alone in 94 (12.5%). Research assistants made 313 attempts at approaching patients in the remaining 242 patient-admissions; however, the patient was unavailable (126), not appropriate per nurse (97), declined to participate (41), did not speak English or Spanish and no caregiver was available (33), did not respond by email when reminded (8), or encountered technical issues (8). The demographic characteristics of the 67.8% (510/752) patient-admissions (480 unique patients) in which a checklist was submitted and the 32.2% (242/752) patient-admissions (238 unique patients) in which the checklist was not submitted are reported in Table 3. In general, those who did not submit a checklist were older, more often Hispanic and non-English speaking, less often privately insured, had higher diagnosis-related group (DRG) weights and longer lengths of stay, and were typically discharged to a destination other than home.

Table 3. Demographics of patient admissions (N=752).

Characteristics	Submitted checklist (n=510)	Did not submit checklist (n=242)	*P* value
Unique patients	**480**	**238**	**—[a]**
1 hospitalization	453	234	
2 or more hospitalizations	27	4	
Age (years), mean (SD)	**58.6 (17.9)**	**62.4 (18.6)**	**.008[b]**
Missing	—	14	—
Gender, n (%)			**.94[c]**
Female	280 (54.9)	126 (52.1)	
Missing	—	14 (5.8)	
Race, n (%)			**.13[c]**
White	340 (66.7)	140 (57.9)	
Nonwhite	162 (31.7)	86 (35.5)	
Missing	8 (1.6)	16 (6.6)	
Ethnicity, n (%)			**<.001**
Non-Hispanic	468 (91.8)	189 (78.1)	
Hispanic	33 (6.5)	36 (14.9)	
Unavailable	9 (1.7)	3 (1.2)	
Missing	—	14 (5.8)	
Primary language, n (%)			**<.001[c]**
English	492 (96.5)	187 (77.3)	
Non-English	11 (2.2)	40 (16.5)	
Missing	7 (1.4)	15 (6.2)	
Median income by ZIP code			**.46[c]**
≤US $47,000	96 (18.8)	49 (20.3)	
US $47,001 to US $63,000	124 (24.3)	48 (19.8)	
Greater than US $63,000	272 (53.3)	131 (54.1)	
Missing	18 (3.5)	14 (5.8)	
Insurance status, n (%)			**<.001[c]**
Private	195 (38.2)	70 (28.9)	
Public (Medicaid, Medicare)	305 (59.8)	134 (55.4)	
Other[c]	10 (2.0)	18 (7.4)	
Missing	—	20 (8.3)	
Primary care physician, n (%)			**.01[c]**
In-network	229 (44.9)	125 (51.7)	
Nonnetwork	280 (54.9)	103 (42.5)	
Missing	1 (0.2)	14 (5.8)	
Elix number of comorbidities, mean (SD)	**4.22 (2.41)**	**4.53 (2.45)**	**.15[b]**
Missing	—	14	
Elix index, comorbidities, n (%)			**.03[c]**
Less or = 0	104 (20.4)	33 (13.6)	
1 to 5	96 (18.8)	29 (12.0)	

Characteristics	Submitted checklist (n=510)	Did not submit checklist (n=242)	*P* value
6 to 10	94 (18.4)	26 (10.7)	
11 or more	216 (42.4)	106 (43.8)	
Missing	—	48 (19.8)	
DRG[d] weight, mean (SD)	**1.83 (1.97)**	**2.26 (2.10)**	**<.001[b]**
Missing	10	19	
Length of stay, mean (SD)	**8.78 (7.93)**	**11.5 (13.7)**	**.02[b]**
Missing	—	14	
Discharge destination, n (%)			**.003[c]**
Home	410 (80.4)	133 (55.0)	
Facility	92 (18.0)	58 (24.0)	
Other	7 (1.4)	2 (0.8)	
Missing	1 (0.2)	49 (20.3)	
Readmissions within 30 days, n (%)			**.18[c]**
Yes	88 (17.3)	39 (16.1)	
No	422 (82.7)	140 (57.9)	
Missing	—	63 (26.0)	

[a]Not applicable.

[b]*P* value calculated by Wilcoxon test.

[c]*P* value calculated via chi-square test.

[d]Nonstandard insurance or self-insured.

[e]DRG: diagnosis-related group.

Usage of each PDTK component for the 510 patient-admissions in which a checklist was submitted is reported in Table 4. Although the patient was enrolled in the acute care patient portal in 173 (33.9%) of the 510 patient-admissions, the patient portal was used to submit the checklist in 53 (10.4%); the remainder were submitted via REDCap. The checklist was submitted once and 2 or more times in 492 and 18 patient-admissions, respectively. The median ($IQR_{25,75}$) days from initial checklist submission to EDD and ADD were 1 (1,2) and 2 (1,5), respectively. On average, 4.24 concerns were reported for each checklist submitted: the most commonly entered concerns by patients were about medications (664/2164, 30.68%) and follow-up (656/2164, 30.31%). The EDD was accurate in 307 (60.2%) of the 510 patient-admissions.

Of the 20 patient experience survey participants, 13 (65%) felt well prepared for discharge, and 16 (80%) stated that they would be willing to self-assess discharge preparedness via a checklist. After viewing how the PDTK components functioned, all (100%) completed the checklist, reporting an average of 5.1 concerns (17 understanding of the plan, 30 medications, 21 self-care management, and 34 follow-up); and 7 (35%) requested secure messaging after discharge; 13 (65%) felt that the checklist facilitated self-identification of potential issues before discharge; 15 (75%) believed that their care team would become aware of these issues via the safety dashboard; 10 (50%) felt more confident about what to do to prevent issues after leaving, and 15 (75%) felt confident that they could quickly communicate with a hospital physician via secure messaging should an issue arise postdischarge.

Of the 20 patients approached for semistructured interviews, 12 participated: 7 (58%) were male, 10 (83%) were white, the median age was 70.5 years, and 8 had public insurance. The most common reason for declining to participate was feeling unwell or tired. We identified two overarching themes about discharge preparation: (1) gaps in communication between patients and their care team resulting in patients feeling inadequately informed about their discharge care plan (eg, *I wasn't informed and kept up to date with what was happening and the reason why. I understood on my own basically what was needed because I've gone through this before, but if it was the first time, I think I would have been very confused*); and (2) despite perceived communication gaps, patients were confident that their care team would address all of their questions and concerns before discharge (eg, *I knew everything was going to be done and things were going to be taken care of, but really, I didn't feel informed, I really didn't*). We also identified key themes regarding patient experiences using the checklist and video components of the PDTK (Table 5).

In total, 22 clinicians (8 physicians, 6 physician assistants, 8 nurses; mean age 36.9 years; 14 (14/22, 64%) female) participated in 1 of 3 focus groups from which we identified 3 major themes regarding the safety dashboard component of the PDTK (Table 5): low awareness and variable workflow, lack of optimization, and inconsistent leadership.

Table 4. Usage of patient-centered discharge toolkit component during 510 patient-admissions (480 unique patients).

Metric	Statistic	Comment
Discharge video watches, n (%)	416 (81.6)	Watched before checklist completion, most often the English version
Discharge checklist version submitted, n (%)		
Patient	497 (97.5)	—[a]
Caregiver	13 (2.5)	Consented if patient preferred or did not have capacity
Electronic workflow used to submit checklist, n (%)		
Web-based REDCap survey	457 (89.6)	Submitted via a mobile device
Patient portal (discharge module)	53 (10.4)	Could submit the checklist via the portal or REDCap
Total number of concerns reported [b]by domain	**164**	**Most frequent items checked *no* or *unsure* by domain**
Understanding the plan, n (%)	355 (16.4)	Understanding the main reason for hospitalization
Medications, n (%)	664 (30.7)	Understanding changes to the medication regimen and how to get and take medications
Self-care, n (%)	437 (20.2)	Understanding *red flag* signs and symptoms
Follow-up, n (%)	656 (30.3)	Time and date of appointments, how to get to them
Other, n (%)	52 (2.4)	Unaddressed clinical concerns, nonmedical barriers
Safety dashboard discharge column		
Viewed by clinical staff during patient-admission, n (%)	210 (41.2)	Accessed safety dashboard's patient-detail view or clicked acknowledgment check-box
Total number of times accessed, n	631	Median (IQR 25,75): 2 (1,4) per patient-admission
RN, n (%)	399 (63.2)	Unit-based bedside nurses
MD, n (%)	180 (28.5)	Attending or resident
Administrative, n (%)	44 (7.0)	Unit clerk
Physician assistant, n (%)	8 (1.3)	Worked on separate nonresident service with attendings
Secure postdischarge messaging (n=422[c])		
Requested by patient, n (%)	141 (33.4)	Patient must have had mobile phone with a mobile web-browser
Initiated by physician, n (%)	3 (2.1)	2 attendings, 1 senior resident

[a]Not applicable.

[b]A discharge checklist item for which the response was *no* or *unsure* was considered a patient-reported concern.

[c]Denominator reflects number of patient admissions in which postdischarge messaging was available.

Table 5. Key themes from patient interviews and clinician focus groups about patient-centered discharge toolkit components.

Theme	Description	Quote
Checklist and video		
Valuable for patients	• The checklist and video increased understanding of self-care needs and follow-up plans and promoted patient engagement and empowerment in the discharge process.	• "I may think of questions I didn't really have. Definitely worth it. It actually makes you think." [Patient] "[The checklist] made the patient feel like a more active participant [in] their care…" [Clinician]
Patient utility dependent on the timing of administration	• The checklist and video were most useful when administered close to discharge but before a detailed discussion of discharge preparation by a care team member.	• "Well, it was a little unclear given that we're not about to leave. It's hard to report on the process because it hasn't actually happened yet." [Patient]
Safety dashboard		
Low awareness, variable workflow	• Although clinicians were generally aware, checklist answers were variably viewed on the safety dashboard. • Reinforcement and reminders to use the safety dashboard to review patient-reported discharge concerns were variable. • The workflow for entering and updating EDD[a] was inconsistent and included both clinical and nonclinical staff.	• "[Discharge checklist responses] on the dashboard?... Did not know that." [Clinician] • "When it first rolled out there was a lot of information about it and then it just dropped off, and then the usage dropped off…" [Clinician] • "[EDD] not really my workflow…I mean we'll put in [the EDD], and it'll get changed by a unit coordinator on a different pod." [Clinician]
Lack of optimization	• Discharge column flag logic was often misinterpreted by different clinicians. • Summarized checklist responses displayed in safety dashboard were too broad and nonspecific. Clinicians could not quickly access the entire checklist.	• "The senior resident did not know really, what green [dashboard flags] meant...are [the patients] ready to be discharged?" [Clinician] • "I would look at [the safety dashboard] sometimes and wonder what [the patient] clicked off [on the checklist], but sometimes I couldn't tell exactly what they had questions about." [Clinician]
Inconsistent leadership	• Usage was dependent on senior-level clinician leadership (attending or senior resident).	• "…when the attendings were into it we were all into it for that week." [Clinician]

[a]EDD: expected discharge date.

Discussion

Principal Findings

We used the RE-AIM framework to evaluate the feasibility and acceptability of a suite of EHR-integrated digital health tools to engage patients, caregivers, and clinicians in discharge preparation. Most patients agreed to watch the discharge video and complete the checklist to self-assess discharge preparedness when coached, and we did not encounter significant technical difficulty in our approach. The patient-facing tools were perceived to be valuable by both patients and clinicians, and most patient-reported concerns submitted via the checklist related to medications and follow-up. Clinician use of the safety dashboard discharge column to view these concerns was modest, mostly due to workflow challenges. A large percentage of patients requested postdischarge messaging, but very few clinicians opted in. Themes identified from our qualitative analysis suggest that timing of administration, additional workflow integration, optimization, and leadership are necessary to promote a more robust adoption of these tools.

We attribute the high rate of patient participation to flexible Web-based workflows and facilitation by research assistants. First, the iterative process to develop and refine the checklist and video incorporated feedback from patient advisers and institutional stakeholders, resulting in a product that was relevant and understandable to patients. Next, research assistants functioned as discharge advocates for study participants, guiding them through the process of viewing the video, and completing the checklist. This encouraged participation from patients who might otherwise have not been comfortable or motivated to do so independently. In addition, we included a video component and incorporated a checklist submission process that did not depend on the acute care patient portal: REDCap's Web-based workflow was useful as a mobile app prototype and also circumvented key barriers to patient portal enrollment and use during hospitalization, such as patients' dislike of having a separate log-in for the acute care patient portal [19,36]. In contrast, the direct hyperlink to the discharge checklist in REDCap offered more streamlined access (eg, no log-in was needed) for patients, and facilitation by research staff mitigated the perceived burden of submitting the checklist electronically on their own.

Although most patient participants perceived that the intervention would facilitate communication regarding discharge concerns to their care team, often this did not occur because of

low uptake of the intervention by clinicians; we attribute this to low awareness of the intervention, inconsistent understanding of its purpose and how to use it, and lack of specificity of patient-reported concerns viewable on the safety dashboard. Specifically, although the safety dashboard discharge column logic was vetted by institutional stakeholders to align with hospital priorities (improving EDD accuracy), clinicians often misinterpreted safety dashboard flag colors (eg, green did not signify *safe for discharge*) and had a different understanding of responsibility for updating the EDD in the EHR because of inconsistent processes. In addition, if clinicians did not access the detailed view for the patient or click on the flag, they did not see the full text of the flag and what it meant, only its color [43]. Finally, variable use of the discharge column by physicians likely led to poor awareness of the link to initiate a secure messaging thread when requested by patients, and many were resistant to using this feature altogether.

As one of the first reported attempts at engaging patients, caregivers, and clinicians in discharge preparation using a suite of EHR-integrated digital health tools, findings from our implementation study, guided by the RE-AIM framework, offer several instructive lessons (Table 6). First, the optimal timing of when to complete the checklist is paramount. If completed too early, patients perceive less utility, anticipating that the care team would eventually address their concerns. If completed too late, concerns identified from the patients' perspective are less likely to be communicated to clinicians, leading to potential deficiencies or delays in addressing them. Second, dashboard flag changes were not linked to relevant EHR data elements because of competing workflows (eg, a newly created process for documenting and escalating discharge barriers to hospital leadership) and technical limitations (eg, lack of a Web service to retrieve readmission risk scores from the EHR). Rather than linking flag changes to EDD documentation and timing, using the safety dashboard to review patient-reported concerns about discharge might be more clinically meaningful in the context of patient-specific readmission risk scores; this would be more consistent with the overall intent of the safety dashboard as a tool for proactively identifying patients at risk for harm [27,44], Third, improving EDD accuracy via this type of intervention is more likely to be achieved if the responsibility for updating EDD resides with clinicians rather than unit clerks and is clear to all parties. Patients and clinicians will then have more confidence in the EDD displayed on the patient portal, bedside display, and safety dashboard. Still, it is noteworthy that EDD accuracy for enrolled patients (307/510, 60.2%) was marginally higher than the EDD accuracy rate for general medicine (965/1702, 56.70%) as a whole. Finally, secure messaging after discharge clearly requires a better understanding of factors predicting whether clinicians will use this feature and how to incentivize its use.

Table 6. Implementation barriers and strategies to promote adoption.

Implementation barriers	Strategies to promote adoption
Discharge video	
Timing and access of video after admission to the unit	• Make videos available via the patient portal, bedside display, and television • Engage nurses to have patients watch videos as EDD[a] approaches
Too generic and impersonal	• Have clinical unit leaders create unit-specific videos • Create videos for each attending, play video for patient's current attending by linking to the treatment team in the EHR[b] • Translate videos into common languages (eg, Spanish) using medical interpreters
Discharge checklist	
Timing and administration	• Determine optimal timing of checklist administration for specific patient categories (eg, admissions for acute on chronic disease exacerbations, awaiting procedures, undifferentiated diagnoses) • Demonstrate impact on key hospital priorities and process metrics (EDD accuracy, early hospital discharges)
Patients' belief that clinicians will address all items	• Encourage patients to review and update the checklist during their hospitalization • Allow patients to update checklist responses as EDD approaches or changes
Checklist responses out-of-date owing to discharge delays	• Identify workflow to update checklist after initial submission (eg, notification via the patient portal, email, or mobile app)
Dashboard discharge column	
Variable EHR data entry of key data elements (EDD, medical, nonmedical barriers)	• Demonstrate how EDD can be viewed by patients (patient portal, bedside display) and clinicians (bedside display, dashboard) • Add a confidence indicator that estimates the likelihood that EDD will equal ADD[c] to manage patient and clinician expectations • Demonstrate the value of structured EHR data entry for driving dashboard logic (flagging red when EDD not entered) • Encourage checklist completion for patients at high risk for readmission by incorporating patient-specific readmission risk scores from EHR into logic • Display barriers to discharge on the dashboard
Competing QI[d] interventions	• Understand current institutional priorities and emerging workflows for identifying and escalating discharge barriers • Propose enhancements based on lessons learned from concurrent QI efforts to explain how the use of a checklist can prepare patients for postdischarge care (increasing patient satisfaction, reducing readmission rates) while maintaining or reducing the length of stay (by proactively identifying and overcoming barriers to timely discharge)
Poor specificity of patient-reported concerns viewed in the dashboard	• Provide a link to discharge checklist questions and patient's responses • Link patient-reported concerns to specific clinical actions (eg, if poor understanding of the main diagnosis, update after visit summary with condition-specific educational materials)
Secure postdischarge messaging	
Physician resistance	• Frame the initiation of secure messaging thread as an opt-in process • Align with value-based incentives for clinical services (readmissions) • Communicate success stories from early adopters to assuage fears (eg, excessive text messages from patients)
Managing patient expectations about whether physicians will initiate secure messaging	• Educate patients about the opt-in process for attendings • Encourage patients to request attendings to use this feature for clearly defined reasons (eg, concern about obtaining a key medication)

[a]EDD: expected discharge date.

[b]EHR: electronic health record.

[c]ADD: actual discharge date.

[d]QI: quality improvement

Limitations

Our study has several limitations. First, it was conducted for general medicine patients at a single institution without a control group; therefore, we could not evaluate the impact of our intervention on clinical outcomes. Similarly, our qualitative analyses were performed on convenience samples of participants, which may limit the generalizability of our findings. Second, we used research assistants to coach patients in submitting the checklist; although dedicated discharge advocates are becoming increasingly common, many institutions lack sufficient personnel [45]. In most hospitals, nursing staff could serve patients in this capacity as they are often the first to identify concerns reported by patients preparing for discharge. Third, we identified disparities among those who submitted and those who did not submit a checklist. Clearly, additional work is needed to address disparities in underrepresented groups to fully evaluate the utility of this intervention in a broader population. Finally, this was a hospital-centric intervention—we did not engage primary care physicians. Although our efforts at postdischarge messaging attempted to bridge the transition from inpatient to ambulatory care, the secure messaging vendor used in this study did not offer the ability to communicate with multiple care team members simultaneously, as we previously described [40]. Nonetheless, seamless communication with key ambulatory clinicians is important during the immediate postdischarge period [46,47].

Conclusions

We believe that EHR-integrated digital health tools such as those we described will become increasingly useful as part of an institutional strategy to engage patients, caregivers, and clinicians in improving discharge safety if they simultaneously address key hospital priorities (eg, improving EDD accuracy and mitigating readmission risk). Currently, we are making further enhancements to the intervention components and their implementation. For example, we are reconfiguring the dashboard discharge column logic to more clearly identify patients at high risk for readmission, which should provide context for the types of concerns patients report after completing the checklist. Exploratory features, such as secure messaging with patients, clearly require further investigation to better characterize patient and clinician perceptions of its value and appropriate use after discharge. However, we believe that many of these features will become increasingly utilized to comply with new regulations (eg, Caregiver Advise, Record, Enable [CARE] Act) [48]. Finally, we plan to conduct rigorously designed studies to evaluate the impact of the PDTK on key outcomes during transitions, such as patient activation at discharge, postdischarge health care resource utilization, and hospital readmissions.

Acknowledgments

This work was supported by a grant from AHRQ (R21-HS024751). AHRQ had no role in the design or conduct of the study; collection, analysis, or interpretation of data; or preparation or review of the manuscript. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of AHRQ.

The authors thank the PSLL research team: Patricia C Dykes, Sarah A Collins, Kumiko O Schnock, Ronen Rozenblum, Lisa S Lehmann, Stuart R Lipsitz, Alexandra Businger, Pamela Garabedian, Eli Mlaver, Brittany Couture, Megan Duckworth, Jenzel Espares, and Kerrin Bersani. The PSLL study was funded by the AHRQ (R21-HS024751).

Authors' Contributions

All authors have contributed sufficiently and meaningfully to the conception, design, and conduct of the study; data acquisition, analysis, and interpretation; or drafting, editing, and revising of the manuscript.

Conflicts of Interest

JS was the recipient of an investigator-initiated grant from Mallinckrodt Pharmaceuticals to study opioid-related adverse drug events and an investigator-initiated grant from Portola Pharmaceuticals to study patients who decline inpatient venous thromboembolism prophylaxis. The remaining authors have no competing interests to declare.

Multimedia Appendix 1
All patient safety learning laboratory enhancements. Enlarged images for viewing the details of the technological enhancements.
[PPTX File , 13848 KB - jmir_v22i4e15573_app1.pptx]

Multimedia Appendix 2
All checklist versions. Three versions of the discharge preparedness checklist were created (English speaking patients and caregivers, as well as Spanish speaking patients), listed in this appendix in that order.
[DOCX File , 440 KB - jmir_v22i4e15573_app2.docx]

References

1. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med 2003 Feb 04;138(3):161-167. [doi: 10.7326/0003-4819-138-3-200302040-00007]
2. Forster AJ, Clark HD, Menard A, Dupuis N, Chernish R, Chandok N, et al. Adverse events among medical patients after discharge from hospital. CMAJ 2004 Feb 03;170(3):345-349 [FREE Full text] [Medline: 14757670]
3. Moore C, Wisnivesky J, Williams S, McGinn T. Medical errors related to discontinuity of care from an inpatient to an outpatient setting. J Gen Intern Med 2003 Aug;18(8):646-651. [doi: 10.1046/j.1525-1497.2003.20722.x] [Medline: 12911647]
4. Tsilimingras D, Schnipper J, Duke A, Agens J, Quintero S, Bellamy G, et al. Post-discharge adverse events among urban and rural patients of an urban community hospital: a prospective cohort study. J Gen Intern Med 2015 Aug;30(8):1164-1171. [doi: 10.1007/s11606-015-3260-3]
5. Kanaan AO, Donovan JL, Duchin NP, Field TS, Tjia J, Cutrona SL, et al. Adverse drug events after hospital discharge in older adults: types, severity, and involvement of Beers Criteria medications. J Am Geriatr Soc 2013 Nov;61(11):1894-1899 [FREE Full text] [doi: 10.1111/jgs.12504] [Medline: 24116689]
6. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. Adverse drug events occurring following hospital discharge. J Gen Intern Med 2005 Apr;20(4):317-323. [doi: 10.1111/j.1525-1497.2005.30390.x] [Medline: 15857487]
7. Rojas-García A, Turner S, Pizzo E, Hudson E, Thomas J, Raine R. Impact and experiences of delayed discharge: a mixed-studies systematic review. Health Expect 2017 Feb;21(1):41-56. [doi: 10.1111/hex.12619] [Medline: 28898930]
8. Arora VM, Prochaska ML, Farnan JM, D'Arcy MJ, Schwanz KJ, Vinci LM, et al. Problems after discharge and understanding of communication with their primary care physicians among hospitalized seniors: a mixed methods study. J Hosp Med 2010 Sep;5(7):385-391 [FREE Full text] [doi: 10.1002/jhm.668] [Medline: 20578045]
9. Howard-Anderson J, Busuttil A, Lonowski S, Vangala S, Afsar-Manesh N. From discharge to readmission: understanding the process from the patient perspective. J Hosp Med 2016 Feb 19;11(6):407-412. [doi: 10.1002/jhm.2560] [Medline: 26895238]
10. Centers for Medicare & Medicaid Services, Health & Human Services. Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and FY 2012 rates; hospitals' FTE resident caps for graduate medical education payment. Final rules. Fed Regist 2011 Aug 18;76(160):51476-51846 [FREE Full text]
11. Howard-Anderson J, Lonowski S, Vangala S, Tseng C, Busuttil A, Afsar-Manesh N. Readmissions in the era of patient engagement. JAMA Intern Med 2014 Nov;174(11):1870-1872. [doi: 10.1001/jamainternmed.2014.4782] [Medline: 25264930]
12. Graham KL, Auerbach AD, Schnipper JL, Flanders SA, Kim CS, Robinson EJ, et al. Preventability of early versus late hospital readmissions in a national cohort of general medicine patients. Ann Intern Med 2018 Jun 05;168(11):766-774 [FREE Full text] [doi: 10.7326/M17-1724] [Medline: 29710243]
13. Soong C, Daub S, Lee J, Majewski C, Musing E, Nord P, et al. Development of a checklist of safe discharge practices for hospital patients. J Hosp Med 2013 Mar;8(8):444-449. [doi: 10.1002/jhm.2032]
14. Halasyamani L, Kripalani S, Coleman E, Schnipper J, van Walraven C, Nagamine J, et al. Transition of care for hospitalized elderly patients—development of a discharge checklist for hospitalists. J Hosp Med 2006 Nov;1(6):354-360. [doi: 10.1002/jhm.129]
15. Okrainec K, Lau D, Abrams HB, Hahn-Goldberg S, Brahmbhatt R, Huynh T, et al. Impact of patient-centered discharge tools: a systematic review. J Hosp Med 2017 Feb;12(2):110-117. [doi: 10.12788/jhm.2692] [Medline: 28182809]
16. Centers for Medicare & Medicaid Services. 2019 Mar. Your Discharge Planning Checklist URL: https://www.medicare.gov/pubs/pdf/11376-discharge-planning-checklist.pdf [accessed 2019-12-01]
17. Agency for Healthcare Research and Quality. 2017. Strategy 4: Care Transitions From Hospital to Home: IDEAL Discharge Planning URL: https://www.ahrq.gov/professionals/systems/hospital/engagingfamilies/strategy4/index.html [accessed 2019-12-01]
18. Greysen SR, Khanna RR, Jacolbia R, Lee HM, Auerbach AD. Tablet computers for hospitalized patients: a pilot study to improve inpatient engagement. J Hosp Med 2014 Jun;9(6):396-399. [doi: 10.1002/jhm.2169] [Medline: 24523051]
19. Dalal AK, Dykes PC, Collins S, Lehmann LS, Ohashi K, Rozenblum R, et al. A Web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation. J Am Med Inform Assoc 2016 Jan;23(1):80-87. [doi: 10.1093/jamia/ocv093] [Medline: 26239859]
20. Greysen S, Harrison JD, Rareshide C, Magan Y, Seghal N, Rosenthal J, et al. A randomized controlled trial to improve engagement of hospitalized patients with their patient portals. J Am Med Inform Assoc 2018 Dec 01;25(12):1626-1633 [FREE Full text] [doi: 10.1093/jamia/ocy125] [Medline: 30346543]
21. Collins S, Dykes P, Bates DW, Couture B, Rozenblum R, Prey J, et al. An informatics research agenda to support patient and family empowerment and engagement in care and recovery during and after hospitalization. J Am Med Inform Assoc 2018 Feb 01;25(2):206-209. [doi: 10.1093/jamia/ocx054] [Medline: 28633483]
22. Dalal AK, Bates DW, Collins S. Opportunities and challenges for improving the patient experience in the acute and postacute care setting using patient portals: the patient's perspective. J Hosp Med 2017 Dec;12(12):1012-1016. [doi: 10.12788/jhm.2860] [Medline: 29073310]

23. Huerta TR, McAlearney AS, Rizer MK. Introducing a patient portal and electronic tablets to inpatient care. Ann Intern Med 2017 Dec 05;167(11):816-817. [doi: 10.7326/M17-1766] [Medline: 29059690]
24. Coleman EA, Berenson RA. Lost in transition: challenges and opportunities for improving the quality of transitional care. Ann Intern Med 2004 Oct 05;141(7):533-536. [doi: 10.7326/0003-4819-141-7-200410050-00009] [Medline: 15466770]
25. Middleton B, Renner K, Leavitt M. Ambulatory practice clinical information management: problems and prospects. Healthc Inf Manage 1997;11(4):97-112. [Medline: 10177282]
26. Bakken S, Ruland CM. Translating clinical informatics interventions into routine clinical care: how can the RE-AIM framework help? J Am Med Inform Assoc 2009;16(6):889-897 [FREE Full text] [doi: 10.1197/jamia.M3085] [Medline: 19717799]
27. Mlaver E, Schnipper JL, Boxer RB, Breuer DJ, Gershanik EF, Dykes PC, et al. User-centered collaborative design and development of an inpatient safety dashboard. Jt Comm J Qual Patient Saf 2017 Dec;43(12):676-685. [doi: 10.1016/j.jcjq.2017.05.010] [Medline: 29173289]
28. Duckworth M, Leung E, Fuller T, Espares J, Couture B, Chang F, et al. Nurse, patient, and care partner perceptions of a personalized safety plan screensaver. J Gerontol Nurs 2017 Apr 01;43(4):15-22. [doi: 10.3928/00989134-20170313-05] [Medline: 28358972]
29. Dalal A, Fuller T, Garabedian P, Ergai A, Balint C, Bates D, et al. Systems engineering and human factors support of a system of novel EHR-integrated tools to prevent harm in the hospital. J Am Med Inform Assoc 2019 Jun 01;26(6):553-560. [doi: 10.1093/jamia/ocz002] [Medline: 30903660]
30. Businger A, Fuller T, Schnipper J, Rossetti S, Schnock K, Rozenblum R. Lessons learned implementing a complex and innovative patient safety learning laboratory project in a large academic medical center. J Am Med Inform Assoc 2020 Feb 01;7(2):301-307. [doi: 10.1093/jamia/ocz193]
31. Dalal AK, Fuller TE, Garabedian P, Breuer D, Lemonias D, Harry E. An interdisciplinary team to support implementation of a 'system-of-systems' to identify, assess, and mitigate threats to patient safety in real-time. 2018 Apr 08 Presented at: Society of Hospital Medical Conference; April 8-11; Orlando, FL URL: https://www.shmabstracts.com/abstract/an-interdisciplinary-team-to-support-implementation-of-a-system-of-systems-to-identify-assess-and-mitigate-threats-to-patient-safety-in-real-time/
32. Dykes PC, Duckworth M, Cunningham S, Dubois S, Driscoll M, Feliciano Z, et al. Pilot testing fall TIPS (Tailoring Interventions for Patient Safety): a patient-centered fall prevention toolkit. Jt Comm J Qual Patient Saf 2017 Aug;43(8):403-413. [doi: 10.1016/j.jcjq.2017.05.002] [Medline: 28738986]
33. Dykes PC, Rozenblum R, Dalal A, Massaro A, Chang F, Clements M, et al. Prospective evaluation of a multifaceted intervention to improve outcomes in intensive care: the Promoting Respect and Ongoing Safety Through Patient Engagement Communication and Technology Study. Crit Care Med 2017;45(8):e806-e813. [doi: 10.1097/ccm.0000000000002449] [Medline: 28471886]
34. Mueller SK, Schnipper JL, Giannelli K, Roy CL, Boxer R. Impact of regionalized care on concordance of plan and preventable adverse events on general medicine services. J Hosp Med 2016 Sep;11(9):620-627. [doi: 10.1002/jhm.2566] [Medline: 26917417]
35. Dykes PC, Stade D, Chang F, Dalal A, Getty G, Kandala R, et al. Participatory design and development of a patient-centered toolkit to engage hospitalized patients and care partners in their plan of care. AMIA Annu Symp Proc 2014;2014:486-495 [FREE Full text] [Medline: 25954353]
36. Schnock KO, Snyder JE, Fuller TE, Duckworth M, Grant M, Yoon C, et al. Acute care patient portal intervention: portal use and patient activation. J Med Internet Res 2019 Jul 18;21(7):e13336. [doi: 10.2196/13336] [Medline: 31322123]
37. Schnipper JN, Dalal A, Magny-Normilus C, Samal L, Bitton A, Thompson R, et al. Effects of a multi-faceted intervention to improve care transitions within a pioneer accountable care organization. 2016 Presented at: Society of Hospital Medicine Annual Meeting; March 6-9; San Diego, CA URL: https://www.shmabstracts.com/abstract/effects-of-a-multi-faceted-intervention-to-improve-care-transitions-within-a-pioneer-accountable-care-organization/
38. Ganguli I, Sikora C, Nestor B, Sisodia RC, Licurse A, Ferris TG, et al. A scalable program for customized patient education videos. Jt Comm J Qual Patient Saf 2017 Nov;43(11):606-610. [doi: 10.1016/j.jcjq.2017.05.009] [Medline: 29056181]
39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
40. Dalal AK, Schnipper J, Massaro A, Hanna J, Mlaver E, McNally K, et al. A Web-based and mobile patient-centered 'microblog' messaging platform to improve care team communication in acute care. J Am Med Inform Assoc 2017 Apr 01;24(e1):e178-e184. [doi: 10.1093/jamia/ocw110] [Medline: 27539201]
41. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2017 Sep 14;52(4):1893-1907. [doi: 10.1007/s11135-017-0574-8]
42. Glaser BG. The constant comparative method of qualitative analysis. Soc Prob 1965;12(4):436-445. [doi: 10.1525/sp.1965.12.4.03a00070]

43. Bersani K, Fuller TE, Garabedian P, Espares J, Mlaver E, Businger A, et al. Use, perceived usability, and barriers to implementation of a patient safety dashboard integrated within a vendor EHR. Appl Clin Inform 2020 Jan 15;11(1):34-45. [doi: 10.1055/s-0039-3402756] [Medline: 31940670]
44. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. JAMA Intern Med 2013 Apr 22;173(8):632-638. [doi: 10.1001/jamainternmed.2013.3023] [Medline: 23529115]
45. Voss R, Gardner R, Baier R, Butterfield K, Lehrman S, Gravenstein S. The Care Transitions Intervention: translating from efficacy to effectiveness. Arch Intern Med 2011 Jul 25;171(14):1232-1237. [doi: 10.1001/archinternmed.2011.278] [Medline: 21788540]
46. van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. J Gen Intern Med 2002 Mar;17(3):186-192. [doi: 10.1046/j.1525-1497.2002.10741.x] [Medline: 11929504]
47. Misky GJ, Wald HL, Coleman EA. Post-hospitalization transitions: examining the effects of timing of primary care provider follow-up. J Hosp Med 2010 Sep;5(7):392-397. [doi: 10.1002/jhm.666] [Medline: 20578046]
48. Coleman EA. Family caregivers as partners in care transitions: the Caregiver Advise Record and Enable Act. J Hosp Med 2016 Dec;11(12):883-885. [doi: 10.1002/jhm.2637]

Abbreviations

ADD: actual discharge date
AHRQ: Agency for Healthcare Research and Quality
DRG: diagnosis-related group
EHR: electronic health record
PDTK: patient-centered discharge toolkit
PSLL: Patient Safety Learning Laboratory
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by R Greysen, P Gazarian, M Duckworth, C Vis; comments to author 21.10.19; revised version received 16.12.19; accepted 04.02.20; published 28.04.20.

Please cite as:
Fuller TE, Pong DD, Piniella N, Pardo M, Bessa N, Yoon C, Boxer RB, Schnipper JL, Dalal AK
Interactive Digital Health Tools to Engage Patients and Caregivers in Discharge Preparation: Implementation Study
J Med Internet Res 2020;22(4):e15573
URL: http://www.jmir.org/2020/4/e15573/
doi:10.2196/15573
PMID:32343248

Original Paper

Parent Perspectives on Family-Centered Pediatric Electronic Consultations: Qualitative Study

Rhea Verma[1]; Tamar Krishnamurti[2], PhD; Kristin N Ray[1], MD, MS

[1]Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[2]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Corresponding Author:
Kristin N Ray, MD, MS
Department of Pediatrics
University of Pittsburgh School of Medicine
3414 Fifth Avenue
Pittsburgh, PA, 15213
United States
Phone: 1 4126926000
Email: kristin.ray@chp.edu

Abstract

Background: Electronic consultations, which use store-and-forward transfer of clinical information between a primary care physician and a specialist, improve access to specialty care. Adoption of electronic consultations is beginning in pediatric health care systems, but little is known about parent perspectives, informational needs, and preferences for interaction with this new model of care.

Objective: This study aimed to examine parent perspectives about electronic consultations, including perceived benefits and risks, anticipated informational needs, and preferences for parent engagement with electronic consultations.

Methods: We recruited caregivers of pediatric patients (aged 0-21 years) attending visits at an academic primary care center. Caregivers were eligible if their child had ever been referred for in-person specialty care. Caregivers participated in a semistructured interview about electronic consultations, including general perspectives, desired information, and preferences for parental engagement. Interviews were transcribed and qualitatively analyzed to identify parent perspectives on electronic consultations in general, information parents would like to receive about electronic consultations, and perspectives on opportunities to enhance parent engagement with electronic consultations.

Results: Interviewees (n=20) anticipated that electronic consultations would reduce the time burden of specialty care on families and that these had the potential to improve the integrity and availability of clinical information, but interviewees also expressed concern about data confidentiality. The most detailed information desired by interviewees about electronic consultations related to data security, including data confidentiality, availability, and integrity. Interviewees expressed concern that electronic consultations could exclude parents from their child's health care decisions. Interviewees saw value in the potential ability to track the consultation status or to participate in the consultation dialogue, but they were more ambivalent about the idea of read-only access to consultation documentation.

Conclusions: Parents identified the potential risks and benefits of pediatric electronic consultations, with implications for communication with families about electronic consultations and for incorporation of features to enhance parent engagement.

(J Med Internet Res 2020;22(4):e16954) doi:10.2196/16954

KEYWORDS

consultation; referral; telemedicine; telehealth; child health; child health services

Introduction

Background

The demand for pediatric specialty care exceeds supply, resulting in challenges such as long wait times for families seeking specialty care [1,2]. One innovative and promising strategy to improve timely access to specialty care is electronic consultations, a store-and-forward type of telemedicine [3]. Electronic consultations, also called eReferrals and eConsults in specific health systems, allow primary care physicians (PCPs)

to communicate with specialists about a specific patient as a way to *right size* the patient's specialty care [4]. In an electronic consultation, the PCP sends a clinical question to a specialist, along with photos, videos, and any other relevant media through a secure electronic platform. The specialist reviews the information at a later time and then sends recommendations to the PCP, including advice for further PCP-driven evaluation, potential management through the PCP, or timeframe for an in-person specialty consultation if indicated. The PCP, consequently, communicates those recommendations back to the patient. This process, intended for nonurgent specialist input, may completely avert the need for the patient to physically attend an in-person specialty consultation, or it may guide interval care so that evaluation and management can be optimized while awaiting in-person specialty consultation.

To date, electronic consultations have been implemented in several health care systems, including the Department of Veterans Affairs, the Mayo Clinic, San Francisco General Hospital, and the Los Angeles County Department of Health Services, and these show initial promise in their ability to efficiently meet the demand for specialty care [5-7]. After electronic consultations became a required preliminary step for all referrals through the LA County Department of Human Services, 25% of the electronic consultations were resolved without a specialist visit, and the percentage of referrals scheduled within 30 days improved from 24% to 30% [8]. Studies of clinician perspectives on electronic consultations identified potential clinician-perceived benefits and risks [9,10]. Studies of the perspectives of adult patients are more limited, but these studies reported general acceptability [11,12]. Patients appear to value the potential for electronic consultations to improve access to specialist expertise and to place PCPs in a more central role [7], but patients raise concerns that the information transmitted may not be comprehensive and that quality of the outcome may depend on patient-PCP relationships [11,13]. In addition, adult patients also expressed a desire to be more informed about and engaged with the electronic consultation process, which often appeared to occur without patient knowledge [13].

Electronic consultations are now also beginning to be used by innovative pediatric referral centers. The Canadian Champlain Building Access to Specialists through eConsultations eConsult service reported on over 1000 pediatric electronic consultations between 2014 and 2016, where 36% of the electronic consultations were resolved without an in-person visit and PCPs reported high satisfaction [14]. At Boston Children's Hospital, a pilot electronic consultation from 2014 to 2016 significantly reduced wait time for specialty appointments and improved appointment completion rates, with PCPs also reporting improved communication and care [15]. For the adaption and implementation of new technology in general, input from the end user is needed to optimize acceptability and impact. When new technology and care models are directed at children, caregiver perspectives and preferences become an important part of the design processes, given the essential role of adult caregivers as integral partners of the pediatric care team [16]. Thus, as additional pediatric systems consider adopting electronic consultations, it will be important to consider parental views on electronic consultations and optimal design of electronic consultation systems, yet little is known about parent perspectives and preferences around this model of care.

Objective

This study aimed to fill this knowledge gap by assessing parent perspectives, anticipated informational needs, and preferences to guide the development of family-centered electronic consultation systems. Specifically, we aimed to answer three questions: (1) what do parents see as the benefits and risks of electronic consultations? (2) what information would parents like to receive before a physician initiates an electronic consultation for their child? and (3) what value do parents perceive in features that could heighten parent engagement with electronic consultation systems? Answering these questions will facilitate a more family-centered design of pediatric electronic consultations to optimize usability and usefulness from parent perspectives.

Methods

Study Design

We performed a qualitative analysis of semistructured interviews to identify family preferences regarding electronic consultations for pediatric specialty care.

Recruitment

During June 2019 and July 2019, we conducted semistructured interviews with caregivers of children attending primary care visits at the University of Pittsburgh Medical Center Children's Hospital of Pittsburgh (CHP) Primary Care Clinic. The CHP Primary Care Center is an academic pediatric primary care center where the majority of patients are insured by Medicaid. Eligible participants were the parents of pediatric patients (aged 0-21 years) attending either well-child visits or acute visits whose child had ever been referred to specialty care. Pediatricians at the clinic identified potentially eligible participants and asked for permission for the research team to provide study information. A research team member then invited parents to participate in a semistructured interview and obtained verbal consent from participants. Participants received a US $25 gift card at the conclusion of the interview to compensate them for their participation. The University of Pittsburgh Institutional Review Board provided ethical review and determined this study to be exempt from formal further review.

Interview Guide

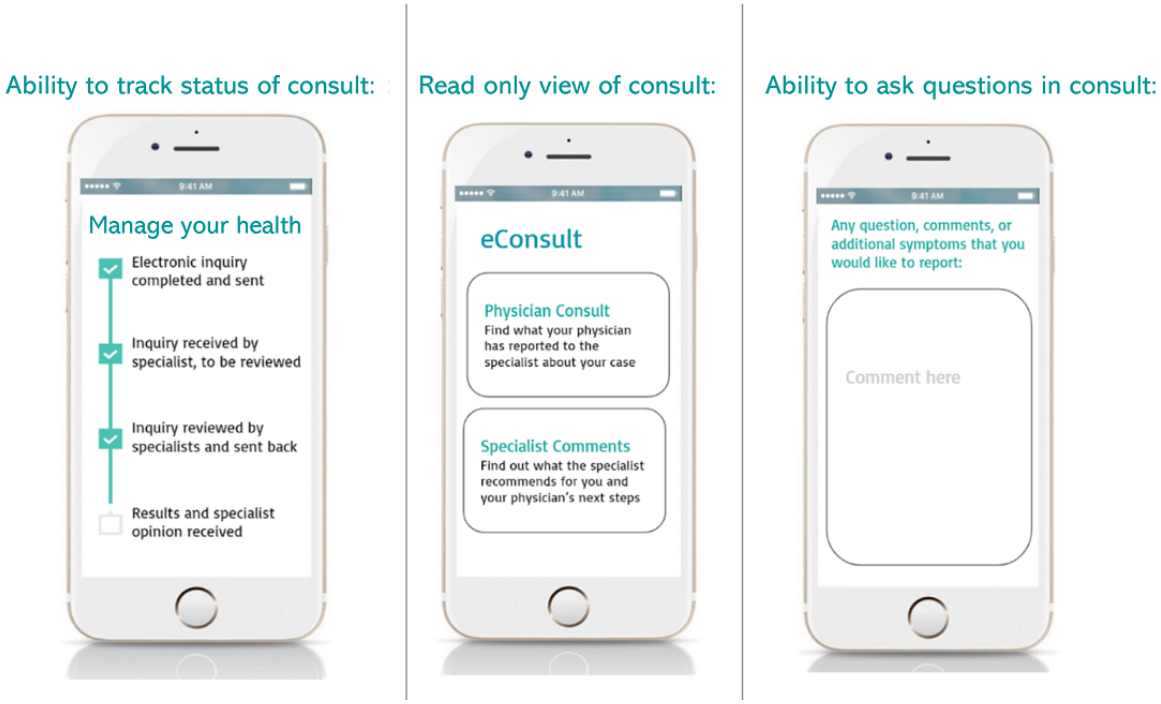
An interview guide was developed to include questions that explored parents' perspectives on electronic consultations, the information they would like to receive if electronic consultations were to be used for their child, and potential features to enhance family engagement. Electronic consultations were described for interviewees using standard language as a process involving the following:

> *A pediatrician summarizing a clinical question about a child in written form and sending it to a specialist. Sometimes, the written information could also include a picture or a video. The specialist then reviews this information, provides their recommendations, and*

sends it back to the pediatrician, who updates the family by phone within a few days.

Interview prompts then asked about circumstances where they would might prefer to use either electronic consultations or in-person visits, information desired about the process, preferences regarding the information transferred within the electronic consultation, and experiences while accessing child health information on the internet (eg, through a patient portal). Finally, we examined parent perspectives on different levels of potential family involvement with the electronic consultation process. For this final question, we developed and presented static prototypes of three hypothetical options for family engagement with electronic consultations within a patient portal (Figure 1), with parents asked to discuss their reactions to each. First, parents were asked to consider an option where they could view the status of the electronic consultation (sent, read, or responded) but not the actual content of the consultations. Second, parents were asked to consider an option where they could read the actual text of the communication between the PCP and the specialist in a read-only view. Finally, parents were asked to consider an option where they could read the text of the communication between the PCP and specialist with the additional ability of adding their own questions and comments to the pediatrician-specialist electronic consultation dialogue.

Figure 1. Prototype options for family engagement with electronic consultation. Three prototype static visual images sequentially presented to interviewees during interviews.



Qualitative Analysis

The interviews were digitally recorded and subsequently transcribed. Transcripts and audio files were stored without personal identifiers on a password-protected server. Two investigators (RV and KR) analyzed interview transcripts using thematic content analysis. After coding the first five interviews separately, a preliminary codebook was developed. Using this preliminary codebook, the two investigators then recoded the first five interviews and coded all subsequent interviews, meeting regularly to discuss discrepancies in coding and to refine the codebook to encompass additional emerging themes. Analysis was performed using qualitative data software (Dedoose, SocioCultural Research Consultants, Los Angeles, California), with recruitment continuing until thematic saturation was reached [17]. The results in this paper are organized around three major overarching themes: perspectives on electronic consultations in general, information parents would like to receive about electronic consultations, and perspectives on opportunities to enhance parent engagement with electronic consultations.

Results

Participants

In total, we interviewed 20 caregivers (17 mothers, 2 fathers, and 1 grandmother) of children referred to specialty care (Table 1). Reflecting clinic demographics, 90% (18/20) of the caregivers reported that their children were insured by Medicaid, and 85% (17/20) of the participants identified as black. All respondents reported owning a smartphone. In addition, 17 respondents reported experience using a web-based patient portal, either for themselves or for their child.

Table 1. Participant demographics.

Characteristics	Value, n (%)
Participant age (years)	
<24	2 (10)
25 to 44	16 (80)
45 to 64	2 (10)
Participant gender	
Female	18 (90)
Male	2 (10)
Participant relationship with patient	
Mother	17 (85)
Father	2 (10)
Grandparent	1 (5)
Participant patient portal experience	
Portal user	17 (85)
Nonuser	3 (15)
Patient visit reason	
Wellness check	18 (90)
Sick visit	2 (10)
Patient insurance	
Commercial insurance	2 (10)
Medicaid/Children's Health Insurance Plan	18 (90)

Perspectives on Positive and Negative Aspects of Electronic Consultations

In discussing electronic consultations as a potential substitute or complement to in-person specialty visits, interviewees discussed several ways in which electronic consultations could alter (1) the time burden of specialty care on families, (2) the transfer of information between the doctors and caregivers, and (3) caregiver involvement with the consultative process (Table 2).

There was general agreement that electronic consultations would reduce the time burden on families in multiple ways. Interviewees noted that if an issue could be resolved without a specialist visit, they would also be able to avoid unnecessary travel and reduce the time spent away from school and work. Interviewees were optimistic that electronic consultations might also reduce the time they currently spend waiting for specialist input in their child's care:

> *I mean, I think that the wait time to get in to see a psychiatrist is like a year and a half. So if you have an issue and you need to get in there now, a digital consult might actually be helpful. That way, at least they can get you on some medicine or whatever.*

Interviewees also expressed hope that if an in-person specialist visit was determined to be necessary, the electronic consultation process might still result in more efficient care if it could simplify scheduling processes or allow for increased efficiency because of previsit communication. For example, one mother suggested the following:

> *It would save time of sitting there and have me go over every [piece of medical history]—if you already had it when I come in, I can give you a brief synopsis of what it is and you can save a little bit of time there.*

Interviewees had more mixed comments across issues related to information transfer. Interviewees expressed concern and hesitancy related to data confidentiality, stating concern that their child's information would "get into the wrong hands." Interviewees raising this concern elaborated that they would be particularly worried if multimedia information (eg, images) were included. Relatedly, some interviewees queried whether parental permission would be required for this digital transfer of information. However, interviewees also predicted a positive effect on data integrity and availability. Specifically, the ability of the specialist to have access to comprehensive health information was a feature in which many interviewees saw value. Interviewees also valued the potential "paper trail" of electronic consultations, which might clarify what information was available and who had access:

> *I think it would help out a whole lot. I think it will eliminate a lot of, you know, confusion and paperwork, too.*

Table 2. Parent perspectives on anticipated benefits and risks of electronic consultations.

Theme	Definition	Example
Time burden: potential benefits		
Less time to hear back	Hearing back electronically would allow faster access to specialist expertise	"I mean, I think that the wait time to get in to see a psychiatrist is like a year and a half. So if you have an issue and you need to get in there now, a digital consult might actually be helpful. That way, at least they can get you on some medicine or whatever."
Save unnecessary visits	Electronic consultation could avoid a need for specialty visits in some circumstances	"Yeah, so if they would've had [electronic consults], that would've saved us trips going to the hospital."
Save travel time	Time saved from not having to drive to specialty care location	"I guess that would be a little bit better for me, so I'm not traveling 20 million miles."
Less missed school/work/other commitments	Avoiding opportunity costs of missed school/work because of appointments	"I think that would make it a lot easier for some of us parents that have to deal with truancy."
Have a specialist agree that a visit is needed before going	Likes the idea of a specialist reviewing and agreeing that visits are needed rather than going in and finding out visits are not needed	"I'm sure she could have probably sent over, you know, the stuff from the bloodwork in conjunction with the growth chart in an e-mail or whatever and said, "Hey, what do you think of this?" And if it was medically necessary, then we'd go to visit."
Previsit communication can save time in overall appointment	Previsit consultation allows doctors to know what to look for	"So yeah, definitely the electronic way would have helped in the past and most likely in the future, as well, with them being able to see the issue prior to us getting there. So that way they can really relay to us exactly what's going on."
Shorter appointments overall	Doctors know what to look for during appointments and have patient information ready	"it would save the time of sitting there and have me go over every—if you already had have it when I come in, I can give you a brief synopsis of what it is and you can save a little bit of time there."
Removes scheduling difficulty	Less stress about getting an appointment in a timely manner, especially if one is not needed at all	"I would have appreciated [an electronic consultation] versus me actually...going and schedule an appointment because it would have been the timeframe, the process..."
Postvisit communication and follow-up are more timely electronically	Keeps postvisit communication concise, less need to keep going back in after the initial visit	"I definitely will like it when my child has an ongoing issue, you know, say, like asthma or something and they just need a refill. I don't feel I should have to come into the office just for them to give me a refill. You know what I mean? Like just something that's ongoing, you know?"
Data availability, integrity, and confidentiality: potential benefits		
More comprehensive transfer of information	More comprehensive transfer of information from the primary care physician to the specialist	"But at least knowing that they have a heads up and they know what to look for and why we're coming would be even more reassuring."
More convenient transfer of information	The transfer of information is more convenient when an electronic consultation is used	"I think it would help out a whole lot. I think it will eliminate a lot of, you know, confusion and paperwork, too."
Better paper trail	Clearer paper trail that parents can refer to as documentation of visit occurrence and visit content	"That'll be kind of like my back-up, you know what I mean? My paper trail."
Data availability, integrity, and confidentiality: potential risks		
Data confidentiality	Data should not be discussed with unauthorized people	"Confidentiality, that's it. That's the only thing...I would be concerned about with stuff being sent electronically."
Incorrect/incomplete information transferred	Possibility of incomplete transfer of information or incorrect interpretation	"I think they have to see up close and personal because maybe there's something that they can see that the picture didn't quite capture, you know?"
Inaccurate decisions made with incomplete information	Electronic consultations do not give the full picture and may result in incorrect diagnosis	"So if I didn't bring her in we might've not found that. And it was good that we came in and we didn't do it electronically."
More uncertainty when decisions are made with incomplete information	The specialist may not be sure how to diagnose based on just the information provided	"I'm like instead of just like 'Oh yeah, you know, this *might* be the problem' like no, I need you to look."
Parent involvement: potential risk		

Theme	Definition	Example
Less parent interaction with specialist	Less direct interaction with specialists	"I need to ask questions, I need to know everything"; "I'm hands-on. I want to see you. I want you to physically see my child."
Reduced quality of communication	Less ability to ensure high-quality communication	"No, I'd rather go to a specialist and then hear it like from the horse's mouth and then instead of being a third party, she can explain everything to me at that time."
Decreased opportunity to ask questions	Decreased parent involvement and opportunity to ask questions	"I would've preferred to just go in and see the specialist because it...it's better. It – I mean, you get more answers that way, I guess, so."
Contextual factors impacting relative risk/benefit		
Family dependent	Interest in use of electronic consultations depends on family situation and preferences	"That one, too, it's different because I guess it all depends on where people's at within the medical field thing, because I feel that I can always ask the questions to them myself, not necessarily in between what they're talking about."
Clinical situation dependent	The use of electronic consultations depends on the immediate clinical situation and the urgency for care	"If its more of an issue that's more in-depth, where it actually has to be seen and they're not too sure, then yes, I would prefer to just go, just to get more so a clear, a better answer to what's going on."
Parent permission dependent	Parents should decide who can assess the patients' records	"Of course, they get parental permission."
Anticipated value relative to current systems	Interest in electronic consultations expressed relative to current systems	"Yeah, that's something I would be willing to use. It sounds like way more easier than the stuff that goes on now."

Interviewees had mixed views on how accurate electronic consultations could be, which related to concerns about data integrity and availability. Specifically, interviewees stated with concern that if the specialists only receive the information that is specifically sent to them, they may miss the broader context of the child's health.

An additional concern raised by some interviewees was that the process of an electronic consultation could exclude caregivers from the clinical conversations and decision making, in contrast to an in-person visit, limiting their ability to provide information and context, as well as their ability to ask questions:

> *I would've preferred to just go in and see the specialist because it...it's better. It – I mean, you get more answers that way, I guess, so.*

Balancing these advantages and disadvantages, overall, interviewees appeared to favor the possibility of using electronic consultations, with most interviewees stating that there are times that they would have preferred the use of an electronic consultation rather than an in-person visit to the specialist:

> *Yeah, that's something I would be willing to use. It sounds like way easier than the stuff that goes on now.*

However, interviewees noted that whether electronic consultations were appropriate in a given circumstance might vary with both the clinical indication as well as with family circumstances and comfort with technology:

> *Plus, I...I'm not computer illiterate, you know what I mean? So it's OK for me.*

Desired Information About Electronic Consultations

When interviewees were asked about what information about an electronic consultation they would like to know when considering its use for their child, most interviewees suggested a desire for relatively detailed information, including what the steps of the consultation are, what information is being sent about their child, and the speed of the expected response:

> *I would like for them to explain the whole process. What it entails, what they're going to – the information that they're going to give the doctor and even the timeframe when they – when we should hear something. So yes, I would expect for them to explain everything before proceeding.*

Some interviewees stated that they would want additional specific information, such as the rationale for use, specifics of information transfer and security, and potential outcomes of the process (Table 3). A small minority of interviewees, in contrast, expressed minimal need for information on the process.

Table 3. Parent perspectives on information they would like to receive at the time of pediatric electronic consultation.

Theme	Definition	Example
Minimal explanation desired	Parents want minimal explanation	"I think the process is pretty clear, and if it wasn't, I'm sure I would be able to ask the questions electronically, so."
Desire to know the speed of the consultation	Want to know when they will hear back from specialist	"I would like for them to explain the whole process. What it entails, what they're going to – the information that they're going to give the doctor and even the timeframe when they – when we should hear something. So yes, I would expect for them to explain everything before actually proceeding."
Desire to know rationale for use	Parents want to understand why an electronic consultation is being used in their case instead of an in-person visit	"Why is it electronic consult versus seeing the person and...you know, actually in person."
Desire to know what information will be transferred	Desire to know what information is being put in the consultation for the specialist to review	"I guess just letting me know everything that she was going to be doing, and you know, just keeping me informed of like, you know, I guess any pictures or videos or anything that's being sent to them."
Desire to know about security/quality of information transfer	Parents want to know what security measures are taken to protect the child's information	"What kind of security is there? You know, if for some reason there would be some type of breach, what are the protocols to let the parents know that pictures of my child are no longer safe, that kind of thing. Those are the – probably the biggest questions in my mind."
Desire to understand the steps/paper trail	Parents want to understand how information is being transferred and the steps to an electronic consultation	"Exactly what they're taking the test on, like I need to go from point A to point B, C D. Every step that they're doing, I need to know, and I need to be broken down."
Explanation of possible diagnoses and management	Parents want possible diagnoses and treatments explained to them while awaiting consultation advice	"The most serious, the most important, and what can be helped – like a solution. So, yeah, that's basically it. Like, the most serious, 'This is this,' like the most important about it, what can it affect, the stuff like that."

Perspectives on Parent Participation in Electronic Consultations

When presented with options that might enhance parent engagement with the electronic consultation process (Figure 1), almost all the respondents responded positively to the idea of a feature allowing tracking of the status of the electronic consultation (eg, sent, read, and replied) to help them know how the process was advancing and to anticipate when they might hear back (Table 4):

> *I guess I don't have to keep calling people and them calling me back or saying, you know, they're busy right now. I guess it would be plain in sight for me to be able to see myself instead of having to go through 50 people.*

Interviewees were more ambivalent about read-only access to the electronic consultation. Some anticipated value in being able to "stay up to date" on the physician conversation, and others valued the potential to assess the quality of the physician-to-physician communication with a read-only feature:

> *Just to know what's going on, and then determine if I want to see the consultant in person.*

However, others were apprehensive that medical terminology would be confusing and reported they prefer to receive information verbally through their PCP after the electronic consultation process:

> *I would prefer they do that in private and then talk to me because I'm not no doctor and half the words that they going to be saying, I don't know.*

Relatedly, some interviewees also mentioned that getting too much information could be overwhelming.

Despite these concerns raised about reviewing electronic consultations through read-only access, interviewee responses were generally positive when asked about the possibility of a feature allowing parents to join the generalist-specialist dialogue by adding their own questions and comments to the consultation. Interviewees thought this feature would improve their ability to get questions answered and would also better approximate an in-person appointment:

> *If I have a question about something, I could ask the doctors directly, you know, see what they're saying. 'Cause that's how it is when you go to appointments: they include you. So I would want to be included.*

Many families also mentioned that if the doctor missed something or had incorrect information, this feature would give them the opportunity to correct them, which might consequently lead to a more accurate diagnosis. Two families, however, were skeptical of whether the parents' comments would add value to a consultation between two doctors ("I don't know on the medical perspective how it will be used"), and one caregiver interviewee suggested that a word limit should be placed to keep parent comments concise.

Table 4. Parent perspectives on parent participation in electronic consultations.

Theme	Description	Example
Tracking electronic consultation status: benefit		
Ability to track consultation status	Benefit of tracking consultation status/reduces uncertainty	"I guess I don't have to keep calling people and them calling me back or saying, you know, they're busy right now. I guess it would be plain in sight for me to be able to see myself instead of have to going through 50 people."
Read-only access to electronic consultation documentation: benefit		
Ability to stay up to date	Ability to follow communication within the consultation	"...because you are in the loop, what's going on, yeah. With this experience with [name] seven years ago, we wanted to know exactly was what going on every day or every time. Yeah, so I am pro to see the communication."
Ability to assess quality of electronic consultation	Ability to decide whether to trust the electronic consultation and the information given	"Just to know what's going on, and then determine if I want to see the consultant in person."
Read-only access to electronic consultation documentation: risk		
May not understand medical terminology	Parents may feel excluded from interaction if they cannot understand terminology	"I would prefer they do that in private and then talk to me because I'm not no doctor and half the words that they going to be saying, I don't know."
Parents would rather hear recommendations through a PCP[a] than read them	Parent prefers for a PCP to explain the problem to them	"So I would prefer them do that in their time and then come explain it to me when they get, you know, all their facts and stuff together."
Ability to comment on electronic consultation dialogue: benefit		
Ability to get questions answered	Parent use of interactive features would increase timely answers	"I mean, once their initial communication is complete...at that time, once I view it, if I have a question or concern, I can type out a message and then send it out to both the doctor and the specialist."
Parent input/comments can improve consultation quality	Parent use of interactive features to add relevant information to improve the consultation quality	"Yes – that's important as well too. Cause the parent has a different perspective. The parent might say, 'OK, well, this is not what's – its more geared toward this, this is the issue more' or something. Year, I think that's important too, to have that little, you know, open communication."
Parent input better approximates in-person visits	Parent use of interactive features would better replicate communication in an in-person visit	"Because if I have a question about something, I could ask the doctors directly, you know, see what they're saying. 'Cause that's how it is when you go to appointments: they include you. So I would want to be included."
Ability to comment on electronic consultation dialogue: risk		
Skeptical that parent comments can add value	Parents worried that their comments may not be accounted for or may not be helpful	"I don't know on the medical perspective how it will be used."

[a]PCP: primary care physician.

Discussion

Using a qualitative analysis of semistructured interviews, we identified caregiver perspectives on potential benefits and risks of electronic consultations in pediatric care, information desired by caregivers about electronic consultations before use, and reactions to potential strategies to enhance parent engagement with electronic consultations. With the adoption of electronic consultations just beginning in pediatric health care systems [14,15], these results are important for envisioning optimal parent engagement and proactively developing approaches to increase the acceptability, uptake, and impact of this emerging model of specialty care.

Caregivers appreciated a range of ways through which electronic consultations could reduce the time spent obtaining specialist expertise through in-person care. They anticipated benefits not only from gaining specialist advice more rapidly through this system but also from avoiding the time burden of scheduling and attending an in-person visit. These perceived benefits are supported by previous studies highlighting resolution of specialty care needs without in-person visits and improved time to appointment when visits are needed [8,15]. Although adult patients and primary care providers tempered similar perceptions with concerns about the electronic consultation process potentially adding delays to definitive care [9,13], parent interviewees did not voice such concerns. Overall, the perceived benefits in access by parents contributed to an overall positive perception of this model of care. Parents' informational needs related to this domain were relatively straightforward, with a desire to know the anticipated timeframe for follow-up communication from their PCP.

Caregiver perceptions of risks and benefits related to data security (including data confidentiality, data integrity, and data availability) associated with electronic consultations were more varied. In general, parents raised concerns about electronic consultation data falling into "the wrong hands," but they perceived benefits in the resulting "paper trail" regarding their child's care. Parents also had mixed perceptions regarding whether the information transmitted between PCPs and specialists would be more or less comprehensive than current processes, with implications for their confidence in the accuracy of the resulting clinical decisions. Much of the information desired by parents about electronic consultations related to these data security concerns, including a desire to know what information will be transferred, who would have access, and details of data transfer security and quality. These concerns were not predominant features of previous studies of adult patients [13], perhaps suggesting a greater drive to be a good steward of data for others than for oneself. Alternatively, adult patient studies focused on individuals with experience using an electronic consultation system, whereas this study discussed a hypothetical system, it may be that these fears become allayed after the experience of using a well-designed system. Regardless, the study's results suggest that, at least during the initial adoption phases, parents desire comprehensive information regarding all domains of data security (eg, data confidentiality, integrity, and availability) at the time of electronic consultation initiation. As a result, systems may wish to develop patient education tools so that the PCPs initiating electronic consultations can share this information accurately and efficiently.

Caregivers also voiced concerns about the idea that they would have less ability to contribute information or to ask questions of the specialist during the electronic consultation process. This relates to ideas of data integrity—information might be missing without caregiver involvement—but extends farther to the ways in which electronic consultations alter decision making in the triad of specialist, PCP, and parent. Adult patients who used electronic consultations appeared to value the strengthening of the role of their PCP relative to the specialist in clinical care [11], but this idea did not emerge from parent respondents considering hypothetical use of electronic consultations, who focused instead on the relative diminishing of their own role. This primarily suggests that centering the family-PCP relationship in the information about the electronic consultation process may improve parent acceptability. This also suggests that the opportunities to increase family involvement in electronic consultations may be particularly valued in pediatric settings.

Regarding specific strategies to increase family involvement, caregivers were generally interested in being able to track the electronic consultation process, with no risks of this strategy raised by parents. Parents were more ambivalent about having read-only access to electronic consultation dialogue, with some valuing the potential to ascertain accuracy but others expressing concern that it could further their sense of being outside of the process because of lack of comprehension or inability to participate. Of note, data from settings sharing clinical notes in general (eg, OpenNotes) suggest that patients do often identify accuracy concerns, but they also benefit from enhanced patient understanding and patient-doctor relationships [18,19]. Such findings suggest that read-only access to electronic consultations may not ultimately result in the disenfranchisement that some parents envision, but these parental concerns warrant consideration of other levels of engagement or actionable patient education to assuage these concerns. Specifically, parents preferred access to notes when access was paired with opportunities to add details and ask questions, effectively creating a three-way dialogue among parent, specialist, and PCP. Although caregivers perceived this to best approximate in-person visits, this strategy could actually be viewed as a step beyond usual care. Instead of sequential dyadic conversations (parent-PCP and parent-specialist), this option could generate an ongoing conversation among all three relevant parties.

This study has several limitations. First, interviews were conducted at a single, urban, and academic primary care center, and the majority of participants were African American and female. In addition, most participants reported experience using a patient portal either for themselves or their child, which differs from national estimates in which one-third of the individuals reported patient portal use [20]. All caregivers had children who were referred to specialty care, but the recency of that referral varied (some referred that day and some in the past). In addition, as formative work to inform the design of an electronic consultation system, this study asked caregivers to consider the hypothetical use of an electronic consultation system, and the findings may differ as parents gain experience with electronic consultations in general or systems with specific features. Finally, as a qualitative analysis, the study's results should be considered hypothesis generating rather than hypothesis confirming.

In conclusion, the study's results suggest that caregivers perceive value in the use of electronic consultations, largely motivated by more timely and efficient access to specialist expertise for their children. Parents wish to receive information about the confidentiality, integrity, and availability of clinical information throughout the electronic consultation process, and systems may also wish to include messaging about electronic consultations that centers the family-PCP relationship. Systems considering electronic consultation should consider developing clear communications that address parents' concerns and informational needs before integration. The inclusion of design features to track the electronic consultation process and to contribute to a three-way dialogue was of interest to families. Incorporating these parent perspectives into the design of pediatric electronic consultations may enhance acceptability and uptake of electronic consultations and optimize their ability to improve upon current processes of specialty care delivery.

Acknowledgments

This study was supported in part by grant funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (K23HD088642, KR) and by the National Institutes of Health (NIH; NIH KL2 TR001856, TK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders 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; or decision to submit the manuscript for publication.

Conflicts of Interest

TK is a cofounder of Naima Health LLC, which develops digital health tools to engage patients in clinical care. The remaining authors declare no conflicts of interest.

References

1. Bethell CD, Kogan MD, Strickland BB, Schor EL, Robertson J, Newacheck PW. A national and state profile of leading health problems and health care quality for US children: key insurance disparities and across-state variations. Acad Pediatr 2011;11(3 Suppl):S22-S33 [FREE Full text] [doi: 10.1016/j.acap.2010.08.011] [Medline: 21570014]
2. Pletcher BA, Rimsza ME, Cull WL, Shipman SA, Shugerman RP, O'Connor KG. Primary care pediatricians' satisfaction with subspecialty care, perceived supply, and barriers to care. J Pediatr 2010 Jun;156(6):1011-5.e1. [doi: 10.1016/j.jpeds.2009.12.032] [Medline: 20227727]
3. Kim Y, Chen AH, Keith E, Yee HF, Kushel MB. Not perfect, but better: primary care providers' experiences with electronic referrals in a safety net health system. J Gen Intern Med 2009 May;24(5):614-619 [FREE Full text] [doi: 10.1007/s11606-009-0955-3] [Medline: 19308334]
4. Chen AH, Murphy EJ, Yee HF. eReferral--a new model for integrated care. N Engl J Med 2013 Jun 27;368(26):2450-2453. [doi: 10.1056/NEJMp1215594] [Medline: 23802515]
5. Vimalananda VG, Gupte G, Seraj SM, Orlander J, Berlowitz D, Fincke BG, et al. Electronic consultations (e-consults) to improve access to specialty care: a systematic review and narrative synthesis. J Telemed Telecare 2015 Sep;21(6):323-330 [FREE Full text] [doi: 10.1177/1357633X15582108] [Medline: 25995331]
6. Liddy C, Moroz I, Mihan A, Nawar N, Keely E. A systematic review of asynchronous, provider-to-provider, electronic consultation services to improve access to specialty care available worldwide. Telemed J E Health 2019 Mar;25(3):184-198. [doi: 10.1089/tmj.2018.0005] [Medline: 29927711]
7. Vimalananda V, Orlander JD, Afable MK, Fincke BG, Solch AK, Rinne ST, et al. Electronic consultations (E-consults) and their outcomes: a systematic review. J Am Med Inform Assoc 2019 Oct 17:pii: ocz185. [doi: 10.1093/jamia/ocz185] [Medline: 31621847]
8. Barnett ML, Yee HF, Mehrotra A, Giboney P. Los Angeles Safety-Net Program eConsult System was rapidly adopted and decreased wait times to see specialists. Health Aff (Millwood) 2017 Mar 1;36(3):492-499. [doi: 10.1377/hlthaff.2016.1283] [Medline: 28264951]
9. Lee MS, Ray KN, Mehrotra A, Giboney P, Yee HF, Barnett ML. Primary care practitioners' perceptions of electronic consult systems: a qualitative analysis. JAMA Intern Med 2018 Jun 1;178(6):782-789 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0738] [Medline: 29801079]
10. Deeds SA, Dowdell KJ, Chew LD, Ackerman SL. Implementing an opt-in eConsult program at seven academic medical centers: a qualitative analysis of primary care provider experiences. J Gen Intern Med 2019 Aug;34(8):1427-1433. [doi: 10.1007/s11606-019-05067-7] [Medline: 31197734]
11. Joschko J, Liddy C, Moroz I, Reiche M, Crowe L, Afkham A, et al. Just a click away: exploring patients' perspectives on receiving care through the Champlain BASETM eConsult service. Fam Pract 2018 Jan 16;35(1):93-98. [doi: 10.1093/fampra/cmx073] [Medline: 28968806]
12. Keely E, Traczyk L, Liddy C. Patients' perspectives on wait times and the referral-consultation process while attending a tertiary diabetes and endocrinology centre: is eConsultation an acceptable option? Can J Diabetes 2015 Aug;39(4):325-329. [doi: 10.1016/j.jcjd.2014.12.010] [Medline: 25797111]
13. Olayiwola JN, Knox M, Dubé K, Lu EC, Woldeyesus T, James IE, et al. Understanding the potential for patient engagement in electronic consultation and referral systems: lessons from one safety net system. Health Serv Res 2018 Aug;53(4):2483-2502 [FREE Full text] [doi: 10.1111/1475-6773.12776] [Medline: 28940495]
14. Lai L, Liddy C, Keely E, Afkham A, Kurzawa J, Abdeen N, et al. The impact of electronic consultation on a Canadian tertiary care pediatric specialty referral system: A prospective single-center observational study. PLoS One 2018;13(1):e0190247 [FREE Full text] [doi: 10.1371/journal.pone.0190247] [Medline: 29320539]
15. Rea CJ, Wenren LM, Tran KD, Zwemer E, Mallon D, Bernson-Leung M, et al. Shared Care: using an electronic consult form to facilitate primary care provider-specialty care coordination. Acad Pediatr 2018;18(7):797-804. [doi: 10.1016/j.acap.2018.03.010] [Medline: 29625232]

16. COMMITTEE ON HOSPITAL CARE and INSTITUTE FOR PATIENT- AND FAMILY-CENTERED CARE. Patient- and family-centered care and the pediatrician's role. Pediatrics 2012 Feb;129(2):394-404. [doi: 10.1542/peds.2011-3084] [Medline: 22291118]
17. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822x05279903]
18. Sarabu C, Pageler N, Bourgeois F. OpenNotes: toward a participatory pediatric health system. Pediatrics 2018 Oct;142(4):pii: e20180601 [FREE Full text] [doi: 10.1542/peds.2018-0601] [Medline: 30228169]
19. 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 2017 Apr;26(4):262-270. [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
20. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]

Abbreviations

CHP: Children's Hospital of Pittsburgh
NIH: National Institutes of Health
PCP: primary care physician

Edited by G Eysenbach; submitted 07.11.19; peer-reviewed by C Rea, F Meza, C Wu; comments to author 15.01.20; revised version received 19.01.20; accepted 26.01.20; published 09.04.20.

Please cite as:
Verma R, Krishnamurti T, Ray KN
Parent Perspectives on Family-Centered Pediatric Electronic Consultations: Qualitative Study
J Med Internet Res 2020;22(4):e16954
URL: https://www.jmir.org/2020/4/e16954
doi:10.2196/16954
PMID:32084626

Original Paper

Patient Perceptions of Video Visits Using Veterans Affairs Telehealth Tablets: Survey Study

Cindie Slightam[1], MPH; Amy J Gregory[1], BA; Jiaqi Hu[1,2], MPH; Josephine Jacobs[1], PhD; Tolessa Gurmessa[1,3], MD; Rachel Kimerling[1,4], PhD; Daniel Blonigen[1], PhD; Donna M Zulman[1,2], MSc, MD

[1]Center for Innovation to Implementation, Veterans Affairs Palo Alto Health Care System, Menlo Park, CA, United States

[2]Division of Primary Care and Population Health, Stanford University School of Medicine, Stanford, CA, United States

[3]Program Evaluation and Resource Center, Veterans Health Administration, Menlo Park, CA, United States

[4]National Center for Post-Traumatic Stress Disorder, Veterans Affairs Palo Alto Health Care System, Menlo Park, CA, United States

Corresponding Author:
Cindie Slightam, MPH
Center for Innovation to Implementation
Veterans Affairs Palo Alto Health Care System
MPD-152
795 Willow Road
Menlo Park, CA, 94025
United States
Phone: 1 650493500 ext 27155
Email: cindie.slightam@va.gov

Abstract

Background: Video-based health care can help address access gaps for patients and is rapidly being offered by health care organizations. However, patients who lack access to technology may be left behind in these initiatives. In 2016, the US Department of Veterans Affairs (VA) began distributing video-enabled tablets to provide video visits to veterans with health care access barriers.

Objective: This study aimed to evaluate veterans' experiences with VA-issued tablets and identify patient characteristics associated with preferences for video visits vs in-person care.

Methods: A baseline survey was sent to the tablet recipients, and a follow-up survey was sent to the respondents 3 to 6 months later. Multivariate logistic regression was used to identify patient characteristics associated with preferences for care, and we examined qualitative themes around care preferences using standard content analysis methods for coding the data collected in the open-ended questions.

Results: Patient-reported access barriers centered around transportation and health-related challenges, outside commitments, and feeling uncomfortable or uneasy at the VA. Satisfaction with the tablet program was high, and in the follow-up survey, approximately two-thirds of tablet recipients preferred care via a tablet (194/604, 32.1%) or expressed that video-based and in-person care were "about the same" (216/604, 35.7%), whereas one-third (192/604, 31.7%) indicated a preference for in-person care. Patients were significantly more likely to report a preference for video visits (vs a preference for in-person visits or rating them "about the same") if they felt uncomfortable in a VA setting, reported a collaborative communication style with their doctor, had a substance use disorder diagnosis, or lived in a place with better broadband coverage. Patients were less likely to report a preference for video visits if they had more chronic conditions. Qualitative analyses identified four themes related to preferences for video-based care: perceived improvements in access to care, perceived differential quality of care, feasibility of obtaining necessary care, and technology-related challenges.

Conclusions: Many recipients of VA-issued tablets report that video care is equivalent to or preferred to in-person care. Results may inform efforts to identify good candidates for virtual care and interventions to support individuals who experience technical challenges.

(J Med Internet Res 2020;22(4):e15682) doi:10.2196/15682

KEYWORDS

veterans; telehealth; telemedicine; eHealth

Introduction

Background

Health care technology is advancing at a rapid pace and with it the opportunities for health care organizations to engage with patients in new ways [1]. This growth includes the expansion of telehealth services and technologies and now encompasses Web-based care solutions that include patient portals with email communication, access to records, technology to track chronic conditions remotely, video visits, and other resources that can improve access to health care [1-3].

The expansion of these Web-based care services includes the use of video visits to treat patients via mobile apps or Web-based applications [2,3]. The 2017 American Hospital Association Annual Survey found that 66% of health systems have adopted video visits, with variation across academic (61%), rural (25%), and community (45%) hospitals [4]. Video visits can help address issues related to health care access [5-7], and evidence shows that they can be effective methods for delivering care, especially for disease management [8-11] and mental health care [12-15]. Patients often report positive feedback and the desire to continue participating in video visits, citing the convenience, cost and time savings, and the benefit of facilitating timely access to care [16-18]. Prior work has identified the benefits of providing mental health care using video visits, including reduced treatment drop out and improvements in patient and provider satisfaction [9,10,13,15]. Mental health patients may also experience increased connectedness and support, improved privacy, and reduced treatment stigma [12], while providers of patients with mental health and chronic conditions may benefit from visual access to the patient's home environment and other contextual information [9,10,12,13]. Some barriers to implementing video visits still exist, including cost and liability, training and support, and providers' willingness to engage [19-22]. Prior research, for instance, has found that some mental health clinicians felt that video visits could disrupt the clinical workflow or could be perceived as impersonal by some patients [22]. Health care systems continue to address these barriers, including recent changes in reimbursement for video visits [23].

With the rapid growth of Web-based care technologies and video visits, there is an interest in understanding patient attitudes toward these services [24,25], including patient experiences with video visits [16,18,26] and drivers of video visit adoption [27-30]. Previous studies have explored patient characteristics (eg, gender, age, education, and rurality) associated with video visit experiences and found that patient perceptions may improve the acceptance of video encounters [26,31]. Prior experience with the internet and technology [27,31] and the presence of health information seeking and socially motivated personality traits are also potentially associated with a greater willingness to participate in video visits [28].

Despite the promise of in-home video care and acceptance by many patients, access is frequently limited to patients who have a suitable device and the capability of accessing the internet. Patients without technology may not have the opportunity to realize any benefits of Web-based care [32]. This disparity may be an especially important issue for veterans, many of whom experience financial challenges that limit access and use despite their interest and willingness to engage in health technology [24,25]. The US Department of Veterans Affairs (VA) has long supported the use of technology to improve access to care and was an early adopter of video teleconferencing [1,33]. However, until recently, video visits were limited to veterans who could travel to community-based outpatient clinics to connect with providers at other facilities. As the VA developed plans to roll out in-home video visits, there were concerns that the technology requirements would generate new disparities for the many veterans with financial challenges, limiting their technology access and use [23,24].

To address this issue, the VA's Offices of Rural Health and Connected Care developed a pilot initiative to distribute video-enabled tablets to veterans who did not have the necessary technology and who had a geographic, clinical, or social barrier to in-person health care access. A previous evaluation of this program suggests that the tablets were largely used for mental health care and that as many as 20% of tablet recipients did not use their tablets [34,35].

Objectives

To inform optimal tablet distribution and technical support, we evaluated patient experiences with the initiative to learn about the characteristics of patients who prefer video visits to in-person appointments. Our primary objectives were to (1) identify the primary health care access barriers among VA-issued tablet recipients, (2) examine patient experiences with tablets and any changes in perceived access to care, and (3) investigate the patient characteristics associated with preferences for video visits vs in-person care.

Methods

Distribution of Tablets Issued by the Department of Veterans Affairs

In 2016, the VA launched a pilot initiative to distribute video-enabled tablets with 4G wireless broadband or Wi-Fi connectivity to veterans with access barriers. Providers could refer patients to the program if they had a clinical need for services but experienced a barrier to accessing VA care in person and if they lacked a device or the necessary internet connectivity to engage in video appointments [34]. Eligible patients received a video-enabled tablet with built-in wireless connectivity and the option to connect peripheral devices (such as a blood pressure cuff or thermometer) if indicated by the provider. Tablets allowed access to VA-supported programs such as the patient portal for managing prescriptions and secure messaging (My Health*e*Vet), mobile apps, and videoconferencing software. Tablets could be used for a wide range of clinical services, and the specific services and scheduling procedures were determined by local facilities. Technical support was available and provided by local VA facility telehealth coordinators and the VA National Telehealth Technology Help Desk.

The implementation of the VA's tablet distribution initiative has been described previously [34]. Briefly, over the 2-year pilot period, 5000 tablets were distributed to 6745 patients at

86 (out of 130) VA health care systems, spanning all 18 geographic regions of the VA's health care network. Approximately half of the tablet recipients lived in rural areas and 75% had a mental illness diagnosis. Tablets were predominantly used for mental health care [35] but also for spinal cord injury care, primary care, palliative care, rehabilitation, and other services [34]. The high rates of tablet use for VA mental health services are consistent with the early adoption of telemental health care in the VA [36] and likely explain the high rates of mental health conditions among tablet recipients compared with the general VA population.

Patient Survey

As part of the program's evaluation, the VA tablet shipment facility (Denver Acquisitions and Logistics Center, DALC) distributed surveys with all tablet shipments between April 1 and September 30, 2017. The paper survey packets included an initial incentive (US $2 or four first class US stamps), and those who completed surveys received US $10 as a thank-you. Participants could opt out at any time by calling or sending in an opt-out card to the evaluation team. Survey recipient information was provided to the evaluation team by the DALC, and the information was merged with administrative data to complete follow-up. Nonresponders received up to two reminder postcards and two additional survey copies as well as up to two follow-up phone calls within 2 months of the survey mail-out. Among the 2120 recipients of the baseline survey, 1321 returned the survey to the evaluation team, a 62% response rate. Similar procedures were used to send a follow-up survey to baseline survey respondents who had valid contact information 3 to 6 months later (n=1298). A total of 36.04% (763/2120) recipients completed both the baseline and follow-up survey and were included in these analyses. The survey-based evaluation of this quality improvement initiative was reviewed and designated as nonresearch by the supporting VA program office, local institutional review board, and VA Research Administration.

Survey Measures

Baseline Characteristics

Patient-reported barriers to accessing health care were assessed using a list of eight potential barriers generated through a literature review and expert recommendations; the 4-point response options ranged from "not a problem" to "big problem" (see survey in Multimedia Appendix 1 [37-40]). The baseline survey also queried patients about characteristics that might influence engagement in video visits, based on a literature review and expert guidance, including the following: demographics, current experience and reliance on VA services (in-person and video), and experience using technology for health-related purposes, via general resources (eg, internet or social media and apps) and VA resources (eg, the VA's patient portal for managing prescriptions and secure messaging, My Health*e*Vet, and telehealth remote monitoring for chronic conditions) [41]. Health literacy was measured using one item ("How confident are you filling out medical forms by yourself?") developed and validated by Chew et al [42]. The collaborative communication style was assessed using questions from a current VA project that is developing a measure to assess veteran health care engagement ("When I see my provider I bring a list of questions or concerns I want to talk about"; "I can make sure my concerns are fully addressed before my appointment ends"; response options ranged from "Not true" to "Mostly true") [43]. VA reliance was measured with questions in which patients indicated where they receive the majority of their primary, mental health, emergency, and hospital care (response options included "Mostly at the VA," "Mostly outside VA," "Half in VA, half outside VA," and "Nowhere"). The survey also assessed demographics including education, household income, and level of economic hardship ("My household can make ends meet") [44].

Outcome Measures

The follow-up survey evaluated overall experience and attitudes with the tablet, including improvements in access, perception of the video appointment, and feedback about technical aspects of the tablet technology. Patients were asked about their preference for future encounters (response options: video, in person, or "about the same") and were provided space to describe the reasons for selecting their preference.

Both baseline and follow-up surveys included measures related to satisfaction with VA care overall, as well as primary care and mental health care evaluated on a 10-point scale (1=very dissatisfied and 10=very satisfied) adapted from the 2013 Customer Satisfaction Index [45].

Additional Data Sources

Administrative data were collected from the VA Corporate Data Warehouse [46] and included age, gender, race, ethnicity, marital status, and the number of chronic condition diagnoses in the year before receiving the tablet (defined using International Statistical Classification of Disease (ICD)-9 and ICD-10 codes) [47]. Distance from primary VA facility, patient's zip code, and rural or urban designation were obtained from the VA's Planning System Support Group. Rurality is defined based on the Rural Urban Commuting Area categories developed by the Department of Agriculture and Health and Human Services' Health Resources and Services Administration [48]. The contractor (Iron Bow Technologies, Inc) of the VA tablet and the wireless internet provider shared information about the percentage of residents with 4G coverage per zip code.

Quantitative Data Analysis

Following data cleaning, the missing data rate was less than 5% for all variables included in the analyses; no adjustments were made for the missing data. To improve the match of the survey respondents (n=764) to the demographics of the entire tablet cohort (n=5981), we performed poststratification survey weighting based on age, rurality, and presence of a mental health condition. To model the covariance among the eight health care access barriers in the survey, we performed exploratory factor analyses, with a predefined factor cutoff of 0.55 [49]. For regressions, Likert scale variables were dichotomized to group together "mostly true and very true," "strongly agree and agree," and "neutral, disagree, and not true" as the reference variable. Other continuous variables were dichotomized or grouped into categories (eg, distance and age). The total number of conditions was a continuous variable. All quantitative analyses were performed using Stata 15.0 (StataCorp, LLC, Texas, USA).

The primary outcome for the analyses was a patient-reported preference for video visits (vs a preference for in-person care or rating in-person and video-based care as equivalent). We used logistic regression models to identify patient characteristics associated with a preference for video visits. Three sensitivity analyses were conducted, the first combined video visits and equivalent ratings (vs in-person) and the following two analyses did not combine the three-item survey response options for care preferences: a multinomial logistic regression (video, in person [base], and "about the same"), and an ordered logistic regression ranking the response options (in person, "about the same", and video).

An additional goal was to describe patient satisfaction with the VA tablet program and to understand changes in satisfaction with VA services by conducting paired *t*-tests among the baseline and follow-up survey respondents.

Qualitative Data Analysis

The follow-up survey included one open-ended question after patients indicated their preference for future care: "Please explain the preference indicated for receiving VA care." We analyzed qualitative responses to identify additional information from the patient perspective that could be used to identify patients who may prefer to use video visits. After removing blank responses, "not applicable" responses, and responses that did not answer the question, we analyzed qualitative data from 638 survey respondents using standard content analysis methods for coding open-ended textual data [50]. An initial codebook was developed using the eight barriers listed in the baseline patient survey and revised with additional codes after an independent review of the first 300 responses by 2 coders (CS and AG). The 2 coders independently coded the responses, and discrepancies were resolved by consensus. The codes were grouped into mutually exclusive themes that highlighted each patient's care preferences. We identified exemplary quotes to demonstrate and highlight examples for each of the themes.

Results

Survey Respondents

A full response breakdown is listed in Figure 1. The survey sample (n=764) was largely representative of the overall tablet cohort (n=5981), but survey respondents were more likely to be young (<65 years) to be non-Hispanic blacks, to live in an urban area, to have a greater number of mental health conditions, and to use the tablet at 6 months compared with the tablet cohort. A comparison of the survey cohort pre-and postweighting is available in Multimedia Appendix 2. Table 1 compares the unweighted characteristics of the baseline and follow-up survey respondents with baseline survey respondents only; the follow-up survey respondents were more likely to be older, to live in a rural area, and to use the tablet at 6 months compared with baseline responders. The mean (SE) age of the survey respondents was 56 (0.20) years, where 81.7% (624/764) of the respondents were men and 54.6% (412/754) lived in a rural location. The common chronic conditions are listed in Table 2, including hypertension (394/764, 51.6%), depression (389/764, 50.9%), and post-traumatic stress disorder (348/764, 45.5%). These condition rates are similar to the overall tablet cohort, and a comparison of the condition counts can be seen in Multimedia Appendix 2 [33]. The patient-reported access barriers are listed in Table 3; the big and small barriers were combined for this analysis owing to small cell sizes in a few barrier categories. The most common barriers were travel time (503/757, 66.4%), travel cost (416/753, 55.2%), health conditions (405/754, 53.7%), bad weather (426/754, 56.5%) and feeling uncomfortable or uneasy at the VA (248/753, 32.9%). The factor analysis created three categories of barriers from the original list: transportation, outside commitments (eg, work, school, and caregiving responsibilities), and a single barrier related to feeling uncomfortable or uneasy at the VA (Table 3).

Figure 1. Survey response flow chart.

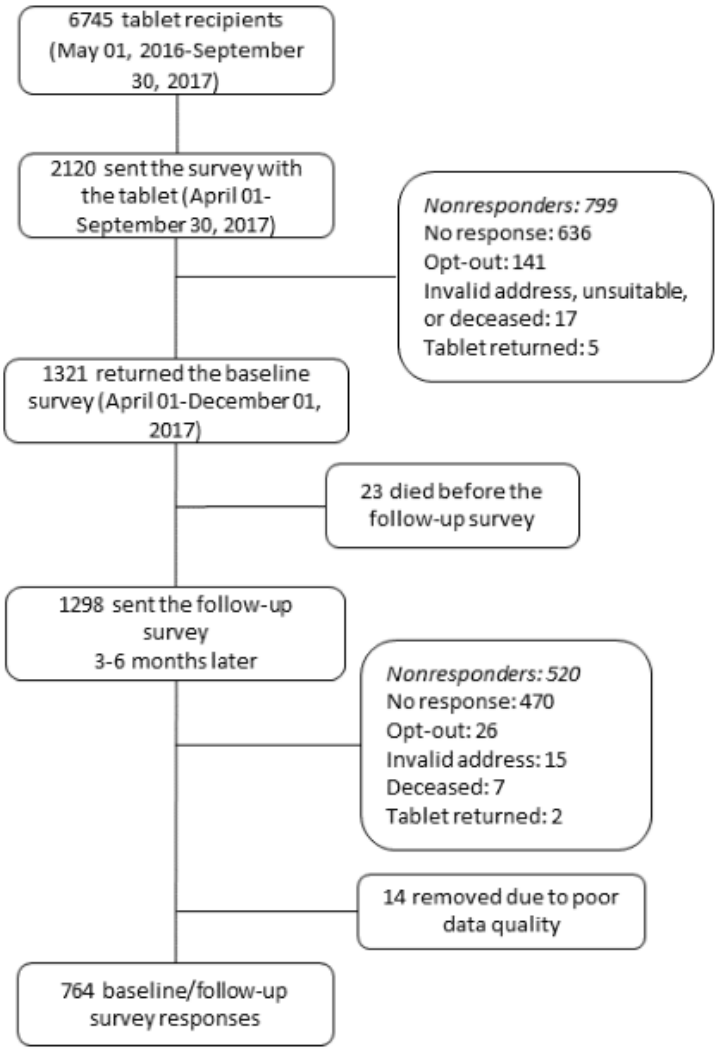


Table 1. Demographics of the survey cohort.

Demographics[a]	Baseline and follow-up survey respondents (n=764)			Baseline survey respondents only (n=530)			*P* value
	n (%)	Mean (SD)	Median (IQR)	n (%)	Mean (SD)	Median (IQR)	
Age (years)[b]		**58.6 (14.5)**			**54.7 (17.1)**		***<.001*** [c]
18-44	137 (18.5)	_[d]	-	167 (32.4)	–	-	–
45-64	306 (41.2)	–	-	173 (33.5)	–	-	–
65-101	299 (40.3)	–	-	176 (34.1)	–	-	–
Male (%)	624 (81.7)	–	-	442 (83.4)	–	-	.42
Marital status[b]							
Married	450 (58.9)	-	-	-	-	-	-
Divorced or never married	287 (37.6)	-	-	-	-	-	-
Widowed	17 (2.2)	-	-	-	-	-	-
Race[b]							**.17**
White or white non-Hispanic	598 (80.4)	-	-	388 (74.6)	-	-	–
Black or African American	90 (12.1)	-	-	83 (16.0)	-	-	–
American Indian, Native Hawaiian, or other	18 (2.4)	-	-	14 (2.7)	-	-	–
Asian	4 (0.5)	-	-	5 (1.0)	-	-	–
Unknown or decline	34 (4.6)	-	-	30 (5.8)	-	-	–
Ethnicity[b]							**.15**
Hispanic or Latino	32 (4.2)	-	-	35 (6.6)	-	-	–
Non-Hispanic/Latino	704 (92.6)	-	-	475 (90.0)	-	-	–
Unknown or decline	24 (3.2)	-	-	18 (3.4)	-	-	–
Rurality[b,e]							***<.01***
Urban	342 (45.4)	-	-	277 (53.2)	-	-	–
Rural or highly rural	412 (54.6)	-	-	244 (46.8)	-	-	–
Education[b,f]							**.75**
Attended or graduated high school or general educational development	227 (29.8)	-	-	157 (30.0)	-	-	–
Some college or 2-year degree	343 (45.1)	-	-	244 (46.7)	-	-	–
4-year college graduate or more	191 (25.1)	-	-	122 (23.3)	-	-	–
Income[b,f] (US $ per year)							**.32**
<25,000	249 (33.6)	-	-	186 (36.0)	-	-	–
25,001-50,000	324 (43.7)	-	-	231 (44.7)	-	-	–
>50,000	169 (22.8)	-	-	100 (19.3)	-	-	–
Economic hardship[b,f]							**.53**
Great difficulty and difficulty	243 (32.1)	-	-	172 (32.7)	-	-	–
Some difficulty	271 (35.8)	-	-	203 (38.6)	-	-	–
Rather easily	149 (19.7)	-	-	97 (18.4)	-	-	–
Easily or very easily	94 (12.4)	-	-	54 (10.3)	-	-	–
Driving distance to primary VA[g] facility (miles)	756 (98.9)	22.9 (22.9)	16 (7-32)	524 (98.9)	21.3 (22.5)	14 (6-30)	.21
Health literacy (out of 4)[f]	755 (98.8)	2.6 (1.3)	3 (2-4)	520 (98.1)	2.5 (1.3)	3 (2-4)	.18

Demographics[a]	Baseline and follow-up survey respondents (n=764)			Baseline survey respondents only (n=530)			*P* value
	n (%)	Mean (SD)	Median (IQR)	n (%)	Mean (SD)	Median (IQR)	
Technology use pretablet[f] (out of 8)	761 (99.6)	2.6 (1.9)	2 (1-4)	529 (99.8)	2.5 (1.9)	2 (1-4)	.29
Percentage of 4G internet coverage per zip code	760 (99.5)	96.5 (10.5)	99.99 (99.1-100)	527 (99.4)	97.4 (9.6)	99.99 (99.4-100)	.16
Number of tablet encounters at 6 months	592 (77.5)	5.7 (5.2)	4 (2-7)	363 (68.5)	4.4 (4.5)	3 (1-6)	*<.001*
Number of mental health tablet encounters at 6 months	383 (50.1)	5.7 (5.0)	4 (2-8)	246 (46.4)	4.7 (4.7)	3 (1-6)	*<.01*

[a]The results of poststratification weighing available in Multimedia Appendix 2 shows changes when weighted on age, rurality, and mental health conditions.

[b]The denominator for the proportions calculated is the total number of individuals with available data

[c]*P* values in italics are statistically significant.

[d]Not applicable.

[e]Rurality provided by the VA's Planning Systems Support Group, which categorizes rural and urban status based on the Rural Urban Commuting Area categories developed by the Department of Agriculture and Health and Human Services' Health Resources and Services Administration [48].

[f]Indicates a survey measure.

[g]VA: Department of Veterans Affairs.

Table 2. Chronic conditions among baseline and follow-up survey respondents.

Chronic conditions	Values		
	n (%)[a]	Mean (SD)	Median (IQR)
Number of chronic conditions		**4.1 (2.3)**	**4 (2-5)**
0-3	297 (38.9)	–	–
4-6	339 (44.4)	–	–
7-14	128 (16.8)	–	–
Conditions[b]			
Acid-related diseases	181 (23.7)	-[c]	-
Alzheimer or Dementia	19 (2.5)	-	-
Arthritis	153 (20)	-	-
Asthma	43 (5.6)	-	-
Cancer	85 (11.1)	-	-
Chronic obstructive pulmonary disease	121 (15.8)	-	-
Diabetes	195 (25.5)	-	-
Heart failure	61 (8.0)	-	-
HIV or AIDS	4 (0.5)	-	-
Headache	97 (12.7)	-	-
Hepatitis C	30 (3.9)	-	-
Hypertension	394 (51.6)	-	-
Ischemic heart disease	114 (14.9)	-	-
Low back pain	279 (36.5)	-	-
Multiple sclerosis	17 (2.2)	-	-
Parkinson disease	25 (3.3)	-	-
Peripheral vascular disease	50 (6.5)	-	-
Prostatic hyperplasia	74 (9.7)	-	-
Renal disease	54 (7.1)	-	-
Spinal cord injury	62 (8.1)	-	-
Stroke	42 (5.5)	-	-
Traumatic brain injury	34 (4.5)	-	-
Mental health conditions			
Any mental health condition	561 (73.1)	-	-
Number of mental health conditions	-	1.3 (1.1)	1 (0-2)
Substance use disorder[d]	**121 (15.9)**	**-**	**-**
Alcohol abuse or dependence	90 (11.8)	-	-
Drug use or dependence	63 (8.2)	-	-
Schizophrenia	21 (2.7)	-	-
Bipolar disorder	50 (6.5)	-	-
Depression	389 (50.9)	-	-
Post-traumatic stress disorder	348 (45.5)	-	-

[a]n (%) represents the unadjusted number of survey respondents; weighted differences are available in Multimedia Appendix 2.

[b]Conditions have been adapted from a list developed by the VA Health Economics Research Center [47].

[c]Not applicable.

[d]Substance use includes alcohol and any drug abuse or dependence diagnosis; individual and combined rates are shown.

Table 3. Self-reported health care access barriers and factor analysis.

Self-reported health care access barriers[a]	Big or small problem, n (%)	Not a problem or don't know, n (%)	Factor 1	Factor 2	Uniqueness	Factor category
Travel time to the VA[b] (n=757)	503 (66.4)	254 (33.6)	0.68	0.24	0.48	Transportation
Difficulty getting transportation to the VA (n=753)	270 (35.9)	483 (64.1)	0.71	0.01	0.49	Transportation
Cost of traveling to the VA (n=753)	416 (55.2)	337 (44.8)	0.60	0.30	0.55	Transportation
Health conditions make it challenging for you to get to the VA (n=754)	405 (53.7)	349 (46.3)	0.67	−0.06	0.55	Transportation
Bad weather conditions (n=754)	426 (56.5)	328 (43.5)	0.58	−0.23	0.61	Transportation
Work or school make it difficult for you to get the health care you need (n=734)	182 (24.8)	552 (75.2)	−0.07	0.84	0.28	Commitments
Family or caregiving responsibilities make it difficult for you to get the health care you need (n=753)	183 (24.3)	570 (75.7)	0.36	0.56	0.56	Commitments
Feeling uncomfortable or uneasy at the VA (n=753)	248 (32.9)	505 (67.1)	0.30	0.45	0.70	Uncomfortable or uneasy

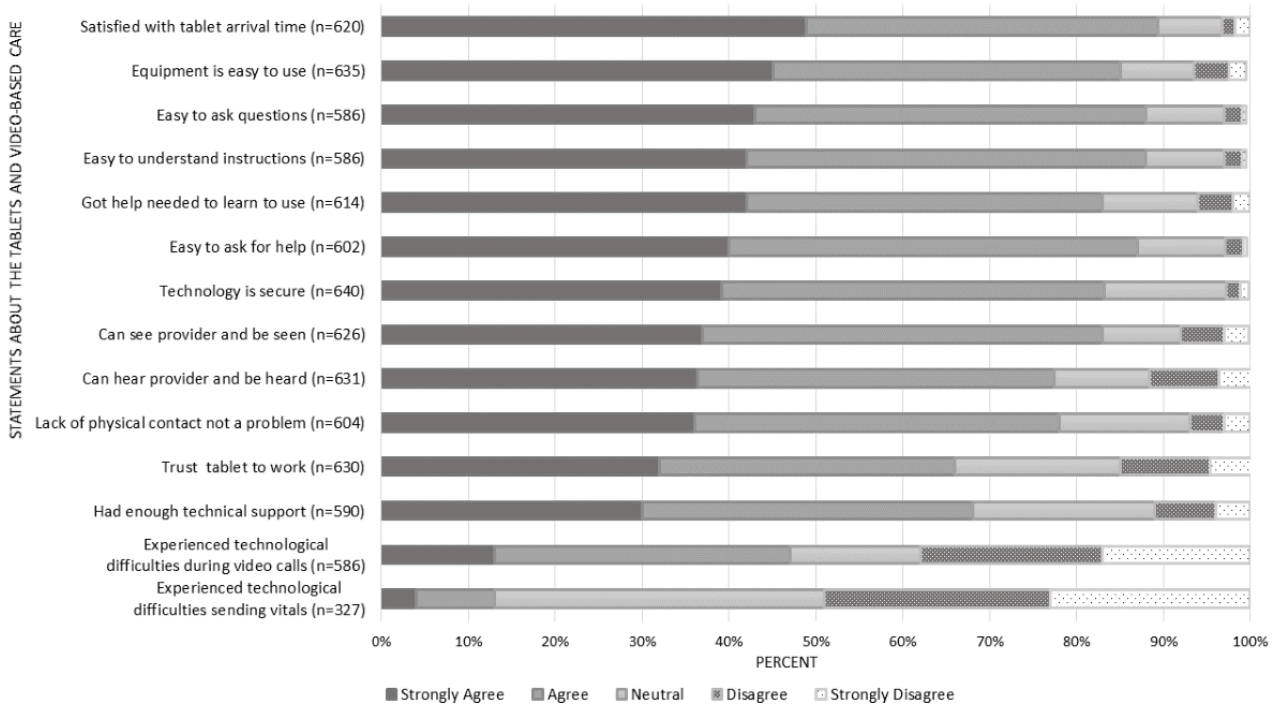
[a]The predefined factor cut-off of .55 was used to group access barriers into categories.

[b]VA: Department of Veterans Affairs.

Satisfaction With Tablets and the Department of Veterans Affairs Health Care

Respondents indicated high levels of satisfaction with their VA health care. Between baseline and follow-up surveys, there were statistically significant increases in patient satisfaction regarding their overall VA care (from mean 7.4, SE 0.10 to mean 7.9, SE 0.08; *P*<.001; n=706), as well as primary care (from mean 7.4, SE 0.1 to mean 7.7, SE 0.1; *P*<.001; n=667) and mental health care (from mean 7.5, SE 0.1 to mean 8.2, SE 0.1; *P*<.001; n=570). In the follow-up survey, 86.0% (523/608) of the respondents indicated that they would recommend video care to others (agree or strongly agree). Satisfaction ratings for the quality of the technology and technical assistance were also high: 86.1% (547/635) agreed or strongly agreed with statements regarding the ease of using the equipment and receiving the help needed to learn the technology (83.5%, 513/614) and that it was easy to ask questions (88.4%, 518/586), ask for help (87.2%, 525/602), and understand instructions (87.4%, 512/586). Satisfaction with the video visits was also high: 84.1% (493/586) agreed or strongly agreed that their provider addressed their concerns during the video visit, 78.1% (472/604) agreed or strongly agreed that the lack of contact was not a problem, and 83.4% (534/640) agreed or strongly agreed that the technology was secure. A breakdown of these results is shown in Figure 2.

Figure 2. Patient-reported experiences using the tablets provided by the Department of Veteran Affairs.



Preferences for Video Visits Versus In-Person Care

In the follow-up survey, 32.1% (194/604) of tablet recipients indicated that they would prefer to conduct their future VA appointments by video, 31.8% (192/604) indicated that they would prefer these visits in person, and 35.7% (216/604) indicated their preference was “about the same”. In the multivariate regression analyses, patients were more likely to report a preference for video-based care (vs in person or “about the same”) if they reported the barrier of feeling uncomfortable or uneasy in the VA setting (adjusted odds ratio [AOR] 2.22, 95% CI 0.88-2.26; *P*<.001), if they indicated mostly/very true to the statement “I can make sure my concerns are fully addressed before my appointment ends” (AOR 1.59, 95% CI 1.02-2.47; *P*=.04), or if they had a substance use disorder in the year before receiving the tablet (AOR 1.91, 95% CI 1.12-3.26; *P*=.02). Patients were less likely to prefer video-based care if they had a greater number of chronic conditions (AOR 0.88, 95% CI 0.78-0.99; *P*=.03). There were trends suggesting that patients were also more likely to prefer video-based care if they had less than college education (AOR 1.52, 95% CI 0.96-2.40; *P*=.08) and if they indicated mostly/very true to the question “When I see my doctor, I bring a list of questions or concerns I want to talk about” (AOR 1.49, 95% CI 0.99-2.26; *P*=.06). The full regression results are listed in Table 4.

Table 4. Characteristics associated with the preference for video visits (N=558).

Preference for video appointments	Adjusted odds ratio[a] (95% CI)	*P* value
VA[b] technology use[c]	1.41 (0.88-2.26)	.16
Other technology use[c]	0.92 (0.60-1.41)	.70
Reliance on VA: medical care[d]	0.93 (0.58-1.48)	.76
Reliance on VA: mental health care[d]	0.70 (0.43-1.14)	.15
Driving distance to primary VA facility (miles; reference: <15 miles)		
16-40	1.24 (0.81-1.89)	.32
>40	1.58 (0.89-2.79)	.12
Access barriers: transportation or travel[e]	1.44 (0.75-2.75)	.27
Access barriers: commitments[e]	1.10 (0.72-1.67)	.66
Access barriers: uncomfortable or uneasy[e]	2.22 (1.43-3.44)	*<.001*
Gender (reference: male)	0.93 (0.56-1.54)	.78
Age (years; reference: 18-44 years)		
45-64	1.01 (0.62-1.66)	.96
65-101	0.72 (0.38-1.37)	.32
Married[a]	1.35 (0.90-2.02)	.14
Verizon coverage (reference: less than 95% coverage)	1.60 (0.84-3.06)	.15
Economic hardship (great and some difficulty making ends meet vs all else)	1.43 (0.94-2.19)	.10
Education (reference: some college or more)		
High school graduate or GED[f]	1.52 (0.96-2.40)	.08
When I see my provider, I bring a list of questions or concerns I want to talk about[d]	1.49 (0.99-2.26)	.06
I can make sure my concerns are fully addressed before my appointment ends[d]	1.59 (1.02-2.47)	*.04*
Health literacy (quite and extremely vs all else)	1.08 (0.71-1.64)	.73
Total number of conditions (continuous)	0.88 (0.78-0.99)	*.03*
Substance use diagnosis[c]	1.91 (1.12-3.26)	*.02*
Depression[c]	0.97 (0.63-1.49)	.89
Post-traumatic stress disorder[c]	1.32 (0.87-2.02)	.19
Schizophrenia or bipolar[c]	1.45 (0.71-2.95)	.31

[a]Multivariate logistic regression comparing characteristics of patients who reported a preference for video visits with those who reported a preference for in-person care or reported a preference for video visits and in-person care "about the same" (reference group). Italicized *P* value indicate significance at $P<.05$.

[b]VA: Department of Veterans Affairs.

[c]Any or yes vs none.

[d]Mostly true and true vs all else.

[e]Big or small problem vs not a problem and don't know.

[f]GED: general educational development.

We conducted several sensitivity analyses to understand the nuances among the survey question response options. The first analysis grouped patient-reported preferences for video visits with rating in-person and video-based care as equivalent and compared this with a preference for in-person care. In this model, patients were more likely to prefer video visits or report that they were equivalent to in-person care if they lived within a driving distance of 16 to 40 miles (AOR 1.65, 95% CI 1.09-2.51; $P=.02$) and were less likely to report these preferences if their age was greater than 65 years (AOR 0.37, 95% CI 0.18-0.72; $P<.01$; Multimedia Appendix 3). Sensitivity analyses that used a 3-category dependent variable (prefer video, prefer

in-person, or "about the same") revealed few differences in predictors of preferences for video visits (Multimedia Appendix 4). An additional predictor for video visits relative to an in-person visit included age older than 65 years and a driving distance of 16 to 40 miles from the VA (relative risk ratios [RRR] 1.66, 95% CI 1.01-2.72; *P*=.04), as well as lower video visit preference among patients age ≥65 years(RRR 0.40, 95% CI 0.18-0.90; *P*=.03). The original model predictors remained significant (feeling uncomfortable in the VA, communicating concerns, number of conditions, and substance use disorder). An ordered logistic regression found similar results; significant predictors for video visits included feeling uncomfortable in the VA, communicating concerns, and substance use disorder, and patients ≥65 years were less likely to prefer video visits (AOR 0.47, 95% CI 0.27-0.82; *P*=.01; Multimedia Appendix 5).

Qualitative analyses revealed four themes underlying patient preferences for video-based vs in-person care: (1) the perceived opportunity to overcome access barriers, (2) the perception of the quality of care provided by video visits versus in-person care, (3) the feasibility of receiving necessary care by video visits versus in-person, and (4) technological issues. Exemplary quotes are presented in Textbox 1.

Textbox 1. Qualitative themes and representative quotes regarding patient care preferences (N=638).

Opportunity to overcome access barriers:

- "Being handicapped & having no transportation, I have to make special arrangements for transportation & pack a lunch for my wife & myself"
- "Sometimes it's nice to have a face to face visit with my psychologist and sometimes it's nice not to have to drive 50 miles one way"
- "I would prefer video because it would expose me less to sick people. This benefits me a lot being a transplant recipient. And my caregiver wouldn't have to take off work to take me to the doctor."

Perceptions of quality of care provided by video visits versus in-person care:

- "I get to see the provider just as if I came to VA in person so to me that is about the same or just as good."
- "The care that I receive is the same in person or by video, excellent"

Feasibility of receiving necessary care by video visits versus in-person:

- "Sometimes doctors need to examine patients. I think it's wonderful for therapy because all I need to do is talk."
- "I prefer a video chat with mental health provider rather than the 2.5-hour commute for a short session. I like to see my medical doctor and orthopedic doctor in person. Video visits are a good way to have questions answered."

Technological issues:

- "Need to give a class on how to use the tablet and make sure the connection & passwords are done right"
- "ept dropping video/calls; it's no longer used because of our location"

Discussion

Principal Findings

This study describes the health care access barriers, experiences, and care preferences for VA patients who received VA-issued tablets for video visits. We identified several patient characteristics that may influence patients' preferences for video visits, including certain diagnoses and number of conditions, comfort in the VA clinic, and communication style.

To our knowledge, this is the first nationally representative survey of VA tablet recipients examining their experiences with VA video visits. The strong satisfaction ratings for tablets and the fact that characteristics such as age, income, health literacy, distance from the closest VA facility, and prior technology use were not significantly associated with tablet preference suggest that engagement in video-based care is possible for many types of patients, including those who are often considered part of the *digital divide* (ie, individuals who are older, have a low income, and have greater health or disability challenges) [51]. Providing tablets to this population may help the VA engage veterans who could otherwise be left behind in technology-focused initiatives. Initiatives that encourage patients to use their own devices are growing rapidly. Distributing devices directly to patients who lack the necessary technology can increase a health system's capacity to reach these patients.

Findings from this survey suggest opportunities to assess potential video-based care patients for specific challenges and preferences, eg, their comfort with technology and desire for in-person encounters. The finding that communication style was associated with preferences for video visits echoes other work that identified patients with certain personality traits (eg, health information seeking and socially motivated) as more comfortable with video visits [28]. The difference in the first sensitivity analyses with the reclassified outcome combining the preference for video visits and "about the same" suggests that driving distance (16-40 miles) and older age (65-101 years) are additional factors that may influence the acceptance of video care. Further research of characteristics or traits may identify additional opportunities to improve patient engagement in video visits.

Although some work has identified that patients may opt to use video care in lieu of in-person primary care [30], the nature of the program we are studying has enabled video visits to be used as an adjunct to in-person care to increase access to providers.

The results indicate the importance of identifying patients who are amenable to using technology for their care and identifying opportunities to improve training for patients and providers who want to conduct video visits. Health care programs could consider patients' chronic conditions and access barriers to identify candidates who may prefer video encounters and review the patients' local broadband capability to ensure connectivity. Previous work has identified that patients sometimes decline telehealth owing to the lack of access or skills needed to engage in video visits [52,53] and that patients with mental health conditions are less likely to have access to the internet and technology [25,54]. Despite the technical challenges that may hinder initial use, once patients participate in a video visit, they often perceive it to be of the same or better quality than in-person care [16,17,27,29]. However, patients acknowledge that video visits cannot fully replace in-person care, particularly when physical examinations are needed for decision making [18]. This study builds on prior work by identifying additional factors such as patient communication style, comfort in the care setting, and health conditions that predict a preference for video visits when the barrier to accessing technology is removed. Telemedicine is more sensitive to patient preferences because it is the mode of health care service delivery [31] rather than a treatment option, and understanding patient preferences will enable health care systems to target this limited resource to ensure it is utilized effectively.

Some limitations of this study include a potential bias introduced by survey nonresponse, despite weighting. Even though the survey respondents were older, they were similar to the population of tablet recipients in most characteristics (Multimedia Appendix 2). Owing to the novel focus of this survey, we included some *de novo* questions, although most of the survey material was derived from validated measures [36-41]. The factor analysis of the access barriers combining big and small groupings could cause us to miss some nuances among some of the factors. However, analyses of the groupings only identified two barriers (uncomfortable or uneasy and travel time) for which the proportion differed significantly between big and small barriers by preference for care. As there were no significant differences noted for the other six factors, and because some factors had relatively small numbers in the *big problem* category, we chose to combine the big and small categories in the analyses. It is to be noted that our evaluation does not include veterans who participate in video visits from their own devices, and providers may have selectively distributed tablets to certain types of patients during this pilot, so our results may not extend to all current or potential veteran patients participating in video visits. Another limitation in interpreting our results is that we cannot attribute changes in satisfaction directly to receipt of the tablet, as surveys were only distributed to tablet recipients and there was no control group with which to compare these outcomes.

Nevertheless, results from this evaluation will inform efforts to improve the reach of this program across participating VA facilities by identifying characteristics associated with preferences for video-based care and the reasons behind these preferences. This information can help VA identify and better engage patients who may be interested in this limited resource as well as address factors that may be limiting tablet use among some populations. This study also clarifies that patients understand that video visits may not be appropriate in all cases, which can be used to inform patient and provider trainings on the appropriateness of offering video visits. For patients who prefer video visits, the VA can utilize the tablets to encourage engagement in programs and services that previously were out of reach owing to access barriers. The role that internet connectivity plays in our findings for patient preference and other research related to health care access underscores the importance of broadband access as a priority in the United States [55]. VA program offices continue to work with broadband carriers on this issue and actively test opportunities to augment this barrier by offering multiple broadband service providers or providing cellular signal boosters to patients in certain areas. The VA health care system serves an older population compared with other US health care systems [56], so the program's success among older veterans (mean age 58.6 years among survey respondents and 54.6 years among all tablet recipients) also provides insights into how best to optimize the use of telehealth and video visits among older adults.

Conclusions

Technology is playing an increasingly important role in enhancing health care access and delivery for patients, especially for those who are geographically isolated or homebound. Although VA has evolved to become both a provider and payer of care, its priority of ensuring access to high-quality care for veterans has not changed. The 2018 Mission Act further expands the role of telemedicine in the VA, including the approved use of video visits in the home and across state lines [57]. Critical issues remain owing to variations in broadband infrastructure that will influence the adoption and use of these technologies. This study provides important information about patient experiences with VA-issued tablets and their preferences for video vs in-person care. The findings may inform the development of assessment and training tools to improve patient targeting and support for tablet recipients as well as opportunities to improve engagement in video visits.

Acknowledgments

The authors would like to thank Leonie Heyworth, MD, and John Peters, MS, from the VA's Office of Connected Care for contributing insights into the VA's nationwide distribution of video-enabled tablets. Liberty Greene, MS, and Pon Su, MS, data analysts at the VA Palo Alto Center for Innovation to Implementation and Health Economics Resource Center, respectively, provided data management support. The views expressed are those of the authors and do not necessarily represent views of the VA.

This work was supported by the VA's Office of Rural Health Enterprise Wide Initiative and the eHealth Partnered Evaluation Initiative, Quality Enhancement Research Initiative (principal investigator: Timothy Hogan, Bedford, Massachusetts).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline and follow-up surveys.
[DOCX File , 720 KB - jmir_v22i4e15682_app1.docx]

Multimedia Appendix 2
Weighted characteristics table.
[DOCX File , 24 KB - jmir_v22i4e15682_app2.docx]

Multimedia Appendix 3
Sensitivity analysis 1.
[DOCX File , 17 KB - jmir_v22i4e15682_app3.docx]

Multimedia Appendix 4
Sensitivity analysis 2: multinomial regression.
[DOCX File , 17 KB - jmir_v22i4e15682_app4.docx]

Multimedia Appendix 5
Sensitivity analysis 3: ordered regression.
[DOCX File , 15 KB - jmir_v22i4e15682_app5.docx]

References

1. Board on Health Care Services, Institute of Medicine, Lustig TA. The Role of Telehealth in an Evolving Health Care Environment: Workshop Summary. Washington, DC: National Academies Press; 2012.
2. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
3. Catalyst N. What Is Telehealth? NEJM Catalyst 2018;4(1).
4. 2017 AHA Annual Survey Information Technology Supplement Database. Chicago, IL. 2017. URL: https://www.ahadata.com/ [accessed 2019-09-25]
5. Fortney JC, Burgess JF, Bosworth HB, Booth BM, Kaboli PJ. A re-conceptualization of access for 21st century healthcare. J Gen Intern Med 2011 Nov;26(Suppl 2):639-647 [FREE Full text] [doi: 10.1007/s11606-011-1806-6] [Medline: 21989616]
6. Der-Martirosian C, Griffin AR, Chu K, Dobalian A. Telehealth at the US Department of Veterans Affairs after Hurricane Sandy. J Telemed Telecare 2019 Jun;25(5):310-317. [doi: 10.1177/1357633X17751005] [Medline: 29384428]
7. Heyworth L. Sharing connections. J Am Med Assoc 2018 Apr 3;319(13):1323-1324. [doi: 10.1001/jama.2018.2717] [Medline: 29614182]
8. Rosen D, McCall JD, Primack BA. Telehealth protocol to prevent readmission among high-risk patients with congestive heart failure. Am J Med 2017 Nov;130(11):1326-1330. [doi: 10.1016/j.amjmed.2017.07.007] [Medline: 28756266]
9. Schreiber SS. Teleneurology for veterans in a major metropolitan area. Telemed J E Health 2018 Sep;24(9):698-701. [doi: 10.1089/tmj.2017.0202] [Medline: 29303679]
10. Sood A, Watts SA, Johnson JK, Hirth S, Aron DC. Telemedicine consultation for patients with diabetes mellitus: a cluster randomised controlled trial. J Telemed Telecare 2018 Jul;24(6):385-391. [doi: 10.1177/1357633X17704346] [Medline: 28406066]
11. Taylor A, Morris G, Pech J, Rechter S, Carati C, Kidd MR. Home telehealth video conferencing: perceptions and performance. JMIR Mhealth Uhealth 2015 Sep 17;3(3):e90 [FREE Full text] [doi: 10.2196/mhealth.4666] [Medline: 26381104]
12. Pruitt LD, Luxton DD, Shore P. Additional clinical benefits of home-based telemental health treatments. Prof Psychol 2014;45(5):340-346. [doi: 10.1037/a0035461]
13. 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]
14. Grubbs KM, Fortney JC, Dean T, Williams JS, Godleski L. A comparison of mental health diagnoses treated via interactive video and face to face in The Veterans Healthcare Administration. Telemed J E Health 2015 Jul;21(7):564-566. [doi: 10.1089/tmj.2014.0152] [Medline: 25799233]

15. Veazie S, Bourne D, Peterson K, Anderson J. Evidence Brief: Video Telehealth for Primary Care and Mental Health Services. Washington, DC: Department of Veterans Affairs (US); 2019.
16. Whealin JM, King L, Shore P, Spira JL. Diverse veterans' pre- and post-intervention perceptions of home telemental health for posttraumatic stress disorder delivered via tablet. Int J Psychiatry Med 2017 Jan;52(1):3-20. [doi: 10.1177/0091217417703291] [Medline: 28486881]
17. Becevic M, Boren S, Mutrux R, Shah Z, Banerjee S. User satisfaction with telehealth: study of patients, providers, and coordinators. Health Care Manag (Frederick) 2015;34(4):337-349. [doi: 10.1097/HCM.0000000000000081] [Medline: 26506296]
18. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient perceptions of telehealth primary care video visits. Ann Fam Med 2017 May;15(3):225-229 [FREE Full text] [doi: 10.1370/afm.2095] [Medline: 28483887]
19. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
20. Lambert D, Gale J, Hartley D, Croll Z, Hansen A. Understanding the business case for telemental health in rural communities. J Behav Health Serv Res 2016 Jul;43(3):366-379. [doi: 10.1007/s11414-015-9490-7] [Medline: 26695645]
21. Perry K, Gold S, Shearer EM. Identifying and addressing mental health providers' perceived barriers to clinical video telehealth utilization. J Clin Psychol 2019 Mar 11:-. [doi: 10.1002/jclp.22770] [Medline: 30859573]
22. Brooks E, Turvey C, Augusterfer EF. Provider barriers to telemental health: obstacles overcome, obstacles remaining. Telemed J E Health 2013 Jun;19(6):433-437. [doi: 10.1089/tmj.2013.0068] [Medline: 23590176]
23. Center for Medicare and Medicaid Services, Department of Health and Human Services. Federal Register. Baltimore, MD: Center for Medicare and Medicaid Services; 2018. Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule and Other Revisions to Part B for CY 2019; Medicare Shared Savings Program Requirements; Quality Payment Program; Medicaid Promoting Interoperability Program; Quality Payment Program-Extreme and Uncontrollable Circumstance Policy for the 2019 MIPS Payment Year; Provisions From the Medicare Shared Savings Program-Accountable Care Organizations-Pathways to Success; and Expanding the Use of Telehealth Services for the Treatment of Opioid Use Disorder Under the Substance Use-Disorder Prevention That Promotes Opioid Recovery and Treatment (SUPPORT) for Patients and Communities Act URL: https://www.federalregister.gov/documents/2018/11/23/2018-24170/medicare-program-revisions-to-payment-policies-under-the-physician-fee-schedule-and-other-revisions [accessed 2020-02-04]
24. Miller CJ, McInnes DK, Stolzmann K, Bauer MS. Interest in use of technology for healthcare among veterans receiving treatment for mental health. Telemed J E Health 2016 Oct;22(10):847-854. [doi: 10.1089/tmj.2015.0190] [Medline: 26982279]
25. Klee A, Stacy M, Rosenheck R, Harkness L, Tsai J. Interest in technology-based therapies hampered by access: a survey of veterans with serious mental illnesses. Psychiatr Rehabil J 2016 Jun;39(2):173-179. [doi: 10.1037/prj0000180] [Medline: 26985680]
26. Morland LA, Wells SY, Glassman LH, Grubbs KM, Mackintosh M, Golshan S, et al. What Do Veterans Want? Understanding Veterans' Preferences for PTSD Treatment Delivery. Mil Med 2019 Dec 1;184(11-12):686-692. [doi: 10.1093/milmed/usz035] [Medline: 30839067]
27. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]
28. 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 4;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]
29. Gardner MR, Jenkins SM, O'Neil DA, Wood DL, Spurrier BR, Pruthi S. Perceptions of video-based appointments from the patient's home: a patient survey. Telemed J E Health 2015 Apr;21(4):281-285 [FREE Full text] [doi: 10.1089/tmj.2014.0037] [Medline: 25166260]
30. Liaw WR, Jetty A, Coffman M, Petterson S, Moore MA, Sridhar G, et al. Disconnected: a survey of users and nonusers of telehealth and their use of primary care. J Am Med Inform Assoc 2019 May 1;26(5):420-428. [doi: 10.1093/jamia/ocy182] [Medline: 30865777]
31. Call VR, Erickson LD, Dailey NK, Hicken BL, Rupper R, Yorgason JB, et al. Attitudes toward telemedicine in urban, rural, and highly rural communities. Telemed J E Health 2015 Aug;21(8):644-651. [doi: 10.1089/tmj.2014.0125] [Medline: 25839334]
32. 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 Oct;37(10):437-445. [doi: 10.1016/s1553-7250(11)37055-9] [Medline: 22013816]
33. Adams SV, Mader MJ, Bollinger MJ, Wong ES, Hudson TJ, Littman AJ. Utilization of interactive clinical video telemedicine by rural and urban veterans in the veterans health administration health care system. J Rural Health 2019 Jun;35(3):308-318. [doi: 10.1111/jrh.12343] [Medline: 30600557]

34. Zulman D, Wong E, Slightam C, Gregory A, Jacobs JC, Kimerling R, et al. Making connections: nationwide implementation of video telehealth tablets to address access barriers in veterans. JAMIA-Open 2019;2(3):323-329. [doi: 10.1093/jamiaopen/ooz024]
35. Jacobs JC, Blonigen DM, Kimerling R, Slightam C, Gregory AJ, Gurmessa T, et al. Increasing mental health care access, continuity, and efficiency for veterans through telehealth with video tablets. Psychiatr Serv 2019 Nov 1;70(11):976-982. [doi: 10.1176/appi.ps.201900104] [Medline: 31378193]
36. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. JAMA Psychiatry 2015 Jan;72(1):58-67. [doi: 10.1001/jamapsychiatry.2014.1575] [Medline: 25409287]
37. Shippee ND, Shippee TP, Hess EP, Beebe TJ. An observational study of emergency department utilization among enrollees of Minnesota Health Care Programs: financial and non-financial barriers have different associations. BMC Health Serv Res 2014 Feb 8;14:62 [FREE Full text] [doi: 10.1186/1472-6963-14-62] [Medline: 24507761]
38. Clement S, Brohan E, Jeffery D, Henderson C, Hatch SL, Thornicroft G. Development and psychometric properties the Barriers to Access to Care Evaluation scale (BACE) related to people with mental ill health. BMC Psychiatry 2012 Jun 20;12:36 [FREE Full text] [doi: 10.1186/1471-244X-12-36] [Medline: 22546012]
39. Center for Studying Health System Change. Health Tracking Household Survey, 2010 [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]; 2012.
40. Washington DL, Bean-Mayberry B, Riopelle D, Yano EM. Access to care for women veterans: delayed healthcare and unmet need. J Gen Intern Med 2011 Nov;26(Suppl 2):655-661 [FREE Full text] [doi: 10.1007/s11606-011-1772-z] [Medline: 21989618]
41. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How can eHealth technology address challenges related to multimorbidity? Perspectives from patients with multiple chronic conditions. J Gen Intern Med 2015 Aug;30(8):1063-1070 [FREE Full text] [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
42. 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]
43. Kimerling RE, Schalet BD, Reise SP, Zulman DM, Lewis ET. Development of a Veteran-Centered Measure of Healthcare Engagement. In: Proceedings of the International Society for Quality of Life Research. 2019 Presented at: ISOQOL'19; October 20-23, 2019; San Diego, CA.
44. United Nations Economic Commission for Europe, United Nations Population Fund. Generations and Gender Programme: Survey Instruments. Switzerland: UNECE Information Unit; 2005.
45. CFI Group. Veterans Affairs. 2014. American Customer Satisfaction Index: Veterans Health Administration. Veterans Affairs - 2013 Customer Satisfaction Outpatient Survey URL: https://www.va.gov/health/docs/VA2013OutpatientACSI.pdf [accessed 2019-06-01]
46. Federal Register. Washington, DC: US Department of Veterans Affairs; 2014. Privacy Act of 1974: VHA Corporate Data Warehouse (172VA10P2) – VA. 79 FR 4377 URL: https://www.federalregister.gov/documents/2014/01/27/2014-01497/privacy-act-of-1974 [accessed 2020-02-04]
47. Yoon J, Chow A. Comparing chronic condition rates using ICD-9 and ICD-10 in VA patients FY2014-2016. BMC Health Serv Res 2017 Aug 17;17(1):572 [FREE Full text] [doi: 10.1186/s12913-017-2504-9] [Medline: 28818082]
48. VIReC. Hines, IL: US Department of Veterans Affairs Information Resource Center VIReC Research User Guides URL: https://www.virec.research.va.gov/Resources/RUGs.asp [accessed 2020-02-04]
49. Tabachnick BG, Fidell LS. Using Multivariate Statistics. Fifth Edition. Boston, MA: Pearson; 2007.
50. Saldana J. The Coding Manual for Qualitative Researchers. Thousand Oaks, CA: SAGE Publications; 2012.
51. Smith A. Pew Research Center. 2014 Apr 3. Older Adults and Technology Use URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [accessed 2020-02-04]
52. Luger TM, Hogan TP, Richardson LM, Cioffari-Bailiff L, Harvey K, Houston TK. Older veteran digital disparities: examining the potential for solutions within social networks. J Med Internet Res 2016 Nov 23;18(11):e296 [FREE Full text] [doi: 10.2196/jmir.6385] [Medline: 27881361]
53. Foster A, Horspool KA, Edwards L, Thomas CL, Salisbury C, Montgomery AA, et al. Who does not participate in telehealth trials and why? A cross-sectional survey. Trials 2015 Jun 5;16:258 [FREE Full text] [doi: 10.1186/s13063-015-0773-3] [Medline: 26044763]
54. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do we still have a digital divide in mental health? A five-year survey follow-up. J Med Internet Res 2016 Nov 22;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
55. Federal Communications Commission. Health IT Answers. Washington, DC; 2017 Jun. FCC's Connect2Health Task Force Announces Data Update for Broadband Health Mapping Platform URL: https://docs.fcc.gov/public/attachments/DOC-344778A1.pdf [accessed 2020-04-04]
56. Howden LM, Meyer JA. United States Census Bureau. Washington, DC: United States Census Bureau; 2011 May. Age and Sex Composition: 2010 Census Briefs URL: https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf [accessed 2020-02-04]

57. Library of Congress. 2018. S.2372 - VA MISSION Act of 2018 URL: https://www.congress.gov/bill/115th-congress/senate-bill/2372 [accessed 2019-05-31]

Abbreviations

AOR: adjusted odds ratio
DALC: Denver Acquisitions and Logistics Center
ICD: International Statistical Classification of Disease
RRR: relative risk ratios
VA: Department of Veterans Affairs

Edited by G Eysenbach; submitted 29.07.19; peer-reviewed by K Grubbs, C Miller, C El Morr, J Hensel, T Jamieson, R Appireddy; comments to author 11.09.19; revised version received 05.11.19; accepted 19.12.19; published 15.04.20.

Please cite as:
Slightam C, Gregory AJ, Hu J, Jacobs J, Gurmessa T, Kimerling R, Blonigen D, Zulman DM
Patient Perceptions of Video Visits Using Veterans Affairs Telehealth Tablets: Survey Study
J Med Internet Res 2020;22(4):e15682
URL: https://www.jmir.org/2020/4/e15682
doi:10.2196/15682
PMID:32293573

Original Paper

Patients' Perspective on Mental Health Specialist Video Consultations in Primary Care: Qualitative Preimplementation Study of Anticipated Benefits and Barriers

Caroline Bleyel[1], MD; Mariell Hoffmann[1], BA, MA; Michel Wensing[2], PhD; Mechthild Hartmann[1], MA; Hans-Christoph Friederich[1], MD, PhD; Markus W Haun[1], BSc, MSc, MD

[1]Department of General Internal Medicine and Psychosomatics, Heidelberg University, Heidelberg, Germany

[2]Department of General Practice and Health Services Research, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Markus W Haun, BSc, MSc, MD
Department of General Internal Medicine and Psychosomatics
Heidelberg University
Im Neuenheimer Feld 410
Heidelberg, D-69120
Germany
Phone: 49 622156 ext 38396
Fax: 49 6221567505
Email: markus.haun@med.uni-heidelberg.de

Abstract

Background: Due to limited access to specialist services, most patients with common mental disorders (depression or anxiety, or both) usually receive treatment in primary care. More recently, innovative technology-based care models (eg, video consultations) have been proposed to facilitate access to specialist services. Against this background, the PROVIDE (Improving Cross-Sectoral Collaboration Between Primary and Psychosocial Care: An Implementation Study on Video Consultations) project aims to improve the provision of psychosocial care through implementing video consultations integrated into routine primary care.

Objective: From the patients' perspective, this qualitative preimplementation study explored (1) anticipated benefits from and (2) barriers to implementing mental health specialist video consultations embedded in primary care services and (3) prerequisites for interacting with therapists via video consultations.

Methods: Using a purposive (ie, stratified) sampling strategy, we recruited 13 patients from primary care practices and a tertiary care hospital (psychosomatic outpatient clinic) for one-off semistructured interviews. In a computer-assisted thematic analysis, we inductively (bottom-up) derived key themes concerning the practicability of mental health specialist video consultations. To validate our results, we discussed our findings with the interviewees as part of a systematic member checking.

Results: Overall, we derived 3 key themes and 10 subthemes. Participants identified specific benefits in 2 areas: the accessibility of mental health specialist care (shorter waiting times: 11/13, 85%; lower threshold for seeking specialist mental health care: 6/13, 46%; shorter travel distances: 3/13, 23%); and the environment in primary care (familiar travel modalities, premises, and employees: 5/13, 38%). The main barriers to the implementation of mental health video consultations from the patients' perspective were the lack of face-to-face contact (13/13, 100%) and technical challenges (12/13, 92%). Notably, participants' prerequisites for interacting with therapists (12/13, 92%) did not seem to differ much from those concerning face-to-face contacts.

Conclusions: Mental health service users mostly welcomed mental health specialist video consultations in primary care. Taking a pragmatic stance, service users, who are often frustrated about uncoordinated care, particularly valued the embedment of the consultations in the familiar environment of the primary care practice. With respect to interventional studies and implementation, our findings underscore the need to minimize technical disruptions during video consultations and to ensure optimal resemblance to face-to-face settings (eg, by training therapists in consistently reacting to nonverbal cues).

Trial Registration: German Clinical Trials Register DRKS00012487; https://tinyurl.com/uhg2one

(J Med Internet Res 2020;22(4):e17330) doi:10.2196/17330

KEYWORDS

telemedicine; remote consultation; implementation; primary health care; mental health services; thematic analysis; integrated behavioral health; health services research

Introduction

Primary Care Mental Health

In most countries, the prevalence of mental health conditions (eg, depression and anxiety) remains high, and the global burden of mental illness accounts for 32.4% of years lived with disability and 13.0% of disability-adjusted life-years. However, access to mental health specialists is precarious, particularly in remote and rural areas [1,2]. Patients with mental health conditions struggle with limited availability of psychosocial services and are exclusively treated in the primary care setting, which has been coined the "de facto mental health care outpatient system" [3] for the United States and other countries [3,4]. To address this care gap, clinicians, researchers, and policy makers have proposed innovative technology-based care models [6-8]. Specifically, remote consultations between patients in the primary care practice and mental health specialists by using videoconferencing have been proposed [9-11]. Given the primary care physicians' decisive role in managing mental health conditions and the fact that most patients are treated in primary care, directly embedding mental health specialist video consultations into primary care services has been suggested [6-8,12-14]. In this regard, several comprehensive reviews indicated fairly robust evidence that interactive video consultations are an effective means to link patients in the primary care practice with specialists [6,7,15-18]. Thus, establishing a sound therapeutic alliance is generally considered to be a key driver for telepsychiatry and of upmost importance for mental health staff, including nurses [19-22]. Nevertheless, little is known about the patients' perspective on the benefits of and barriers to mental health specialist video consultations in primary care.

The PROVIDE Project: Fostering Cross-Sectoral Collaborations Between Primary Care Physicians and Mental Health Care Providers Through Video Consultations

This study was the first (preimplementation) phase of a larger implementation project on fostering cross-sectoral collaborations between primary care physicians and mental health care providers through video consultations (Improving Cross-Sectoral Collaboration Between Primary and Psychosocial Care: An Implementation Study on Video Consultations [PROVIDE] [23]). While we have reported results from the preimplementation phase for the primary care physicians elsewhere [24], in this paper we present findings on how to optimally tailor a video-based integrated mental health care model from the patients' perspective. The PROVIDE service delivery model follows the integrated care approach, which virtually colocates primary care teams and specialists and has been proposed by Hilty and colleagues [7,25]. Leveraging the expertise at a distance, specialists located at their office or private practice (or a suitable, designated room at home) provide high-quality video consultations to patients presenting with depression or anxiety, or both, in the primary care practice. The intervention itself is described elsewhere in detail [26]. Recently, the PROVIDE service delivery model has been applied in a feasibility study (German Clinical Trials Register, registration no. DRKS00015812).

Video Consultations in Mental Health From the Patients' Perspective

There is promising evidence for the effectiveness of videoconferencing both in secondary mental health care [7,10,27-29] and in primary care [15,30]. However, with respect to crucial outcomes applied in implementation science [31], studies on the applicability of and prerequisites for rolling out mental health specialist video consultations in primary care services from the perspective of the actual stakeholders involved are rare. Indeed, it has been argued that it is necessary to involve potential users and target groups at the earliest stage possible to foster the adoption and uptake of new service delivery models [32]. Concerning the patients' perspective, a few studies primarily evaluated patients' satisfaction with video consultations in mental health services [33-36]. However, only 2 qualitative studies have investigated the patients' perspective on mental health specialist video consultations embedded into primary care services [37,38]. First, Simpson [37] focused on patients' satisfaction after a 1-year course of video consultation therapy. Second, Swinton and colleagues [38] explored the acceptability of mental health specialist video consultations prior to their implementation by investigating alliance aspects both between patients and therapists, and between therapists and primary care physicians. In sum, while studies on patients' perspective on mental health specialist video consultations in primary care focused on the patients' acceptance of and satisfaction with such services, they did not provide information on why patients engaged in video consultations, what kind of barriers they experienced, and whether patients had specific expectations of therapists who engage in such consultations.

Rationale and Objectives

The purpose of this exploratory qualitative study was to take the patients' perspective and investigate (1) anticipated benefits from and (2) barriers to implementing mental health specialist video consultations embedded in primary care services and (3) prerequisites for interacting with therapists via video consultations. Specifically, when considering anticipated barriers and benefits, we focused on both tangible (eg, journey to the primary care practice or handling technical devices) and more general aspects (eg, impact on the availability of mental health services) of the proposed mental health specialist video consultation model. With the therapeutic relationship being the main predictor for treatment outcomes in mental health care [39], we also investigated prerequisites for interaction between patients and therapists in this setting. Our findings provide in-depth information for the prospective large-scale implementation of practical and sustainable mental health

specialist video consultations tailored to the needs of patients attending primary care practices.

Methods

Study Design and Conceptual Framework

In a naturalistic preimplementation qualitative explorative study, we conducted one-off semistructured telephone interviews to assess the patients' perspective on implementing mental health specialist video consultations in primary care. We took a critical realist position when designing the study, analyzing the data, and interpreting the results [40]. Subsequently, we applied thematic analysis to identify shared meanings concerning anticipated benefits of and barriers to mental health specialist video consultations, along with prerequisites for interacting with therapists via video consultations [41]. All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. This study was approved by the Ethics Committee of the Medical Faculty at Heidelberg University (no. S-197/2017) and preregistered with the German Clinical Trials Register (registration no. DRKS00012487).

Participants and Recruitment

We applied stratified sampling following a purposive strategy. Given the resources at our disposal, we applied sex and age as the basis for the sample stratification. Specifically, accounting for sex distributions of patient populations in mental health care, we aimed for a recruitment ratio of women to men of 2:1. To ensure appropriate representation of older people, we aimed for a 1:1 recruitment ratio of people aged 18 to 49 years to people aged 50 years or older. We consecutively recruited patients at various primary care practices in Heidelberg, Germany, and our psychosomatic outpatient clinic at Heidelberg University Hospital and asked them to participate in the study. We personally approached patients before or after their scheduled consultations. Inclusion criteria were age 18 years or older, German language proficiency, and informed consent for study participation. Exclusion criteria were poor German language proficiency, lack of informed consent, and cognitive incapability to respond to interview questions. We offered a nonadvertised individual monetary compensation of €30 (about US $35) for each interview.

Data Collection

To capture participants' descriptions systematically, we designed a semistructured question interview guide (Multimedia Appendix 1) that was reviewed after the first 3 interviews. Since the interview guide proved to be coherent and comprehensive, no changes were made. During March and April 2018, the interviews were conducted by the second author (MH, a PhD student focusing on video consultations in primary care, master's degree in sociology, expertise in qualitative research), who had no contact or relationship with any participant prior to the study. We obtained written informed consent from all individual participants prior to study enrollment and guaranteed the absence of nonparticipants during the interview. All interviews were audio recorded. In the interviews, we first asked participants about their experience with the health care system, particularly about seeking help for mental health conditions. Second, we verbally described the mental health specialist video consultation model, including that the patient would be located in the primary care practice while the mental health specialist would consult from her or his office or private practice or a suitable, designated room at home (Multimedia Appendix 2). Third, we discussed benefits, barriers, and prerequisites for interacting with therapists via video consultations. Fourth, we collected sociodemographic data from the participants and supplemented the audio data with field notes produced during the interviews. In between the interviews, MH discussed the progress of sampling and data collection with the first author (CB, MD, resident in psychosomatic medicine) and the last author (MWH, MD, internal medicine specialist and master's degree in clinical psychology), for example, with respect to the sample composition and the level of data saturation. We did not repeat any interviews.

Data Analysis

After the audio recordings were transcribed verbatim by a professional transcription service (Transkripto, Rotterdam, Netherlands), we anonymized the data. We did not return the transcripts to the participants for comments. Two coders (CB and MH) independently conducted a computer-assisted thematic analysis in MAXQDA 12 (VERBI GmbH) of 3 individual interviews. At this stage, the coders developed the code system following an inductive (bottom-up) approach by paraphrasing, generalizing, and abstracting the original data [41]. We subsequently applied this code system to analyze the remaining 9 transcripts top-down. To ensure that all key aspects were represented, we discussed newly derived themes and modified the codes when necessary. Theme saturation was reached when the analyzed data did not provide any new themes or meaning of themes, that is, when the inductively developed themes represented and covered all the data [42]. Both coders compared their analyses and resolved disagreements in a final code system. Multimedia Appendix 3 summarizes the key themes, including definitions and supporting quotes. We followed the consolidated criteria for reporting qualitative research (COREQ) guidelines for reporting qualitative study results [43]. All materials were translated from German to English for this paper by CB and MWH.

Member Checking

To support the credibility of our analyses and review data saturation, we conducted member checking with the participants [44]. To this end, the participants received an anonymized written summary of the interview findings and evaluated to what extent these findings reflected their statements. We contacted all participants via telephone to receive their feedback on the summary that we afterward accounted for when discussing the results.

Results

Sample

We approached 37 people (34 from outpatient clinics, 3 from primary care practices) with an initial interest in scheduling an individual telephone interview. We could not reach 6 patients at the indicated phone number or email address. A total of 18 patients declined to be interviewed after having received further study details, most frequently due to a lack of interest (n=9) and lack of time (n=6). In sum, we conducted individual telephone interviews (range 29-66 minutes, mean 38 minutes) with 13 participants consenting to study participation.

Table 1 [45] presents the sociodemographic characteristics of the 13 participants. Table 2 depicts the cross-tabulated table for the sample stratification by sex and age.

Table 1. Sample description (N=13).

Characteristic	Values
Age (years)	
Mean (SD)	48.7 (17.0)
Range	21-77
Female sex, n (%)	8 (62)
Education (years), n (%)	
>9	5 (39)
≤9	8 (62)
Married, n (%)	8 (62)
Degree of urbanization of the place of residence[a], n (%)	
Cities (densely populated areas)	2 (15)
Towns and suburbs (intermediate-density areas)	9 (69)
Rural areas (thinly populated areas)	2 (15)
Duration of interviews (minutes), mean (range)	39 (29-66)

[a]Stratified according to the degree of urbanization classification, 2018 version, of the European Commission [45].

Table 2. Cross-tabulated table for the sample stratification by sex and age.

Age range (years)	Female	Male
18-49	5	2
≥50	3	3

Themes and Subthemes

We identified 3 key themes and 10 subthemes. Within the key themes, we stratified findings along the corresponding subthemes. In the presentation of the results below, we first describe participants' anticipated benefits from the integration of mental health specialist video consultations into primary care services. Second, we discuss the barriers anticipated by the participants. Third, we discuss related prerequisites for interacting with providers via mental health specialist video consultations in greater detail. Finally, we review our findings considering the member checking results.

Participants' Anticipated Benefits

Participants identified several benefits in 2 specific areas: in the accessibility of mental health specialist care (shorter waiting times: 11/13, 85%; lower threshold for seeking specialist mental health care: 6/13, 46%; and shorter travel distances: 3/13, 23%); and in the primary care environment (familiar travel modalities, premises, and employees: 5/13, 38%). These benefits with corresponding quotes are presented in the following.

Shorter Waiting Times

Participants (11/13, 85%) expected more rapid access to specialist care for mental health conditions due to the coordinating role of the referring primary care physician. They also envisioned a more seamless treatment trajectory, particularly that primary care physicians and mental health specialists would work more closely with each other. Moreover, patients anticipated better treatment outcomes than they would have had with the current practice in connection with shorter waiting times:

> *They [the mental health specialist video consultations] will save time and offer immediate aid. If you are affected and you suffer from depressions and you don't know what to do and they give you a list [of therapists] and you call and don't get any answer, the depression gets worse. So, this solution [the mental health specialist video consultation] is phenomenal, in my opinion...So, you go to your primary care physician and shortly at least you have*

someone, who shows you the way. [Participant 11, female, in her 50s]

Shorter Travel Distances

Some participants (3/13, 23%) expected a shorter travel distance as a further aspect of an improved access to specialist mental health care. They argued that patients could save additional time and effort, as in most cases the primary care practice would be in their hometown or nearby. Thus, long-distance travel or additional driving to a mental health specialist who maintains a practice in a larger city would not be necessary:

And yes, as already mentioned, you are in a familiar environment and you know how to get there. And in case of an emergency, you can be brought by someone, or picked up, because for us [in the countryside], this is easier than driving into the city or being brought there by someone. [Participant 08, female, in her 30s]

Lower Threshold for Seeking Specialist Mental Health Care

Besides the abovementioned more tangible benefits of improved accessibility of mental health specialist care, some participants (6/13, 46%) described a psychological effect of the model:

I think there are many people who may benefit from this [mental health specialist video consultations]. As mentioned, this—the problem with mental health conditions—is steadily increasing. Maybe for some people, those 4 or 5 consultations are already enough...And maybe people who would not start actual therapy would make use of it...I think for those people, it would be somewhat facilitated. [Participant 06, female, in her 30s]

The participants highlighted that the mental health specialist video consultations could decrease the threshold for accessing mental health care and therefore help to overcome the stigma of seeking professional psychosocial support for persons (1) with mild mental health conditions, (2) with no experiences with psychosocial care, and (3) unwilling to attend a face-to-face consultation with a mental health specialist.

Familiar Primary Care Environment

Several participants (5/13, 38%) appreciated that the proposed model would leverage the familiarity of the primary care environment. When engaging in mental health specialist consultations, patients would already know the practice staff, the premises of the practice, and the travel modalities to the practice:

The fact that it [the mental health specialist video consultation] takes place in the primary care practice, I think, this is a familiar environment for the patient. Therefore, one already knows the way to get there; you know what it looks like in there. For some people, it is probably an obstacle to go somewhere unfamiliar. Here [when engaging in mental health specialist video consultations] you already know the practitioner; you have already known the medical assistant for a longer time. [Participant 08, female, in her 30s]

The staff in the primary care practice could encourage the patient to engage in the video consultations and assist him or her with handling the videoconferencing platform. The participants viewed the primary care physician's involvement in the model as crucial, since the participants assigned him or her a key role in (1) managing their health problems (eg, knowing the patient and their medical history and social circumstances in detail) and (2) referring them to medical specialists.

Anticipated Barriers

Despite their generally positive attitude toward video consultations, participants also anticipated barriers, some of which were specifically tied to the proposed model itself (lack of face-to-face contact: 13/13, 100%; technical challenges, such as an unstable internet connection: 12/13, 92%; organizational challenges: 3/13, 23%), while other barriers were related to the acceptance of psychosocial services as such the stigma of seeking mental health care (7/13, 54%).

Lack of Face-to-Face Contact

All participants (13/13, 100%) mentioned the lack of personal face-to-face contact with the mental health specialist as a major concern. Participants imagined "talking to a screen" as impersonal and uncomfortable, especially when discussing sensitive issues:

It [therapy] often becomes painful, doesn't it? Therefore, depression comes with emotions. How can we deal with emotions while sitting in front of a computer? [Participant 11, female, in her 50s]

Physical contact, for example, through common gestures such as the mental health specialist handing a handkerchief to the patient, would be impossible. Referring to their own personal experiences with using Skype, FaceTime, etc, some participants argued that the mental health specialist would miss several aspects of the body language, even with facial expressions and gestures being visible:

Well, I experience it myself when conducting Skype calls with familiar people. Although you speak to someone you know well, it still feels different. The distance between the person and you is different from personal encounters. In addition, the whole gesture and body perception—how you pose questions and so on—how does one express oneself? Facial expressions and other things are not conveyed during such a Skype call. [Participant 07, male, in his 50s]

Although participants valued the possibility of using the mental health specialist video consultations as an initial consultation with a mental health specialist or for tackling emergencies, it was clear to some participants (5/13, 39%) that they would insist on having a regular face-to-face consultation with the specialist at some point:

Let me say, for emergencies, initial contacts, or coordinating steps clarifying what to do next, this [mental health specialist video consultation] is okay, but at some point, I must speak to the doctor in person. [Participant 11, female, in her 50s]

Technical Challenges

Almost all participants (12/13, 92%) related several challenges to the specific technology used for conducting mental health specialist video consultations. Some participants brought up technical limitations, such as an unstable internet connection, insufficient technical features at the office, insecure data protection, or high demands on potential users who are still unfamiliar with the use of computers and videoconferencing (the so-called digital immigrants, that is, individuals born before the widespread adoption of digital technology [46]):

> *Concerning the technical aspect, technologies like Skype and video calls have not truly reached the elderly. In addition, for older people, it may simply be strange to talk to a screen. For younger people, I think, this is no problem at all.* [Participant 08, female, in her 30s]

Organizational Challenges

Some participants (3/13, 23%) also expressed concerns that the implementation of mental health specialist video consultations would require time and spatial resources and therefore profoundly impact on the workflow of the primary care practice. However, the descriptions of the organizational challenges were rather general:

> *I have my doubts, because I believe the primary care physicians are the biggest problem. They need time for this. They need additional rooms. They must deal with this whole thing. This is an additional task that doctors then must handle.* [Participant 02, female, in her 60s]

Stigma of Seeking Mental Health Care

Several participants (7/13, 54%) contemplated that, independent of the acceptance of technology, people may not participate in mental health specialist video consultations due to their negative stance on mental health care as such. Participants elaborated that personal preferences and the societal view on both mental health conditions and psychosocial support may impede the acceptance of the model:

> *In my opinion, this [the refusal of mental health care] is not necessarily related to video consultations. I think that this is a general stance on mental health care.* [Participant 13, female, 40s]

Besides benefits of and barriers to the integration of mental health specialist video consultations in primary care, participants also elaborated on prerequisites for the therapeutic relationship in the context of mental health specialist video consultations. We address this aspect in the following subsection.

Prerequisites for Interacting With Providers in Video Consultations

Notably, participants' prerequisites for interacting with mental health specialists in video consultations did not seem to differ much from those concerning face-to-face contacts. In fact, none of the participants related the prerequisites for the mental health specialist to the setting of the mental health specialist video consultation model, that is, speaking to the mental health specialist via videoconferencing. In contrast, all participants (13/13, 100%) emphasized aspects concerning the general therapeutic relationship independent of the therapeutic setting:

> *The doctor should be competent, charismatic, and sensitive. He should give me the feeling that I matter and that someone is listening to me. He should know what he is talking about and give me the feeling that I am being looked after. Trust is very important. I need mutual trust. If I do not trust the doctor, I will not see him.* [Participant 07, male, in his 50s]

Participants repeatedly named empathy, being taken seriously, and feeling appreciated as necessary prerequisites establishing a trustful relationship with the mental health specialist. Two interviewees wished for the possibility of choosing either a female or male mental health specialist. All in all, participants' primary concern was to feel comfortable with the mental health specialist and to be able to establish a good relationship.

Member Checking

Of the 13 participants, 7 (54%) agreed to take part in member checking. Of these 7 participants, 6 confirmed that their personal stance expressed in the interviews was adequately reflected in the final consolidation of the results. One participant had developed a more negative view of the mental health specialist video consultation model since the initial interview. He had started entertaining privacy concerns with engaging in video consultations in the familiar primary care environment (eg, concerns that medical assistants might breach confidentiality).

Discussion

Principal Findings

We investigated the patients' perspective on mental health specialist video consultations in office-based primary care accounting for both tangible and more general factors. In sum, participants viewed the mental health specialist video consultation model positively and anticipated several fairly strictly defined benefits: (1) improved accessibility of mental health specialist care by shorter waiting times, shorter travel distances, and a lower threshold for seeking specialist mental health care; and (2) the familiar environment in the primary care office with familiar practice staff, as well as familiar travel modalities and premises. Notably, all these benefits were specifically linked to the specific feature of the mental health specialist video consultation model: the integration of mental health specialist video consultations directly into the primary care practice. However, participants also mentioned some barriers, namely the lack of face-to-face contact, technical challenges, and organizational challenges for the practice staff. Somewhat different from the professional stance that clinicians conducting video consultations require special skills, such as to be trained with the technology [47], participants' prerequisites for the mental health specialist conducting video consultations did not differ from the prerequisites applied to face-to-face contacts, namely a trustful and appreciative therapeutic relationship. Taken together, our findings suggest that the integration of mental health specialist video consultations into

primary care practices may be a promising care model that warrants further investigation in interventional studies.

Limitations

First, at this stage, our main goal was to identify and address potential threats to the successful adoption of mental specialist consultations in primary care in a qualitative preimplementation study [48-51]. Due to the nature of the qualitative approach we applied, the generalizability of these findings may be limited. However, qualitative methods have been promoted as a particularly effective means to involve potential users of a new service in developing interventions for mental health conditions, such as when considering the implementation in routine practice [52-55]. Moreover, we aimed at enhancing the transferability by thoroughly describing the research context and methods following established standards for the reporting of qualitative studies.

Second, as is common in preimplementation studies, we relied on self-reports in order to tailor the intervention for planned feasibility and effectiveness trials. Nevertheless, self-reported intentions and practice, which the participants reported in the interviews, may, to some extent, differ from participants' actual behavior, due to issues of social desirability, low self-awareness, comprehension, and accurate recall or prediction [56]. In this respect, it is somewhat likely that the benefits, barriers, and prerequisites mentioned in our study differ to some degree from those that participants would observe after having experienced mental health specialist video consultations. Hence, to avoid an attitudinal fallacy, we must be very cautious in inferring situated behavior from the verbal accounts described in our study [57]. In fact, in a real-life situation, people may have different or more difficulties with mental health specialist video consultations than those reported in this study. While we have tried to minimize the constraints of self-reports, such as by firmly reassuring participants of the confidential nature of their participation, given the limited validity and reliability of self-reported behavior, it is both necessary and beneficial to study actual behavior in future interventional studies. To this end, we have embarked on a feasibility study (PROVIDE-B, German Clinical Trials Register registration no. DRKS00015812), with a protocol that was informed by the results presented in this paper [26].

Third, the anticipated benefits and barriers noted by our participants may not be representative of the larger population. Even by qualitative standards, our sample size was rather moderate. While it has recently been shown that 9 interviews are sufficient for code saturation, our sample size bears some risk that we did not reach meaning saturation, in that we may have missed more subtle conceptual issues [58-60]. However, with respect to our goal to identify specific benefits and barriers, we were primarily concerned with code saturation, which ensures "a comprehensive understanding of explicit concrete issues in data" [59]. Moreover, our sampling frame was somewhat limited to patients from our outpatient clinic, which may have introduced volunteer bias. Since these patients already had at least the experience in seeking mental health care at the outpatient clinic, they might have evaluated the mental health specialist video consultation model from a more experienced viewpoint. Thus, our findings may be biased to an overall positive attitude toward the model, as it may improve the accessibility of mental health specialist care, which is the main challenge in the current system. Future studies could apply higher incentives to stratified sampled service users presenting to general practices, for which the mental health specialist video consultations are directly intended.

Fourth, the interviewer was a young (ie, a so-called digital native), technologically open-minded PhD student whose dissertation is on implementing video consultations in primary care. While we cannot fully rule out that this constellation might have biased the study findings in favor of the benefits from video consultations, we continuously tried to ensure self-reflexivity by using field notes before and member checking after the data analysis. Moreover, by adhering to the COREQ reporting standards, we aimed at maximum transparency in documenting the research process.

Fifth, we did not assess whether and, if applicable, to what extent the study participants had any experience with using video chats (eg, FaceTime or Skype). While it is known that 57% of people aged 10 years or older in the German population use the internet to make video calls [61], we cannot determine whether familiarity with video calls in general may have had an impact on the study participants' attitudes toward mental health video consultations.

Sixth, to support the credibility of our analyses, we conducted member checking with the participants. Notably, 1 of the participants altered her view of the model and expressed a more negative attitude toward mental health specialist video consultations. This is well known as a difficulty in achieving credibility through member checking, since participants may provide different accounts of their experiences or opinions during the interview and the member checking [62]. Since all other participants whom we contacted for member checking approved our summary of their interview statements, the member checking underpinned the credibility of our analyses.

Comparison With Prior Work

Benefits: Improved Accessibility and Familiar Primary Care Environment

Our findings support the growing consensus that incorporating telehealth into primary care services will likely allow patients to access their usual health care providers more easily [27]. Clearly, our work adds that—from the patients' perspective—the 2 key arguments for embedding mental health specialist video consultations into primary care practices are the improved accessibility of mental health care services and the familiar environment of the practice. First, our participants stated that patients would not have to travel long distances to get in touch with a mental health specialist, as the primary care physician's office is often in the local community or at least nearby. This is in line with previous findings [63-65]. Second, participants anticipated that integrating mental health specialist video consultations might also save patients from long waiting times, since immediate scheduling of consultations after the first contact with the primary care physician would be possible. This assessment is consistent with findings from a clinical trial on

joint teleconsultations [66]. Trial participants emphasized the convenience of joint consultation appointments and that punctuality was better in the teleconsultations than in hospital outpatient clinics. However, the problem of the pervasive limited availability of providers seems to be still evident with the mental health specialist video consultation model, as specialists would employ time slots previously reserved for in-person contacts. Hence, waiting times could remain a challenge for both modes of delivery. Nevertheless, it seems highly plausible that, at least for some patients, early or crisis intervention through mental health specialist video consultations could reduce the need for further specialist care in the long run and eventually save resources. Finally, participants in our study argued that familiarity with the primary care practice staff and the localities of the practice provide an encouraging environment, in which the patient feels comfortable to engage in mental health care. While such an engagement still depends on the perceived stigma [67,68], the anonymity and flexibility that the presented model allows bear great potential to reduce this very stigma for patients reluctant to reach out to a specialist in person [69,70].

Barriers: Challenges for the Practice and Lack of Personal Contact

First, the articulated barriers (eg, unstable data connections and insufficient technical equipment in the individual practice) underline that a high technical usability and comprehensive training should be ensured prior to the implementation of video consultations in primary care practices [52]. Concerning the anticipated difficulties in certain patient groups (eg, digital immigrants), 1 study showed that age may be a significant factor in whether a patient would accept having video consultations [70]. This study examined patients' interest in a telehealth model in which they would meet a health care provider from their home via a video call. In this regard, our model could facilitate handling unfamiliar technology, because patients would be instructed and supported by specifically trained practice staff.

Second, the participants had reservations about the organizational resources needed in the primary care practice when integrating the mental health specialist video consultations into routine care (eg, spatial requirements and additional workload for the practice staff). In this respect, some researchers have considered it essential to have in-depth familiarization with the organizational workflows of the respective primary care practice (eg, staff working patterns, room requirements, and practice management) prior to the development and implementation of video consultation services [71,72].

Third, the participants were concerned with the lack of face-to-face contact when reverting to video consultations—an observation that several previous studies of videoconferencing conducted in specialty medical settings have reported [38,70]. Obviously, most participants upheld the personal encounter as a core value, preparing the ground for a trustworthy doctor-patient relationship. Nevertheless, some patients may appreciate the physical separation introduced by the mode of delivery as described in a qualitative study exploring young people's perspectives on receiving psychiatric services via video consultations [73]. Specifically, participants in the study suggested that the new format may alleviate patients' anxiety before a meeting with a specialist. They regarded the format as less intimidating than being in the same room with the therapist. At any rate, a recent randomized clinical trial has shown that, despite the physical separation, patient-provider communication in telemedicine is not inferior to communication during face-to-face consultations [74].

In sum, despite the anticipated barriers, 11 of our 13 participants were willing to engage in the proposed mental health specialist video consultation model. In contrast to other studies, in our study none of the participants perceived security and privacy concerns related to video consultations as a barrier [65,73,75]. This might have resulted from our assurance that data security and end-to-end encryption were mandatory features when describing the model to the participants.

Prerequisites for Interacting With Providers in Video Consultations

While the participants consistently highlighted the importance of the therapeutic relationship for mental health specialist video consultation models, prerequisites for interacting with specialists in video consultations did not differ from those relevant for face-to-face consultations. Namely, patients expected that therapists should prepare the groundwork for a respectful, appreciative, and trustful therapeutic relationship. In contrast to previous findings [38], in our study, participants did not indicate that a personal encounter prior to the first video consultations would be required to establish such a relationship. One explanation for this observation could be that the participants considered the primary care physician to be a "trusted gatekeeper" who facilitates the encounter with the therapist.

Conclusions

Our study showed that mental health service users mostly welcomed innovative technology-based care models such as mental health specialist video consultation–integrated care. Taking a pragmatic stance, service users, often frustrated with uncoordinated care, may consider such models to bear great potential to increase access to mental health care. Specifically, service users value their embedment in the familiar environment of the primary care practice. At the same, they demand minimal technical disruptions during video consultations and optimal resemblance to face-to-face settings by therapists consistently capturing and addressing nonverbal cues. While this main observation may be relevant for clinicians already conducting video consultation as part of routine care, our work contributes to preparing the ground for feasibility and effectiveness trials accounting for our preimplementation results during the development of interventions. In perspective, interventional studies will (1) shed light on the actual behavior of service users and professionals engaging in technology-based care models and (2) determine target patient groups who would maximally benefit from such models.

Acknowledgments

This research was supported by a grant from the German Federal Ministry of Education and Research (grant no. 01GY1612). The funder was not actively involved in the conduct of the study. The authors had full access to all the data in this study and take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured guide for telephone interview.
[DOCX File , 22 KB - jmir_v22i4e17330_app1.docx]

Multimedia Appendix 2
Guideline for verbal presentation of the mental health specialist video consultation model.
[DOCX File , 27 KB - jmir_v22i4e17330_app2.docx]

Multimedia Appendix 3
Summary of themes and subthemes.
[DOCX File , 27 KB - jmir_v22i4e17330_app3.docx]

References

1. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. Psychiatr Serv 2009 Oct;60(10):1323-1328. [doi: 10.1176/ps.2009.60.10.1323] [Medline: 19797371]
2. 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]
3. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
4. Gamm LG, Stone S, Pittman S. Mental health and mental disorders — a rural challenge: a literature review. In: Rural Healthy People 2010: A Companion Document to Healthy People 2010. College Station, TX: The Texas A&M University System Health Science Center, School of Rural Public Health, Southwest Rural Health Research Center; 2003:97-113.
5. Katon WJ. Epidemiology and treatment of depression in patients with chronic medical illness. Dialogues Clin Neurosci 2011;13(1):7-23 [FREE Full text] [Medline: 21485743]
6. Fortney JC, Pyne JM, Turner EE, Farris KM, Normoyle TM, Avery MD, et al. Telepsychiatry integration of mental health services into rural primary care settings. Int Rev Psychiatry 2015;27(6):525-539. [doi: 10.3109/09540261.2015.1085838] [Medline: 26634618]
7. Hilty DM, Rabinowitz T, McCarron RM, Katzelnick DJ, Chang T, Bauer AM, et al. An update on telepsychiatry and how it can leverage collaborative, stepped, and integrated services to primary care. Psychosomatics 2018;59(3):227-250. [doi: 10.1016/j.psym.2017.12.005] [Medline: 29544663]
8. Yellowlees P, Shore JH. Telepsychiatry and Health Technologies: A Guide for Mental Health Professionals. 1st edition. Arlington, VA: American Psychiatric Association Publishing; 2018.
9. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
10. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
11. Richardson LK, Frueh BC, Grubaugh AL, Egede L, Elhai JD. Current directions in videoconferencing tele-mental health research. Clin Psychol (New York) 2009 Sep 01;16(3):323-338 [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01170.x] [Medline: 20161010]
12. Liddy C, Maranger J, Afkham A, Keely E. Ten steps to establishing an e-consultation service to improve access to specialist care. Telemed J E Health 2013 Dec;19(12):982-990 [FREE Full text] [doi: 10.1089/tmj.2013.0056] [Medline: 24073898]
13. Liddy C, Rowan MS, Afkham A, Maranger J, Keely E. Building access to specialist care through e-consultation. Open Med 2013;7(1):e1-e8 [FREE Full text] [Medline: 23687533]
14. Sigel P, Leiper R. GP views of their management and referral of psychological problems: a qualitative study. Psychol Psychother 2004 Sep;77(Pt 3):279-295. [doi: 10.1348/1476083041839394] [Medline: 15355581]
15. Hilty DM, Sunderji N, Suo S, Chan S, McCarron RM. Telepsychiatry and other technologies for integrated care: evidence base, best practice models and competencies. Int Rev Psychiatry 2018 Dec;30(6):292-309. [doi: 10.1080/09540261.2019.1571483] [Medline: 30821540]

16. Ratzliff A, Sunderji N. Tele-behavioral health, collaborative care, and integrated care: learning to leverage scarce psychiatric resources over distance, populations, and time. Acad Psychiatry 2018 Dec;42(6):834-840. [doi: 10.1007/s40596-018-0984-5] [Medline: 30338461]
17. Sanchez Gonzalez ML, McCord CE, Dopp AR, Tarlow KR, Dickey NJ, McMaughan DK, et al. Telemental health training and delivery in primary care: a case report of interdisciplinary treatment. J Clin Psychol 2019 Feb;75(2):260-270. [doi: 10.1002/jclp.22719] [Medline: 30589440]
18. Waugh M, Calderone J, Brown Levey S, Lyon C, Thomas M, DeGruy F, et al. Using telepsychiatry to enrich existing integrated primary care. Telemed J E Health 2019 Aug;25(8):762-768. [doi: 10.1089/tmj.2018.0132] [Medline: 30394851]
19. Hilty DM, Randhawa K, Maheu MM, McKean AJS, Pantera R. Therapeutic relationship of telepsychiatry and telebehavioral health: ideas from research on telepresence, virtual reality and augmented reality. Psychol Cogn Sci Open J 2019 Aug 26;5(1):14-29. [doi: 10.17140/pcsoj-5-145]
20. Losa-Iglesias ME, López López D, Rodriguez Vazquez R, Becerro de Bengoa-Vallejo R. Relationships between social skills and self-esteem in nurses: a questionnaire study. Contemp Nurse 2017 Dec;53(6):681-690. [doi: 10.1080/10376178.2018.1441729] [Medline: 29451080]
21. Maheu MM, Drude KP, Hertlein KM, Lipschutz R, Wall K, Hilty DM. Correction to: An interprofessional framework for telebehavioral health competencies. J Technol Behav Sci 2018 Mar 5;3(2):108-140. [doi: 10.1007/s41347-018-0046-6]
22. Pileño ME, Morillo J, Morillo A, Losa-Iglesias M. The mental health team: evaluation from a professional viewpoint. A qualitative study. Arch Psychiatr Nurs 2018 Apr;32(2):206-214. [doi: 10.1016/j.apnu.2017.11.003] [Medline: 29579514]
23. PROVIDE. ImPROving cross-sectoral collaboration between primary and psychosocial care: an implementation study on VIDEo consultations (PROVIDE). Heidelberg, Germany: BMBF-Nachwuchsgruppe PROVIDE, Klinik für Allgemeine Innere Medizin und Psychosomatik, Universitätsklinikum Heidelberg URL: https://www.provide-project.de/ziel-konzept/?lang=en [accessed 2020-03-16]
24. Hoffmann M, Hartmann M, Wensing M, Friederich HC, Haun MW. Potential for integrating mental health specialist video consultations in office-based routine primary care: cross-sectional qualitative study among family physicians. J Med Internet Res 2019 Aug 19;21(8):e13382 [FREE Full text] [doi: 10.2196/13382] [Medline: 31429419]
25. Hilty DM, Johnston B, McCarron RM. How e-mental health adds to traditional outpatient and newer models of integrated care for patients, providers, and systems. In: Mucic D, Hilty DM, editors. e-Mental Health. Cham, Switzerland: Springer International Publishing; 2016:129-149.
26. Tönnies J, Hartmann M, Wensing M, Szecsenyi J, Icks A, Friederich HC, et al. Mental health specialist video consultations for patients with depression or anxiety disorders in primary care: protocol for a randomised controlled feasibility trial. BMJ Open 2019 Sep 04;9(9):e030003 [FREE Full text] [doi: 10.1136/bmjopen-2019-030003] [Medline: 31488484]
27. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]
28. Chakrabarti S, Shah R. Telepsychiatry in the developing world: whither promised joy? Indian J Soc Psychiatry 2016;32(3):273. [doi: 10.4103/0971-9962.193200]
29. Fatehi F, Armfield NR, Dimitrijevic M, Gray LC. Clinical applications of videoconferencing: a scoping review of the literature for the period 2002-2012. J Telemed Telecare 2014 Oct;20(7):377-383. [doi: 10.1177/1357633X14552385] [Medline: 25399998]
30. Fortney JC, Pyne JM, Edlund MJ, Williams DK, Robinson DE, Mittal D, et al. A randomized trial of telemedicine-based collaborative care for depression. J Gen Intern Med 2007 Aug;22(8):1086-1093 [FREE Full text] [doi: 10.1007/s11606-007-0201-9] [Medline: 17492326]
31. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
32. Grol R. Personal paper. Beliefs and evidence in changing clinical practice. BMJ 1997 Aug 16;315(7105):418-421 [FREE Full text] [doi: 10.1136/bmj.315.7105.418] [Medline: 9277610]
33. Bishop JE, O'Reilly RL, Maddox K, Hutchinson LJ. Client satisfaction in a feasibility study comparing face-to-face interviews with telepsychiatry. J Telemed Telecare 2002;8(4):217-221. [doi: 10.1258/135763302320272185] [Medline: 12217104]
34. Hilty DM, Luo JS, Morache C, Marcelo DA, Nesbitt TS. Telepsychiatry: an overview for psychiatrists. CNS Drugs 2002;16(8):527-548. [doi: 10.2165/00023210-200216080-00003] [Medline: 12096934]
35. Richardson L, Reid C, Dziurawiec S. “Going the extra mile”: satisfaction and alliance findings from an evaluation of videoconferencing telepsychology in rural western Australia: alliance in telepsychology. Aust Psychol 2015 Jul 20;50(4):252-258. [doi: 10.1111/ap.12126]
36. Trondsen MV, Bolle SR, Stensland G, Tjora A. Video-confidence: a qualitative exploration of videoconferencing for psychiatric emergencies. BMC Health Serv Res 2014 Oct 31;14:544 [FREE Full text] [doi: 10.1186/s12913-014-0544-y] [Medline: 25359404]
37. Simpson S. The provision of a telepsychology service to Shetland: client and therapist satisfaction and the ability to develop a therapeutic alliance. J Telemed Telecare 2001;7 Suppl 1:34-36. [doi: 10.1177/1357633X010070S114] [Medline: 11576484]

38. Swinton JJ, Robinson WD, Bischoff RJ. Telehealth and rural depression: physician and patient perspectives. Fam Syst Health 2009 Jun;27(2):172-182. [doi: 10.1037/a0016014] [Medline: 19630458]
39. Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. Psychotherapy (Chic) 2011 Mar;48(1):9-16. [doi: 10.1037/a0022186] [Medline: 21401269]
40. Madill A, Jordan A, Shirley C. Objectivity and reliability in qualitative analysis: realist, contextualist and radical constructionist epistemologies. Br J Psychol 2000 Feb;91 (Pt 1):1-20. [Medline: 10717768]
41. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology. Volume 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, DC: American Psychological Association; 2012:57-71.
42. Bowen GA. Naturalistic inquiry and the saturation concept: a research note. Qual Res 2008 Feb;8(1):137-152. [doi: 10.1177/1468794107085301]
43. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
44. Meadows L, Morse J. Constructing evidence within the qualitative project. In: Morse J, Swanson J, Kuzel A, editors. The Nature of Qualitative Evidence. Thousand Oaks, CA: SAGE Publications, Inc; 2001:188-202.
45. Eurostat. Degree of urbanisation.: European Commission URL: https://ec.europa.eu/eurostat/web/degree-of-urbanisation/background [accessed 2020-03-16]
46. Prensky M. Digital natives, digital immigrants part 1. On Horiz 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
47. Johansson AM, Lindberg I, Söderberg S. The views of health-care personnel about video consultation prior to implementation in primary health care in rural areas. Prim Health Care Res Dev 2014 Apr;15(2):170-179 [FREE Full text] [doi: 10.1017/S1463423613000030] [Medline: 23402617]
48. Hwang Y, Chung JY, Shin DH, Lee Y. An empirical study on the integrative pre-implementation model of technology acceptance in a mandatory environment. Behav Inf Technol 2017 Mar 29;36(8):861-874. [doi: 10.1080/0144929x.2017.1306751]
49. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015;350:h1258 [FREE Full text] [Medline: 25791983]
50. Or C, Wong K, Tong E, Sek A. Private primary care physicians' perspectives on factors affecting the adoption of electronic medical records: a qualitative pre-implementation study. Work 2014;48(4):529-538. [doi: 10.3233/WOR-131808] [Medline: 24346272]
51. Walker EM, Mwaria M, Coppola N, Chen C. Improving the replication success of evidence-based interventions: why a preimplementation phase matters. J Adolesc Health 2014 Mar;54(3 Suppl):S24-S28 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.11.028] [Medline: 24560072]
52. De Weger E, Macinnes D, Enser J, Francis SJ, Jones FW. Implementing video conferencing in mental health practice. J Psychiatr Ment Health Nurs 2013 Apr;20(5):448-454. [doi: 10.1111/j.1365-2850.2012.01947.x] [Medline: 22762416]
53. Palinkas LA. Qualitative and mixed methods in mental health services and implementation research. J Clin Child Adolesc Psychol 2014;43(6):851-861 [FREE Full text] [doi: 10.1080/15374416.2014.910791] [Medline: 25350675]
54. Peters S. Qualitative research methods in mental health. Evid Based Ment Health 2010 May;13(2):35-40. [doi: 10.1136/ebmh.13.2.35] [Medline: 21856603]
55. Sale JEM, Thielke S. Qualitative research is a fundamental scientific process. J Clin Epidemiol 2018 Oct;102:129-133. [doi: 10.1016/j.jclinepi.2018.04.024] [Medline: 29793008]
56. Baumeister RF, Vohs KD, Funder DC. Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? Perspect Psychol Sci 2007 Dec;2(4):396-403. [doi: 10.1111/j.1745-6916.2007.00051.x] [Medline: 26151975]
57. Jerolmack C, Khan S. Talk is cheap: ethnography and the attitudinal fallacy. Sociol Methods Res 2014 Mar 09;43(2):178-209. [doi: 10.1177/0049124114523396]
58. Guest G. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
59. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
60. Namey E, Guest G, McKenna K, Chen M. Evaluating bang for the buck. Am J Eval 2016 Jul 09;37(3):425-440. [doi: 10.1177/1098214016630406]
61. German Federal Statistical Office. Häufigste Internetaktivität: Informationssuche über Waren und Dienstleistungen. 2019. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2019/01/PD19_028_639.html [accessed 2020-02-11]
62. Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. Intl J Qual Methods 2016 Nov 30;1(2):13-22. [doi: 10.1177/160940690200100202]
63. Deen TL, Fortney JC, Schroeder G. Patient acceptance of and initiation and engagement in telepsychotherapy in primary care. Psychiatr Serv 2013 Apr 01;64(4):380-384 [FREE Full text] [doi: 10.1176/appi.ps.201200198] [Medline: 23370530]

64. Hartmann M, Finkenzeller C, Boehlen FH, Wagenlechner P, Peters-Klimm F, Herzog W. [Psychosomatic specialist consultation in the general practitioner's office - a new model of integrated care]. Psychother Psychosom Med Psychol 2019 Jun;69(6):231-236. [doi: 10.1055/a-0668-1019] [Medline: 30326535]
65. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient perceptions of telehealth primary care video visits. Ann Fam Med 2017 Dec;15(3):225-229 [FREE Full text] [doi: 10.1370/afm.2095] [Medline: 28483887]
66. Harrison R, Macfarlane A, Murray E, Wallace P. Patients' perceptions of joint teleconsultations: a qualitative evaluation. Health Expect 2006 Mar;9(1):81-90 [FREE Full text] [doi: 10.1111/j.1369-7625.2006.00368.x] [Medline: 16436164]
67. Barley EA, Murray J, Walters P, Tylee A. Managing depression in primary care: a meta-synthesis of qualitative and quantitative research from the UK to identify barriers and facilitators. BMC Fam Pract 2011 Jun 09;12:47 [FREE Full text] [doi: 10.1186/1471-2296-12-47] [Medline: 21658214]
68. Givens JL, Houston TK, Van Voorhees BW, Ford DE, Cooper LA. Ethnicity and preferences for depression treatment. Gen Hosp Psychiatry 2007;29(3):182-191. [doi: 10.1016/j.genhosppsych.2006.11.002] [Medline: 17484934]
69. Callahan EJ, Hilty DM, Nesbitt TS. Patient satisfaction with telemedicine consultation in primary care: comparison of ratings of medical and mental health applications. Telemed J 1998;4(4):363-369. [doi: 10.1089/tmj.1.1998.4.363] [Medline: 10220477]
70. Gardner MR, Jenkins SM, O'Neil DA, Wood DL, Spurrier BR, Pruthi S. Perceptions of video-based appointments from the patient's home: a patient survey. Telemed J E Health 2015 Apr;21(4):281-285 [FREE Full text] [doi: 10.1089/tmj.2014.0037] [Medline: 25166260]
71. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform 2001 Dec;64(2-3):143-156. [Medline: 11734382]
72. Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care--an interactive sociotechnical analysis. J Am Med Inform Assoc 2007;14(5):542-549 [FREE Full text] [doi: 10.1197/jamia.M2384] [Medline: 17600093]
73. Boydell KM, Volpe T, Pignatiello A. A qualitative study of young people's perspectives on receiving psychiatric services via televideo. J Can Acad Child Adolesc Psychiatry 2010 Feb;19(1):5-11 [FREE Full text] [Medline: 20119561]
74. Agha Z, Schapira RM, Laud PW, McNutt G, Roter DL. Patient satisfaction with physician-patient communication during telemedicine. Telemed J E Health 2009 Nov;15(9):830-839. [doi: 10.1089/tmj.2009.0030] [Medline: 19919189]
75. Mielonen ML, Ohinmaa A, Moring J, Isohanni M. Psychiatric inpatient care planning via telemedicine. J Telemed Telecare 2000;6(3):152-157. [doi: 10.1258/1357633001935248] [Medline: 10912333]

Abbreviations

COREQ: consolidated criteria for reporting qualitative research
PROVIDE: Improving Cross-Sectoral Collaboration Between Primary and Psychosocial Care: An Implementation Study on Video Consultations

Edited by G Eysenbach; submitted 06.12.19; peer-reviewed by D Hilty, D López López, T Muto; comments to author 06.01.20; revised version received 29.02.20; accepted 02.03.20; published 20.04.20.

Please cite as:
Bleyel C, Hoffmann M, Wensing M, Hartmann M, Friederich HC, Haun MW
Patients' Perspective on Mental Health Specialist Video Consultations in Primary Care: Qualitative Preimplementation Study of Anticipated Benefits and Barriers
J Med Internet Res 2020;22(4):e17330
URL: http://www.jmir.org/2020/4/e17330/
doi:10.2196/17330
PMID:32310139

Original Paper

Benefits of Teledermatology for Geriatric Patients: Population-Based Cross-Sectional Study

Mara G Bianchi[1*], MD, PhD; Andre Santos[1*], MSc; Eduardo Cordioli[1*], MD

Telemedicine, Hospital Israelita Albert Einstein, São Paulo, Brazil

[*]all authors contributed equally

Corresponding Author:
Mara G Bianchi, MD, PhD
Telemedicine
Hospital Israelita Albert Einstein
Avenida Albert Einstein 627
São Paulo
Brazil
Phone: 55 11 21513381
Email: marahgbianchi@gmail.com

Abstract

Background: Teledermatology is a health care tool that has been increasingly used around the world, mostly because dermatology has an emphasis on visual diagnosis. Many studies have shown that access to specialized care improves using teledermatology, which provides accurate diagnosis and reduces the time taken for treatment, with high patient satisfaction. As the population around the world grows old, there will be even more demand for dermatologists in years to come. It is essential to know which are the most prevalent skin conditions in the primary care population and if they can be addressed through teledermatology.

Objective: Our main goal was to evaluate the proportion of lesions in individuals aged 60 years and older that could be managed using teledermatology in conjunction with primary care physicians. Second, we aimed to assess the most frequent skin lesions, the most common treatments provided to patients, and the distribution and causes of referrals made by the teledermatologists.

Methods: This was a retrospective cohort study from July 2017 to July 2018 in São Paulo, Brazil. We included 6633 individuals aged 60 years and older who presented with 12,770 skin lesions. Teledermatologists had three options to refer patients: (1) to undergo biopsy directly, (2) to an in-person dermatologist visit, and (3) back to the primary care physician with the most probable diagnosis and treatment.

Results: Teledermatology managed 66.66% (8408/12614) of dermatoses with the primary care physician without the need for an in-presence visit; 27.10% (3419/12614) were referred to dermatologists, and 6.24% (787/12614) directly to biopsy. The most frequent diseases were seborrheic keratosis, solar lentigo, onychomycosis, melanocytic nevus, benign neoplasms, actinic keratosis, epidermoid cyst, xerosis, leucoderma, and wart, with significant differences between sexes. Malignant tumors increased with age and were the leading cause for biopsies, while infectious skin conditions and pigmentary disorders decreased. Emollient was the most frequent treatment prescribed, in 31.88% (909/2856) of the cases.

Conclusions: Teledermatology helped to treat 67% of the dermatoses of older individuals, addressing cases of minor complexity quickly and conveniently together with the primary care physician, thus optimizing dermatological appointments for the most severe, surgical, or complex diseases. Teledermatology does not aim to replace a face-to-face visit with the dermatologist; however, it might help to democratize dermatological treatment access for patients and decrease health care expenses.

(J Med Internet Res 2020;22(4):e16700) doi:10.2196/16700

KEYWORDS

access to and use of services; decision making; epidemiology; economics; health care systems and management (telehealth); management; technology; teledermatology; geriatric population

Introduction

Background

Teledermatology is a health care tool that has been increasingly used around the world, mostly because dermatology has an emphasis on visual diagnosis. Studies from all over the world have been performed on this subspecialty [1]. Real-time or store-and-forward are the most common types of delivering images. In store-and-forward teledermatology, patient data and images are collected and sent to a dermatologist to analyze at a later time. In real-time teledermatology, patients and physicians exchange data and images from separate locations in real time [2]. In hybrid teledermatology, both types of images and data are used together. Many studies have shown that access to specialized care is improved by using store-and-forward teledermatology, which provides accurate diagnosis and reduces the time taken for treatment, with high patient satisfaction [3]. A metanalysis with 21 studies showed that the diagnostic accuracy (defined as agreement with histopathology for excised lesions or clinical diagnosis for nonexcised lesions) is still higher for the determination of skin cancer in face-to-face dermatologist, ranging from 67% to 85% depending on the study, than teledermatology, which varied from 51% to 85%. However, some studies do report higher accuracy of teledermatology diagnosis [1].

Typical skin conditions such as mild atopic dermatitis, acne, fungal infections, xerosis, and others are common diseases that may be manageable within the primary care attention service. However, this is not often observed because health professionals are not well trained in diagnosing or triaging skin diseases. The lack of well-trained primary care physicians can lead to unnecessary referral of patients to dermatologists. When there are not enough dermatologists to meet demand, appointments can be filled by patients who do not need specialist care limiting availability of visits for those who do need them.

The city of São Paulo has nearly 12 million inhabitants [4], and 58% of them depend exclusively on the public health care system [5]. The public municipal health care system provides most primary care services. By July 2017, 57,832 individuals were waiting for an appointment with a dermatologist, which could take up to 1 year to occur. As the population around the world grows old, there will be even more demand for dermatologists in years to come, as many skin conditions appear or worsen in older patients. It is essential to know which are the most prevalent skin conditions in the primary care population. If they can be addressed through teledermatology, we may be able to optimize the public health system to manage individuals aged 60 years and older properly. This study had the advantage of including a large number of individuals with many types of skin disease. In contrast, most teledermatology articles focus on one disease (melanoma) or a class of diseases (malignant tumors) [6-8].

Rationale of the Teledermatology Project

Due to this high unmet demand, the municipal health department established a teledermatology project in partnership with Hospital Israelita Albert Einstein, a large private hospital in the city. The aim was to assist patients in primary care, avoiding unnecessary consultations with dermatologists and accelerating in-presence visits and biopsies for those who have more complex, surgical, or even lethal conditions.

Objectives of the Study

The primary goal was to determine the proportion of dermatosis in the older population (aged 60 years and older) that could be managed in primary care through teledermatology, but the project included patients of all ages. Second, we assessed the distribution of referrals and frequency (according to sex and age), treatment, and causes of the most common skin diseases for patients assisted in the project. Both objectives have been achieved in this study.

Methods

Design

This study was approved by the Hospital Israelita Albert Einstein and Municipal Ethics Committees (CAAE: 97126618.6.3001.0086), and it is in accordance with ethical standards on human experimentation and the Declaration of Helsinki. Data were fully anonymized before being accessed, and the institutional review board waived the requirement for informed consent.

As this teledermatology project included a large number of individuals, it was divided into topics of interest to better analyze the subpopulations. The study design and method are similar to ones that have been previously published [9,10]. In summary, it was a retrospective cohort conducted in the city of Sao Paulo, where 10,545 individuals aged 60 years and older were waiting for an appointment with a dermatologist in July 2017. The municipal health department in conjunction with Hospital Israelita Albert Einstein developed a platform and mobile app to be used by health technicians. Photographs were taken with a digital camera and uploaded along with a short clinical history and patient data. The standard protocol for taking pictures was one photo with enough distance to include the entire part of the body in question (face, arm, leg, trunk), a second one in close-up, around 15 cm away from the lesion, and a third one in a lateral view to capture the volume of the lesion. All data were collected and uploaded to a platform accessed only by dermatologists recruited for this project using a secured online process.

Patients on the waiting list for dermatological assistance were phoned by the public health care service and scheduled to go to an appointment in one of three public city hospitals participating in the project. From July 2017 to July 2018, 13 teledermatologists worked to triage patients, first deciding whether the photographs of the lesions and essential clinical history were satisfactory for diagnostic purposes. If not, they would mark the “bad photo” box on the platform and refer the patient for a face-to-face appointment with a dermatologist. If the photos and essential clinical information were of good quality, the triage dermatologists would formulate the most probable diagnostic hypothesis and choose among three referral options for each lesion assessed: (1) directly to biopsy (after which the patient would return to an in-presence dermatologist appointment with the result); (2) to a dermatologist consultation;

or (3) back to the primary care physician with the most probable diagnosis and recommended treatment or guidance on how to proceed with the investigation or management of the lesion. If the same patient had more than one lesion with different referrals, a biopsy referral would prevail over dermatologist referral, which would prevail over a back to primary care physician referral. All patients who attended the project were included in this study.

To better analyze the population of interest, patients aged 60 years and older, we divided them into four categories: 60 to 69 years, 70 to 79 years, 80 to 89 years, and 90 years and older. Only dermatologists certified by the Brazilian Board of Dermatology participated in the project to decrease the chance of diagnostic error by teledermatology. In our municipal basic health unit, a direct exam to search for fungal infection is not done. For this reason, suspicious cases of superficial fungal infections received antifungal treatment as a therapeutic test and were reevaluated afterward. If deep fungal infections were suspected, the teledermatologist would refer the patient to biopsy or a dermatologist and not treat the patient before confirmation of the pathogen.

Statistical Analysis

Missing data were reported when the patient had the photographs taken. If for any reason, such as problems with the platform, the photographs did not make their way to the teledermatologists or the reports made by the teledermatologists were not uploaded to the platform (this was more frequent at the beginning of the project, mostly for technical reasons), those cases were referred to face-to-face dermatologists. All calculations and frequency analyses were done using only the available teledermatologist reports, meaning that missing data and bad photos were not included. For differences between groups (sex), a statistical calculation was done using a 2-tail chi-square with Yates correction test using Prism 6.0 (GraphPad Software). $P<.05$ was considered significant.

Results

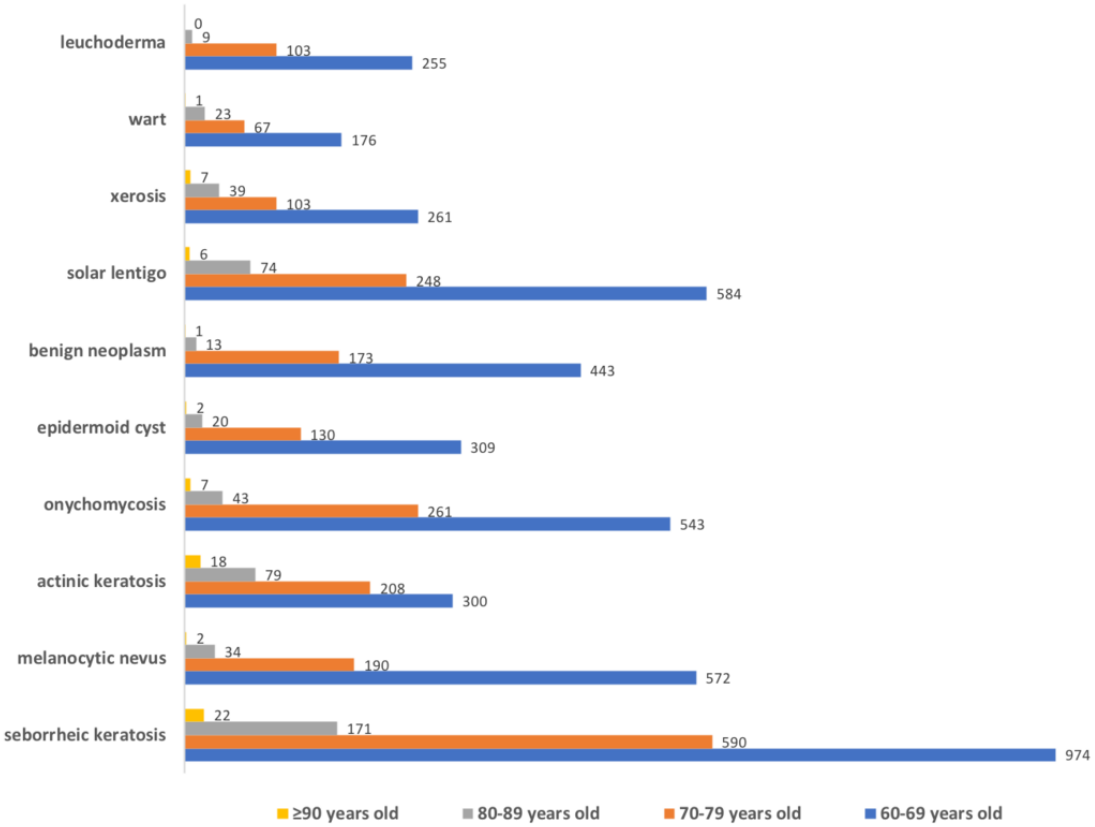
From a population of 10,545 individuals aged 60 years and older waiting for dermatologist consultations, 6633 patients participated in this project (62.90%); 6320 referrals were made, and 313 participants were lost due to technical problems (4.7%). Multimedia Appendix 1 shows the patient demographic data. There were more females waiting for consultation with a dermatologist across all the ages studied. As the age increased, the percentage of patients who responded to the phone call and participated in the project decreased slightly: 65.14% (4276/6564) at age 60 to 69 years and 55.23% (525/1004) at age 80 years and older. A total of 2.23% (148/6633) of patients were referred to a dermatologist due to the poor quality of photographs.

The total number of photographed lesions was 13,432, and 12,614 diagnoses were made. The mean number of photographed lesions per person was 2. Bleeding was present in only 8.09% (982/12,138) of the lesions, but pruritus was reported in 40.62% (4930/12,138) of them, being more frequent in lesions on individuals aged 90 years and older. Photographs with poor quality were calculated at 1.22% (156/12,770; Multimedia Appendix 1), and patient records lost due to technical problems was calculated at 4.9%.

A total of 49.81% (3148/6320) of patients were referred back to their primary physicians, 42.10% (2661/6320) to consultation with an in-presence dermatologist, and 8.09% (511/6320) directly to biopsy; 66.66% (8408/12,614) of all lesions were referred back to the primary care physician, 27.10% (3419/12,614) to the in-presence dermatologist, and 6.24% (787/12,614) directly to biopsy. The mean waiting time for a face-to-face dermatologist was 6.7 months before the project, which dropped to 1.5 months during the project (reduction of 78%).

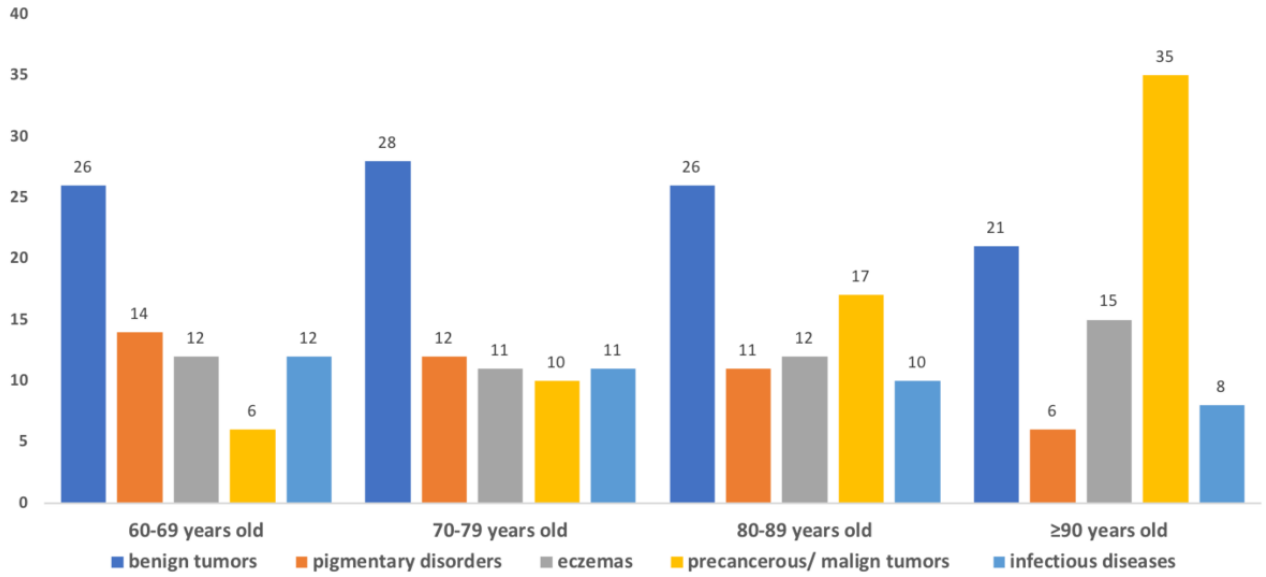
The most common causes of consultation for individuals aged 60 years and older are shown in Figure 1. Seborrheic keratosis was the leading cause for teletriage consultation, totaling 13.93% (1757/12,614) of all complaints. Other benign tumors, such as melanocytic nevus, benign neoplasms, and epidermoid cysts, accounted for 14.96% (1889/12,614). Pigmentary disorders (solar lentigo and leucoderma) were also very frequent, accounting for 10.38% (1309/12,614) of the lesions. Actinic keratosis alone was responsible for 4.80% (605/12,614) of the diagnoses, infectious diseases such as onychomycosis and warts were fairly common (6.77% [854/12,614] and 2.12% [267/12,614] of the cases, respectively), and xerosis was responsible for 3.25% (410/12,614) of all complaints. There were significant differences in the frequency of lesions according to sex. Epidermoid cyst ($P<.001$), wart ($P=.04$), and actinic keratosis ($P=.008$) were more frequent in men. Solar lentigo ($P<.001$), onychomycosis ($P=.02$), leucoderma ($P=.01$), and melanocytic nevus ($P<.001$) were statistically more frequently diagnosed in women. Seborrheic keratosis, xerosis, and benign neoplasms showed no difference in frequency between the sexes.

Figure 1. Most frequent diseases (n) in patients aged 60 years and older who participated in the teledermatology project according to sex from July 2017 to July 2018, in São Paulo, Brazil.



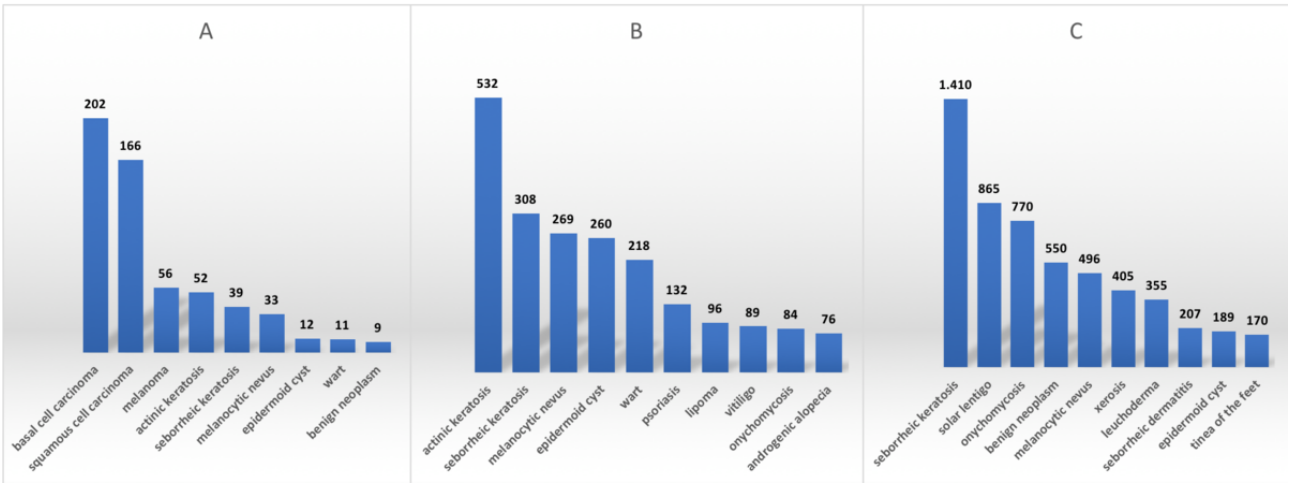
We assessed the most common diseases according to age and divided them into disease groups (Multimedia Appendix 2). Benign tumors were the most frequent complaint for patients up to age 89 years, accounting for about 28.17% (3551/12,614) of the cases. Pigmentary disorders were the second cause for consultations, but this slightly decreased with age. Eczemas were the third cause, very steady in percentage over the years (mean frequency of 12.28% [1549/12,614]. Precancerous and malignant tumors were the fourth cause of visits in individuals aged from 60 to 69 years. Still, the rate increased a lot with age, reaching 34.65% (44/127), and was the leading cause of teletriage consultations in individuals aged 90 years and older. Infectious skin diseases occupied fifth place at age 60 to 69 years old, and the rate gradually decreased with age (Figure 2).

Figure 2. Most frequent group of diseases (%) in patients aged 60 years and older who participated in the teledermatology project by group age from July 2017 to July 2018 in São Paulo, Brazil.



The most frequent lesions referred to biopsy, dermatologists, and back to the primary care physician are shown in Figure 3. Biopsy was indicated in 6.24% (787/12,614) of all lesions. Malignant tumors were the most frequent cause: basal cell carcinoma, squamous cell carcinoma, and melanoma being the top 3 indications, in 83.79% (202/244), 90.71% (166/183), and 76.68% (56/74) of the cases with these diseases, respectively. Only 8.60% (52/605) of actinic keratosis cases were sent to biopsy. Seborrheic keratosis, melanocytic nevus, epidermoid cyst, warts, and benign neoplasms such as soft fibroma and acrochordon accounted for 65 cases that were referred to biopsy, representing less than 0.27% (104/3913) of all lesions with this diagnosis.

Figure 3. Most frequent lesions (n) sent to biopsy, dermatologists, and back to physician in patients aged 60 years and older who participated in the teledermatology project from July 2017 to July 2018 in São Paulo, Brazil.



Referrals to dermatologists occurred in 27.10% (3419/12,614) of the lesions. Actinic keratosis was the lesion which most frequently sent the patient to an in-presence appointment (in 87.93% [532/605] of the times), followed by warts (218/267, 81.65%), epidermoid cyst (260/461, 56,40.%), lipomas (96/245, 39.18%), melanocytic nevus (269/798, 33.71%), and seborrheic keratosis (308/1757, 17.53%). Chronic skin diseases that frequently require follow-ups, such as psoriasis, vitiligo, onychomycosis, and androgenic alopecia, were also in the top 10 causes for referrals to dermatologists. Frequency of referral was 60.27% (132/219), 74.17% (89/120), 9.84.% (84/854), and 32.34% (76/235), respectively.

Teledermatologists sent the patients back to their primary care physicians in 66.71% (8408/12614) of cases. Seborrheic keratosis could be treated at primary care in 80.25% (1410/1757) of cases, solar lentigo in 94.85% (865/912), benign neoplasms in 87.30% (550/630), leuchoderma in 96.73% (355/367), melanocytic nevus in 62.16% (496/798), and epidermoid cyst in 41.00% (189/461). Onychomycosis was managed by

teledermatology in 90.16% (770/854) of cases, xerosis in 98.78% (405/410), seborrheic dermatitis in 90.79% (207/228), and tinea on the feet in 93.92% (170/181) of cases.

We also searched for the most frequent treatments prescribed by teledermatologists. Emollients were the most common prescription, in 31.88% (909/2856) of all cases. The other top classes of drugs prescribed were as follows: topical antifungal in 29.52% (843/2856), sunscreen in 27.87% (796/2856), topical corticosteroids (low and high potency) in 24.40% (697/2856), oral antifungals in 7.18% (205/2856), and hydroquinone in 1.33% (38/2856) of cases.

Discussion

Principal Findings

The vast majority of the dermatoses assessed did not require an in-presence evaluation by a dermatologist (73%), suggesting that most skin conditions in our primary care setting are of low complexity and, therefore, can be addressed appropriately without referral to more specialized centers. This finding also reinforces the feasibility and importance of teledermatology in this context, helping primary care physicians to manage such conditions and optimizing patient access to care for more severe, surgical or complex illnesses, which do require a dermatologist, especially in the public system. The reduction in the mean waiting time for a face-to-face dermatologist during the project is a great advantage, especially considering that many of these cases could be skin cancers. During the study, the patient was always treated by a physician, which gave us more confidence that potentially severe, life-threatening, or rare diseases were not being overlooked. This work shows that teledermatology is a well-defined, structured, scalable process with standardized collection and fairness of care, which might promote the democratization of access to dermatology for underprivileged patients.

Most of the cases referred for biopsy were patients with malignant lesions, especially basal cell and squamous cell carcinomas. Melanoma, although rarer, was the third cause of biopsy referrals, which is of great concern, since it can lead to death if not promptly diagnosed and treated. The fact that biopsy was indicated in 74% and not 100% of the melanoma cases, as expected, could reflect the teledermatologists' fear about the type of biopsy that would be performed in those cases: incisional versus excisional. They probably sent the melanoma patients to the dermatologist aiming to ensure the indication of excisional biopsy, which is the gold standard treatment for this disease [11]. Only a fraction of patients with common benign tumors were sent to biopsy (less than 0.5%).

Referrals to dermatologists were made mostly for treatment of premalignant lesions (actinic keratosis) and warts, lesions that usually require procedures such as cryotherapy with liquid nitrogen, excision, curettage, and electrocoagulation. A small portion of patients with benign tumors, such as melanocytic nevi and seborrheic keratosis, were sent to an in-presence dermatologist consultation. Still, these numbers could be even smaller if access to dermoscopic images were granted [10]. Dermoscopy devices permit a 20× augmentation of skin lesions and may help in the diagnosis of pigmented lesions and skin cancer, especially in incipient lesions. Vitiligo, psoriasis, onychomycosis, especially if treated with oral antifungal medicines, and androgenic alopecia cases were also sent to dermatologists, mostly because they are chronic conditions that require follow-ups.

Regarding the frequency of lesions, seborrheic keratosis was the most common complaint found in the study. It is one of the most common skin tumors, but due to its benign nature, treatment is not mandatory. The second condition in frequency was solar lentigo, a benign lesion associated with aging, constituting a general aesthetic and social concern. It appears mostly on chronically sun-exposed surfaces (face and scalp, dorsum of the hands, neckline, and upper back) and is frequently present among the elderly population. Estimates are that more than 90% of white patients aged older than 50 years are affected [12]. Onychomycosis was the third most frequent cause of complaints. Its worldwide prevalence is estimated at 5.5% of the population, and it is more common in older individuals [13]. Nevus was the fourth most frequent concern, about which there are very little data in the older population. Benign neoplasms, such as acrochordon and soft fibromas, were very common complaints in our research. Actinic keratosis was next in prevalence, corresponding to 5% of all lesions. Actinic keratoses are common lesions representing a step in the development of squamous cell carcinoma. Induced by ultraviolet radiation, actinic keratoses increase in number with age and were found to be the most common reason to visit a dermatologist [14]. Epidermoid cysts are one of the common benign subcutaneous tumors, and males are more affected than females [15]. Xerosis (dry skin) was the primary cause of consultation in 3% of our cases, but one study in nursing homes with older persons (mean age 75 years) observed a prevalence of 56% [16]. Classifying diseases in groups and assessing their prevalence according to age, we observed that benign tumors and eczematous diseases showed stable rates. Precancerous and malignant tumors increased 6-fold with age, from age 60 years to older than 90 years, becoming the most prevalent group disease in the oldest population; our findings were similar to study in Turkey [17].

Regarding treatment, a list of drugs available to the population studied is found in Multimedia Appendix 3. All physicians involved in the project were encouraged, as much as possible, to use only medications from this list, since they were available free for patients, who would be unable to buy most of them otherwise. The results showed the importance of emollients in skin conditions, prescribed in one-third of the cases. Topical antifungals were the second most prescribed medication due to the high rate of fungal diseases. Sunscreen was third, mainly to prevent new cases of actinic keratosis and malignant tumors. Low-potency and high-potency corticosteroids also played an important role, rounding out the top 5 treatments list.

Limitations

One significant limitation of our work is the chance of error and bias in teledermatology diagnosis. However, many articles have attested to a high agreement rate between teledermatology and in-presence dermatology diagnosis [6,18]. The fact that teledermatologists receive multiples photographs of parts of the

body and head but are not able to examine the body as a whole makes the diagnosis more challenging. Also, some critical impressions that would help to corroborate the diagnosis, such as feeling the texture of the skin, or proceeding with easy tests (pe vitrocompression) cannot be done. However, in a previous study we have shown that the inability to palpate the lesion is not a major issue for teledermatologists, who became much more confident in teledermatology after working with it [10]. As this teledermatology project was designed to be a triage program, patient follow-ups with the teledermatologists were not included in the study, making it impossible to evaluate the efficacy of the dispensed treatment and gather more information about further referrals made to these patients.

Comparison With Prior Work

Only one study including 500 patients analyzed the proportion of in-person consultations that could be avoided using teledermatology and showed that in 73% of patients, no referral to an in-person dermatologist was needed [19].

Conclusions

The current teledermatology initiative in São Paulo managed most of the patients' dermatoses without the need for an in-presence appointment with a dermatologist, keeping them within the primary care setting. The mean waiting time reduction for an in-presence visit was a definite benefit for the population who presented with more complex diseases. Although it does not aim to replace a face-to-face visit with a dermatologist, teledermatology proved to be an efficient tool for triage and management of less complicated skin conditions quickly and conveniently, and, in our project, it has promoted the democratization of access to dermatology for underprivileged patients. Future works should focus on the cost effectiveness of the project, on the accuracy of teledermatology in inflammatory and tumoral dermatoses, and on the efficacy of the management proposed by teledermatologists to general physicians.

The frequency of skin diseases differed according to age and sex. Recognizing the prevalence and distribution of the most common diseases affecting individuals aged 60 years and older is crucial to address social and health policies that could orient treatment or prevention through popular campaigns and advertisements.

Acknowledgments

The authors would like to thank Raquel Machado de Sousa from the Department of Telemedicine Hospital Israelita Albert Einstein for her research assistance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Individuals aged 60 years and older waiting for a dermatologist consultation who participated or not in the teledermatology project and the characteristics of their lesions according to age and sex from July 2017 to July 2018 in São Paulo, Brazil.
[DOCX File , 14 KB - jmir_v22i4e16700_app1.docx]

Multimedia Appendix 2
List of most common diseases included in the disease groups.
[DOCX File , 13 KB - jmir_v22i4e16700_app2.docx]

Multimedia Appendix 3
Table of medications available for the teledermatologists from the municipal health system in São Paulo, Brazil, from July 2017 to July 2018.
[DOCX File , 13 KB - jmir_v22i4e16700_app3.docx]

References

1. Finnane A, Dallest K, Janda M, Soyer HP. Teledermatology for the diagnosis and management of skin cancer: a systematic review. JAMA Dermatol 2017 Mar 01;153(3):319-327. [doi: 10.1001/jamadermatol.2016.4361] [Medline: 27926766]
2. Heffner VA, Lyon VB, Brousseau DC, Holland KE, Yen K. Store-and-forward teledermatology versus in-person visits: a comparison in pediatric teledermatology clinic. J Am Acad Dermatol 2009 Jun;60(6):956-961. [doi: 10.1016/j.jaad.2008.11.026] [Medline: 19362751]
3. O'Connor DM, Jew OS, Perman MJ, Castelo-Soccio LA, Winston FK, McMahon PJ. Diagnostic accuracy of pediatric teledermatology using parent-submitted photographs: a randomized clinical trial. JAMA Dermatol 2017 Dec 01;153(12):1243-1248 [FREE Full text] [doi: 10.1001/jamadermatol.2017.4280] [Medline: 29141082]
4. Instituto Brasileiro de Geografia e Estatística. Panorama São Paulo URL: https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama [accessed 2019-02-02]

5. Rede Brasil Atual. No estado de São Paulo, 25 milhões de pacientes dependem do SUS URL: https://tinyurl.com/wvj5hpt [accessed 2019-02-02]
6. Kroemer S, Frühauf J, Campbell TM, Massone C, Schwantzer G, Soyer HP, et al. Mobile teledermatology for skin tumour screening: diagnostic accuracy of clinical and dermoscopic image tele-evaluation using cellular phones. Br J Dermatol 2011 May;164(5):973-979. [doi: 10.1111/j.1365-2133.2011.10208.x] [Medline: 21219286]
7. Ferrándiz L, Ruiz-de-Casas A, Martin-Gutierrez FJ, Peral-Rubio F, Mendez-Abad C, Rios-Martin JJ, et al. Effect of teledermatology on the prognosis of patients with cutaneous melanoma. Arch Dermatol 2012 Sep;148(9):1025-1028. [doi: 10.1001/archdermatol.2012.778] [Medline: 22986852]
8. Bruce AF, Mallow JA, Theeke LA. The use of teledermoscopy in the accurate identification of cancerous skin lesions in the adult population: a systematic review. J Telemed Telecare 2018 Feb;24(2):75-83 [FREE Full text] [doi: 10.1177/1357633X16686770] [Medline: 28056600]
9. Giavina Bianchi M, Santos AP, Cordioli E. The majority of skin lesions in pediatric primary care attention could be managed by teledermatology. PLoS One 2019;14(12):e0225479 [FREE Full text] [doi: 10.1371/journal.pone.0225479] [Medline: 31790453]
10. Giavina Bianchi M, Santos A, Cordioli E. Dermatologists' perceptions on the utility and limitations of teledermatology after examining 55,000 lesions. J Telemed Telecare 2019 Aug 14:1357633X19864829. [doi: 10.1177/1357633X19864829] [Medline: 31409225]
11. Pavri SN, Clune J, Ariyan S, Narayan D. Malignant melanoma: beyond the basics. Plast Reconstr Surg 2016 Aug;138(2):330e-340e. [doi: 10.1097/PRS.0000000000002367] [Medline: 27465194]
12. Arginelli F, Greco M, Ciardo S, Josse G, Rossi AB, Le Digabel J, et al. Efficacy of D-pigment dermocosmetic lightening product for solar lentigo lesions of the hand: a randomized controlled trial. PLoS One 2019;14(5):e0214714 [FREE Full text] [doi: 10.1371/journal.pone.0214714] [Medline: 31042723]
13. Lipner SR, Scher RK. Onychomycosis: clinical overview and diagnosis. J Am Acad Dermatol 2019 Apr;80(4):835-851. [doi: 10.1016/j.jaad.2018.03.062] [Medline: 29959961]
14. Warino L, Tusa M, Camacho F, Teuschler H, Fleischer AB, Feldman SR. Frequency and cost of actinic keratosis treatment. Dermatol Surg 2006 Aug;32(8):1045-1049. [doi: 10.1111/j.1524-4725.2006.32228.x] [Medline: 16918567]
15. Nigam JS, Bharti JN, Nair V, Gargade CB, Deshpande AH, Dey B, et al. Epidermal cysts: a clinicopathological analysis with emphasis on unusual findings. Int J Trichology 2017;9(3):108-112 [FREE Full text] [doi: 10.4103/ijt.ijt_16_17] [Medline: 28932061]
16. Augustin M, Kirsten N, Körber A, Wilsmann-Theis D, Itschert G, Staubach-Renz P, et al. Prevalence, predictors and comorbidity of dry skin in the general population. J Eur Acad Dermatol Venereol 2019 Jan;33(1):147-150. [doi: 10.1111/jdv.15157] [Medline: 29953684]
17. Yaldiz M. Dermatological diseases in the geriatric age group: retrospective analysis of 7092 patients. Geriatr Gerontol Int 2019 Apr 05;19(7):582-585. [doi: 10.1111/ggi.13665]
18. Pak H, Triplett CA, Lindquist JH, Grambow SC, Whited JD. Store-and-forward teledermatology results in similar clinical outcomes to conventional clinic-based care. J Telemed Telecare 2007;13(1):26-30. [doi: 10.1258/135763307779701185] [Medline: 17288655]
19. McAfee JL, Vij A, Warren CB. Store-and-forward teledermatology improves care and reduces dermatology referrals from walk-in clinics: a retrospective descriptive study. J Am Acad Dermatol 2020 Feb;82(2):499-501. [doi: 10.1016/j.jaad.2019.08.006] [Medline: 31404572]

Edited by G Eysenbach; submitted 15.10.19; peer-reviewed by H Mehdizadeh, S Ziehfreund , D Juzwishin; comments to author 02.11.19; revised version received 21.12.19; accepted 27.01.20; published 21.04.20.

Please cite as:
G Bianchi M, Santos A, Cordioli E
Benefits of Teledermatology for Geriatric Patients: Population-Based Cross-Sectional Study
J Med Internet Res 2020;22(4):e16700
URL: http://www.jmir.org/2020/4/e16700/
doi:10.2196/16700
PMID:32314966

Original Paper

Wirelessly Observed Therapy to Optimize Adherence and Target Interventions for Oral Hepatitis C Treatment: Observational Pilot Study

Maurizio Bonacini[1], MD; Yoona Kim[2], PharmD; Caroline Pitney[3], PharmD; Lee McKoin[3], BA; Melody Tran[2], PharmD; Charles Landis[3], MD

[1]Mission Gastroenterology and Hepatology, San Francisco, CA, United States

[2]Proteus Digital Health, Redwood City, CA, United States

[3]Paul G Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, United States

Corresponding Author:
Maurizio Bonacini, MD
Mission Gastroenterology and Hepatology
2300 Sutter St, Suite 207
San Francisco, CA, 94115
United States
Phone: 1 4153530212
Email: bonacim08@gmail.com

Abstract

Background: A fixed-dose combination of ledipasvir/sofosbuvir (LDV/SOF) is efficacious in treating chronic hepatitis C virus (HCV) infection; however, objective adherence to prescribed regimens in real-world clinical settings has not been well studied.

Objective: This study aimed to evaluate adherence and virologic outcomes in patients with chronic HCV infection treated with LDV/SOF using a novel digital medicine program that directly measures drug ingestion adherence.

Methods: This prospective, observational, open-label, single-arm pilot study was conducted at 2 clinical research sites and followed patients with HCV infection who were prescribed LDV/SOF along with an ingestible sensor. Patients were treated for 8 or 12 weeks. The main outcomes were ingestion adherence, medical interventions, virologic response, safety, and patient satisfaction.

Results: Of the 28 patients (mean 59 years, SD 7), 61% (17/28) were male, 61% (17/28) were non-Caucasian, and 93% (26/28) were treatment naïve. All 28 had genotype 1 HCV, and of these, 27 completed an 8- or 12-week treatment. Patients used the digital medicine program for 92% of the expected days; the overall mean ingestion adherence rate was 97%. Providers used the digital medicine program data for same-day medication therapy management in 39% (11/28) of patients. End-of-treatment response was achieved in all the available 21 of 28 patients. Sustained virologic response at 12 weeks or more was achieved in 26 of 28 patients; of the 2 patients who relapsed, one had less than 90% adherence and the other had greater than or equal to 95% adherence, lending insights into reasons for treatment failure. A total of 4 subjects reported nonserious adverse events, which were resolved.

Conclusions: The findings of this study suggest that digital medicines can be used for wirelessly observed therapy to support adherence to antiviral HCV therapy, reduce unnecessary medication wastage and retreatment costs, and potentially optimize sustained virologic response rates, especially in populations at high risk for nonadherence.

(J Med Internet Res 2020;22(4):e15532) doi:10.2196/15532

KEYWORDS

chronic hepatitis C; HCV; compliance; sustained virologic response; antivirals; sofosbuvir

Introduction

Background

The World Health Organization estimated that in 2015, 71 million people worldwide were living with chronic hepatitis C virus (HCV) infection. Although recent studies report that the incidence of HCV has decreased during the past 50 years [1-3], estimates obtained from modeling suggest that in 2015, there were still 1.75 million new HCV infections worldwide—a global incidence rate of 23.7 per 100,000 [4]. It is estimated that up to

20% to 30% of individuals with chronic HCV infection will develop liver cirrhosis with associated complications, including hepatocellular carcinoma, liver decompensation, and increased mortality [4,5]. In addition, there has been a recent steep increase in HCV diagnoses in the United States, owing, in part, to the opioid epidemic [6]. Treatment for chronic HCV infection has evolved rapidly over the past 5 years, with new all-oral, single-tablet, direct-acting antiviral (DAA) agents. Among the currently available medications is a once-daily, fixed-dose combination of ledipasvir and sofosbuvir (LDV/SOF) for the treatment of chronic HCV genotype 1 (GT-1) infection [7].

A recent clinical trial investigating the efficacy of an LDV/SOF combined formulation, conducted in a real-world setting, showed high response rates in the per-protocol population following 8 weeks (98.2%) and 12 weeks (98.3%) but notably lower response rates in the intent-to-treat (ITT) population following 8 weeks (84.6%) and 12 weeks (85.1%) [8]. The differences between per-protocol and ITT outcomes highlight the importance of adherence and patient engagement to HCV therapy as an essential element in achieving sustained virologic response (SVR) or cure. In addition, in special populations, such as people who inject drugs, SVR rates are lower and are associated with better compliance [9].

A novel digital medicine program, Proteus Discover (Proteus Digital Health), evaluates medication adherence through wirelessly observed therapy and addresses these limitations [10]. The digital medicine program directly measures medication ingestion adherence, heart rate, physical activity, and other biometrics. It then provides real-time feedback to patients and health care providers via mobile devices and a dedicated Web portal to support patient self-management and facilitate therapy optimization by the provider.

Objective

This study aimed to evaluate medication ingestion adherence and virologic outcomes in patients with chronic HCV infection who were prescribed LDV/SOF along with the digital medicine program for wirelessly observed therapy. Safety and patient satisfaction related to digital medicine program were also assessed.

Methods

Study Design

This was a 24-week, single-arm, prospective, open-label, pilot study designed to assess real-world adherence to LDV/SOF treatment among patients with HCV infection using the digital medicine program. The study was conducted at 2 US clinical research sites from August 25, 2015, to November 9, 2017. Potential patients who met the inclusion criteria were identified through electronic medical records and contacted to assess their interest in study participation.

An independent investigational review board (E&I) approved the study protocol. All aspects of this study were conducted in accordance with the US Food and Drug Administration (FDA) regulations, the International Council for Harmonisation E6 (R1) guideline for Good Clinical Practice, and applicable local, state, and federal laws.

Patients

The study enrolled adult patients diagnosed with chronic HCV infection to be initiated on fixed-dose LDV/SOF. Patients were included if they had HCV viremia greater than 50,000 IU/mL, capacity to use a smartphone or tablet (assessed by the investigators), and adequate data connectivity at home via cellular service or a secure wireless internet network. Patients were excluded from the study if any of the following criteria were met: history of skin sensitivity to adhesives, history of acute or chronic dermatitis, decompensated cirrhosis, liver transplant candidate, lack of insurance coverage for fixed-dose LDV/SOF, BMI greater than 40 kg/m^2, and currently known to be pregnant or nursing an infant. For women of childbearing potential, the following exclusion criteria were considered: not using an acceptable form of contraception for at least 3 months before the start of the study and throughout the study, lactating, current participation in another clinical trial, terminal illness (≤1 year of life anticipated), inability to swallow pills, or any condition that in the investigator's opinion could preclude safe participation in the study. All patients signed a written informed consent form before entering the study.

Digital Medicine Program

The digital medicine program consists of 3 components that function together: (1) LDV/SOF that was individually repackaged with FDA-approved ingestible sensors by a specialty pharmacy; (2) an FDA-approved wearable sensor patch, which is worn by the patient over the left upper quadrant and collects time-stamped medication ingestion events and physiologic metrics (eg, steps, activity, rest, and heart rate); (3) a mobile app and software that calculates and summarizes adherence patterns, physical activity, rest, and other self-entered clinical data that patients can view through a mobile device and providers can view through a secure Web portal. After being swallowed, the ingestible sensor activates after 2 to 3 min, sending a brief signal to the patch before passing through the body naturally. The digital medicine program directly measures medication ingestion adherence, providing objective data to patients and providers.

Procedures

At the screening visit, patients' medical records were reassessed to confirm eligibility, baseline laboratory values were taken (per usual care), signed informed consent was obtained, and a urine sample for pregnancy testing was collected. Patients received hands-on training on using the digital medicine program and were instructed to change the sensor patch every 7 days or sooner, as determined by the sensor. Patients were taught to review the app daily to obtain feedback on their medication adherence. Patients used the digital medicine program for the duration of their 8- or 12-week treatment period. The determination of treatment duration was based on the patient's baseline liver severity and HCV viral load as well as prior therapy.

Throughout the treatment period, a provider reviewed patients' adherence data from a secure Web portal at their discretion. If a missed dose was detected, the patient was contacted by phone to account for the missing dose and was provided adherence

counseling if needed. Nonadherent patients were asked to return to the clinic at weeks 4 or 6 for laboratory assessment and adherence counseling. Patients with high levels of adherence were not required to return to the clinic unless the provider wanted to assess other laboratory examinations, such as liver function tests. HCV viral load was assessed at the end of the treatment (8 or 12 weeks, depending on patient treatment plan) and at 12 weeks or more after the treatment ended. Patients also completed a satisfaction survey at the end of the treatment. All clinic procedures were performed per usual care except for the digital medicine program intervention, which prompted adherence interventions. Any technical issues were handled by Proteus Digital Health, and clinical issues were forwarded to the site for evaluation.

Outcomes

The primary end point of the study was the percentage of subjects with greater than or equal to 95% medication ingestion adherence. Ingestion nonadherence was defined as a failure to trigger iPad (Apple Inc) activation within the 4-hour window surrounding the scheduled dose, that is, the number of medication ingestions measured by the digital medicine program plus same-day patient confirmation of ingestions reported to the health care provider over the total expected number of ingestions during the treatment period. Other outcomes included virologic response, medical interventions made, safety of the digital medicine program, and patient satisfaction.

Statistical Analysis

Owing to the pilot nature of the study, a priori sample sizes were determined. All analyses were performed on the ITT population, defined as all patients who received at least one dose of the study medication. All end points were presented descriptively. Categorical variables were described as proportions, whereas continuous variables were described as mean (SD) or median (IQR) where appropriate. Multivariate linear regressions were conducted and used to assess demographic, clinical, and behavioral predictors of adherence using R version 3.5.1 (The R Foundation for Statistical Computing).

Results

Treatment Cohort

A total of 31 patients were screened, and of these patients, 28 met the inclusion criteria and were enrolled in the study (ITT population). All 28 patients (mean 59 years, SD 7; 61% [17/28] male; 61% [17/28] non-Caucasians; 93% [26/28] treatment naïve) had HCV GT-1. Moreover, 1 patient withdrew from the study 21 days after initiating digital medicine program use, and 27 patients completed digital medicine program use for 8 weeks (n=10) or 12 weeks (n=17) as prescribed. In addition, 13 (46%) patients had psychiatric comorbidities, and 9 (32%) of the patients had a history of drug abuse. Table 1 presents the baseline characteristics of enrolled patients.

Table 1. Baseline demographics and characteristics (N=28).

Characteristic	Value
Male, n (%)	17 (61)
Age (years), mean (SD)	59 (7)
Genotype, n (%)	
1a	25 (89)
1b	3 (11)
Race, n (%)	
Caucasian	11 (39)
African American	10 (36)
Hispanic/Latino	4 (14)
American Indian/Alaska Native and other	3 (11)
Cirrhosis, n (%)	1 (4)
Prior treatment history, n (%)	
Treatment naïve	26 (93)
Previous treatment relapse	2 (7)
Frequency of mobile phone use per week, n (%)	
<1 time	3 (11)
2-4 times	4 (14)
5-7 times	21 (75)
Education level, n (%)	
Less than high school	2 (7)
Some high school or high school graduate	12 (43)
Some college or college graduate	13 (46)
Postgraduate education	1 (4)
Income level (US $/year), n (%)	
<25,000	23 (82)
≥25,000	5 (18)
Psychiatric comorbidity, n (%)	13 (46)
Prior or current drug abuse history, n (%)	9 (32)

Adherence Outcomes

High rates of medication ingestion adherence were observed, with 89% of patients achieving greater than or equal to 95% adherence. Patients were connected to the digital medicine program for 92% of expected days, with an overall mean adherence rate of 97%. Using a stepwise approach to select variables to include in the model, multivariate linear regression showed that being African American, having high school education or less, and having a psychiatric comorbidity were found to be consistently significant predictors of nonadherence (Table 2). Mean adherence was lower in African Americans than others (88% vs 96%), those with high school education vs those with less (91% vs 96%), and individuals with a psychiatric comorbidity vs those without a psychiatric comorbidity (90% vs 96%). There was no difference in adherence between patients treated for 8 weeks vs 12 weeks.

Table 2. Predictors of adherence using stepwise multivariate linear regression.

Factors	Model coefficient[a]
Race African American vs other	−8.8
Education high school graduate or less	−5.7
Comorbidities psychiatric comorbidity	−5.4

[a]Only statistically significant factors (P<.05) are included.

Efficacy Outcomes

Virologic results at the end of the treatment were available for 21 patients, all of whom achieved an end-of-treatment response. All 28 subjects were assessed at 12 weeks or more posttreatment. Among those subjects, 26 (93%) achieved SVR, including 2 treatment-experienced patients who had previously failed treatment. Of the 2 subjects who did not achieve SVR, one had documented suboptimal adherence (<90%) and the other had greater than or equal to 95% adherence, suggesting viral resistance as the cause of relapse.

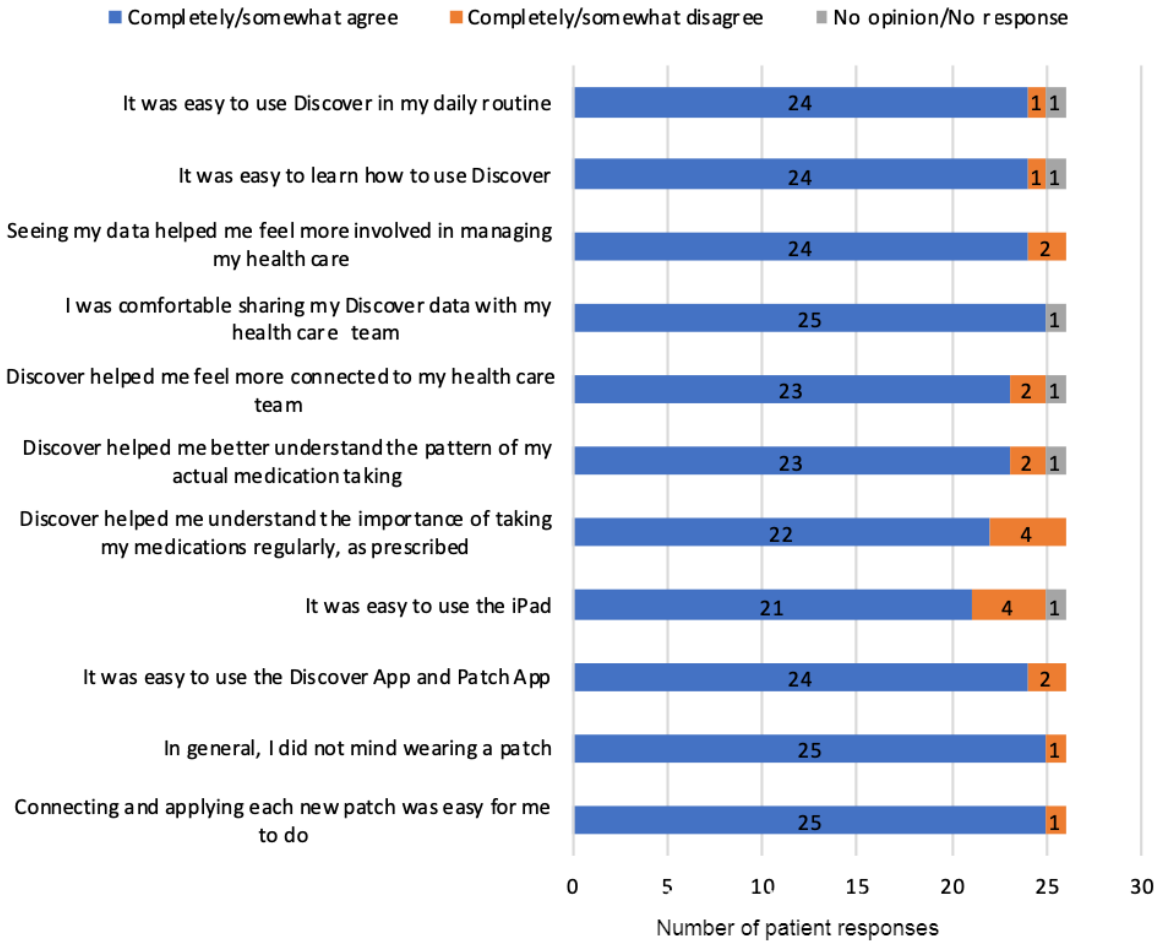
Provider Medication Interventions Made

Health care providers used the digital medicine program data for timely (same-day) assessment of medication ingestion adherence. A total of 130 documented initial missed doses occurred among 11 (39%) patients. Interventions made to mitigate nonadherence among these patients included 75 follow-up phone calls, 21 adherence counseling events, and 5 follow-up visits; some patients required multiple interventions per missed dose. No action was performed for the remaining 35 events (eg, patient unreachable).

Patient Satisfaction

Overall satisfaction with the use of the device was high among study subjects (Figure 1). Most (92%) patients completely or somewhat agreed that the DMP was easy to use in their daily routine and helped them feel more involved in managing their health care, 96% did not mind wearing the patch and sharing their data with their health care team, and 85% agreed that the DMP helped them understand the importance of taking their medications regularly.

Figure 1. Patient satisfaction survey results (n=26).



Safety Findings

A total of 4 nonserious, device-related adverse events were reported in 4 subjects during the study. These adverse events were reflective of commonly reported symptoms of irritant contact dermatitis (rash). Of these, 1 reported adverse event (itching) was assessed by the investigator as severe and resolved without any action taken with the DMP device. The remaining 3 reported adverse events of rash were assessed by the investigator as mild in severity and resolved.

Discussion

Principal Findings

This real-world pilot study in a cohort of patients with chronic HCV infection demonstrated that the use of the DMP with same-day interventions (when needed) prompted high rates of medication adherence (97%), despite the presence of multiple risk factors for nonadherence, such as race, psychiatric comorbidities, and substance use or abuse. More importantly, SVR at 12 weeks or more was achieved in 92.8% (26/28) of

the patients. Patients were connected to the DMP 92% of the time, indicating high patient engagement, and the system was easy to use even in those patients with historically low mobile phone use. In addition, the DMP had a favorable safety profile and provided data to facilitate timely adherence interventions. Thus, the unique ability of the DMP to provide objective adherence data enabled timely medication interventions aimed at improving adherence and optimizing outcomes, even in high-risk patients who have previously failed therapy.

Large clinical trials have demonstrated that treatment with once-daily, combined, fixed-dose LDV/SOF therapy is effective in achieving SVR in a vast majority of patients with chronic HCV infection [8,11-13]. However, adherence to prescribed treatment regimens is critical to achieving treatment success [14]. Nonadherence to HCV DAA regimens has been shown to be associated with a lower rate of SVR [12], particularly in patients with psychiatric illness [15]. Current indirect assessment methods (eg, medication possession ratios, patient self-report, pill counts, and electronic pill bottles) may overestimate actual medication adherence [16]. In contrast, DMP addresses the inherent limitations of these methods through direct objective measurement of medication ingestion. In addition, and perhaps more importantly, it allows real-time digital intervention by patients and/or clinicians when doses are not recorded as *ingested*.

High levels of adherence have been reported in large clinical trials using various measurement assessments, including electronic medication event monitoring (MEM) caps, pill counts, and patient reports [12,13]. However, the accuracy of the current measurement methods may be questionable because of the limitations of each. For example, patients may purposefully mislead the MEM system by accidentally actuating the container without taking the medication [17]. Intentional manipulation of pill counts (*pill dumping*) and falsified patient reports to appear adherent must also be considered when measuring medication adherence [17]. None of these methods allow the patient or provider an opportunity to respond quickly or proactively to perceived missed doses [18,19]. Although measurement of the proportion of days covered is commonly used to assess long-term adherence in patients with chronic illnesses [20], the use of prescription refill data does not allow for timely detection of nonadherence, and it does not necessarily indicate that all the medications have been taken [21].

Medication adherence is of particular concern among patients with HCV treated with fixed-dose LDV/SOF therapy, many of whom are at high risk for nonadherence because of transient living situations, depression, neurocognitive impairment, psychiatric comorbidities, and concurrent alcohol and other substance abuse [22,23]. As successful treatment is dependent on strict adherence to the prescribed medication regimen, nonadherence may result in the same poor clinical outcomes (eg, liver cirrhosis, hepatocellular carcinoma, liver decompensation, and increased mortality) associated with nontreatment [4,5].

The DMP has been studied in multiple therapeutic areas, including cardiovascular and infectious diseases. In a randomized controlled trial of the DMP compared with usual care in patients with hypertension and diabetes, those using the DMP with their prescribed medications experienced greater reductions in systolic blood pressure (BP; mean change −24.6 vs −15.2; mean difference −9.4; SE 2.7; 95% CI −14.6 to −4.2 mm Hg) and glycohemoglobin (mean change −0.08% vs 0.28%; mean difference −0.57%; 95% CI −1.53 to 0.39) at 12 weeks [24]. In addition, a significantly higher number of subjects reached their BP goal in the DMP group compared with those in usual care after 12 weeks (98% vs 52%; mean difference 46%; 95% CI 7.1% to 84.5%). The DMP was also studied as wirelessly observed therapy in the setting of tuberculosis treatment, demonstrating a positive detection accuracy for wirelessly observed therapy of 98.4% (95% CI 97.5% to 99%) and confirming 54% more doses than even the previous gold standard of directly observed therapy [25]. The accuracy and objectivity of this DMP at quantifying and recording medication adherence have been firmly established in the literature [26].

Our results are similar to those reported in recent adherence and efficacy studies [8,11-13]. However, deriving correlations between adherence and treatment effect is problematic. For example, Petersen et al [13] reported 97.6% adherence as measured by an MEM system but provided no data regarding treatment effect. Conversely, neither of the 2 efficacy studies reported medication adherence data [8,11]. In addition, a recent study highlighted the ineffectiveness of low-cost reminder devices in improving adherence among nonadherent patients with chronic conditions, suggesting the need for objectivity, effectual interventions, and personalized feedback [27].

Strengths and Limitations

A key strength of our study was the ability to directly measure medication ingestion, providing objective adherence data to better inform medical decision making. Although the small study population does not allow for definitive assessment of correlations between adherence and treatment effect, positive associations can be inferred from the timely interventions on missed doses, thereby improving adherence, which is expected to result in SVR. By extension, another positive point is the ability to improve patient care by contacting patients in real time and proactively assisting with missed doses.

Our study is limited by the single-arm design that precludes direct comparison with a usual care group. Furthermore, because of the small sample size, our results may not accurately reflect the broader population of patients with chronic HCV infection [28]. A larger multicenter clinical trial is currently underway to confirm and further elucidate these findings.

As reported by Moreno et al [29], improving patient access to HCV treatment will likely yield significant cost reductions for payers, accrued from the long-term reduction in prevalent and incident cases, mortality, and medical costs. Technology such as DMP could ensure that more patients with risk factors for nonadherence can complete therapy to ensure maximal real-world SVR rates. As reported in this study, although 6 patients failed to attend the end-of-treatment assessment for unknown reasons, all 28 patients were assessed at the 12-week or more evaluation period, where 26 (93%) achieved SVR. Patients were connected to the digital medicine program 92% of the expected days, indicating high patient engagement, and

the system was easy to use even in those patients with historically low mobile phone use. In addition, the digital medicine program had a favorable safety profile and provided data to facilitate timely adherence interventions.

Conclusions

These data suggest that the digital medicine program can support adherence to therapy, enhance patient engagement, reduce unnecessary medication wastage, and optimize SVR rates in patients with chronic HCV infection, including those with multiple risk factors for nonadherence and in those who have previously failed therapy.

Acknowledgments

This study was funded by Proteus Digital Health.

Authors' Contributions

MB, YK, and CL contributed to the conception and design of the study. All authors contributed to data acquisition and analysis, critical review, scientific input, and final approval, and approved the final version of the paper.

Conflicts of Interest

YK and MT are employees of Proteus Digital Health. MB has received research support from AbbVie, Allergan, Assembly Biosciences, Boehringer-Ingelheim, Intercept, Genfit, Gilead, Novartis, and Pfizer and has received speaker fees from AbbVie, Gilead, and Intercept. CL received research funding from Gilead, AbbVie, and Proteus. CP and LM report no disclosures.

References

1. Bruggmann P, Berg T, Øvrehus AL, Moreno C, Mello CE, Roudot-Thoraval F, et al. Historical epidemiology of hepatitis C virus (HCV) in selected countries. J Viral Hepat 2014 May;21(Suppl 1):5-33. [doi: 10.1111/jvh.12247] [Medline: 24713004]
2. Saraswat V, Norris S, de Knegt RJ, Sanchez Avila JF, Sonderup M, Zuckerman E, et al. Historical epidemiology of hepatitis C virus (HCV) in select countries - volume 2. J Viral Hepat 2015 Jan;22(Suppl 1):6-25. [doi: 10.1111/jvh.12350] [Medline: 25560839]
3. Liakina V, Hamid S, Tanaka J, Olafsson S, Sharara AI, Alavian SM, et al. Historical epidemiology of hepatitis C virus (HCV) in select countries - volume 3. J Viral Hepat 2015 Dec;22(Suppl 4):4-20. [doi: 10.1111/jvh.12475] [Medline: 26513445]
4. World Health Organization. Global Hepatitis Report, 2017. Geneva: World Health Organization; 2017.
5. Sangiovanni A, Prati GM, Fasani P, Ronchi G, Romeo R, Manini M, et al. The natural history of compensated cirrhosis due to hepatitis C virus: A 17-year cohort study of 214 patients. Hepatology 2006 Jun;43(6):1303-1310. [doi: 10.1002/hep.21176] [Medline: 16729298]
6. Zibbell JE, Asher AK, Patel RC, Kupronis B, Iqbal K, Ward JW, et al. Increases in acute hepatitis C virus infection related to a growing opioid epidemic and associated injection drug use, United States, 2004 to 2014. Am J Public Health 2018 Feb;108(2):175-181. [doi: 10.2105/AJPH.2017.304132] [Medline: 29267061]
7. Harvoni (R) (ledipasvir and sofosbuvir) package insert. Accessed June 11, 2019 URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/212477s000lbl.pdf
8. Buggisch P, Vermehren J, Mauss S, Günther R, Schott E, Pathil A, et al. Real-world effectiveness of 8-week treatment with ledipasvir/sofosbuvir in chronic hepatitis C. J Hepatol 2018 Apr;68(4):663-671. [doi: 10.1016/j.jhep.2017.11.009] [Medline: 29133244]
9. Rosenthal E, Hill K, Nussdorf L, Mathur P, Gross C, Silk R, et al. Collocation of Buprenorphine with HCV treatment to improve adherence and reduce harm in PWID with HCV: Preliminary data from the ANCHOR study. J Hepatology 2018;68:S51-S13 [FREE Full text] [doi: 10.1016/s0168-8278(18)30322-2]
10. Browne SH, Peloquin C, Santillo F, Haubrich R, Muttera L, Moser K, et al. Digitizing medicines for remote capture of oral medication adherence using co-encapsulation. Clin Pharmacol Ther 2018 Mar;103(3):502-510 [FREE Full text] [doi: 10.1002/cpt.760] [Medline: 28597911]
11. Curry MP, Tapper EB, Bacon B, Dieterich D, Flamm SL, Guest L, et al. Effectiveness of 8- or 12-weeks of ledipasvir and sofosbuvir in real-world treatment-naïve, genotype 1 hepatitis C infected patients. Aliment Pharmacol Ther 2017 Sep;46(5):540-548 [FREE Full text] [doi: 10.1111/apt.14204] [Medline: 28691377]
12. Younossi ZM, Stepanova M, Henry L, Nader F, Younossi Y, Hunt S. Adherence to treatment of chronic hepatitis C: from interferon containing regimens to interferon and ribavirin free regimens. Medicine (Baltimore) 2016 Jul;95(28):e4151 [FREE Full text] [doi: 10.1097/MD.0000000000004151] [Medline: 27428205]

13. Petersen T, Townsend K, Gordon LA, Sidharthan S, Silk R, Nelson A, et al. High adherence to all-oral directly acting antiviral HCV therapy among an inner-city patient population in a phase 2a study. Hepatol Int 2016 Mar;10(2):310-319 [FREE Full text] [doi: 10.1007/s12072-015-9680-7] [Medline: 26612014]
14. McHutchison JG, Poynard T, Esteban-Mur R, Davis GL, Goodman ZD, Harvey J, International Hepatitis Interventional Therapy Group. Hepatic HCV RNA before and after treatment with interferon alone or combined with ribavirin. Hepatology 2002 Mar;35(3):688-693. [doi: 10.1053/jhep.2002.31870] [Medline: 11870385]
15. Altstadt AR, Holznecht C, Rahman SI, Rein L, Szabo A, Saeian K. Non-adherence diminishes real-world SVR12 in an inter-disciplinary care clinic. Gastroenterology 2016;150(4):S1039 [FREE Full text] [doi: 10.1016/s0016-5085(16)33512-0]
16. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. Annu Rev Pharmacol Toxicol 2012;52:275-301. [doi: 10.1146/annurev-pharmtox-011711-113247] [Medline: 21942628]
17. Lam WY, Fresco P. Medication adherence measures: an overview. Biomed Res Int 2015;2015:217047 [FREE Full text] [doi: 10.1155/2015/217047] [Medline: 26539470]
18. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med 2006 May 16;144(10):705-714. [doi: 10.7326/0003-4819-144-10-200605160-00004] [Medline: 16702586]
19. Veterans Administration Office of Public Health/Population Health and VA Pharmacy Benefits Management Services. 2018 Aug. Report: Treating Hepatitis C in Veterans Affairs (VA)arly Experience with Sofosbuvir-based Regimes URL: https://www.hepatitis.va.gov/pdf/treatment-considerations-2018-08-27.pdf
20. Zhu VJ, Tu W, Rosenman MB, Overhage JM. A comparison of data driven-based measures of adherence to oral hypoglycemic agents in Medicaid patients. AMIA Annu Symp Proc 2014;2014:1294-1301 [FREE Full text] [Medline: 25954441]
21. Weiss JJ, Bräu N, Stivala A, Swan T, Fishbein D. Review article: adherence to medication for chronic hepatitis C - building on the model of human immunodeficiency virus antiretroviral adherence research. Aliment Pharmacol Ther 2009 Jul;30(1):14-27 [FREE Full text] [doi: 10.1111/j.1365-2036.2009.04004.x] [Medline: 19416131]
22. Wagner GJ, Ryan GW. Hepatitis C virus treatment decision-making in the context of HIV co-infection: the role of medical, behavioral and mental health factors in assessing treatment readiness. AIDS 2005 Oct;19(Suppl 3):S190-S198. [doi: 10.1097/01.aids.0000192089.54130.b6] [Medline: 16251817]
23. Martín-Santos R, Díez-Quevedo C, Castellví P, Navinés R, Miquel M, Masnou H, et al. De novo depression and anxiety disorders and influence on adherence during peginterferon-alpha-2a and ribavirin treatment in patients with hepatitis C. Aliment Pharmacol Ther 2008 Feb 1;27(3):257-265 [FREE Full text] [doi: 10.1111/j.1365-2036.2007.03568.x] [Medline: 17988237]
24. Frias J, Virdi N, Raja P, Kim Y, Savage G, Osterberg L. Effectiveness of digital medicines to improve clinical outcomes in patients with uncontrolled hypertension and type 2 diabetes: prospective, open-label, cluster-randomized pilot clinical trial. J Med Internet Res 2017 Jul 11;19(7):e246 [FREE Full text] [doi: 10.2196/jmir.7833] [Medline: 28698169]
25. Browne S, Moser K, Low J, Vaida F, Tucker-Maytom A, Gonzalez-Garcia J. Wirelessly observed therapy is accurate and confirms more TB medication doses than directly observed therapy. Int J Tuberc Lung Dis 2017;21(11, Suppl2):S65 [FREE Full text]
26. Hafezi H, Robertson TL, Moon GD, Au-Yeung K, Zdeblick MJ, Savage GM. An ingestible sensor for measuring medication adherence. IEEE Trans Biomed Eng 2015 Jan;62(1):99-109. [doi: 10.1109/TBME.2014.2341272] [Medline: 25069107]
27. Choudhry NK, Krumme AA, Ercole PM, Girdish C, Tong AY, Khan NF, et al. Effect of reminder devices on medication adherence: The REMIND randomized clinical trial. JAMA Intern Med 2017 May 1;177(5):624-631 [FREE Full text] [doi: 10.1001/jamainternmed.2016.9627] [Medline: 28241271]
28. McHutchison JG, Manns M, Patel K, Poynard T, Lindsay KL, Trepo C, International Hepatitis Interventional Therapy Group. Adherence to combination therapy enhances sustained response in genotype-1-infected patients with chronic hepatitis C. Gastroenterology 2002 Oct;123(4):1061-1069. [doi: 10.1053/gast.2002.35950] [Medline: 12360468]
29. Moreno GA, Wang A, González YS, Espinosa OD, Vania DK, Edlin BR, et al. Value of comprehensive HCV treatment among vulnerable, high-risk populations. Value Health 2017 Jun;20(6):736-744 [FREE Full text] [doi: 10.1016/j.jval.2017.01.015] [Medline: 28577690]

Abbreviations

BP: blood pressure
DAA: direct-acting antiviral
FDA: Food and Drug Administration
GT-1: genotype 1
HCV: hepatitis C virus
ITT: intent-to-treat
LDV/SOF: ledipasvir/sofosbuvir
MEM: medication event monitoring

SVR: sustained virologic response

Edited by G Eysenbach; submitted 17.07.19; peer-reviewed by R Hart-Malloy, D Smith, J Opoku; comments to author 18.09.19; revised version received 07.10.19; accepted 20.10.19; published 30.04.20.

Please cite as:
Bonacini M, Kim Y, Pitney C, McKoin L, Tran M, Landis C
Wirelessly Observed Therapy to Optimize Adherence and Target Interventions for Oral Hepatitis C Treatment: Observational Pilot Study
J Med Internet Res 2020;22(4):e15532
URL: https://www.jmir.org/2020/4/e15532
doi:10.2196/15532
PMID:32352385

Original Paper

Use of Telephone and Digital Channels to Engage Socioeconomically Disadvantaged Adults in Health Disparities Research Within a Social Service Setting: Cross-Sectional Study

Kassandra I Alcaraz[1], PhD, MPH; Rhyan N Vereen[1], MPH; Donna Burnham[2], BA

[1]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, United States

[2]United Way of Greater Atlanta, Atlanta, GA, United States

Corresponding Author:
Kassandra I Alcaraz, PhD, MPH
Behavioral and Epidemiology Research Group
American Cancer Society
250 Williams Street NW
Atlanta, GA, 30303
United States
Phone: 1 4044178019
Email: kassandra.alcaraz@cancer.org

Abstract

Background: Engaging socioeconomically disadvantaged populations in health research is vital to understanding and, ultimately, eliminating health-related disparities. Digital communication channels are increasingly used to recruit study participants, and recent trends indicate a growing need to partner with the social service sector to improve population health. However, few studies have recruited participants from social service settings using multiple digital channels.

Objective: This study aimed to recruit and survey 3791 adult clients of a social service organization via telephone and digital channels. This paper aimed to describe recruitment outcomes across five channels and compare participant characteristics by recruitment channel type.

Methods: The Cancer Communication Channels in Context Study recruited and surveyed adult clients of 2-1-1, a social service–focused information and referral system, using five channels: telephone, website, text message, web-based live chat, and email. Participants completed surveys administered either by phone (if recruited by phone) or on the web (if recruited from digital channels, ie, website, text message, Web-based live chat, or email). Measures for the current analysis included demographic and health characteristics.

Results: A total of 3293 participants were recruited, with 1907 recruited by phone and 1386 recruited from digital channels. Those recruited by phone had a moderate study eligibility rate (42.23%) and the highest survey completion rate (91.24%) of all channels. Individuals recruited by text message had a high study eligibility rate (94.14%) yet the lowest survey completion rate (74.0%) of all channels. Sample accrual goals were achieved for phone, text message, and website recruitment. Multivariable analyses found differences in participant characteristics by recruitment channel type. Compared with participants recruited by phone, those recruited from digital channels were younger (adjusted odds ratio [aOR] 0.96, 95% CI 0.96-0.97) and more likely to be female (aOR 1.52, 95% CI 1.23-1.88), married (aOR 1.52, 95% CI 1.22-1.89), and other than non-Hispanic black (aOR 1.48, 95% CI 1.22-1.79). Those recruited via phone also were more likely to have more than a high school education (aOR 2.17, 95% CI 1.67-2.82), have a household income ≥US $25,000 a year (aOR 2.02, 95% CI 1.56-2.61), and have children living in the home (aOR 1.26, 95% CI 1.06-1.51). Additionally, participants recruited from digital channels were less likely than those recruited by phone to have public health insurance (aOR 0.75, 95% CI 0.62-0.90) and more likely to report better overall health (aOR 1.52, 95% CI 1.27-1.83 for good-to-excellent health).

Conclusions: Findings indicate the feasibility and utility of recruiting socioeconomically disadvantaged adults from the social service sector using multiple communication channels, including digital channels. As social service–based health research evolves, strategic recruitment using a combination of traditional and digital channels may be warranted to avoid underrepresentation of highly medically vulnerable individuals, which could exacerbate disparities in health.

(J Med Internet Res 2020;22(4):e16680) doi:10.2196/16680

KEYWORDS

cross-sectional studies; electronic mail; health status disparities; health care disparities; internet; mobile phone; telephone; text messaging; social services

Introduction

Digital Communication and Disparities

Socioeconomically disadvantaged populations bear a disproportionate burden of disease for both infectious and chronic conditions [1-4], yet they remain underrepresented in health research [5,6]. Persistent underrepresentation of populations suffering from disparities in health hinders progress in understanding and eliminating these disparities [7]. Increasingly, digital communication channels such as text message or social media are used to engage individuals in health research [8-11]. Recruitment using digital channels can overcome some of the limitations of traditional recruitment channels (eg, telephone) and has the potential to narrow disparities in health [12,13]. However, evidence-informed strategies are needed to maximally leverage digital technologies for health disparities reduction [13].

Digital communication technologies such as smartphones are increasingly accessible across sociodemographic groups [14-16], including homeless adults [17]. Despite this growing access to digital technology, inequities in technology access continue to be documented among medically vulnerable populations such as low-income individuals [15] and residents of rural communities [16]. Ensuring representation of medically vulnerable populations when recruiting research participants using digital channels, therefore, presents an ongoing challenge.

Use of Digital Channels in Health Research

Some research suggests socioeconomically disadvantaged populations can be successfully recruited using digital channels, while other findings indicate bias in samples recruited using digital channels. For example, results from a trial of Quit4Baby (a text message–based smoking cessation intervention for pregnant women) demonstrated the feasibility of recruiting high proportions of low-income, unemployed, and publicly insured participants via text message [18]. However, a study comparing characteristics of 12,280 *eCohort* participants recruited on the web to the US population found that participants were more likely to have a college education, less likely to be from racial or ethnic minority groups, and more likely to be in excellent general health [19]. Similarly, a health study employing multichannel recruitment (eg, flyer, email, Facebook, website) found that none of the channels were successful in recruiting individuals of a low socioeconomic status, those from racial or ethnic minority groups, or men [20].

Several trends suggest digital tools will continue to be used in diverse settings to improve population health. First, recent evidence supports the acceptability, feasibility, and efficacy of digital interventions for behavior change [21-27]. Second, also documented is the promise of digital technologies to reduce health care disparities [28]. Third, in 2019, the National Academies of Sciences, Engineering, and Medicine released a report reflecting the growing integration of social care into health care delivery [29]. These trends indicate an emerging demand not only for more health disparities research using digital tools but also for more engagement with the social service sector to enhance health. In recent years, a growing number of researchers have successfully reached and recruited socioeconomically disadvantaged adults for health disparities research through social service organizations [30-37]. However, to our knowledge, scant research to date has employed multiple digital channels to recruit or survey study participants from these types of settings.

Study Aims

As availability and use of digital channels are increasing, evaluations of web-based recruitment strategies are needed to better understand their effectiveness and potential biases for use in research [38]. Given current trends and needs, understanding how digital channels can be used to engage individuals for health disparities research can enhance research planning. The objective of this study was to recruit and survey a community-based sample of 3791 socioeconomically disadvantaged adults from a social service setting using telephone and digital channels. We also sought to examine recruitment outcomes by channel and participant characteristics by recruitment channel type. We hypothesized that recruitment success would vary across channels and that participant characteristics would vary across the two recruitment channel types (ie, telephone vs digital channels). This paper aimed to present recruitment outcomes for the study and discuss implications for reaching medically vulnerable populations in a social service setting.

Methods

Setting

This study was conducted in partnership with 2-1-1, a nationally designated, locally administered information and referral system that connects individuals with resources to meet their basic human and social needs (eg, food, safety). As of 2019, 2-1-1 has been made available to 94.6% of the US population [39] and throughout most of Canada [40]. Where available, individuals can dial 2-1-1 from their phone to request and obtain referrals for services in their local community. Data indicate callers to 2-1-1 are disproportionately low income, uninsured, and unemployed and have high health needs such as for smoking cessation or cancer screening [36,41]. In some communities, 2-1-1 can be reached using digital communication channels such as email or text message.

Study Overview

This study has reported data from the Cancer Communication Channels in Context (4C) Study, a cross-sectional study that administered a survey to clients of 2-1-1. Survey data from the 4C Study will inform targeted strategies for connecting socioeconomically disadvantaged populations with health- and cancer-related information, programs, and resources. Participants

were recruited from United Way 2-1-1 of Greater Atlanta, which was the first 2-1-1 established in the United States. This 2-1-1 Contact Center receives more than 590,000 contacts annually. Individuals can access this Contact Center via telephone, text message, the 2-1-1 website, web-based live chat, email, or a mobile app to request referrals such as a telephone number or website for a community resource. For this study, channels of interest were *telephone* (calling 2-1-1 to request referrals), *website* (searching the self-service 2-1-1 web database), *text message* (texting a referral request to 2-1-1), *web-based live chat* (chatting in real time with 2-1-1 staff via the internet), and *email* (emailing a referral request to 2-1-1). The 4C Study sought to recruit and survey 1895 participants via telephone and 474 via each of the four digital channels (3791 total). These target sample sizes were selected to provide adequate statistical power for primary 4C Study analyses. On the basis of 2-1-1 client volume data, we projected that a 9-month recruitment period would be needed to reach accrual goals.

Participants

Individuals were eligible for the 4C Study if they were accessing 2-1-1 for referral assistance via 1 of the 5 channels of interest (ie, telephone, website, text message, web-based live chat, email); accessing 2-1-1 from within United Way 2-1-1 of Greater Atlanta's 13-county primary service area; aged ≥21 years; and able to speak or read English. Exclusion criteria were the following: experiencing an acute crisis (eg, imminent eviction, natural disaster); accessing 2-1-1 on behalf of another person; accessing 2-1-1 in error; or performing a non-English search on the 2-1-1 website.

Recruitment

Participants were recruited from January to November 2016 by designated 2-1-1 staff who were trained to recruit for the study. Initially, 11 recruiters were designated; 6 recruiters were added in April 2016 to accelerate sample accrual. Individuals were screened for interest and eligibility for the study after receiving standard 2-1-1 service. All individuals searching for referrals on the 2-1-1 website were screened for eligibility; for the other four channels, only those individuals interacting with designated 2-1-1 staff were screened. Screening and recruitment occurred 24 hours per day, 7 days per week.

Recruitment procedures and survey administration mode were based on the communication channel an individual initially utilized to access social services through 2-1-1. Therefore, individuals who contacted 2-1-1 via phone were screened for interest and eligibility during the call. If eligible, informed consent and 4C Study survey administration were conducted immediately after providing the requested 2-1-1 referrals, that is, during the same phone call. Those accessing 2-1-1 using the web-based database received an on-screen notification asking if they were interested in a health survey (yes/no). Those who responded yes received a survey in a new tab where they were screened for eligibility; those who were eligible were directed to a web-based consent page followed by a web-based survey. Individuals who contacted 2-1-1 via chat, text message, or email were sent (via the corresponding channel they used to contact 2-1-1) a statement informing them about a health survey as well as the screener/consent/4C Study survey link. The same survey was used across both survey modes (ie, phone and the web).

Participants were mailed a US $15 gift card incentive after completing the study survey. Participants were also mailed a free resource guide listing free or low-cost health-related cancer prevention services available in their community. Study procedures were approved by the Institutional Review Board at Morehouse School of Medicine.

Measures

Demographic Characteristics

Standard demographic measures included age, sex, educational attainment, marital status, and annual household income. Presence of any children under the age of 18 years living in the home and self-reported race and ethnicity were also assessed. Due to response distribution, race and ethnicity were combined and dichotomized as non-Hispanic black vs other (Hispanic; white; Asian, Native Hawaiian, or Other Pacific Islander; American Indian or Alaskan Native; or other).

Health Characteristics

Self-rated health was measured using a standard item: "In general, would you say your health is: excellent, very good, good, fair, or poor?" [42]. To assess health insurance type, respondents were asked to choose which health insurance best describe(s) what they have to help pay their medical bills today. Participants could select more than one response, and responses were recoded into four categories for analysis: uninsured; private; government/public (Medicare, Medicaid, State Children's Health Insurance Program, Military health care, and/or another government program); or a combination of public and private insurance.

Statistical Analysis

Analyses aimed to describe recruitment outcomes by recruitment channel (ie, telephone, website, text message, web-based live chat, or email) and participant characteristics by recruitment channel type (ie, telephone or digital channels). First, we examined accrued frequencies and percentages across recruitment channels. For each channel, we computed channel efficiency as the total number of surveys completed divided by the total number of individuals encountered. Second, we compared demographic and health characteristics by recruitment channel type. Means, standard deviations, frequencies, and percentages are presented, along with results of chi-square tests or *t* tests as appropriate. Third, we conducted multivariable binary logistic regression to assess differences in demographic and health characteristics by recruitment channel type while controlling for other characteristics, using phone as the reference group. Adjusted odds ratios and 95% confidence intervals have been presented. All analyses were conducted using SAS software, version 9.4 (SAS Institute Inc).

Results

Recruitment

Sample accrual goals for the telephone and text message channels were reached in June 2016, and the accrual goal for

the website was reached in July 2016. Due to funding constraints, recruitment via web-based chat and email ended in November 2016. Figure 1 depicts the number of participants recruited by month across channels.

Figure 2 summarizes sample accrual by recruitment channel. After the exclusion of 129 duplicates, recruiters encountered a total of 100,391 2-1-1 clients. Of these, 10.74% (10,777/100,391) were eligible to participate in the study, 35.63% (3840/10,777) consented to participate, and 94.71% (3637/3840) started the survey. A total of 85.76% (3293/3840) of individuals who consented completed the survey, with 1907 recruited by telephone and 1386 recruited from digital channels.

A wide range of recruitment outcomes was observed across channels (Figure 2). Only 1.85% (1578/85,234) of individuals who were recruited via the website were eligible for the study, compared with 94.14% (3084/3276) and 91.4% (427/467) of individuals who were recruited by text message and email, respectively. Participants who were recruited by phone had the highest survey completion rate (1907/2090, 91.24%), followed by those recruited from web-based live chat (371/421, 88.1%), those recruited by email (28/35, 80%), those recruited from the website (493/626, 78.8%), and those recruited by text message (494/668, 74.0%). Only 28 participants were recruited by email, compared with 371 to 494 for the three other digital channels. Additionally, individuals recruited by email had the lowest consent rate (8.2% vs up to 54.70% for other channels).

The five recruitment channels had a wide range of channel efficiency, with recruitment by phone producing the highest proportion of completed surveys relative to individuals encountered. Specifically, channel efficiency was 21.08% (1907/9047) for phone, 15.67% (371/2367) for web-based live chat, 15.08% (494/3276) for text message, 6.0% (28/467) for email, and 0.58% (493/85,234) for the website. These findings indicate that, to ultimately obtain one completed survey, encounters with 5 individuals on average were required if recruiting by phone; encounters with 7 individuals were required if recruiting from web-based live chat or text message; encounters with 17 individuals were required if recruiting from email; and encounters with 173 individuals were required if recruiting from the website.

Figure 1. Number of completed Cancer Communication Channels in Context Study surveys per month by recruitment channel, January-November 2016 (accrual goals for phone and text message recruitment were reached in June; accrual goal for website recruitment was reached in July).

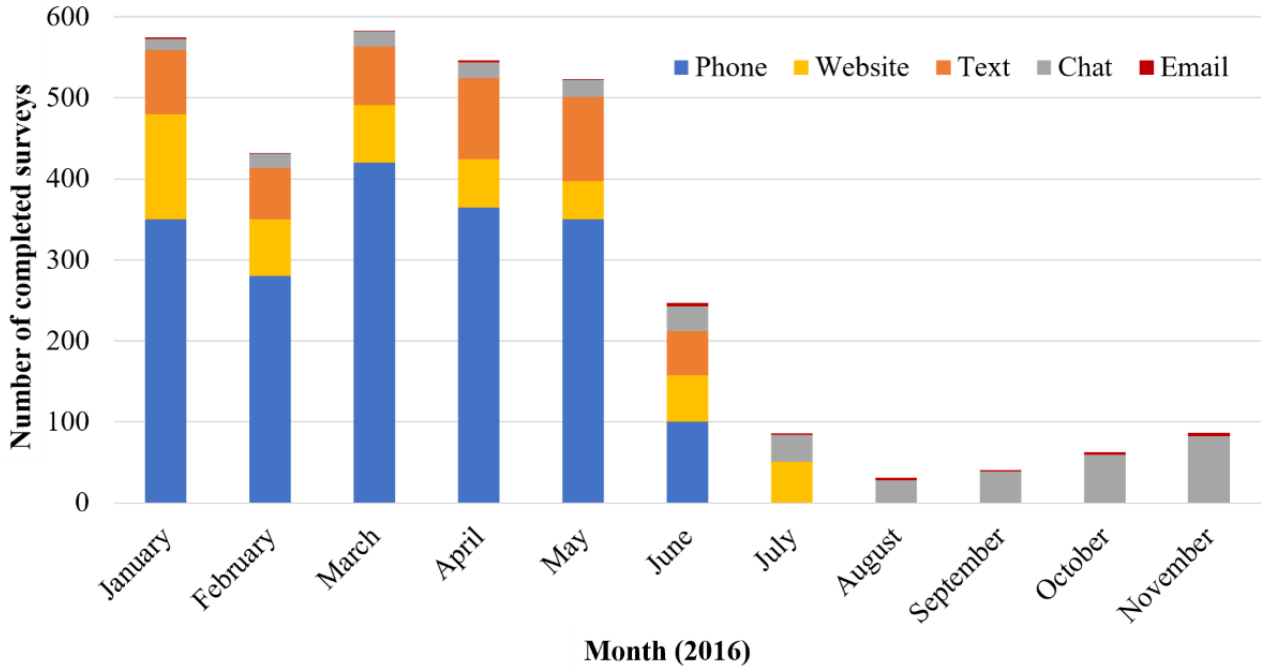
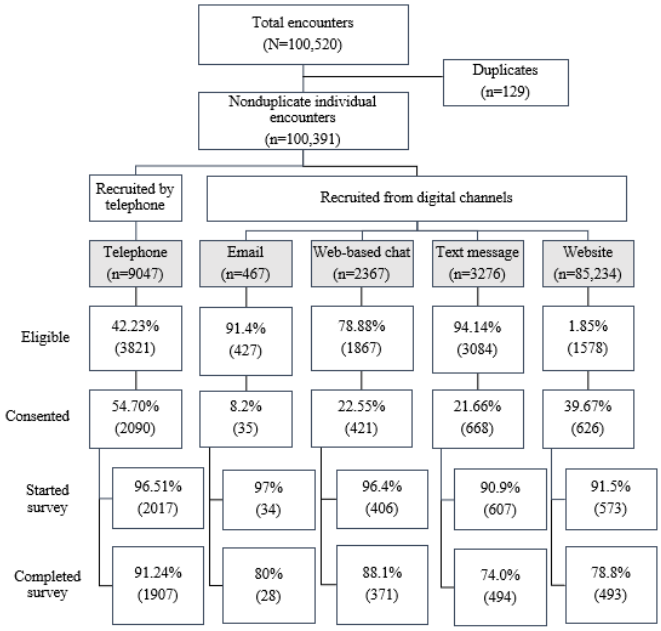


Figure 2. Cancer Communication Channels in Context Study accrual by recruitment channel, January-November 2016.



Participant Characteristics

Participant characteristics by recruitment channel type are reported in Table 1. Respondents were predominately female (2662/3293, 80.84%) and non-Hispanic black (2543/3293, 77.22%). Overall, 43.15% (1421/3293) had a high school education or less, and 37.11% (1222/3293) had an annual household income less than US $5000. About half (1566/3293, 47.56%) were never married and 54.02% (1779/3293) had children living in the home. The majority of respondents either had public insurance (1466/3293, 44.52%) or were uninsured (993/3293, 30.15%). About a third of the respondents (996/3293, 30.25%) rated their health as fair or poor.

Demographic characteristics significantly differed by recruitment channel type (Table 1). For example, 16.26% (310/1907) of participants who were recruited by phone had less than a high school education, compared with 8.66% (120/1386) of those who were recruited from digital channels (P<.001). Additionally, 41.85% (798/1907) of respondents recruited by phone had a household income less than US $5000, compared with 30.59% (424/1386) of respondents recruited from digital channels (P<.001).

Health characteristics also differed by recruitment channel type (Table 1), where 49.66% (947/1907) of respondents recruited by phone had public insurance, compared with 37.45% (519/1386) of respondents recruited from digital channels (P<.001). More than a third (706/1907, 37.02%) of respondents recruited by phone rated their health as fair or poor, compared with 20.92% (290/1386) of respondents who were recruited from digital channels (P<.001).

Table 1. Characteristics of Cancer Communication Channels in Context Study participants by recruitment channel type.

Characteristic	Total (N=3293)[a]	Recruited by telephone (n=1907)[a]	Recruited from digital channels (n=1386)[a,b]	*P* value
Age (years), mean (SD)	42.1 (12.72)	44.8 (13.26)	38.4 (10.87)	<.001
Sex, n (%)				<.001
Female	2662 (80.84)	1479 (77.56)	1183 (85.35)	
Male	631 (19.16)	428 (22.44)	203 (14.65)	
Race and ethnicity, n (%)				<.001
Non-Hispanic black	2543 (77.22)	1549 (81.23)	994 (71.72)	
Other	688 (20.89)	345 (18.09)	343 (24.75)	
Educational attainment , n (%)				<.001
Less than high school	430 (13.06)	310 (16.26)	120 (8.66)	
High school graduate or equivalent	991 (30.09)	683 (35.82)	308 (22.22)	
More than high school	1859 (56.45)	912 (47.82)	947 (68.33)	
Annual household i ncome (US $), n (%)				<.001
Less than 5000	1222 (37.11)	798 (41.85)	424 (30.59)	
5000 to 14,999	874 (26.54)	543 (28.47)	331 (23.88)	
15,000 to 24,999	546 (16.58)	290 (15.21)	256 (18.47)	
25,000 or more	500 (15.18)	196 (10.28)	304 (21.93)	
Marital status, n (%)				<.001
Never married	1566 (47.56)	906 (47.51)	660 (47.62)	
Divorced, widowed, or separated	1071 (32.52)	712 (37.34)	359 (25.90)	
Married or have a partner	631 (19.16)	284 (14.89)	347 (25.04)	
Has any children in the home, n (%)				<.001
No	1499 (45.52)	992 (52.02)	507 (36.58)	
Yes	1779 (54.02)	914 (47.93)	865 (62.41)	
Health insurance type, n (%)				<.001
Uninsured	993 (30.15)	546 (28.63)	447 (32.25)	
Public	1466 (44.52)	947 (49.66)	519 (37.45)	
Private	710 (21.56)	348 (18.25)	362 (26.12)	
Public and private	86 (2.61)	55 (2.88)	31 (2.24)	
Self-rated health, n (%)				<.001
Poor	233 (7.08)	177 (9.28)	56 (4.04)	
Fair	763 (23.17)	529 (27.74)	234 (16.88)	
Good	1078 (32.73)	560 (29.37)	518 (37.37)	
Very good	750 (22.78)	378 (19.82)	372 (26.84)	
Excellent	463 (14.06)	260 (13.63)	203 (14.65)	

[a]Column percentages may not total 100% due to missing data.

[b]Digital channels were website, text message, web-based live chat, and email.

Characteristics Associated With Recruitment Channel Type

Table 2 presents a multivariable logistic regression model comparing demographic and health characteristics of participants who were recruited from digital channels compared with those who were recruited by phone. Respondents who were recruited from digital channels were more likely than respondents recruited by phone to be younger, female, other than non-Hispanic black, have more than a high school education, have higher incomes, be married or have a partner, or have children in the home (all *P*<.05). Additionally, compared with

respondents who were recruited by phone, respondents who were recruited from digital channels were less likely to have public health insurance and more likely to report better self-rated health (all $P<.05$).

Table 2. Logistic regression model for characteristics associated with recruitment channel type, using phone as the reference category.

Demographics	Recruited from digital channels[a], aOR[b] (95% CI)
Age (years)	0.96 (0.96-0.97)[c]
Sex	
Male	1.00 (reference)
Female	1.52 (1.23-1.88)[c]
Race and ethnicity	
Non-Hispanic black	1.00 (reference)
Other	1.48 (1.22-1.79)[c]
Educational attainment	
Less than high school graduate	1.00 (reference)
High school graduate or equivalent	1.06 (0.80-1.39)
More than high school graduate	2.17 (1.67-2.82)[c]
Annual household income (US $)	
Less than 5000	1.00 (reference)
5000 to 14,999	1.21 (0.99-1.48)
15,000 to 24,999	1.48 (1.18-1.85)[c]
25,000 or more	2.02 (1.56-2.61)[c]
Marital status	
Never married	1.00 (reference)
Divorced, widowed, or separated	0.98 (0.81-1.20)
Married or have a partner	1.52 (1.22, 1.89)[c]
Has any children in the home	
No	1.00 (reference)
Yes	1.26 (1.06-1.51)[c]
Health insurance type	
Uninsured	1.00 (reference)
Public	0.75 (0.62-0.90)[c]
Private	0.87 (0.69-1.09)
Public and private	0.77 (0.46-1.30)
Self-rated health	
Poor or fair	1.00 (reference)
Good, very good, or excellent	1.52 (1.27-1.83)[c]

[a]Digital channels were website, text message, web-based live chat, and email.

[b]aOR: adjusted odds ratio.

[c]Statistically significant; $P<.05$.

Discussion

Principal Findings

Numerous studies have compared the effectiveness of digital channel-based recruitment with traditional recruitment methods, yet few have examined recruitment outcomes across multiple digital channels, particularly in social service settings. To our knowledge, the 4C Study is the first study of social service clients recruited using multiple digital communication channels. The study aimed to recruit 3791 socioeconomically disadvantaged adults across five channels within a social service setting. Recruitment goals were met for 3 of the 5 channels—all except email and web-based live chat, although the latter had moderate recruitment success. The highest channel efficiency was achieved from recruiting by phone. Among the digital channels, recruitment from the website resulted in the largest number of individual encounters and a high number of completed surveys (despite low channel efficiency). Recruitment by text message produced a comparable number of completed surveys despite fewer individuals encountered (demonstrating higher channel efficiency). In contrast, email recruitment resulted in both a low number of individual encounters and a low number of completed surveys.

Comparison With Previous Work

Findings can enhance the literature on the use of digital channels in diverse populations for research planning, as the appropriateness of a particular recruitment strategy is influenced by technology preferences among the target population [43,44]. Importantly, the number of individual encounters observed by channel in this study reflects the naturalistic use of these channels by individuals accessing United Way of Greater Atlanta's (UWGA) 2-1-1, the social service setting in which study recruitment occurred. Within UWGA 2-1-1, most requests for referrals occur via phone (ie, calling the 2-1-1 Contact Center). Recruitment for the study was dependent on individuals employing the selected channels to reach or use 2-1-1. For example, recruiters encountered only 467 individuals through the email channel during the entire recruitment period, making this channel less suitable for reaching a large volume of clients quickly. When recruitment for this study was implemented, UWGA 2-1-1 had recently implemented text message as a new communication channel option for clients. In recent years, requests for referrals received via email have declined as options to use other digital channels to request referrals have become more popular among UWGA 2-1-1 clients. Nevertheless, the wide variability in rates of study eligibility, informed consent, and survey completion suggest variability in reach across populations using these channels.

The overall sample recruited reflects the client population served at the recruitment site, which is predominantly female, racial and ethnic minority adults. However, similar to this study, previous research found variation in characteristics of socioeconomically disadvantaged populations across recruitment channel types. A comparison of in-person vs web-based recruitment of adults of low socioeconomic status found that 45% of those recruited in person had annual incomes of <US $10,000 compared with only 16% of those recruited through the web [45]. Thus, even in a targeted recruitment effort, proactively identifying potential bias in a recruitment channel is important for research planning. As noted by Safi et al [45], although different types of recruitment channels may reach socioeconomically disadvantaged participants generally, the channels may differ in the *extent* of disadvantage among participants recruited by each.

In this study of social service clients—a largely socioeconomically disadvantaged group overall—multichannel recruitment resulted in potentially important demographic and health differences between samples from each channel type. The study found that certain channels were more or less likely to recruit participants representative of the local social service client population. The sample recruited from digital channels generally was younger and comprised higher proportions of individuals who were female, married, other than non-Hispanic black, had higher education and income, and had children living in the home compared with the sample recruited by phone. Participants recruited by phone were generally less healthy than those recruited from digital channels and comprised a larger proportion of publicly insured individuals. Similar to these findings, previous research has found that multichannel recruitment is advantageous for recruiting a demographically heterogeneous sample and, in particular, for ensuring representation of underserved populations [44,46-48].

Findings have implications for future health disparities research in social service settings. The findings of this study suggest that future studies may need to recruit across multiple channels (as available in the social service setting) to ensure participants reflect the broader client population. Conversely, for studies requiring targeted recruitment, some channels may provide better access to the target population than others in terms of client volume and/or characteristics. In this study's setting, findings suggest that targeted recruitment of married individuals or adults with higher educational attainment may be more efficient using digital channels, whereas recruitment by phone may be more efficient for recruiting older adults or individuals with poorer health. Channel type is just one possible strategy to consider for targeted recruitment planning. Other data, such as an individual's social service needs [49] can be used to profile prospective participant subgroups. Additional research is needed *within* socioeconomically disadvantaged populations and across diverse recruitment venues, such as social service settings, to optimize recruitment outcomes for health disparities research.

Understanding barriers and facilitators to adoption of digital tools across diverse populations can inform research planning. The social service and health care sectors are expected to become more integrated [29] in the immediate future. Digital channels are likely to be used increasingly in both sectors—not only for research recruitment but also for intervention (although more economic research is needed to support, for example, the use of mobile health behavioral interventions [50,51]). One factor that can hinder the impact of digital tools is any channel's utilization rate in a population, which was observed in this study for email-based recruitment. It is unclear whether the low email engagement rates among clients in the study reflect the low use of email generally or the low use of email for interacting with the 2-1-1 system specifically. Additional research is needed to

better understand the factors driving digital technology use in socioeconomically disadvantaged groups including social service clients. It also is unclear whether the reasons for low email channel use are because of preference or access. Inequitable access to technology is another factor than can hinder the reach of digital tools [52]. Some evidence suggests that lack of consistent internet access may present a barrier to certain communication channels among socioeconomically disadvantaged adults [35,53]. Evidence is needed to inform strategies that reduce inequitable access and use of digital tools.

Limitations

Several potential study limitations must be considered. First, the sample was limited to a single site. In addition, awareness of the availability of the 2-1-1 system may make the 2-1-1 client population different from other socioeconomically disadvantaged adults. Therefore, results may not be generalizable to other populations or settings. However, findings provide some insight into recruitment of socioeconomically disadvantaged populations from the social service sector using multiple communication channels. Second, the cross-sectional nature of the study might not reflect current trends in use of digital channels by UWGA 2-1-1 clients or other social service client populations. As digital technology is ever-evolving, future research is needed to provide evidence on temporal trends in availability and use of digital communication in specific groups and settings. Third, the requirement that individuals who were recruited by phone were required to complete the survey during the call could have biased the sample due to some otherwise eligible individuals not having time to complete the survey immediately. Nevertheless, the demographics of the sample are generally similar to the client population of the recruitment site. Finally, recruitment by channel was dependent on incoming referral requests to UWGA 2-1-1, where client volume per channel varies. However, because recruitment occurred 24 hours per day and 7 days per week, the patterns of encounters observed in the study generally reflect per-channel client volume at UWGA 2-1-1 during the recruitment period.

Conclusions

Digital communication is increasingly ubiquitous. Concomitant with this trend is the growing availability of digital communication in social services provision [54], offering clients and service providers an array of channels for communicating and accessing or providing services. This study had varying degrees of recruitment success using digital channels to recruit socioeconomically disadvantaged clients of a social service organization over an 11-month period. Recruitment success, in part, reflects patterns of channel use among clients. Accordingly, client volume by channel should be considered in recruitment planning. Recruitment planning also can be informed by understanding the likelihood of a given recruitment channel to engage prospective participants with specific demographic or health characteristics. Overall, findings demonstrated the feasibility of recruiting a sample of socioeconomically disadvantaged adults from a social service setting using digital communication channels, particularly when a channel is well utilized among clients.

Difficulty engaging underserved populations for health research is widely reported in the literature [55]. Despite some channel-specific limitations, the 4C Study recruited and surveyed thousands of socioeconomically disadvantaged adults within a social service setting for a health disparities research study. Recommended strategies for reaching populations underrepresented in research include having direct or derived rapport with potential participants [48] and engaging community organizations or other trusted sources relevant to the population of interest [12,56,57]. Partnering with a social service organization trusted by the study population likely contributed to the study's overall recruitment success. Ongoing multisector collaboration, coupled with a more nuanced understanding of populations suffering disparities in health, can help overcome persistent recruitment challenges and, ultimately, help eliminate health-related disparities.

Acknowledgments

The 4C Study was funded by the Intramural Research Department of the American Cancer Society. The authors are grateful to UWGA and the Community Connection Specialist Team. The authors also thank Kimberly Perkins-Heywood, Training and Team Manager, and Marioly Botero, Director, Resource Database Management. They are especially grateful to the 2-1-1 clients who completed the 4C Study survey.

Authors' Contributions

KA was responsible for the study concept and design, acquisition of data, analysis and interpretation of the data, and drafting and revising the paper. RV contributed to data acquisition and analysis, interpretation of data, and drafting and revising the paper. DB contributed to study design, data acquisition, and drafting and revising the paper.

Conflicts of Interest

None declared.

References

1. Braveman PA, Cubbin C, Egerter S, Williams DR, Pamuk E. Socioeconomic disparities in health in the United States: what the patterns tell us. Am J Public Health 2010 Apr 1;100(Suppl 1):S186-S196. [doi: 10.2105/AJPH.2009.166082] [Medline: 20147693]
2. Singh GK, Azuine RE, Siahpush M. Widening socioeconomic, racial, and geographic disparities in HIV/AIDS mortality in the United States, 1987-2011. Adv Prev Med 2013;2013:657961 [FREE Full text] [doi: 10.1155/2013/657961] [Medline: 23738084]
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019 Jan;69(1):7-34 [FREE Full text] [doi: 10.3322/caac.21551] [Medline: 30620402]
4. National Academies of Sciences Engineering and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Community-Based Solutions to Promote Health Equity in the United States, Negussio Y, Baciu A, et al. Communities in Action: Pathways to Health Equity. Washington, DC: National Academies Press (US); 2017.
5. Oh SS, Galanter J, Thakur N, Pino-Yanes M, Barcelo NE, White MJ, et al. Diversity in clinical and biomedical research: a promise yet to be fulfilled. PLoS Med 2015 Dec;12(12):e1001918 [FREE Full text] [doi: 10.1371/journal.pmed.1001918] [Medline: 26671224]
6. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
7. Erves JC, Mayo-Gamble TL, Malin-Fair A, Boyer A, Joosten Y, Vaughn YC, et al. Needs, priorities, and recommendations for engaging underrepresented populations in clinical research: a community perspective. J Community Health 2017 Jun;42(3):472-480 [FREE Full text] [doi: 10.1007/s10900-016-0279-2] [Medline: 27812847]
8. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
9. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
10. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
11. 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 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
12. Alcaraz KI, Sly J, Ashing K, Fleisher L, Gil-Rivas V, Ford S, et al. The ConNECT Framework: a model for advancing behavioral medicine science and practice to foster health equity. J Behav Med 2017 Feb;40(1):23-38 [FREE Full text] [doi: 10.1007/s10865-016-9780-4] [Medline: 27509892]
13. Robinson L, Cotten SR, Ono H, Quan-Haase A, Mesch G, Chen W, et al. Digital inequalities and why they matter. Inf Commun Soc 2015 Mar 16;18(5):569-582. [doi: 10.1080/1369118X.2015.1012532]
14. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
15. Anderson M, Kumar M. Pew Research Center. 2019 May 7. Digital Divide Persists Even as Lower-Income Americans Make Gains in Tech Adoption URL: https://www.pewresearch.org/fact-tank/2019/05/07/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/ [accessed 2019-09-19]
16. Perrin A. Pew Research Center. 2019 May 31. Digital Gap Between Rural and Nonrural America Persists URL: https://www.pewresearch.org/fact-tank/2019/05/31/digital-gap-between-rural-and-nonrural-america-persists/ [accessed 2019-09-13]
17. Rhoades H, Wenzel SL, Rice E, Winetrobe H, Henwood B. No digital divide? Technology use among homeless adults. J Soc Distress Homeless 2017;26(1):73-77 [FREE Full text] [doi: 10.1080/10530789.2017.1305140] [Medline: 31097900]
18. Leavitt L, Abroms L, Johnson P, Schindler-Ruwisch J, Bushar J, Singh I, et al. Recruiting pregnant smokers from Text4baby for a randomized controlled trial of Quit4baby. Transl Behav Med 2017 Jun;7(2):157-165 [FREE Full text] [doi: 10.1007/s13142-016-0450-4] [Medline: 27909881]
19. Guo X, Vittinghoff E, Olgin JE, Marcus GM, Pletcher MJ. Volunteer participation in the health eHeart study: a comparison with the US population. Sci Rep 2017 May 16;7(1):1956 [FREE Full text] [doi: 10.1038/s41598-017-02232-y] [Medline: 28512303]
20. Nolte MT, Shauver MJ, Chung KC. 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]
21. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]

22. 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 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
23. 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]
24. 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]
25. 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]
26. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
27. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
28. Alcaraz KI, Wiedt TL, Daniels EC, Yabroff KR, Guerra CE, Wender RC. Understanding and addressing social determinants to advance cancer health equity in the United States: a blueprint for practice, research, and policy. CA Cancer J Clin 2020 Jan;70(1):31-46 [FREE Full text] [doi: 10.3322/caac.21586] [Medline: 31661164]
29. National Academies of Sciences, Engineering, and Medicine, Board on Health Care Services, Committee on Integrating Social Needs Care into the Delivery of Health Care to Improve the Nation's Health, Health and Medicine Division. Integrating Social Care into the Delivery of Health Care: Moving Upstream to Improve the Nation's Health. Washington, DC: National Academies Press (US); 2019.
30. Eddens KS, Kreuter MW. Proactive screening for health needs in United Way's 2-1-1 information and referral service. J Soc Serv Res 2011;37(2):113-123 [FREE Full text] [doi: 10.1080/01488376.2011.547445] [Medline: 21566702]
31. Kreuter MW, Eddens KS, Alcaraz KI, Rath S, Lai C, Caito N, et al. Use of cancer control referrals by 2-1-1 callers: a randomized trial. Am J Prev Med 2012 Dec;43(6 Suppl 5):S425-S434 [FREE Full text] [doi: 10.1016/j.amepre.2012.09.004] [Medline: 23157761]
32. Mullen PD, Savas LS, Bundy Ł, Haardörfer R, Hovell M, Fernández ME, et al. Minimal intervention delivered by 2-1-1 information and referral specialists promotes smoke-free homes among 2-1-1 callers: a Texas generalisation trial. Tob Control 2016 Oct;25(Suppl 1):i10-i18 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053045] [Medline: 27697943]
33. Kegler MC, Bundy L, Haardörfer R, Escoffery C, Berg C, Yembra D, et al. A minimal intervention to promote smoke-free homes among 2-1-1 callers: a randomized controlled trial. Am J Public Health 2015 Mar;105(3):530-537. [doi: 10.2105/AJPH.2014.302260] [Medline: 25602863]
34. Williams RS, Stollings JH, Bundy Ł, Haardörfer R, Kreuter MW, Mullen PD, et al. A minimal intervention to promote smoke-free homes among 2-1-1 callers: North Carolina randomized effectiveness trial. PLoS One 2016;11(11):e0165086 [FREE Full text] [doi: 10.1371/journal.pone.0165086] [Medline: 27806060]
35. Alcaraz KI, Riehman K, Vereen R, Bontemps-Jones J, Westmaas JL. To text or not to text? Technology-based cessation communication preferences among urban, socioeconomically disadvantaged smokers. Ethn Dis 2018;28(3):161-168 [FREE Full text] [doi: 10.18865/ed.28.3.161] [Medline: 30038477]
36. Purnell JQ, Kreuter MW, Eddens KS, Ribisl KM, Hannon P, Williams RS, et al. Cancer control needs of 2-1-1 callers in Missouri, North Carolina, Texas, and Washington. J Health Care Poor Underserved 2012 May;23(2):752-767 [FREE Full text] [doi: 10.1353/hpu.2012.0061] [Medline: 22643622]
37. Savas LS, Fernández ME, Jobe D, Carmack CC. Human papillomavirus vaccine: 2-1-1 helplines and minority parent decision-making. Am J Prev Med 2012 Dec;43(6 Suppl 5):S490-S496 [FREE Full text] [doi: 10.1016/j.amepre.2012.09.003] [Medline: 23157770]
38. 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]
39. United Way Worldwide. 211. 2019. About 2-1-1 URL: http://www.211.org/pages/about [accessed 2019-09-20]
40. United Way Centraide Canada. 211 Canada. 2019. About 211 URL: http://211.ca/about-211/ [accessed 2019-09-21]
41. Kreuter MW. Reach, effectiveness, and connections: the case for partnering with 2-1-1 to eliminate health disparities. Am J Prev Med 2012 Dec;43(6 Suppl 5):S420-S421 [FREE Full text] [doi: 10.1016/j.amepre.2012.09.024] [Medline: 23157759]
42. Stewart AL, Hays RD. Health perceptions, energy\fatigue, and health distress measures. In: Stewart AL, Ware JE, editors. Measuring Functioning and Well-being: The Medical Outcomes Study Approach. Durham, NC: Duke University Press; 1992:143-172.
43. 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]

44. Rockliffe L, Chorley AJ, Marlow LA, Forster AS. It's hard to reach the 'hard-to-reach': the challenges of recruiting people who do not access preventative healthcare services into interview studies. Int J Qual Stud Health Well-being 2018 Dec;13(1):1479582 [FREE Full text] [doi: 10.1080/17482631.2018.1479582] [Medline: 29912650]
45. Safi AG, Reyes C, Jesch E, Steinhardt J, Niederdeppe J, Skurka C, et al. Comparing in person and internet methods to recruit low-SES populations for tobacco control policy research. Soc Sci Med 2019 Dec;242:112597. [doi: 10.1016/j.socscimed.2019.112597] [Medline: 31670216]
46. Tabriz AA, Fleming P, Shin Y, Resnicow K, Jones R, Flocke S, et al. Challenges and opportunities using online portals to recruit diverse patients to behavioral trials. J Am Med Inform Assoc 2019 Dec 1;26(12):1637-1644. [doi: 10.1093/jamia/ocz157] [Medline: 31532482]
47. Ibarra JL, Agas JM, Lee M, Pan JL, Buttenheim AM. Comparison of online survey recruitment platforms for hard-to-reach pregnant smoking populations: feasibility study. JMIR Res Protoc 2018 Apr 16;7(4):e101 [FREE Full text] [doi: 10.2196/resprot.8071] [Medline: 29661751]
48. Temple EC, Brown RF. A comparison of internet-based participant recruitment methods: engaging the hidden population of cannabis users in research. J Res Prac 2011;7(2):1-20 [FREE Full text]
49. Alcaraz KI, Arnold LD, Eddens KS, Lai C, Rath S, Greer R, et al. Exploring 2-1-1 service requests as potential markers for cancer control needs. Am J Prev Med 2012 Dec;43(6 Suppl 5):S469-S474 [FREE Full text] [doi: 10.1016/j.amepre.2012.08.019] [Medline: 23157767]
50. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
51. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
52. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
53. Côté M, Lapointe A, Laramée C, Lemieux S, Desroches S, Belanger-Gravel A, et al. Beliefs related to participation in a large web-based prospective survey on diet and health among individuals with a low socioeconomic status: qualitative study. JMIR Form Res 2019 Dec 10;3(4):e13854 [FREE Full text] [doi: 10.2196/13854] [Medline: 31821149]
54. Bullock AN, Colvin AD. Communication technology integration into social work practice. Adv Soc Work 2015 Jul 27;16(1):1-14. [doi: 10.18060/18259]
55. Ejiogu N, Norbeck JH, Mason MA, Cromwell BC, Zonderman AB, Evans MK. Recruitment and retention strategies for minority or poor clinical research participants: lessons from the Healthy Aging in Neighborhoods of Diversity across the Life Span study. Gerontologist 2011 Jun;51(Suppl 1):S33-S45 [FREE Full text] [doi: 10.1093/geront/gnr027] [Medline: 21565817]
56. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant: solutions for recruiting hidden, hard-to-reach, and vulnerable populations. Int J Qual Method 2015 Dec 17;14(5):160940691562142-160940691562110. [doi: 10.1177/1609406915621420]
57. Matsuda Y, Brooks JL, Beeber LS. Guidelines for research recruitment of underserved populations (EERC). Appl Nurs Res 2016 Nov;32:164-170 [FREE Full text] [doi: 10.1016/j.apnr.2016.07.009] [Medline: 27969022]

Abbreviations

4C: Cancer Communication Channels in Context
aOR: adjusted odds ratio
UWGA: United Way of Greater Atlanta

Edited by G Eysenbach; submitted 14.10.19; peer-reviewed by S Badawy, A Meyer; comments to author 11.11.19; revised version received 13.02.20; accepted 29.02.20; published 01.04.20.

Please cite as:
Alcaraz KI, Vereen RN, Burnham D
Use of Telephone and Digital Channels to Engage Socioeconomically Disadvantaged Adults in Health Disparities Research Within a Social Service Setting: Cross-Sectional Study
J Med Internet Res 2020;22(4):e16680
URL: https://www.jmir.org/2020/4/e16680
doi:10.2196/16680
PMID:32234699

Original Paper

The Use of Technology for Communicating With Clinicians or Seeking Health Information in a Multilingual Urban Cohort: Cross-Sectional Survey

Elaine C Khoong[1,2], MD, MS; Natalie A Rivadeneira[1,2], MPH; Robert A Hiatt[3,4], MD, PhD; Urmimala Sarkar[1,2], MD, MPH

[1]Division of General Internal Medicine, Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, CA, United States

[2]Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, CA, United States

[3]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States

[4]Hellen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Elaine C Khoong, MD, MS
Division of General Internal Medicine
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 6282063188
Email: elaine.khoong@ucsf.edu

Abstract

Background: Technology is being increasingly used to communicate health information, but there is limited knowledge on whether these strategies are effective for vulnerable populations, including non–English speaking or low-income individuals.

Objective: This study assessed how language preferences (eg, English, Spanish, or Chinese), smartphone ownership, and the type of clinic for usual source of care (eg, no usual source of care, nonintegrated safety net, integrated safety net, private or community clinic, academic tertiary medical center, or integrated payer-provider) affect technology use for health-related communication.

Methods: From May to September 2017, we administered a nonrandom, targeted survey to 1027 English-, Spanish-, and Chinese-speaking San Francisco residents and used weighted multivariable logistic regression analyses to assess predictors of five technology use outcomes. The three primary predictors of interest—language preference, smartphone ownership, and type of clinic for usual care—were adjusted for age, gender, race or ethnicity, limited English proficiency, educational attainment, health literacy, and health status. Three outcomes focused on use of email, SMS text message, or phone apps to communicate with clinicians. The two other outcomes were use of Web-based health videos or online health support groups.

Results: Nearly one-third of participants watched Web-based health videos (367/1027, 35.74%) or used emails to communicate with their clinician (318/1027, 30.96%). In adjusted analyses, individuals without smartphones had significantly lower odds of texting their clinician (adjusted odds ratio [aOR] 0.27, 95% CI 0.13-0.56), using online health support groups (aOR 0.14, 95% CI 0.04-0.55), or watching Web-based health videos (aOR 0.31, 95% CI 0.15-0.64). Relative to English-speaking survey respondents, individuals who preferred Chinese had lower odds of texting their clinician (aOR 0.25, 95% CI 0.08-0.79), whereas Spanish-speaking survey respondents had lower odds of using apps to communicate with clinicians (aOR 0.34, 95% CI 0.16-0.75) or joining an online support group (aOR 0.30, 95% CI 0.10-0.92). Respondents who received care from a clinic affiliated with the integrated safety net, academic tertiary medical center, or integrated payer-provider systems had higher odds than individuals without a usual source of care at using emails, SMS text messages, or apps to communicate with clinicians.

Conclusions: In vulnerable populations, smartphone ownership increases the use of many forms of technology for health purposes, but device ownership itself is not sufficient to increase the use of all technologies for communicating with clinicians.

Language preference impacts the use of technology for health purposes even after considering English proficiency. Health system factors impact patients' use of technology-enabled approaches for communicating with clinicians. No single factor was associated with higher odds of using technology for all health purposes; therefore, existing disparities in the use of digital health tools among diverse and vulnerable populations can only be addressed using a multipronged approach.

(J Med Internet Res 2020;22(4):e16951) doi:10.2196/16951

KEYWORDS

vulnerable populations; health information technology; physician patient relations; consumer health information; digital divide; social media; internet

Introduction

Inequities in Technology-Enabled Health Communication Strategies

Technology is increasingly being used to communicate health information [1]. Given the known disparities in internet use and broadband access among rural, older, lower socioeconomic status, nonwhite populations [2], reliance on technology to disseminate health information may exacerbate health inequities [3,4]. In response to the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act, in the last decade, health systems have joined the broader trend of using technology for communication. As part of the HITECH Act, health systems received financial incentives to provide patients increased access to their health care team and clinical records, measured through patient portal use. Owing to these incentive payments, patient portals have been the primary manner in which health systems and providers have utilized technology for communicating with patients. Unfortunately, studies have consistently shown that patient portals are used less frequently by racial and ethnic minorities, persons of a lower socioeconomic status, and those without neighborhood broadband internet access [5-8].

Gaps in National Surveys About Health Information Trends

To gain insight on how individuals seek information, the National Cancer Institute (NCI) administers the Health Information National Trends Survey (HINTS) [9]. However, HINTS has several limitations for vulnerable populations; it is distributed only in English and Spanish, lacks the assessment of health literacy, and has poor representation of nonwhite, less affluent populations [10]. As part of an effort to inform local cancer communication strategies, the NCI awarded supplemental funding to Cancer Center Support Grants to administer a modified HINTS tailored for vulnerable populations. In San Francisco, this survey (also known as SFHINTS, San Francisco Health Information National Trends Survey) was administered in English, Spanish, and Chinese to increase data collection from non-English speakers. We also targeted low-income, nonwhite populations who experience cancer outcome disparities (eg, breast and prostate cancer in African Americans and liver cancer in Asians) [11]. Given the proximity to the Silicon Valley, in SFHINTS we included additional questions about the use of technology for exchanging information with a health care professional and the use of social media for health-related purposes.

We previously reported the health information–seeking behaviors and preferences of our SFHINTS cohort and found that participants who preferred English or owned smartphones were more likely to use the internet for health information or prefer emails for provider-distributed health information [12]. To build on these findings, in this report, we explored how both language preference and smartphone ownership impacted the use of technology for communication with health care clinicians or the use of social media for health-related purposes.

Technology Use for Health Communication in Vulnerable Populations

Despite high levels of interest, prior studies have shown that individuals of a lower socioeconomic status were less likely to use emails to communicate with clinicians [13-17]. However, some studies have shown an increased interest and use in young, nonwhite populations [13,18]. Less is known about other electronic means (eg, SMS text messages and smartphone apps) to communicate with clinicians, particularly among non-English speakers. Similarly, prior studies have shown frequent use of social media for health-related purposes [19-21], but studies have rarely included non–English speaking, low socioeconomic populations.

Given the local nature of SFHINTS, we were able to explore an additional contributor to participants' use of technology for clinician-directed communication: the type of clinic that participants use for their usual source of care (eg, safety net clinic or academic tertiary medical center). Prior literature has shown usability challenges for electronic health records (EHRs) [22] and that safety net EHRs are less likely to have patient engagement features (including patient portal–related features), potentially exacerbating inequities in patient-clinician communication [23,24]. Consequently, safety net clinics that disproportionally serve groups experiencing health disparities may also lack the technology infrastructure to facilitate technology-based approaches to increase communication and access to health care clinicians.

The multilingual, vulnerable population in the SFHINTS cohort provided an opportunity to explore the impact of three relatively understudied factors—language preference, smartphone ownership, and type of clinic for usual care (eg, safety net clinic or academic tertiary medical center) as a proxy for the digital infrastructure or patient portal usability—on participants' use of technology to communicate with their health care team and participants' use of social media for health-related purposes. We hypothesized that language preference and smartphone ownership would impact all technology use, but type of clinic

would only impact technology use for communication with clinicians.

Methods

Research Setting

The 2017 American Community Survey estimates that San Francisco has a minority-majority: 34% Asians, 15% Latinx, and 5% African American. Nearly 45% of residents speak a non-English language, with Chinese being the most common; one-fifth of San Francisco residents have limited English proficiency [25]. There are several health care systems in San Francisco that deliver primary care, and each system uses its own EHR system. There is one tertiary academic center (University of California San Francisco) as well as two integrated payer-provider systems (Kaiser and Veterans Affairs). The primary care clinics in these systems have used EHR systems with English-language patient portals for over 5 years. There are two larger networks of safety net clinics within San Francisco. One group of clinics, which uses the same EHR, is run by the Department of Public Health, which also operates the county hospital. The patient portal within these clinics had been active for approximately 2.5 years at the time of survey administration and was only available in English. The other group of safety net clinics is a consortium of loosely affiliated clinics. Each clinic has independently chosen an EHR system and, therefore, the patient engagement features of the EHR at each clinic are variable. Similarly, the remaining private and community clinics within San Francisco vary in terms of the EHR system and availability of patient access or engagement features. At the time of the survey, only Kaiser offered a non–English language patient portal (Spanish).

Survey Development

We used English and Spanish HINTS questions [9] as well as validated health care access [26] and health literacy [27] questions to create our survey (SFHINTS: Multimedia Appendix 1, questions relevant to this report are in sections B-D, G, and H.) We used a standard dual-reviewer process [28] to translate questions for the Spanish and Chinese surveys. The details of survey development and administration are described in prior papers [29].

Sampling Procedure, Recruitment, and Survey Administration

Using community-based snowball sampling with prespecified language and race and ethnicity targets to reach populations with known cancer disparities, we aimed for half of the surveys to be completed in English (with half of the participants identifying as African American) and non-English surveys to be equally divided between Spanish and Chinese (Mandarin or Cantonese) participants [11,29].

As previously reported in greater detail [12,29], from May to September 2017, bilingual staff administered the survey in-person on tablet devices at community establishments and events as well as small businesses and street locations in specific neighborhoods to target our populations of interest. (Surveys were administered via REDCap [Research Electronic Data Capture] electronic data capture tools hosted at our institution. REDCap [30,31] is a secure, Web-based software platform designed to support data capture for research studies.) The staff explained the survey's purpose, acquired verbal consent, and then administered the survey in the participants' preferred language. A US $25 incentive was provided; our institution's institutional review board approved this study.

Conceptual Model and Predictor Variables

We used an information-seeking behavior and use model described by Longo (Textbox 1) to identify a complete list of potential predictors that could explain the variation in participants' use of technology for health purposes [32]. Longo [32] described both contextual and personal factors that impact behavior. The SFHINTS survey included more personal than contextual factors. Ultimately, we included a total of 10 predictor variables. We had three predictor variables of primary interest: two contextual factors—smartphone ownership (an information environment factor) and type of clinic (a health care structure)—and one personal factor (ie, language preference). Guided by prior literature, we included seven additional personal factors (ie, age, gender, race or ethnicity, health literacy, education, English proficiency, and current health status) as predictors of technology use for health purposes [33-36].

Smartphone ownership was a binary variable (ie, yes vs no). We also dichotomized health status (ie, poor or fair vs good or very good or excellent), English proficiency, and health literacy. English proficiency and health literacy were reported as limited if participants reported speaking English less than *well* or if participants felt less than *quite a bit* comfortable completing medical forms independently [27]. Within the race or ethnicity and language variables, non-Hispanic white and English served respectively as the reference categories for analyses. We categorized age (ie, 18-34 as the reference category, 35-49, 50-64, and ≥65 years) and education (ie, less than high school; high school or equivalent; some college or vocational training; at least college graduate as the reference category) into four groups. Clinic type was organized into six categories: no usual source of care (reference category), nonintegrated safety net, integrated safety net, private or community clinic, tertiary academic, and fully integrated payer and provider (Kaiser Permanente or Veterans Affairs).

Textbox 1. A conceptual model of the factors that impact information-seeking behaviors.

Contextual

- Health status
- Health care structure
- Delivery of care
- Information environment factors
- Information seeking for self, family, or friend at risk or with a current medical problem

Personal

- Demographic factors
- Socioeconomic factors
- Health history
- Family medical history
- Education
- Culture
- Languages
- Attitudes, intentions, behaviors
- Current health status

Outcome Variables

We reported the use of technology for health communication with clinicians or peers (questions B4 and B5 in the SFHINTS Survey in Multimedia Appendix 1) as a binary variable (yes or no) for the five sources used by at least 9.5% of participants (ie, use of an email, an SMS text message, or an app with a health care provider and use of online support groups or health-related videos). Each survey respondent was able to answer yes or no for each type of technology use. On the basis of prior literature showing differences in the use of these types of technology, we reported these as independent outcomes rather than an aggregated outcome [13,18]. Therefore, five different regression models have been reported for outcomes in this report.

Analysis

The relationships between predictor variables and technology use were assessed using bivariate logistic regressions. In addition, we conducted weighted, multivariable logistic regression analyses to identify factors associated with technology use for each of the five outcomes. Weights were computed using iterative proportional fitting (raking). This technique is used for nonprobability samples and involves iteratively adjusting over a set of variables (ie, age, gender, and education) within each race or ethnicity group to reweight the respondent population to match the distribution of the reference population (ie, San Francisco) [37]. We determined no significant collinearity (tolerance >0.10) between predictor variables. All logistic regressions were done using the PROC SURVEYLOGISTIC procedure of SAS 9.4 statistical software (Cary, North Carolina). All regressions were performed using complete case analyses, which totaled 944 observations (944/1027, 91.92% of respondents). The data analyzed for this study is available from the senior investigators of the SFHINTS study (RH and US).

Results

Participant Characteristics

The 1027 participants (514 English surveys with 242 non-Hispanic black, 115 Latinx, and 43 non-Hispanic white participants; 256 Spanish surveys; and 257 Chinese surveys) have been previously described (Multimedia Appendix 2: participants' sociodemographic traits) [12,29]. Our cohort had more limited English-proficient participants (344/1027, 33.50%) than the 2017 national HINTS cohort (2%) [9]. In our cohort, 440/1027 (42.84%) participants had limited health literacy, whereas 791/1027 (77.02%) owned smartphones. Nearly one-fifth (178/1027, 17.33%) reported no usual source of care. Over 50% (148 nonintegrated safety net clinics and 378 integrated safety nets) received care in the safety net systems.

Use of Technology

As detailed in Table 1, approximately one-third (318/1027, 30.96%) of the participants used an email to communicate with clinicians. Fewer used an SMS text message (218/1027, 21.23%) or an app (136/1027, 13.24%) for communications with clinicians. The use of Web-based videos to learn about health information was common (367/1027, 35.74%). Across all language groups, at least one-quarter of the population reported watching Web-based videos about health. Online support groups were used by the lowest portion of respondents (99/1027, 9.64%). Figure 1 shows the use of technology by language and smartphone ownership. Smartphone owners across all language preferences had higher rates of using all forms of technology.

Table 2 reports results of the bivariate logistic regression analyses for the three predictors of interest. In the unadjusted analyses, all 10 variables impacted the odds of using at least one form of technology (see Multimedia Appendix 3). For the

three predictors of particular interest in our study, in unadjusted analyses, individuals without smartphones had lower odds of using all forms of technology. We also found that individuals who preferred Spanish had lower odds of using an email to communicate with clinicians, whereas individuals who preferred Chinese had lower odds of using an email, an SMS text message, or apps to communicate with clinicians as well as watching Web-based health videos. Relative to respondents with no usual source of care, participants who received care at an integrated safety net, academic tertiary medical center, or integrated payer-provider clinic had higher odds of using technology (ie, emails, SMS text messages, or apps) to communicate with clinicians. Respondents who received care at a private clinic or an integrated payer-provider clinic had higher odds of watching Web-based health videos.

In the multivariable analyses, when holding all other variables constant, race or ethnicity, English proficiency, and health literacy no longer significantly predict any of the outcomes, but younger females with more education had higher odds of using at least one form of technology (adjusted odds for all variables are in Multimedia Appendix 4). Of the three variables of primary interest for this report, we found that individuals without smartphones had lower odds of using text to communicate with clinicians, joining online support groups, or watching Web-based health videos (Table 3). Individuals who preferred Chinese had lower odds of using SMS text messages to communicate with clinicians, whereas Spanish-speakers had lower odds of using apps to communicate with clinicians and joining online support groups. Individuals who received care at an integrated safety net, academic tertiary medical center, or integrated payer-provider clinic had higher odds than respondents with no usual care of using emails, SMS text messages, or apps to communicate with clinicians. Respondents that received care at the 2-clinic systems which adopted patient portals the earliest in San Francisco (academic tertiary system and integrated payer-provider) had higher odds of using apps to communicate with clinicians. Participants with care at the private clinic or community hospital had higher odds of watching Web-based health videos.

Table 1. The use of technology to communicate with clinicians or the use of social media for health purposes.

Types of technology use	All (N=1027), n (%)	English (n=514), n (%)	Spanish (n=256), n (%)	Chinese (n=257), n (%)
Technology use to communicate with clinicians				
Email	318 (30.96)	199 (38.7)	76 (29.7)	43 (16.7)
SMS text message	218 (21.23)	142 (27.6)	56 (21.9)	20 (7.8)
App	136 (13.24)	92 (17.9)	29 (11.3)	15 (5.8)
Social media use for health purposes				
Online support group	99 (9.64)	59 (11.5)	19 (7.4)	21 (8.2)
Web-based health videos (eg, YouTube)	367 (35.74)	189 (36.8)	109 (42.6)	69 (26.8)

Figure 1. The use of technology for health purposes by smartphone ownership and language preference. Percentage is based on the subset of participants without smart phones (N=236) or with smartphones (N=791) for the left and right panel respectively.

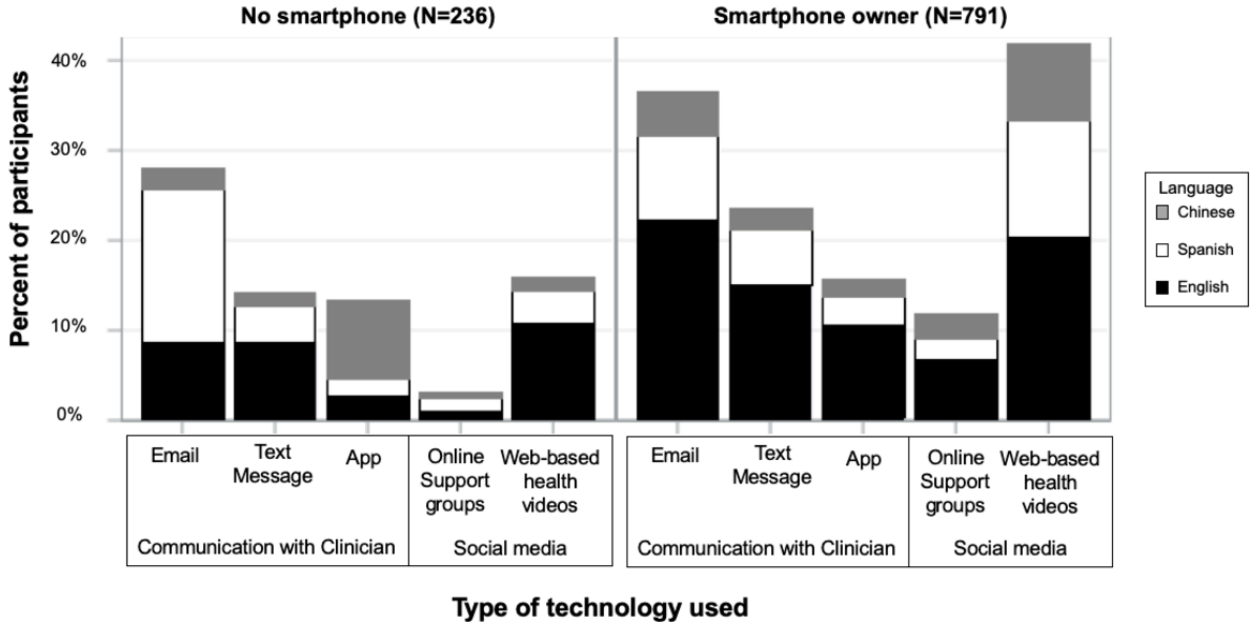


Table 2. Unadjusted odds of using technology for health purposes.

Predictor	Email with clinician, uOR[a,b] (95% CI)	SMS with clinician, uOR[a] (95% CI)	App with clinician, uOR[a] (95% CI)	Support group, uOR[a] (95% CI)	Web-based videos, uOR[a] (95% CI)
No smartphone	0.25 (0.12-0.52)[c]	0.41 (0.17-0.99)[c]	0.325 (0.13-0.84)[c]	0.10 (0.04-0.26)[c]	0.14 (0.07-0.25)[c]
Language[d]					
Spanish	0.48 (0.27-0.85)[c]	0.85 (0.46-1.57)	0.51 (0.25-1.04)	0.46 (0.20-1.06)	0.76 (0.42-1.36)
Chinese	0.29 (0.17-0.49)[c]	0.23 (0.12-0.44)[c]	0.22 (0.10-0.47)[c]	0.64 (0.29-1.42)	0.52 (0.32-0.84)[c]
Type of clinic for usual source of care[d]					
Nonintegrated safety net	1.10 (0.43-2.80)	0.56 (0.25-1.27)	0.74 (0.27-2.00)	0.54 (0.19-1.52)	0.91 (0.42-1.96)
Integrated safety net	2.03 (1.04-3.98)[c]	3.26 (1.44-7.38)[c]	1.11 (0.44-2.83)	0.83 (0.27-2.50)	1.59 (0.83-3.05)
Private clinic or community hospital	1.15 (0.45-2.94)	0.69 (0.26-1.83)	1.17 (0.42-3.29)	0.37 (0.11-1.25)	2.85 (1.13-7.19)[c]
Academic tertiary medical center	14.10 (4.58-43.38)[c]	4.54 (1.03-20.02)[c]	9.09 (2.05-40.28)[c]	1.36 (0.21-8.72)	3.51 (0.90-13.72)
Integrated payer and provider	4.46 (1.71-11.64)[c]	0.83 (0.30-2.28)	3.92 (1.22-12.58)[c]	2.14 (0.57-8.08)	2.66 (1.10-6.43)[c]

[a]All odds ratios are weighted but unadjusted within this table.

[b]uOR: unadjusted odds ratio.

[c]$P<.05$.

[d]The reference categories for the following variables are as follows: language (English), type of clinic (no usual source of care).

Table 3. Adjusted odds of using technology for health purposes.

Predictor	Email with clinician, aOR[a,b] (95% CI)	SMS with clinician, aOR[a] (95% CI)	App with clinician, aOR[a] (95% CI)	Support group, aOR[a] (95% CI)	Web-based videos, aOR[a] (95% CI)
No smartphone	0.61 (0.25-1.48)	0.27 (0.13-0.56)[c]	1.12 (0.42-2.99)	0.14 (0.04-0.55)[c]	0.31 (0.15-0.64)[c]
Language[d]					
Spanish	0.69 (0.28-1.69)	0.51 (0.16-1.62)	0.34 (0.16-0.75)[c]	0.30 (0.10-0.92)[c]	0.66 (0.21-2.02)
Chinese	0.97 (0.41-2.30)	0.25 (0.08-0.79)[c]	0.32 (0.10-1.03)	2.01 (0.50-8.06)	1.00 (0.41-2.41)
Type of clinic for usual source of care[d]					
Nonintegrated safety net	1.02 (0.39-2.66)	0.64 (0.25-1.63)	1.12 (0.36-3.54)	0.55 (0.17-1.81)	0.68 (0.27-1.69)
Integrated safety net	2.36 (1.08-5.12)[c]	2.96 (1.25-7.04)[c]	1.60 (0.57-4.54)	0.47 (0.14-1.57)	1.35 (0.65-2.83)
Private clinic or community hospital	1.02 (0.33-3.15)	0.72 (0.22-2.33)	1.41 (0.31-6.42)	0.19 (0.03-1.08)	2.65 (1.08-6.51)[c]
Academic tertiary medical center	9.08 (2.46-33.5)[c]	3.39 (0.57-20.19)	12.41 (2.76-55.89)[c]	0.70 (0.11-4.28)	2.50 (0.75-8.29)
Integrated payer and provider	2.68 (0.99-7.27)	0.78 (0.27-2.26)	4.81 (1.44-16.05)[c]	1.20 (0.30, 4.80)	1.79 (0.74-4.32)

[a]All odds ratios are weighted and adjusted for age, gender, race or ethnicity, English proficiency, education, health literacy, health status, smartphone ownership, language preference, and type of clinic for usual source of care.

[b]aOR: adjusted odds ratio.

[c]$P<.05$.

[d]The reference categories for the following variables are as follows: language (English) and type of clinic (no usual source of care).

Discussion

Principal Findings

Even in vulnerable, diverse populations, nearly 4 in 5 individuals own a smartphone, with a large portion using mobile technologies to communicate with clinicians. Just over one-third of individuals are watching health-related videos online, including over 1 in 4 participants regardless of language preference. For the most part, many of our findings are consistent with prior literature that females who are younger and better educated are more likely to use digital tools to engage

in their health care [13,16,36]. Importantly, no variable was a significant predictor for all five technology use outcomes, which highlights the importance of distinguishing among different types of digital tools when devising communication strategies for diverse populations. Digital communication strategies may need to be tailored to reach one specific population versus a different population.

Although we had anticipated that smartphone ownership and English language preference would be associated with higher odds of using all types of technology, both factors were significantly associated with only a subset of the outcomes. We had also anticipated that type of clinic should only be associated with the use of technology to communicate with clinicians but found that it was also associated with whether respondents watched Web-based health videos. Unfortunately, we do not have data to explore potential explanations for these findings. However, we can use the technology acceptance model (TAM) [38] as a conceptual model to try to explain why certain populations may be more likely to use technology for each purpose. In the TAM, whether or not an individual adopts a technology is impacted by two main factors: perceived ease of use or perceived usefulness of a given technology.

Smartphone Possession Alone Does Not Increase the Use of All Types of Technology

For example, smartphone ownership was a significant predictor only for the use of online support groups, Web-based health videos, and SMS text messaging with clinicians. Using TAM to guide our thinking, we can hypothesize that smartphone owners perceive a greater ease (or usefulness) of only texting a clinician, watching an online health video, or using an online support group. Smartphone ownership was not a significant predictor of using emails or apps with clinicians perhaps because the perceived ease or usefulness of these activities was not as different between smartphone owners and nonowners. We do not have data to assess this perception and propose that future research in digital health equity should explicitly evaluate how provision or ownership of a smartphone alone impacts the use of digital tools to engage with clinicians or Web health resources.

Another potential explanation is that not all smartphone owners have the digital literacy to use all the functions of the smartphone. Studies have shown that digital health literacy, a distinct concept from health literacy, poses a barrier for using digital health tools in underserved populations [39]. A digital equity survey of more than 1000 San Francisco residents found that although 93% owned smartphones, 5% did not have a data plan. Among those with an income less than US $25,000, 79% owned smartphones, but 14% of these smartphone users did not have a data plan [40]. Without data access, users are inherently limited in the number of activities that can be performed on their smartphones. This same survey also found that nearly 25% to 30% of internet users who were non–English speaking, older than 65 years, or had an income less than US $25,000 did not possess basic digital literacy (defined as the ability to search for information, find a website, send an email, or fill out an online form) [40]. Digital literacy may be a mediator or moderator, which explains why smartphone ownership was not found to be an important predictor for all technology outcomes.

Surprisingly, smartphone ownership was not an important predictor of using apps to communicate with clinicians, despite apps necessitating the ownership of a smartphone or tablet. Our survey did not allow us to explore if respondents answered this question while envisioning a patient portal app or an alternative communication app (ie, Facebook messenger, WhatsApp, etc). Regardless, these findings suggest that the apps currently available for communicating with clinicians are not adequately useful or easy to use (TAM constructs) such that smartphone owners are more likely to use apps than nonsmartphone owners. If participants were considering patient portal apps when answering these questions, it further supports assertions that health care system digital communications are not mobile friendly [41].

Language Preference Predicts Technology Use Patterns Beyond English Proficiency

We also anticipated that language preference, similar to smartphone ownership, would be associated with all technology use behaviors given the known barriers to communicating with health care clinicians for limited English-proficient patients as well as the higher quantity of English-language health content in the internet [42-44]. Although English proficiency and language preference are often correlated, they are distinct concepts [45,46]. Specifically, studies have found language preference to be associated with acculturation even after accounting for English proficiency measures, and acculturation has been found to impact health information–seeking behaviors [34,47-49].

With this in mind, it is worth noting that English proficiency was not a significant predictor for any of the studied outcomes. On the contrary, both Spanish and Chinese preferences were associated with lower odds of using at least one form of technology. Of note, neither using emails to communicate with clinicians nor watching Web-based videos was significantly impacted by language preference. These findings may suggest that the usefulness or ease of using emails or Web-based health videos is not significantly different for individuals who prefer English vs non-English languages and, therefore, that both these approaches may be potential avenues for communication, which will avoid significantly exacerbating communication disparities already experienced by non-English speakers. Although Web-based health videos have been found in some studies to be an effective means of disseminating information to non–English speaking populations [50-52], earlier studies have suggested differences in email use based on the language (though earlier studies did not consider both English proficiency and language preference) [24].

Health System Factors May Impact Electronic-Based Communication With Clinicians

Owing to differences in the digital and EHR infrastructure within different San Francisco health care systems, we anticipated that the type of clinic would impact behaviors surrounding communication habits with clinicians. Moreover, as the reference variable was no usual source of care, we anticipated

that having any usual source of care should result in higher odds of communicating with clinicians. We did find that respondents who received usual care at clinics affiliated with the academic tertiary medical center or the integrated payer-provider health care systems—systems that have had the longest, most established patient portal systems—had higher odds of using apps to communicate with clinicians, potentially through more mature patient portal apps. Notably, care at either of the safety net systems or private clinics not clearly affiliated with large health care systems was not associated with higher odds of using apps. This supports the literature that safety net EHRs are less likely to have usable patient engagement features [23]. It also reinforces findings from a recent study that found patients at safety net systems are less likely to use a patient portal, and patients receiving care at nonacademic medical centers and small health care systems are less likely to access their medical records [53].

The odds of emailing your clinician were higher for individuals whose primary source of care was a clinic affiliated with the integrated safety net or an academic medical center. One possible explanation for this is that many of the clinicians who provide care in the integrated safety net are faculty at the main academic medical center in San Francisco. There may be behaviors or attitudes about patient engagement that are common to these group of clinicians which results in their patients perceiving higher ease or usefulness of emailing their clinician. This is similar to patient portal usage studies that show patient-clinician relationships and clinician attitudes and behaviors about patient portals impact their patients' use of the patient portal [54,55].

Limitations

This study is limited by its reliance on participant self-reporting and sampling of a single city and county, which was inherent in the design to inform local communication efforts. However, we surveyed more than 1000 individuals from groups underrepresented in health information–seeking studies. We did not collect information on who declined to participate in the survey and, therefore, could not report a response rate. This may have resulted in a sampling bias, but weighting our sample should reduce bias, and in-person surveys generally show higher response rates than other survey methods [56]. A small number of observations within the levels of independent variables (eg, college graduates and patients at academic tertiary medical care clinics) resulted in some estimates with wide confidence intervals.

Conclusions

We found that even after controlling for other known factors, smartphone ownership, language preference, and type of delivery care system for usual care changed the odds of using technology for health purposes. Smartphone ownership was important for some behaviors and, therefore, ensuring technology and communications are optimized for mobile devices is important. However, none of the studied patient or contextual factors was significantly associated with all behaviors, suggesting that health care systems and public health messages may have to utilize a variety of approaches when using technology-enabled communications to reach a broad population. These communication strategies must be delivered with an eye on equity for diverse, underserved populations to ensure any intervention does not exacerbate the existing disparities. No single solution alone—including the provision of smartphones, creation of non-English communications and workflows, or development of better patient engagement digital infrastructure—is likely to fully address the existing inequities in the use of digital health tools.

Acknowledgments

The authors would like to acknowledge Cynthia Cheung, Corina Liew, Alejo Perez-Stable, Francisco Quintanilla, Pamela Williams, Ching Wong, as well as the San Francisco Cancer Initiative Steering Committee and the SFHINTS steering committee (Anna Napoles, Tung Nguyen, and Rena Pasick). They also want to acknowledge community partners including Angela Sun at the Chinese Community Health Resource Center, Monique LaSarre at the Rafiki Health Center, and Brenda Storey at Mission Neighborhood Health Center.

This work was supported by the NCI through an Administrative Supplement to the University of California, San Francisco Helen Diller Family Comprehensive Cancer Center Support Grant (grant number P30CA082103-18S7), the National Institutes of Health's (NIH's) National Research Service Award (grant number T32HP19025), the National Heart Lung and Blood Institute of the NIH under Award Number K12HL138046, and the National Center for Advancing Translational Sciences of the NIH under Award Number KL2TR001870. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors' Contributions

RH and US made substantial contributions to the conception or design of the work. EK and NR made substantial contributions to the analysis and interpretation of data for the work. All authors drafted or critically revised the work, provided final approval of the submitted version, and agreed to be accountable for the accuracy and integrity of all parts of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
San Francisco Health Information National Trends Survey English questionnaire.
[PDF File (Adobe PDF File), 104 KB - jmir_v22i4e16951_app1.pdf]

Multimedia Appendix 2
Sociodemographic traits of San Francisco Health Information National Trends Survey participants.
[DOCX File , 24 KB - jmir_v22i4e16951_app2.docx]

Multimedia Appendix 3
Unadjusted odds of technology for health purposes.
[DOCX File , 18 KB - jmir_v22i4e16951_app3.docx]

Multimedia Appendix 4
Adjusted odds of technology for health purposes.
[DOCX File , 15 KB - jmir_v22i4e16951_app4.docx]

References

1. Crotty BH, Tamrat Y, Mostaghimi A, Safran C, Landon BE. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. Health Aff (Millwood) 2014 Oct;33(10):1817-1822 [FREE Full text] [doi: 10.1377/hlthaff.2013.1145] [Medline: 25288428]
2. Pew Research Center. 2019 Jul 12. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-02-10]
3. Lyles C, Schillinger D, Sarkar U. Connecting the dots: health information technology expansion and health disparities. PLoS Med 2015 Jul;12(7):e1001852 [FREE Full text] [doi: 10.1371/journal.pmed.1001852] [Medline: 26172977]
4. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
5. Perzynski AT, Roach MJ, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc 2017 Sep 1;24(5):927-932 [FREE Full text] [doi: 10.1093/jamia/ocx020] [Medline: 28371853]
6. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
7. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
8. Rodriguez JA, Lipsitz SR, Lyles CR, Samal L. Association between patient portal use and broadband access: a national evaluation. J Gen Intern Med 2020 Jan 10:-. [doi: 10.1007/s11606-020-05633-4] [Medline: 31925739]
9. Health Information National Trends Survey. About HINTS URL: https://hints.cancer.gov/about-hints/learn-more-about-hints.aspx [accessed 2018-09-23]
10. Wigfall LT, Friedman DB. Cancer information seeking and cancer-related health outcomes: a scoping review of the health information national trends survey literature. J Health Commun 2016 Sep;21(9):989-1005 [FREE Full text] [doi: 10.1080/10810730.2016.1184358] [Medline: 27466828]
11. Hiatt RA, Sibley A, Fejerman L, Glantz S, Nguyen T, Pasick R, et al. The San Francisco Cancer Initiative: a community effort to reduce the population burden of cancer. Health Aff (Millwood) 2018 Jan;37(1):54-61. [doi: 10.1377/hlthaff.2017.1260] [Medline: 29309234]
12. Khoong EC, Le GM, Hoskote M, Rivadeneira NA, Hiatt RA, Sarkar U. Health information-seeking behaviors and preferences of a diverse, multilingual urban cohort. Med Care 2019 Jun;57(Suppl 6 Suppl 2):S176-S183. [doi: 10.1097/MLR.0000000000001050] [Medline: 31095058]
13. 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 2016 Jan;31(1):45-51 [FREE Full text] [doi: 10.1007/s11606-015-3374-7] [Medline: 26105675]
14. Singh H, Fox SA, Petersen NJ, Shethia A, Street RL. Older patients' enthusiasm to use electronic mail to communicate with their physicians: cross-sectional survey. J Med Internet Res 2009 Jun 16;11(2):e18 [FREE Full text] [doi: 10.2196/jmir.1143] [Medline: 19632972]
15. Virji A, Yarnall KS, Krause KM, Pollak KI, Scannell MA, Gradison M, et al. Use of email in a family practice setting: opportunities and challenges in patient- and physician-initiated communication. BMC Med 2006 Aug 15;4:18 [FREE Full text] [doi: 10.1186/1741-7015-4-18] [Medline: 16911780]

16. 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 Mar 6;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
17. Dudas RA, Crocetti M. Pediatric caregiver attitudes toward email communication: survey in an urban primary care setting. J Med Internet Res 2013 Oct 23;15(10):e228 [FREE Full text] [doi: 10.2196/jmir.2738] [Medline: 24152542]
18. Baptist AP, Thompson M, Grossman KS, Mohammed L, Sy A, Sanders GM. Social media, text messaging, and email-preferences of asthma patients between 12 and 40 years old. J Asthma 2011 Oct;48(8):824-830. [doi: 10.3109/02770903.2011.608460] [Medline: 21864099]
19. Iftikhar R, Abaalkhail B. Health-seeking influence reflected by online health-related messages received on social media: cross-sectional survey. J Med Internet Res 2017 Nov 16;19(11):e382 [FREE Full text] [doi: 10.2196/jmir.5989] [Medline: 29146568]
20. Taggart T, Grewe M, Conserve D, Gliwa C, Isler MR. Social media and HIV: a systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov 2;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
21. 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]
22. Ratwani RM, Reider J, Singh H. A decade of health information technology usability challenges and the path forward. J Am Med Assoc 2019 Feb 26;321(8):743-744. [doi: 10.1001/jama.2019.0161] [Medline: 30715133]
23. Adler-Milstein J, Holmgren A, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital 'advanced use' divide. J Am Med Inform Assoc 2017 Nov 1;24(6):1142-1148. [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
24. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
25. American FactFinder - Census Bureau. American FactFinder URL: https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml [accessed 2019-11-06]
26. Centers for Disease Control and Prevention. About the National Health Interview Survey URL: http://www.cdc.gov/nchs/nhis/about_nhis.htm [accessed 2018-02-09]
27. 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]
28. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. J Adv Nurs 2004 Oct;48(2):175-186. [doi: 10.1111/j.1365-2648.2004.03185.x] [Medline: 15369498]
29. Rivadeneira NA, Hoskote M, Le GM, Nguyen TT, Nápoles AM, Pasick RJ, et al. Advancing cancer control in San Francisco: cancer screening in under-represented populations. Am J Prev Med 2020 Jan;58(1):e1-e9. [doi: 10.1016/j.amepre.2019.08.024] [Medline: 31862104]
30. 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]
31. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
32. Longo DR. Understanding health information, communication, and information seeking of patients and consumers: a comprehensive and integrated model. Health Expect 2005 Sep;8(3):189-194 [FREE Full text] [doi: 10.1111/j.1369-7625.2005.00339.x] [Medline: 16098149]
33. Rowley J, Johnson F, Sbaffi L. Gender as an influencer of online health information-seeking and evaluation behavior. J Assn Inf Sci Technol 2017 Jan;68(1):36-47. [doi: 10.1002/asi.23597]
34. 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 Jul 4;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]
35. Clayman ML, Manganello JA, Viswanath K, Hesse BW, Arora NK. Providing health messages to Hispanics/Latinos: understanding the importance of language, trust in health information sources, and media use. J Health Commun 2010;15(Suppl 3):252-263 [FREE Full text] [doi: 10.1080/10810730.2010.522697] [Medline: 21154097]
36. 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 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]

37. Iachan R, Berman L, Kyle TM, Martin KJ, Deng Y, Moyse DN, et al. Weighting nonprobability and probability sample surveys in describing cancer catchment areas. Cancer Epidemiol Biomarkers Prev 2019 Mar;28(3):471-477 [FREE Full text] [doi: 10.1158/1055-9965.EPI-18-0797] [Medline: 30642842]
38. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
39. 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]
40. San Francisco Mayor's Office of Housing and Community Development. Digital Equity URL: https://sfmohcd.org/digital-equity [accessed 2019-11-04]
41. Cheng C, Dunn M. How well are health information websites displayed on mobile phones? Implications for the readability of health information. Health Promot J Austr 2017 Mar;28(1):15-20. [doi: 10.1071/HE15127] [Medline: 27250916]
42. Statista. 2019. Most Common Languages Used on the Internet 2019 URL: https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/ [accessed 2019-11-04]
43. Chaet AV, Morshedi B, Wells KJ, Barnes LE, Valdez R. Spanish-language consumer health information technology interventions: a systematic review. J Med Internet Res 2016 Aug 10;18(8):e214 [FREE Full text] [doi: 10.2196/jmir.5794] [Medline: 27511437]
44. Diamond L, Izquierdo K, Canfield D, Matsoukas K, Gany F. A systematic review of the impact of patient-physician non-English language concordance on quality of care and outcomes. J Gen Intern Med 2019 Aug;34(8):1591-1606. [doi: 10.1007/s11606-019-04847-5] [Medline: 31147980]
45. 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]
46. Gee GC, Walsemann KM, Takeuchi DT. English proficiency and language preference: testing the equivalence of two measures. Am J Public Health 2010 Mar;100(3):563-569. [doi: 10.2105/AJPH.2008.156976] [Medline: 19696376]
47. Housten AJ, Hoover DS, Correa-Fernández V, Strong LL, Heppner WL, Vinci C, et al. Associations of acculturation with English- and Spanish-language health literacy among bilingual Latino adults. Health Lit Res Pract 2019 Apr;3(2):e81-e89 [FREE Full text] [doi: 10.3928/24748307-20190219-01] [Medline: 31294309]
48. Wang W, Yu N. Coping with a new health culture: acculturation and online health information seeking among Chinese immigrants in the United States. J Immigr Minor Health 2015 Oct;17(5):1427-1435. [doi: 10.1007/s10903-014-0106-8] [Medline: 25228489]
49. Massey PM, Langellier BA, Sentell T, Manganello J. Nativity and language preference as drivers of health information seeking: examining differences and trends from a US population-based survey. Ethn Health 2017 Dec;22(6):596-609 [FREE Full text] [doi: 10.1080/13557858.2016.1244745] [Medline: 27766894]
50. Zheng X, Woo BK. E-mental health in ethnic minority: a comparison of YouTube and talk-based educational workshops in dementia. Asian J Psychiatr 2017 Feb;25:246-248. [doi: 10.1016/j.ajp.2016.12.002] [Medline: 28262161]
51. Lam NH, Tsiang JT, Woo BK. Exploring the role of YouTube in disseminating psychoeducation. Acad Psychiatry 2017 Dec;41(6):819-822. [doi: 10.1007/s40596-017-0835-9] [Medline: 29022242]
52. Chung K, Augustin F, Esparza S. Development of a Spanish-language hospice video. Am J Hosp Palliat Care 2017 Sep;34(8):737-743. [doi: 10.1177/1049909116658022] [Medline: 27380929]
53. Lin SC, Lyles CR, Sarkar U, Adler-Milstein J. Are patients electronically accessing their medical records? Evidence from national hospital data. Health Aff (Millwood) 2019 Nov;38(11):1850-1857. [doi: 10.1377/hlthaff.2018.05437] [Medline: 31682494]
54. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, Moffet HH, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study. J Am Med Inform Assoc 2013;20(6):1128-1131 [FREE Full text] [doi: 10.1136/amiajnl-2012-001567] [Medline: 23676243]
55. Ronda MC, Dijkhorst-Oei LT, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
56. Sitzia J, Wood N. Response rate in patient satisfaction research: an analysis of 210 published studies. Int J Qual Health Care 1998 Aug;10(4):311-317. [doi: 10.1093/intqhc/10.4.311] [Medline: 9835247]

Abbreviations

aOR: adjusted odds ratio
EHR: electronic health record
HINTS: Health Information National Trends Survey
HITECH: Health Information Technology for Economic and Clinical Health
NCI: National Cancer Institute

NIH: National Institutes of Health
REDCap: Research Electronic Data Capture
SFHINTS: San Francisco Health Information National Trends Survey
TAM: technology acceptance model

Edited by G Eysenbach; submitted 10.11.19; peer-reviewed by M Peltola, B Woo, R Lee; comments to author 22.01.20; revised version received 28.01.20; accepted 28.01.20; published 06.04.20.

Please cite as:
Khoong EC, Rivadeneira NA, Hiatt RA, Sarkar U
The Use of Technology for Communicating With Clinicians or Seeking Health Information in a Multilingual Urban Cohort: Cross-Sectional Survey
J Med Internet Res 2020;22(4):e16951
URL: https://www.jmir.org/2020/4/e16951
doi:10.2196/16951
PMID:32250280

Original Paper

Re-Enactment as a Method to Reproduce Real-World Fall Events Using Inertial Sensor Data: Development and Usability Study

Kim Sarah Sczuka[1], MSc; Lars Schwickert[1], PhD; Clemens Becker[1], MD; Jochen Klenk[1,2,3], PhD

[1]Department of Clinical Gerontology, Robert-Bosch-Hospital, Stuttgart, Germany

[2]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany

[3]IB University for Applied Science Berlin, Study Center Stuttgart, Stuttgart, Germany

Corresponding Author:
Kim Sarah Sczuka, MSc
Department of Clinical Gerontology
Robert-Bosch-Hospital
Auerbachstraße 110
Stuttgart, 70376
Germany
Phone: 49 711 8101 6078
Fax: 49 711 8101 3194
Email: kim.sczuka@rbk.de

Abstract

Background: Falls are a common health problem, which in the worst cases can lead to death. To develop reliable fall detection algorithms as well as suitable prevention interventions, it is important to understand circumstances and characteristics of real-world fall events. Although falls are common, they are seldom observed, and reports are often biased. Wearable inertial sensors provide an objective approach to capture real-world fall signals. However, it is difficult to directly derive visualization and interpretation of body movements from the fall signals, and corresponding video data is rarely available.

Objective: The re-enactment method uses available information from inertial sensors to simulate fall events, replicate the data, validate the simulation, and thereby enable a more precise description of the fall event. The aim of this paper is to describe this method and demonstrate the validity of the re-enactment approach.

Methods: Real-world fall data, measured by inertial sensors attached to the lower back, were selected from the Fall Repository for the Design of Smart and Self-Adaptive Environments Prolonging Independent Living (FARSEEING) database. We focused on well-described fall events such as stumbling to be re-enacted under safe conditions in a laboratory setting. For the purposes of exemplification, we selected the acceleration signal of one fall event to establish a detailed simulation protocol based on identified postures and trunk movement sequences. The subsequent re-enactment experiments were recorded with comparable inertial sensor configurations as well as synchronized video cameras to analyze the movement behavior in detail. The re-enacted sensor signals were then compared with the real-world signals to adapt the protocol and repeat the re-enactment method if necessary. The similarity between the simulated and the real-world fall signals was analyzed with a dynamic time warping algorithm, which enables the comparison of two temporal sequences varying in speed and timing.

Results: A fall example from the FARSEEING database was used to show the feasibility of producing a similar sensor signal with the re-enactment method. Although fall events were heterogeneous concerning chronological sequence and curve progression, it was possible to reproduce a good approximation of the motion of a person's center of mass during fall events based on the available sensor information.

Conclusions: Re-enactment is a promising method to understand and visualize the biomechanics of inertial sensor-recorded real-world falls when performed in a suitable setup, especially if video data is not available.

(J Med Internet Res 2020;22(4):e13961) doi:10.2196/13961

KEYWORDS

falls; simulation; inertial sensor; method

Introduction

Falls are a common health problem that can lead to serious physical consequences such as fractures, reduced quality of life, loss of independence, and institutionalization. Furthermore, fall-related injuries seriously increase mortality in older persons [1]. One-third of community-dwelling people older than 65 years fall at least once a year, and half of them fall more than once [2]. Besides the individuals' health burden, falls also have a major social and economic effect with annual costs accounting for 0.85% to 1.5% of the total health care expenditures [3].

Meta-analyses identified about 30 fall risk factors in community-dwelling older persons [4]. These risks include fall history as well as balance and gait problems. Nevertheless, fall risk prediction models show limited performance, which suggests that we do not fully understand the complex interplay of factors triggering fall events. This may be because information about fall events are mainly derived from subjective reports by fallers or proxies, which can be biased in many ways [5]. Lack of reporting or false reporting can be related to cognitive impairment of the subjects, shame of reporting and fear of consequences, or simply due to difficulties in defining a fall [6]. Objective information is rare and, therefore, many aspects including fall-related activities, environmental factors and movement patterns before, during, and after the falling phase remain unclear. Body-worn sensor technology might enhance our understanding of falls and thereby also lead to more effective methods for fall prevention, fall risk assessment, and fall detection. With the rapid development of eHealth, small wearable devices such as body-worn sensor technology can provide objective measures of physical activity and the kinematics of human movement [7].

Although falls are common, it is challenging to capture real-world fall signals due to the long observation period and the limited recording duration of sensor devices. The Fall Repository for the Design of Smart and Self-Adaptive Environments Prolonging Independent Living (FARSEEING) consortium, funded by the seventh European Union Framework Program for Research, has been able to capture and validate real-world falls of older people who have an increased risk of falling, measured by body-worn sensor technology [8]. Analysis of these fall signals showed, for example, that the characteristics presented by inertial sensor measurements are relevant to improve the understanding of the postimpact phase [9]. Movement patterns during the ground phase were different between fall events with and without successful recovery to a standing position. These findings are important for redesigning emergency response processes after falls to better support individuals in cases of an unrecovered fall.

Even though the signals provide precise measures for acceleration, angular velocity, and magnetic north, it is not possible to directly derive visualization and interpretation of movements during a fall event. In contrast, video data facilitates the possibility of estimating the kinematics of falls as well as the fall-related movements before and after the fall event [10]. However, analyzing the complex movement patterns of fallers from planar video data is challenging, due to motions of body segments that are out of the plane or occluded [11]. Furthermore, due to privacy issues, it is usually not possible to capture video data during everyday life. In the absence of well-described real-world fall recordings and due to the huge effort in recording objective fall data, researchers have tried to bridge the knowledge gaps by simulating fall events. However, comparisons between acceleration signals of simulated and real-world fall events has shown that there are considerable differences [12]. This might be due to a lack of real-world data for designing a more suitable and realistic simulation protocol. For example, the preimpact phase was excluded from simulations, but analysis of this phase could identify protective movements or provide a better understanding of the circumstances that lead to a fall. Furthermore, if the simulated fall was self-initiated, the movement pattern differed a lot from the movement pattern of the real-world fall, because the volunteers did not know how to fall in a realistic way. It was shown that the acceleration values were closer to those of real-world fall events when the subjects were forced to fall by releasing them suddenly from a backward lean with the instruction to avoid a fall [12]. To perform more suitable simulations, it is essential to find methods to create a realistic experimental protocol that can be easily reproduced. The obtained information could be of great value and give insight into the causes of falls as well as what happened during the fall. To the best of our knowledge there is no current method to obtain this kind of information based on sensor signals without having another information source such as video. However, such information could help better predict falls and develop new fall prevention interventions as well as fall detection approaches.

Connell and Wolf [13] previously proposed a re-enactment method to validate subjective fall reports. Participants were interviewed and asked to re-enact in detail (if they felt comfortable) all activities, body movements, body part placements, and interactions with the environment at the location of the incident to obtain more precise information. This method of re-enactment is a promising approach to improve simulation protocols and to produce more realistic fall simulations. We adapted the re-enactment method to visualize and enhance the interpretation of sensor signals. The aim of this study was to describe this adapted method and to demonstrate the validity of the re-enactment approach by means of a selected common fall example.

Methods

Real-World Fall Data

Real-world fall data measured by inertial sensors during everyday life were obtained from subjects in different settings (eg, community-dwelling, geriatric rehabilitation) and different populations with moderate to high risk of falling (eg, Parkinson disease, cerebellar and sensory ataxia). All fall events were stored in the FARSEEING database [8]. The process of data collection by combining different sources was approved by the Ethics Committee of the University of Tübingen (495/2012BO2) and the data protection office of the Federal State of Baden-Württemberg, Germany (T 1500/231). The large number

of real-world fall events within the FARSEEING database facilitated the comparisons of sensor signals that represented similar curve progressions as well as the confirmation of the feasibility to apply the re-enactment method for several diverse fall paradigms. For the purposes of exemplification, we selected one fall event of a female patient (42 years of age, height=154 cm, weight=60 kg, Montreal Cognitive Assessment=27 [14], Timed Up-and-Go=17.72 seconds [15], Short Physical Performance Battery=9 [16]) with ataxia that presented a reliable fall report and sensor signal. The corresponding fall report described the event as a forward fall initiated by stumbling over the entrance door sill. Analysis and interpretation of the triaxial sensor signal concerning movement patterns and curve progression confirmed the fall description. Based on the collected signals from the repository, the selected fall signal represents a common fall paradigm that corresponds with everyday life situations.

Data Processing

Data acquisition was performed during the patient monitoring as well as during the re-enactment experiments using the Samsung Galaxy (SG) S3 smartphone worn on the lower back at the lumbar position (L5) with a belt, close to the center of mass. The smartphone includes a triaxial accelerometer (2 g SGS3) sampled at 100 Hz. Data were stored for off-line analysis on the smartphone. Orientation was defined as follows: z=vertical, y=mediolateral, and x=sagittal. Additionally, the re-enactment experiments were captured with a video camera (SGS8, 200-Hz sampling rate) to analyze the movements in detail. For this purpose, the video data and the sensor signal were synchronized to assign specific postures, as seen from the video tape, to each frame of the acceleration signal.

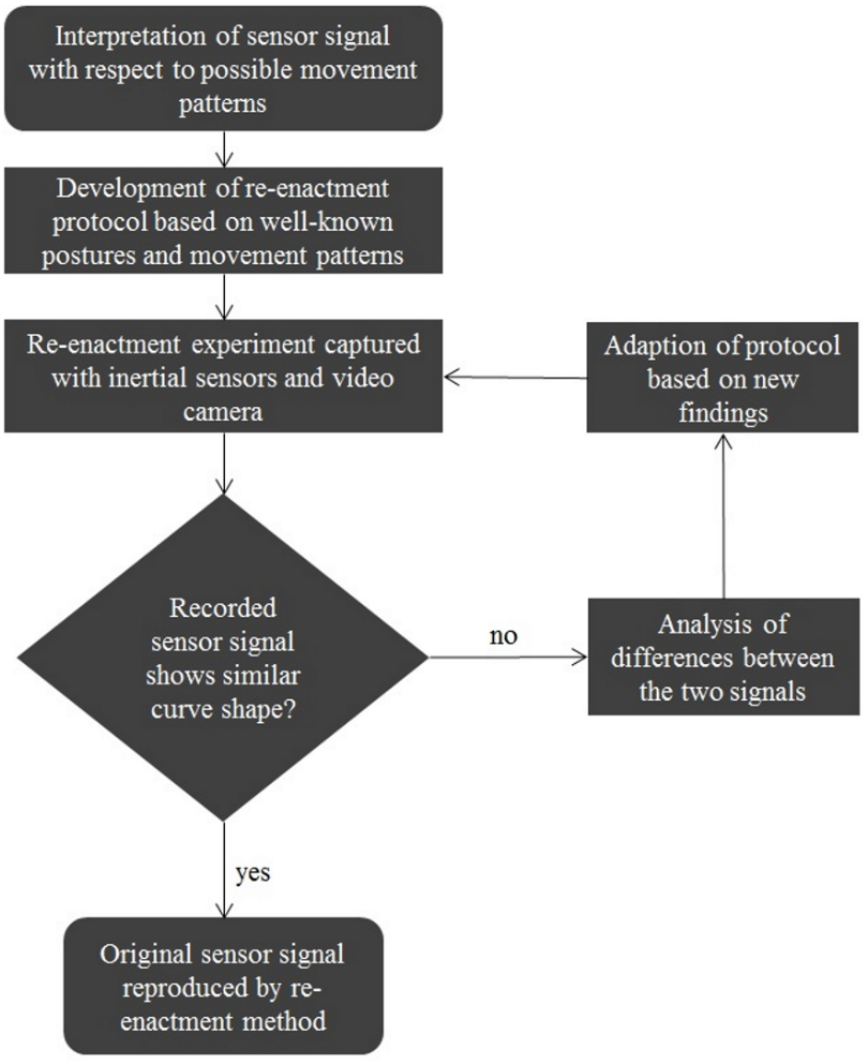
Re-Enactment Protocol

Aiming to establish a simulation protocol, the selected fall signal was analyzed with regard to the movement patterns during the prefall phase (5-10 seconds before impact), falling phase, and impact phase as well as the resting and recovery phase with the faller achieving an upright standing position [17]. Based on the signal interpretation, a simulation protocol including identified postures and movement sequences was established. The protocol was conducted by an expert (woman, 28 years of age, height=168 cm, weight=61 kg, healthy, and physically active) in analyzing sensor signals of real-world fall events. Simulations started with the prefall activity and ended with the person standing upright subsequent to the impact phase. The re-enactment method was conducted under safe conditions in a laboratory setting using protective layers of mattresses to reduce the impact and avoid injuries. With the aim of producing a simulated acceleration signal similar to the real-world fall signal the protocol was performed several times. Subsequently, the re-enacted signals were compared to the real-world fall signal, and the protocol was adapted based on the findings after the first re-enactment experiment. Special attention was paid to the z component, as it indicated the motion in vertical direction as well as bending movements of the trunk section. The adapted protocol was conducted again and the newly recorded signals were compared to the original signal. In cases of new findings, the protocol was adapted again as a basis for a further trial. This re-enactment process is visualized in Figure 1. The adaptation of the protocol was repeated until the experiment led to satisfying results that showed a similar curve shape compared to the original signal from the real-world fall event.

Validation of Re-Enactment Method

The resemblance between the simulated and the real-world fall signals was analyzed using a dynamic time warping (DTW) algorithm, which enables the comparison of two temporal sequences varying in speed and timing. The DTW algorithm compensated the temporal differences between the sensor signals. Both time series were aligned by stretching the two vectors to minimize the sum of the Euclidean distances between the corresponding points. DTW alignment was processed in R-3.4.2 (R Foundation for Statistical Computing, Vienna, Austria) with the slope-constrained step pattern “asymmetricP1” published by Sakoe and Chiba [18] with open start and open end to achieve time normalization by transforming the time axis of the real-world fall signal pattern (query) onto that of the re-enacted one (template). The slope constraining factor P=1 was chosen due to its best recognition performance, which was also shown in the study of Sakoe and Chiba [18]. Similarity between the curves was calculated by the normalized distance as defined by the Euclidean distance divided by the number of samples in the query.

Figure 1. Flowchart of the re-enactment method.

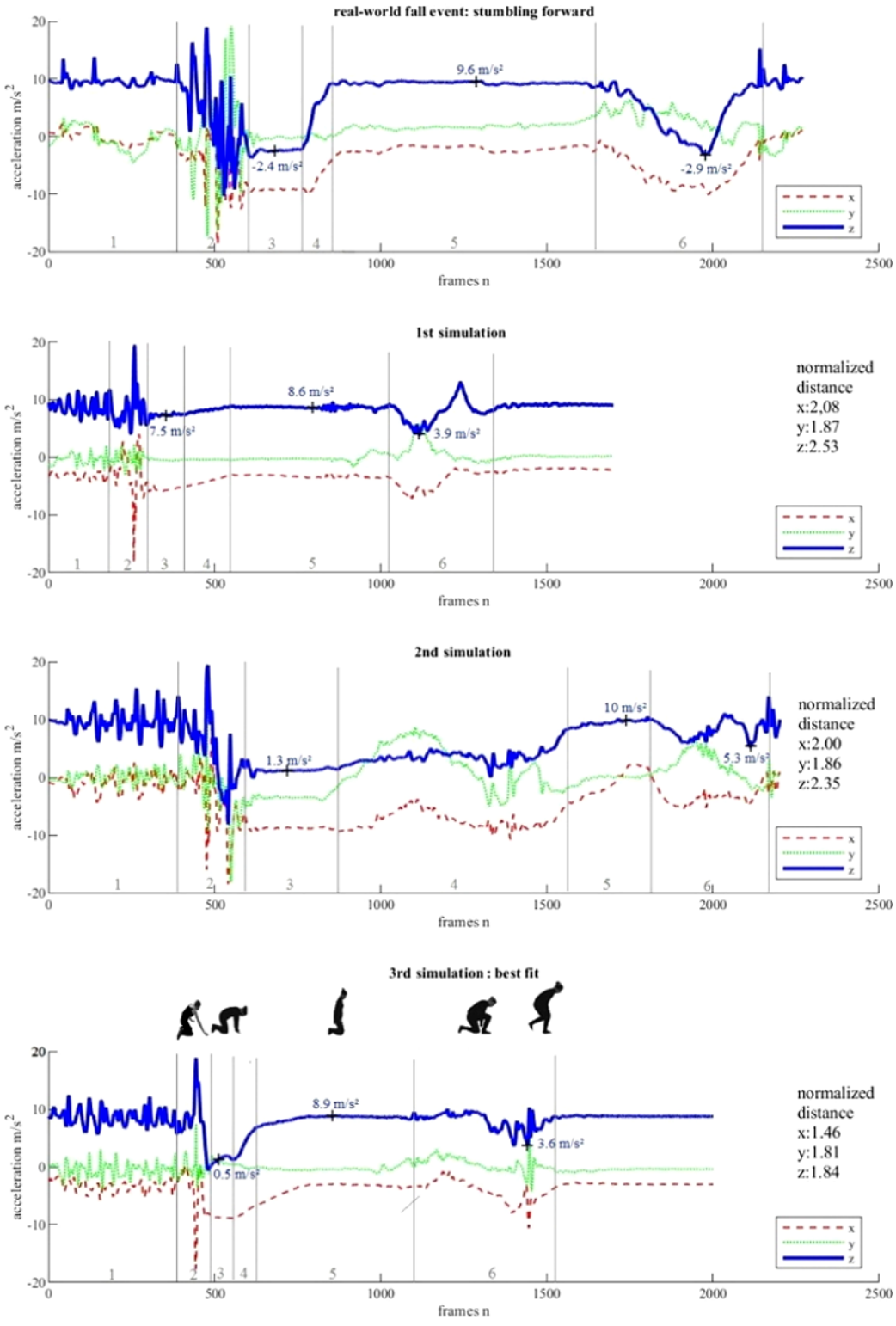


Results

Figure 2 shows the triaxial acceleration signal of the real-world fall event at the top, and the three acceleration signals recorded during the re-enactment experiments below. The movement sequence of the chosen real-world fall event was as follows: standing unsteady, stumbling over doorsill, falling forward on the knees, and standing up again. The first simulation followed these rough instructions and the result was not satisfying, demonstrated by the second plot from the top. The impact in section 2 could not be simulated realistically due to the soft mattress, and the frequency of the steps in section 1. Focusing on the z component (vertical), the original fall signal showed a steeper slope after the impact compared to the signal of the first simulation. The acceleration values of the simulated signal in section 3 were higher as well as the value of the local minimum in section 6; whereas the values in section 5 were lower compared to the signal of the real-world fall event. Furthermore, the lifting of the upper body in section 4 was barely visible. However, the curve shape in section 3, 5, and 6 was analogous to those of the real-world fall signal. The normalized distance was 2.53, which was the highest value in all experiments. The x component of the first simulated sensor signal showed a similar curve progression in sections 3, 5, and 6 but differed regarding the acceleration values in comparison to the real-world fall signal. For the y component, the curve shape was quite similar compared with the real-world fall signal with slight deviations in section 6. Subsequently, the protocol was adapted: more pronounced steps (section 1), bending forward while kneeling (section 3), and resting in this posture for a short time period.

The second simulation produced more similar acceleration values but was still too high for the z component in sections 3 and 6. The forward bending of the upper body while standing up was not pronounced enough, which was illustrated by the minimal value of the z component in section 6. Furthermore, the z component in section 6 showed two local minima instead of one as shown in the signal of the real-world fall. Section 5 especially showed a time period that was too short. Noticeable was section 4, which was about 7 times longer and included a lot more motion in all three axes compared to the real-world fall event. The x component improved within section 3 and 6, but section 4 and 5 remained different in acceleration values and curve progression. For the y component the curve shape had to be improved, especially in section 4. However, the normalized Euclidean distance was lower for all three axes, indicating a general improvement of the simulation.

Figure 2. Triaxial acceleration signals of a real-world fall event and signals stepwise derived during the re-enactment experiments. Numbers 1 to 6 indicate the particular phases of the fall event (1: prefall phase with steps, 2: stumbling, falling, and impact, 3: resting, 4: raising upper body, 5: resting, 6: straighten up into standing position).



The findings were added to the protocol, and the experiment was repeated with the following instructions: scuttling or less pronounced steps (section 1), stumbling and falling forward on the knees (section 2), bending the upper body forward and touching the ground with the hands for 1 to 2 seconds (section 3), sudden raising (section 4), 4 seconds of resting on the knees with upper body upright (section 5), placing the left foot on the floor and standing up without the upper body swaying (section 6).

The resulting signal of the third simulation experiment is the lowermost subplot in Figure 2. With exception of section 1 and 2, all three axes showed very similar curve shapes compared with the real-world fall signal. The visual similarity was confirmed by the decreased normalized Euclidean distance for

all three axes. The lowest distance was at 1.46, calculated for the x component. The x and z component showed a nearly 30% reduction of the normalized Euclidean distance; although the bending in section 6 could have been more pronounced for the z and x component, and the whole simulated signal was about 6 seconds shorter.

Figure 3 illustrates the alignment of the vertical acceleration (z) component of the real-world fall acceleration signal and the signal obtained during the third simulation. The appropriate movement sequences were linked, but there was a temporal delay resulting from the shorter duration of the re-enacted fall signal.

Figure 3. Time series alignment of the vertical acceleration signals for the real-world fall event as well as the best-fit re-enacted signal.

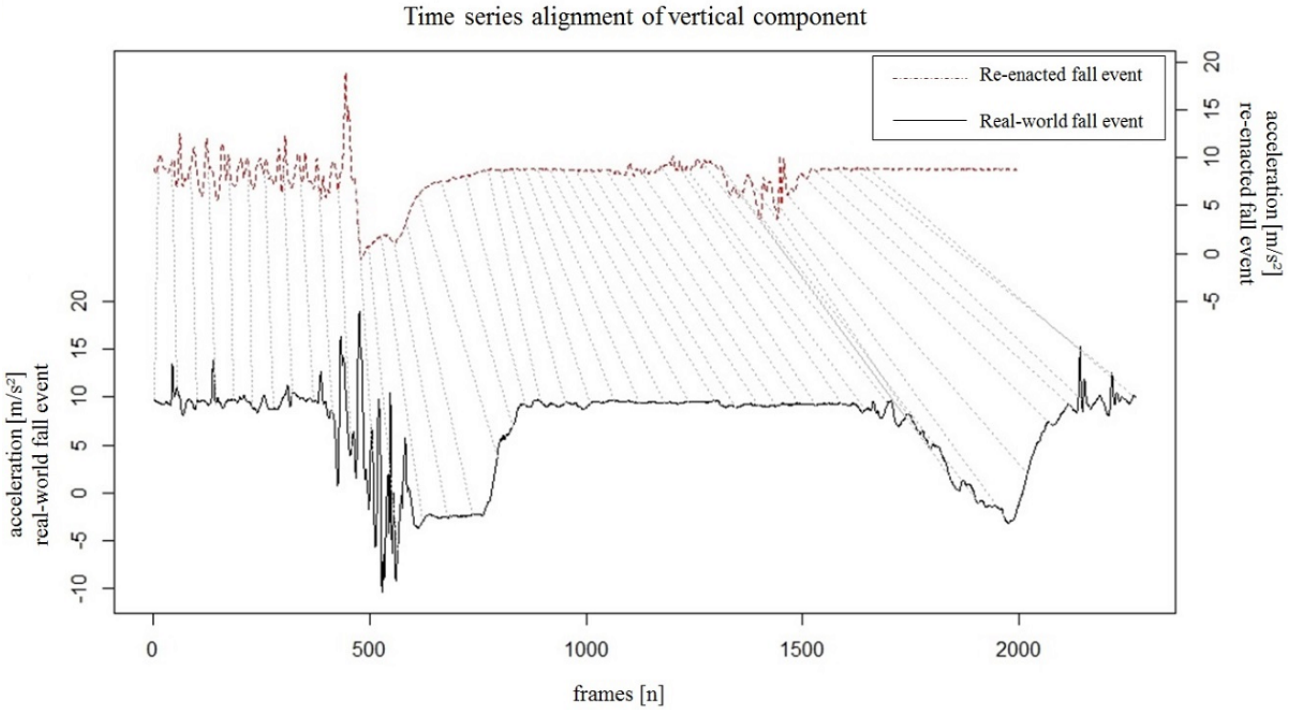
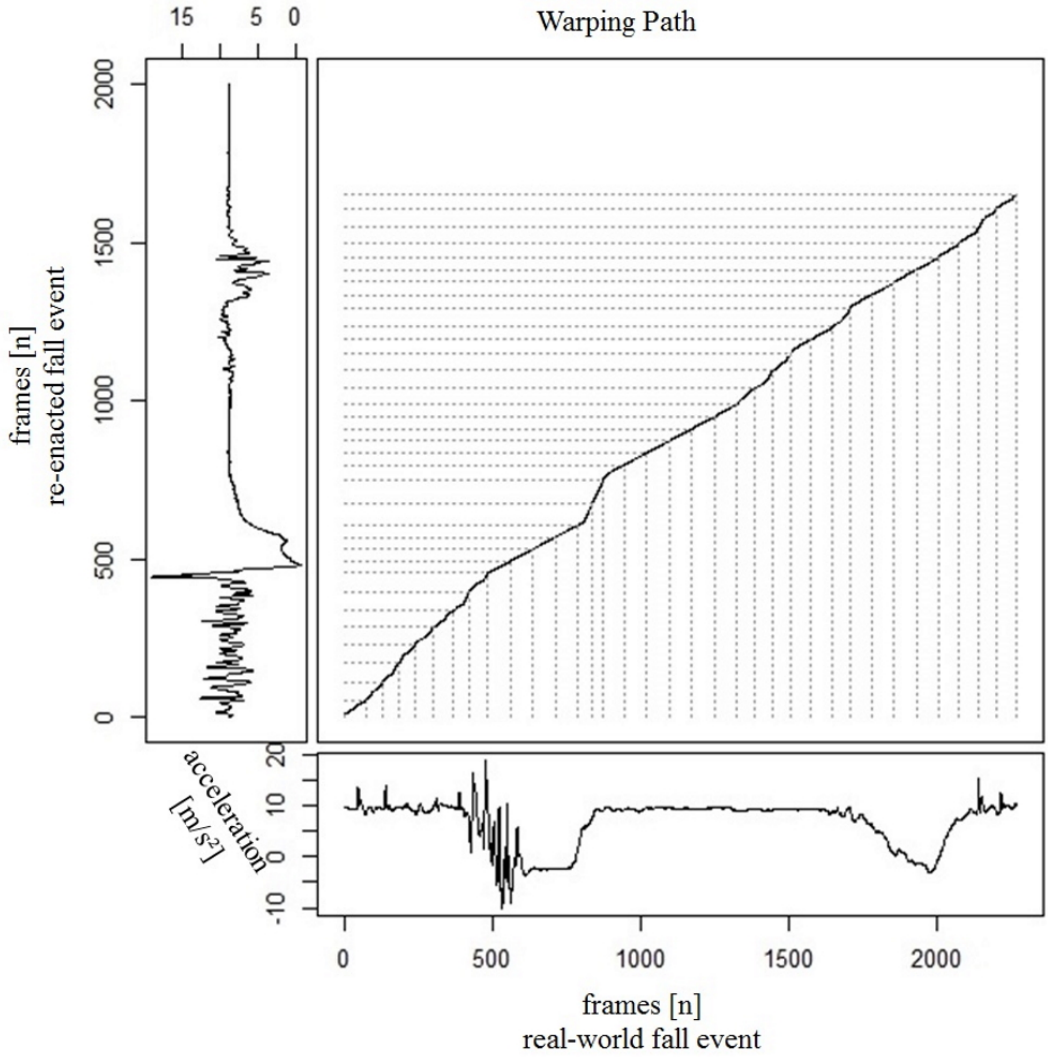


Figure 4 displays the warping path for the vertical (z) component of the real-world fall signal and the re-enacted fall signal derived from accelerometer data. Only one part of the real-world fall signal located around frame 800 showed a dominant discrepancy and led to a sharp bend in the line because of the shorter resting phase (section 5, Figure 2) in the re-enacted signal. The DTW algorithm compensated the resulting differences in length within this section by stretching this part of the signal in the re-enacted fall event by repeating each element as many times as necessary.

Figure 4. Time series alignment and warping path calculated using a dynamic time warping algorithm for the vertical component of the real-world and re-enacted fall signal.



Discussion

Principle Findings

Re-enactment was demonstrated as a suitable approach to provide new insight into real-world fall events recorded with inertial sensors, as well as fall events in general. It was possible to simulate fall events more realistically and thereby verify the interpretation even though falls were heterogeneous and showed a high variability due to the 3D movements of the fallers.

To the best of our knowledge, the introduced method is a new approach to reproduce with good approximation the inertial sensor signals of real-world fall events, which allows for more valid fall simulations under safe laboratory conditions. Prior re-enactment studies already showed that this method is adequate to conduct an examination of behavioral and environmental circumstances associated with falls [13]. We modified this approach by enhancing the written fall report with the corresponding real-world fall signal. Reproducing the sensor signal by re-enactment and additional videotaping of the simulation seems to be a suitable method to prove the interpretation of a real-world fall event when video data is missing. By synchronizing video and sensor data, every change in curve progression of the sensor signal can be associated with a specific posture or movement pattern and the circumstances can be retraced more easily.

Comparison between real-world and re-enacted fall signals was performed by applying a DTW algorithm. This method was suitable to measure the similarity between two events of different lengths in the time series. This was an important finding since our experiments demonstrated that it was impossible to simulate fall events chronologically with enough precision in the exact timing of each specific posture. Most other correlation approaches cannot handle this problem. It was further shown that the applied DTW algorithm was able to stretch the signals in a way that enables the characteristic patterns to be assigned to each other. There is no natural threshold for the normalized distance to show an appropriate fit. However, previous studies already indicated that DTW was a suitable approach with regard to pattern recognition in human motion regardless of thresholds [19-22]. Our results also suggest that the DTW approach might be a promising method to detect entire fall patterns or at least parts of the events from accelerometer data. This could also help to further improve fall detection algorithms.

Our results show that the re-enactment method provides a tool to understand fall-related movements. Based on the FARSEEING database, it will be possible to build up a database with sequences of sensor signals that represent specific movement patterns and link the corresponding pictures or video data that show the re-enacted movement. Using such a database in combination with the method of DTW could help to identify movement sequences of newly acquired fall signals and thereby improve the understanding of falls, including the causes and consequences. With this new knowledge it will be possible to develop more reliable fall risk models [23].

The FARSEEING database contains about 200 well-described fall events [8]; however, realistic simulations are still necessary. It seems unrealistic to collect a sufficient number of real-world falls for all open research questions and data intensive analytic methods such as automated machine learning approaches for fall detection. Currently, fall detection is insufficient, such as when implemented in home alarm systems. The main reason seems to be unrealistic simulated falls by younger subjects with insufficient knowledge of real-world falls [12]. Applying such fall detection algorithms to real-world situations resulted in high rates of false positive or false negative alarms [24]. The application of data on real-world falls from the FARSEEING database in the algorithm development highly improved the detection performance [25]. However, there is still an unacceptable high false alarm rate. Machine-learning approaches already show promising results in activity recognition based on data from waist-worn inertial sensors [26,27] and might further improve the results, but would need additional realistic fall data.

The re-enactment method will facilitate and enhance the quality of fall simulation to provide more realistic data input for algorithm developers. Volunteers could be trained to simulate real-world falls that were sufficiently similar using re-enacted fall signal data. As a next step, a study is planned to systematically re-enact the falls provided by the FARSEEING database and to analyze the repeatability of the results when re-enactment is performed by different persons.

Strengths and Limitations

This new approach was developed and validated to broaden the knowledge of real-world fall events and close the information gap in cases of missing or fragmentary fall description. A remarkable strength of this study is the development of the re-enactment method based on real-world fall data derived from the FARSEEING meta database, which is currently the largest collection of real-world fall events recorded with inertial sensors [8]. Although fall events are heterogeneous, it was possible to compare several real-world data sets with a similar fall scenario and to identify fall paradigm-specific patterns, which could be replicated within the re-enactment protocol and lead to realistic simulations. Furthermore, it seems that re-enactment can be performed by any person able to simulate the fall event without matching the clinical characteristics of the original faller. Shawen et al [28] even demonstrated that a fall detection approach based on the inertial sensor signals of healthy participants recorded with a smartphone can be used to resemble characteristics of other populations, such as individuals using prostheses. Even though we introduced the re-enactment method by means of only one exemplary real-world fall event, this new approach is feasible for several real-world fall data with similar results as shown. Further examples can be found in the Multimedia Appendicies 1-3.

However, this method has several limitations. Due to safety conditions the re-enactment experiments had to be performed with constraints on the impact phase. A soft mattress was used to lessen the impact during re-enactment experiments. Therefore, the acceleration values caused by the impact differ from those of real-world fall signals. Nevertheless, the reliable alignment of both signals can be seen in Figure 3. The DTW algorithm that was used is able to match the signals on the basis of a similar curve progression, despite the values in the original and re-enacted signals differing in the relatively short part of the impact phase. However, when focusing on the impact phase it will be necessary to use signals derived from real-world falls with sufficient ranges of acceleration. It was previously shown that signals derived from simulated impacts were not acceptable, at least when fall detection algorithms were evaluated [29].

Even though the simulation of fall events, including all phases, is feasible in general, less dominant motions might have been overlooked and not reproduced by re-enactment. Due to the fact that no video data were available from the real-world fall events, the re-enacted movements were impossible to verify. In addition, the fall reports were often limited and provided little to no information concerning movement behavior and circumstances.

Furthermore, we only performed the re-enactment method for common fall paradigms such as stumbling over a doorsill or falling backwards while opening a door. It seems that these paradigms can be reliably simulated in a laboratory. However, every paradigm might not be transferable in an experimental setting (eg, falling due to dizziness or collision with another person).

Conclusions

The re-enactment approach provides a possibility to generate data that is similar to those of real-world fall events. This method could help to better understand real-world falls and further improve the simulation of fall events to increase the available realistic fall data for algorithm development.

Acknowledgments

The research leading to these results has been funded by the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement n. 288940 (FARSEEING project) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 401741517).

Authors' Contributions

KS wrote the manuscript. KS, CB, and JK designed and conceptualized the study. KS and LS collected data. KS, LS, and JK analyzed the data. KS, CB, and JK interpreted the data. LS, CB, and JK critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Triaxial acceleration signals and timeseries alignment of the vertical component of a further real-world fall event (example 1) and a very similar signal derived during re-enactment experiment.
[PNG File , 1064 KB - jmir_v22i4e13961_app1.png]

Multimedia Appendix 2
Triaxial acceleration signals and timeseries alignment of the vertical component of a further real-world fall event (example 2) and a very similar signal derived during re-enactment experiment.
[PNG File , 834 KB - jmir_v22i4e13961_app2.png]

Multimedia Appendix 3
Triaxial acceleration signals and timeseries alignment of the vertical component of a further real-world fall event (example 3) and a very similar signal derived during re-enactment experiment.
[PNG File , 736 KB - jmir_v22i4e13961_app3.png]

References

1. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. Age Ageing 2006 Sep;35 Suppl 2:ii37-ii41. [doi: 10.1093/ageing/afl084] [Medline: 16926202]
2. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988 Dec 29;319(26):1701-1707. [doi: 10.1056/NEJM198812293192604] [Medline: 3205267]
3. Heinrich S, Rapp K, Rissmann U, Becker C, König H. Cost of falls in old age: a systematic review. Osteoporos Int 2010 Jun;21(6):891-902. [doi: 10.1007/s00198-009-1100-1] [Medline: 19924496]
4. Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis. Epidemiology 2010 Sep;21(5):658-668. [doi: 10.1097/EDE.0b013e3181e89905] [Medline: 20585256]
5. Hauer K, Lamb SE, Jorstad EC, Todd C, Becker C, PROFANE-Group. Systematic review of definitions and methods of measuring falls in randomised controlled fall prevention trials. Age Ageing 2006 Jan;35(1):5-10. [doi: 10.1093/ageing/afi218] [Medline: 16364930]
6. Zecevic AA, Salmoni AW, Speechley M, Vandervoort AA. Defining a fall and reasons for falling: comparisons among the views of seniors, health care providers, and the research literature. Gerontologist 2006 Jun;46(3):367-376. [doi: 10.1093/geront/46.3.367] [Medline: 16731875]
7. de Bruin ED, Hartmann A, Uebelhart D, Murer K, Zijlstra W. Wearable systems for monitoring mobility-related activities in older people: a systematic review. Clin Rehabil 2008;22(10-11):878-895. [doi: 10.1177/0269215508090675] [Medline: 18955420]
8. Klenk J, Schwickert L, Palmerini L, Mellone S, Bourke A, Ihlen EAF, FARSEEING Consortium. The FARSEEING real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls. Eur Rev Aging Phys Act 2016;13:8 [FREE Full text] [doi: 10.1186/s11556-016-0168-9] [Medline: 27807468]
9. Schwickert L, Klenk J, Zijlstra W, Forst-Gill M, Sczuka K, Helbostad JL, et al. Reading from the Black Box: what sensors tell us about resting and recovery after real-world falls. Gerontology 2018;64(1):90-95. [doi: 10.1159/000478092] [Medline: 28848150]
10. Robinovitch SN, Feldman F, Yang Y, Schonnop R, Leung PM, Sarraf T, et al. Video capture of the circumstances of falls in elderly people residing in long-term care: an observational study. Lancet 2013 Jan 05;381(9860):47-54 [FREE Full text] [doi: 10.1016/S0140-6736(12)61263-X] [Medline: 23083889]
11. Choi WJ, Wakeling JM, Robinovitch SN. Kinematic analysis of video-captured falls experienced by older adults in long-term care. J Biomech 2015 Apr 13;48(6):911-920. [doi: 10.1016/j.jbiomech.2015.02.025] [Medline: 25769730]
12. Klenk J, Becker C, Lieken F, Nicolai S, Maetzler W, Alt W, et al. Comparison of acceleration signals of simulated and real-world backward falls. Med Eng Phys 2011 Apr;33(3):368-373. [Medline: 21123104]

13. Connell BR, Wolf SL. Environmental and behavioral circumstances associated with falls at home among healthy elderly individuals. Atlanta FICSIT Group. Arch Phys Med Rehabil 1997 Feb;78(2):179-186. [doi: 10.1016/s0003-9993(97)90261-6] [Medline: 9041900]
14. 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]
15. Wall JC, Bell C, Campbell S, Davis J. The Timed Get-up-and-Go test revisited: measurement of the component tasks. J Rehabil Res Dev 2000;37(1):109-113 [FREE Full text] [Medline: 10847578]
16. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
17. Becker C, Schwickert L, Mellone S, Bagalà F, Chiari L, Helbostad JL, FARSEEING Consortium, FARSEEING Meta Database Consensus Group. Proposal for a multiphase fall model based on real-world fall recordings with body-fixed sensors. Z Gerontol Geriatr 2012 Dec;45(8):707-715. [doi: 10.1007/s00391-012-0403-6] [Medline: 23184296]
18. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust, Speech, Signal Process 1978 Feb;26(1):43-49. [doi: 10.1109/TASSP.1978.1163055]
19. Barth J, Oberndorfer C, Kugler P, Schuldhaus D, Winkler J, Klucken J, et al. Subsequence dynamic time warping as a method for robust step segmentation using gyroscope signals of daily life activities. Conf Proc IEEE Eng Med Biol Soc 2013;2013:6744-6747. [doi: 10.1109/EMBC.2013.6611104] [Medline: 24111291]
20. Wang X, Kyrarini M, Ristic-Durrant D, Spranger M, Graser A. Monitoring of gait performance using dynamic time warping on IMU-sensor data. 2016 Presented at: 2016 IEEE International Symposium on Medical Measurements and Applications; 15-18 May 2016; Benevento, Italy. [doi: 10.1109/MeMeA.2016.7533745]
21. Al-Jawad A, Adame MR, Romanovas M, Hobert M, Maetzler W, Traechtler M, et al. Using multi-dimensional dynamic time warping for TUG test instrumentation with inertial sensors. 2012 Presented at: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI); 13-15 Sept 2012; Hamburg, Germany. [doi: 10.1109/MFI.2012.6343011]
22. Paiyarom S, Tungamchit P, Keinprasit R, Kayasith P. Activity monitoring system using Dynamic Time Warping for the elderly and disabled people. 2009 Presented at: 2nd International Conference on Computer, Control and Communication; 17-18 Feb 2009; Karachi, Pakistan. [doi: 10.1109/IC4.2009.4909158]
23. Klenk J, Becker C, Palumbo P, Schwickert L, Rapp K, Helbostad JL, et al. Conceptualizing a dynamic fall risk model including intrinsic risks and exposures. J Am Med Dir Assoc 2017 Nov 01;18(11):921-927. [doi: 10.1016/j.jamda.2017.08.001] [Medline: 28916290]
24. Bagalà F, Becker C, Cappello A, Chiari L, Aminian K, Hausdorff JM, et al. Evaluation of accelerometer-based fall detection algorithms on real-world falls. PLoS One 2012;7(5):e37062 [FREE Full text] [doi: 10.1371/journal.pone.0037062] [Medline: 22615890]
25. Palmerini L, Bagalà F, Zanetti A, Klenk J, Becker C, Cappello A. A wavelet-based approach to fall detection. Sensors (Basel) 2015 May 20;15(5):11575-11586 [FREE Full text] [doi: 10.3390/s150511575] [Medline: 26007719]
26. Lonini L, Gupta A, Deems-Dluhy S, Hoppe-Ludwig S, Kording K, Jayaraman A. Activity recognition in individuals walking with assistive devices: the benefits of device-specific models. JMIR Rehabil Assist Technol 2017 Aug 10;4(2):e8 [FREE Full text] [doi: 10.2196/rehab.7317] [Medline: 28798008]
27. Li K, Habre R, Deng H, Urman R, Morrison J, Gilliland FD, et al. Applying multivariate segmentation methods to human activity recognition from wearable sensors' data. JMIR Mhealth Uhealth 2019 Feb 07;7(2):e11201 [FREE Full text] [doi: 10.2196/11201] [Medline: 30730297]
28. Shawen N, Lonini L, Mummidisetty CK, Shparii I, Albert MV, Kording K, et al. Fall detection in individuals with lower limb amputations using mobile phones: machine learning enhances robustness for real-world applications. JMIR Mhealth Uhealth 2017 Oct 11;5(10):e151 [FREE Full text] [doi: 10.2196/mhealth.8201] [Medline: 29021127]
29. Schwickert L, Becker C, Lindemann U, Maréchal C, Bourke A, Chiari L, FARSEEING Consortium, FARSEEING Meta Database Consensus Group. Fall detection with body-worn sensors: a systematic review. Z Gerontol Geriatr 2013 Dec;46(8):706-719. [doi: 10.1007/s00391-013-0559-8] [Medline: 24271251]

Abbreviations

DTW: dynamic time warping
FARSEEING: Fall Repository for the Design of Smart and Self-Adaptive Environments Prolonging Independent Living
L: lumbar
SG: Samsung Galaxy.

Edited by G Eysenbach; submitted 03.05.19; peer-reviewed by H Sievänen, M Immonen; comments to author 05.09.19; revised version received 24.10.19; accepted 03.02.20; published 03.04.20.

Please cite as:
Sczuka KS, Schwickert L, Becker C, Klenk J
Re-Enactment as a Method to Reproduce Real-World Fall Events Using Inertial Sensor Data: Development and Usability Study
J Med Internet Res 2020;22(4):e13961
URL: https://www.jmir.org/2020/4/e13961
doi:10.2196/13961
PMID:32242825

Original Paper

Toward Continuous Social Phenotyping: Analyzing Gaze Patterns in an Emotion Recognition Task for Children With Autism Through Wearable Smart Glasses

Anish Nag[1]; Nick Haber[2], PhD; Catalin Voss[3], MSc; Serena Tamura[4], BA; Jena Daniels[5], BSc; Jeffrey Ma[6], BSc; Bryan Chiang[6], BSc; Shasta Ramachandran[6]; Jessey Schwartz[6], BSc; Terry Winograd[3], PhD; Carl Feinstein[6], MD; Dennis P Wall[6], PhD

[1]University of California, Berkeley, CA, United States

[2]Graduate School of Education, Stanford University, Stanford, CA, United States

[3]Department of Computer Science, Stanford University, Stanford, CA, United States

[4]UC San Francisco, San Francisco, CA, United States

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

[6]Departments of Pediatrics, Biomedical Data Science, Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States

Corresponding Author:
Dennis P Wall, PhD
Departments of Pediatrics, Biomedical Data Science, Psychiatry and Behavioral Sciences
Stanford University
1265 Welch Road
Stanford, CA, 94305
United States
Phone: 1 (650) 497 9214
Email: dpwall@stanford.edu

Abstract

Background: Several studies have shown that facial attention differs in children with autism. Measuring eye gaze and emotion recognition in children with autism is challenging, as standard clinical assessments must be delivered in clinical settings by a trained clinician. Wearable technologies may be able to bring eye gaze and emotion recognition into natural social interactions and settings.

Objective: This study aimed to test: (1) the feasibility of tracking gaze using wearable smart glasses during a facial expression recognition task and (2) the ability of these gaze-tracking data, together with facial expression recognition responses, to distinguish children with autism from neurotypical controls (NCs).

Methods: We compared the eye gaze and emotion recognition patterns of 16 children with autism spectrum disorder (ASD) and 17 children without ASD via wearable smart glasses fitted with a custom eye tracker. Children identified static facial expressions of images presented on a computer screen along with nonsocial distractors while wearing Google Glass and the eye tracker. Faces were presented in three trials, during one of which children received feedback in the form of the correct classification. We employed hybrid human-labeling and computer vision–enabled methods for pupil tracking and world–gaze translation calibration. We analyzed the impact of gaze and emotion recognition features in a prediction task aiming to distinguish children with ASD from NC participants.

Results: Gaze and emotion recognition patterns enabled the training of a classifier that distinguished ASD and NC groups. However, it was unable to significantly outperform other classifiers that used only age and gender features, suggesting that further work is necessary to disentangle these effects.

Conclusions: Although wearable smart glasses show promise in identifying subtle differences in gaze tracking and emotion recognition patterns in children with and without ASD, the present form factor and data do not allow for these differences to be reliably exploited by machine learning systems. Resolving these challenges will be an important step toward continuous tracking of the ASD phenotype.

(J Med Internet Res 2020;22(4):e13810) doi:10.2196/13810

KEYWORDS

autism spectrum disorder; translational medicine; eye tracking; wearable technologies; artificial intelligence; machine learning; precision health; digital therapy

Introduction

Background

Autism Spectrum Disorder (ASD) continues to be one of the most important public health challenges we face today, with 1 in 59 American children affected by it [1-3]. Children with autism are well known to differ from neurotypical controls (NCs) in their emotion recognition and facial processing patterns [4-9]. There are several leading theories about facial processing in ASD, and the underlying biological mechanisms are not fully understood [10-12]. However, children with autism exhibit many observable symptoms in facial attention, such as a lack of eye fixation, increased fixation on mouths [13], and requiring more time to extract emotions from faces [14]. Prior studies have found that individuals with autism have particular trouble recognizing certain emotions [15], such as happiness, neutrality [16], surprise [13,17-19], and fear [16,20]. At a more abstract level, they have been shown to struggle with making complex social judgements about trustworthiness, shame, and approachability [8]. These eye contact and facial affect recognition skills are important to improve social functioning [5], but the methods currently used to measure and track such skills are delivered in clinical settings or via trained administrators [21,22] outside of the social context where these skills are practiced.

Measuring emotion recognition and eye gaze in children with autism through mobile and wearable machine learning platforms has the potential to fill this gap for continuous phenotyping [23-28] during natural social interactions. Thus far, emotion recognition [8,16,20] and social attention [29,30] in autism have mostly been studied in isolation. Both have separately been proposed as indicators for diagnosis and quantification of autism. We hypothesized that they are deeply linked and studied them together as potential markers for phenotyping using wearable smart glasses and eye tracking.

Objectives

In this study, we compared gaze and emotion recognition pattern data from 16 children with ASD to 17 NCs participating in an in-lab computer-based emotion recognition task to determine if gaze differences exist between ASD and NC children. Participating children were tasked with identifying emotions of standardized faces on a computer screen. During the task, they wore an early prototype of a Google Glass–based emotion recognition learning aid named *Superpower Glass* [9,23,31-34], fitted with a custom-built eye tracker that followed children's gaze looking at emotional stimuli or distractors. This prototype is one of several attempts [35-37] to use Google Glass as a learning aid for children with ASD. A prior analysis focusing only on the emotion recognition data obtained from this study showed that ASD and NC participants differed only subtly in emotion recognition accuracies but that participants from the two groups showed noticeably different patterns in their emotion responses [31]. Children were eager to engage with the smart glasses, showing promise for the form factor. In analyzing gaze-tracking data from this study, we aimed to explore whether combining gaze and emotion recognition data may yield better distinguishing features for the two groups. We hypothesized that (1) the NC and ASD groups differ in gaze attention patterns and (2) this difference enables us to design an interpretable machine learning classifier distinguishing the two groups on our wearable platform.

Methods

Participants

Families were recruited from February 25, 2015, to January 26, 2016, at Stanford University via the Autism and Developmental Disabilities Research Registry, referrals to the Autism and Developmental Disabilities Clinic, the Developmental Behavioral Unit of Lucile Packard Children's Hospital, and through academic presentations. ASD participants were included if they were between the ages of 6 to 17 years and if they provided an official autism diagnosis confirmed via medical record. We assessed parent reports of each child's diagnosis via the Social Communication Questionnaire (SCQ) [38]. Participants are screened positively for ASD if they scored >16 on the SCQ. ASD participants were excluded if they had (1) evidence of a genetic, metabolic, or infectious etiology for their autism (in other words, had *syndromic autism*) based on medical record; (2) history of seizures or other neurologic disorders; (3) vision impairment; and/or (4) history of personality or bipolar disorder. NC participants were excluded if they had any of the following: (1) a score >14 on the SCQ, (2) a history of mood or personality disorder confirmed via parent report or medical record, (3) a sibling diagnosed with ASD or schizophrenia, (4) a history of seizures and/or other neurologic disorder, or (5) vision impairment.

Procedure

Eligible participants (both parents and children) provided written informed consent under an approved institutional review board (IRB) protocol. Following consent, a trained research assistant delivered the Stanford Binet Intelligence Scales, Fifth Edition, Abbreviated Battery Intelligence Quotient (ABIQ) [39] to each child participant, and parents completed the Social Responsiveness Scale (SRS)-2 [40]. Demographic and evaluation results are demonstrated in Table 1.

Participants wore Google Glass fitted with an eye tracker over the course of a 20-min computer task in which they were asked to identify the emotion (ie, happy, sad, angry, scared, disgust, surprised, and calm) portrayed by child actors on a screen. Before the task began, participants were familiarized with the list of facial emotions they would be asked to choose from. Participants were seated approximately 25 inches from the 24-inch screen (1920×1200 resolution) so that stimuli were presented at the eye level. Researchers conducted three successive trials using 125 images selected from the Child

Affective Facial Expression (CAFE) dataset [41], balanced in each trial for race, gender, and emotion expression (T1 N=41, T2 N=42, and T3 N=41). The CAFE dataset is a set of diverse faces of children aged 2 to 8 years (mean 5.3 years, SD 1.5; range 2.7-8.7) depicting seven emotional facial expressions (ie, sad, happy, surprise, anger, disgust, fear, and neutral). The full set of 154 images includes 90 female and 64 male children that represent an even balance of African American, Asian, white, Latino, and South Asian racial groups [41]. Along with each facial affect image, each frame during all three trials displayed two nonsocial images of high autism interest (ie, Legos, train, and car) that have been previously validated [42] to its right and left to create an opportunity for distraction from the center emotion expression image (see Figure 1). Each facial stimulus covered approximately 49% of the width and 87% of the height of the screen. The two distractors were each displayed at approximately 17% screen width and 31% screen height.

Facial images and corresponding distractor images were displayed for 6 seconds before the participant was prompted to choose from a list of the seven possible emotions. The list of emotions was displayed until the participant verbally responded. The glasses were deactivated during the first and third trial. In the second trial, after 3 seconds of displaying the image, the glasses played an audio cue, speaking out the correct labeled emotion for the displayed image, emulating the emotion recognition functionality of the glasses. The first 4 participants received visual feedback in the form of a word shown on the heads-up display indicating the emotion, but this was found to be distracting, with reading ability strongly affecting behavior, and was replaced by auditory cues [9,31]. Between each trial, the eye tracker was recalibrated. Between every eight images within each trial, a dot displaying at its center a dancing Santa Claus appeared in the middle of the screen for 5 seconds to draw the child's attention to the middle of the screen for validation of the eye tracking calibration. The sessions were all video recorded, and a researcher accompanied the participant through the task, recording the response option. Emotion classification responses were confirmed via the video session recording. During the second trial, participants received feedback from the glasses indicating the correct emotion classification.

Table 1. Cohort composition after excluding study failures. Medication/comorbidity surveys were not completed by 5 participants from the autism spectrum disorder cohort.

Demographic and phenotypic characteristics	Autism spectrum disorder (N=16)	Neurotypical controls (N=17)
Gender, n (%)		
Males	13 (81)	9 (53)
Females	3 (19)	8 (47)
Age (years), mean (SD; range)	12.13 (3.31; 6-17)	11.53 (2.48; 8-17)
Social Communication Questionnaire score, mean (SD; range)	18.86 (6.43; 7-31)	1.82 (1.07; 0-4)
Abbreviated Battery Intelligence Quotient standard score, mean (SD; range)	102.75 (19.54; 55-133)	108.94 (9.58; 91-129)
Social Responsiveness Scale Total score, mean (SD; range)	78.85 (11.13; 58->90)	44.41 (8.11; 36-64)
Comorbid psychological conditions, n (%)		
Anxiety disorder/depression	1 (9)[a]	0 (0)
Attention-deficit/hyperactivity disorder	1 (9)[a]	1 (5)
Current medication, n (%)		
Methylphenidate	3 (27)[a]	0 (0)
Arginine vasopressin	1 (9)[a]	0 (0)
Guanfacine extended release	1 (9)[a]	0 (0)
Sertraline	2 (18)[a]	0 (0)
Carbamazepine	1 (9)[a]	0 (0)
Aripiprazole	1 (9)[a]	0 (0)
Dexmethylphenidate	1 (9)[a]	0 (0)
Allergy medication (unspecified)	0 (0)[a]	1 (5)
Other (unspecified)	0 (0)[a]	1 (5)
No medication	4 (36)[a]	15 (88)

[a]N=11.

Figure 1. Study setup: (a) Study screen displaying facial affect stimuli and nonsocial distractors displayed for 6 seconds. (b) The screen displaying the list of emotions that the participant is asked to classify the face from. (c) A nonparticipant child wearing a Google Glass with a custom-built eye tracker fitted using a 3D-printed mount in a dry-run of the study protocol.

Gaze-Tracking Apparatus

The Google Glass worn by the participant was fitted with a custom 3D-printed mount that slid onto the unit's prism, holding a repurposed Microsoft LifeCam HD-6000 Webcam acting as an eye tracker (Figure 1). The webcam was modified by breaking and removing its infrared filter and replacing two indicator light emitting diodes with infrared emitters to produce a low-cost pupil recording device. The mount and modified webcam were adapted from the open source Pupil Project prototype [43]. A number of different mounts were 3D-printed such that at the beginning of each session, a mount could be chosen that provided the best view of the participant's eye. The webcam was connected to the computer, displaying the facial stimuli, where outward-facing video from the Glass unit was recorded and synchronized with the inward-facing eye video. The total hardware cost of the eye tracker add-on was approximately US $35.

Gaze-Tracking Data

In what follows, we have briefly outlined the procedure for obtaining gaze-tracking estimates. Details are given in Multimedia Appendix 1.

Pupil Tracking

We employed hybrid pupil identification methods, including maximally stable extremal regions [44], a gradient-based method [45] and an optical flow approach [46].

Calibration for World-Gaze Spatial Correspondence

Each participant performed four calibration processes, one at the beginning and one after each trial, yielding pupil coordinates (relative to the field of view of the eye-facing camera) and corresponding coordinates of the direction of gaze (referred to as *world-gaze coordinates* relative to the field of view of the front-facing camera). The true-value of the direction of each eye gaze event was assumed to be the direction of the stimulus provided on-screen, with no verification procedure. For each calibration, a separate polynomial regression was performed, yielding four candidate gaze prediction models. The gaze prediction models each provided, for each frame (image captured by camera), a predicted direction of eye gaze in world-gaze coordinates.

Each participant's eye tracking session was then manually inspected and partitioned into time intervals for which of the four calibration models, or none, visually appeared to apply.

Gaze Clusters

Using the world-gaze coordinate estimate, we coded each frame as 1 of 4 categories: (F) on the facial stimulus, (L) on the left distractor stimulus, (R) on the right distractor stimulus, or (N) "nowhere in particular." Calibration and gaze cluster labeling were semiautomated processes that were performed by independent labelers. Further details and interlabeler reliability data are provided in Multimedia Appendix 1.

Outlier Exclusion

In some instances, various failures (eg, camera slipping out of the mount made gaze tracking infeasible) led to unusable data. Trials with such data were discarded as outliers upon visual inspection. Detailed criteria are available in Multimedia Appendix 1.

Hypothesis 1: Gaze Pattern Analysis

To test hypothesis 1, we calculated the following distraction ratio for each participant p and each facial stimulus s in each trial t:

We plotted a histogram to visualize the differences between ASD and NC in gaze patterns throughout the task. We averaged the distraction ratios for multiple trials for each participant to arrive at a per-participant distraction ratio $d(p)$. We then performed 1-tailed t tests on the distraction ratio aggregated across groups (ASD vs NC) to test the primary hypothesis. We also performed exploratory t tests on the $d(p,s,t)$s across groups. However, this assumes what is likely an unreasonable noise model in which the $d(p,s,t)$s are independent within the same participant.

To make a more direct comparison between attention to face and attention to distractions, we further computed the given p, s, and t as mentioned earlier:

This excludes frames coded as N from the denominator. We then averaged d^{FLR} over all stimuli in trial t corresponding to true-value emotion e to obtain $d_e^{FLR}(p,t)$.

Hypothesis 2: Machine Learning Classification

To test hypothesis 2, we designed a machine learning classifier predicting binary ASD vs NC from the following features:

- Emotion confusion matrices (cm; 49-dimensional): 7×7 confusion matrices were computed for each participant across the entire trial, defined as the square matrix with rows and columns corresponding to the possible emotion responses, with entry in row *r* and column *c* to be the number of frames for which the true depicted emotion corresponds to *r*, but the participant inputted the emotion corresponding to *c* The confusion matrices were then normalized such that all rows summed to 1. Each element of the resulting normalized confusion matrix was extracted as an independent feature.
- Emotion confusion details (conf; 41- or 42-dimensional, depending on the trial): every face in the trial was assigned a binary value indicating whether the participant correctly identified the emotion, encapsulating the performance over time.
- Gaze patterns (gaze; 123- or 126-dimensional, depending on the trial): For the 6-second duration of eye tracking corresponding to a face in the trial, several features were extracted: the percentage of frames spent looking directly at the face as opposed to either of the distractors, percentage of changes of gaze fixations directed toward the face, and the number of frames elapsed before looking directly at the face. Each face contributed three of these additional features to the large pool of features.
- Participant metadata (pat; 2-dimensional): The participant's age and gender were considered meta-features.

This yielded a total of 219 features per trial for consideration. As this number far exceeds the number of participants, regularization was important to prevent model overfitting. An elastic net model was chosen as the base classifier, as this model incorporates both lasso and ridge regularization. Our primary model, then, was an elastic net classifier trained and evaluated on all available trial data using all features, concatenated by trials. For various ablations, we trained elastic net classifiers and standard logistic regression classifiers on subsets of the features and the three trials individually. We evaluated accuracy across all classifiers using leave-one-participant-out cross-validation. We performed Monte Carlo shuffling tests to assess the statistical significance of classifier predictions. For logistic regression, we identified hyperparameters, including the type of regularization (l_1 or l_2), automatically using grid search on the training sets. Note that the logistic regression models are a special case of elastic net in which one of the regularizing terms is set to 0. In practice, optimizers for elastic net can perform unstably in these edge cases, and so we included the logistic regression model as a way of better optimizing over hyperparameters.

Results

Overview

Between February 25, 2015, and January 26, 2016, we enrolled 43 (ASD=23; NC=20) participants at Stanford University under a Stanford University IRB-approved protocol. We were unable to use data from 10 participants because of the following technical errors: 4 NC participants were excluded because they had received an early version of the Superpower Glass intervention, which included a visual display of the correct word describing the emotion displayed in the second trial but were unable to read those cues or complete the computer task. After realizing this failure, the prototype was adapted such that the remaining participants received only audio cues. Five ASD participants were excluded because of an image order randomization error as they received visual stimuli in a different order than all other participants. Finally, 1 ASD participant experienced a health issue during the study that was unrelated to the study procedures and was unable to complete study procedures.

The following analysis was conducted with 16 ASD participants (mean age 12.13, SD 3.3 years) of whom 81% (13/16) were male and 17 NC participants (mean age 11.53, SD 2.5 years) of whom 53% (9/17) were male. See Table 1 for additional participant demographics including SCQ, SRS, and ABIQ scores.

Hypothesis 1: Gaze Pattern Analysis

Children with ASD showed a higher mean distraction ratio (mean 0.0433, SD 0.0911) than NCs (mean 0.0139, SD 0.0215), but this difference was not significant across the 33 participants in a 1-tailed *t* test (*P*=.12).

When per-facial-stimuli distraction ratios *d(p,s,t)* were compared (1792 ASD samples and 1756 NC samples), a significant difference between ASD and NC datapoints was observed, making an independence assumption for the data within each of the two groups (uncorrected *P*<.01). A histogram of distraction ratios is shown in Figure 2 and reveals that distraction was higher and more inconsistent in participants with autism than in NCs.

Similarly, we plotted N-frame-excluding distraction counts $d_e^{FLR}(p,t)$, averaged over all participants in the ASD and NC groups, keeping each emotion separate (Figure 3). Averaging over each emotion, we found no significant differences in the means between groups (*P*=.11 in trial 3, higher in others), owing largely to the high variance in the ASD group.

Figure 2. A histogram of the distraction ratio of autism spectrum disorder (ASD; red) and neurotypical control (NC; blue) participants on a logarithmic scale. On average, the ASD group looked at facial stimuli for less time than NCs. However, there is also considerable overlap between the groups that reduces the predictiveness of gaze features in the individual diagnosis prediction task.

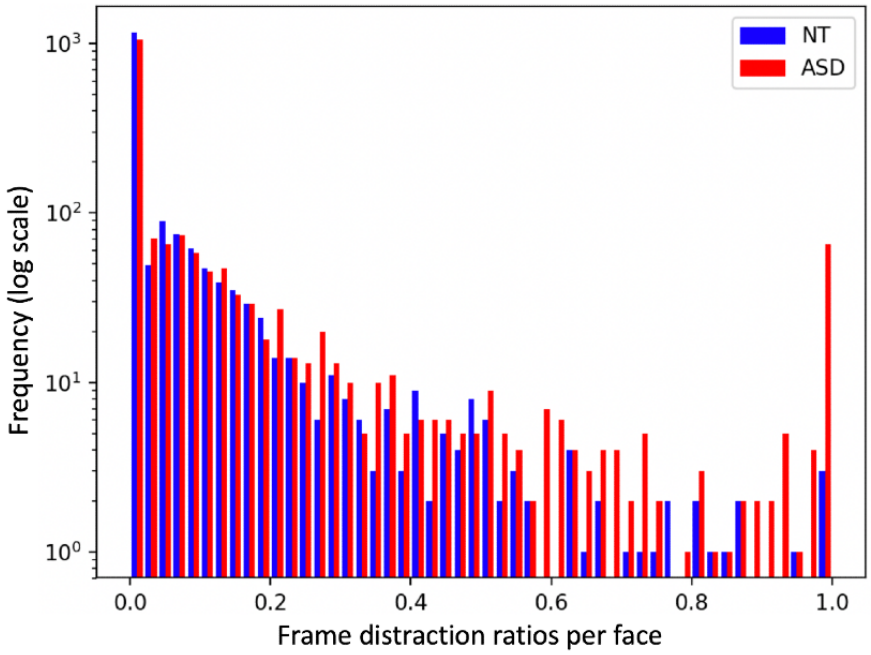
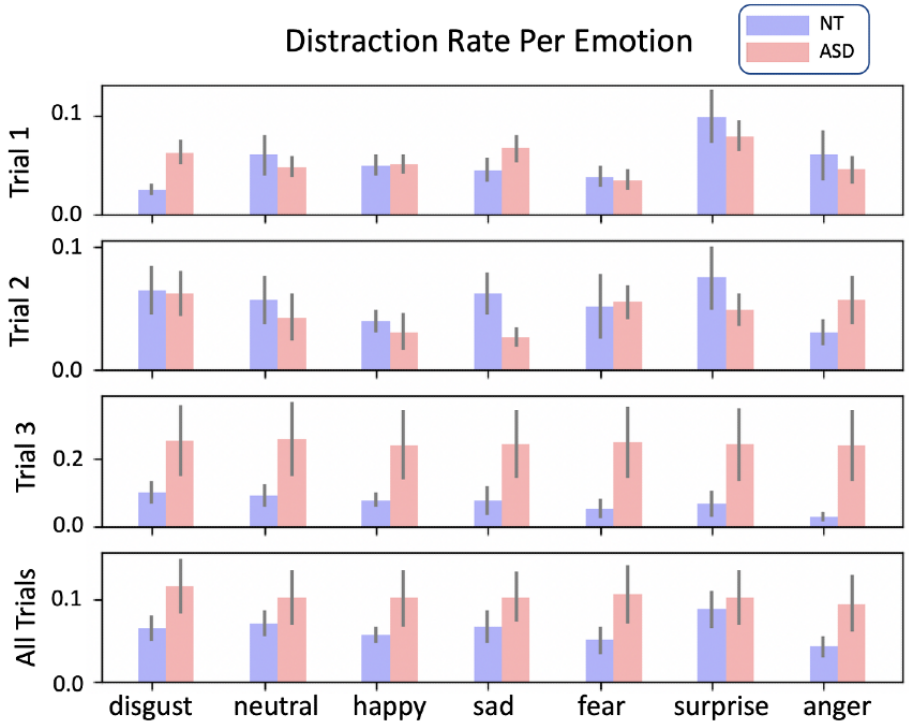


Figure 3. Histograms of (N-frame excluding) distraction ratio deFLR(p,t) of autism spectrum disorder (red) and neurotypical control (blue) participants, averaged over participants and broken down by emotion.



Hypothesis 2: Machine Learning Classification

Cross-validation confusion matrices for an elastic net classifier trained on all features are presented in Figure 4. Across all trials, the model achieved a classification accuracy of 0.71 (P=.52 using Monte Carlo shuffling; see Figure 5).

Elastic net classification accuracies and uncorrected significance tests for all feature combinations are presented in Table 2. Classifiers trained on some gaze and expression recognition features were able to outperform those trained on combinations of pat, conf, and cm features in some cases, with significant P values on a shuffling test achieving significance before (but not after) a Bonferroni correction. Performance was best on all-trial features as well as data restricted to trial 1. However, given the relatively high performance on participant metadata (age and gender), we could not conclude that gaze and expression recognition features enhance classification. Elastic net models

outperformed logistic regression ablation models in most tasks, including the primary pat-gaze-cm-conf model, suggesting a need for regularization in this dataset.

Figure 4. Cross-validation confusion matrices for the elastic net classifier trained on all features (pat, gaze, conf, and cm) for trials 1, 2, and 3, and features from all trials concatenated (accuracies in parenthesis). Autism spectrum disorder and neurotypical control participants are most distinguishable in trial 1, the first trial conducted which was before receiving any feedback or adjusting to the task. cm=Emotion confusion matrices; conf=Emotion confusion details; gaze=Gaze patterns; pat=Participant metadata.

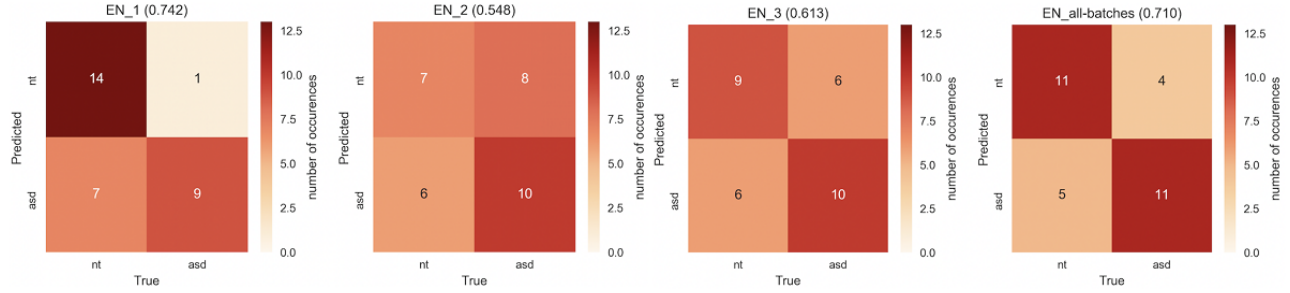


Figure 5. The shuffle test visualization for the elastic net classifier trained on all features (pat, gaze, conf, cm) concatenated for all trials yielding *P*=.05. cm=Emotion confusion matrices; conf=Emotion confusion details; gaze=Gaze patterns; pat=Participant metadata.

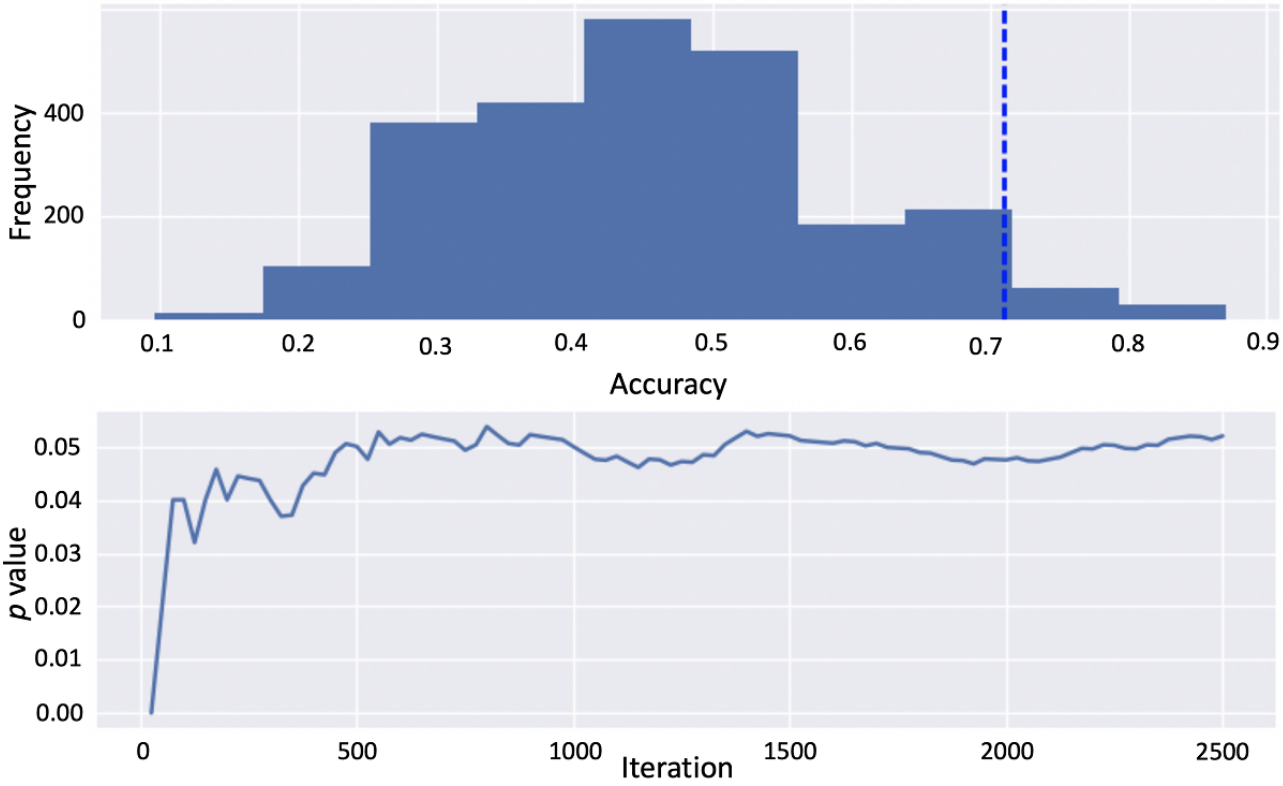


Table 2. Classification accuracies and significance tests for the elastic net classifier trained on different feature combinations and trials. All shuffle tests were performed for 2500 iterations and checked for convergence.

Data tested	All trials		Trial 1		Trial 2		Trial 3	
	Accuracy, %	*P* value	Accuracy, %	*P* value	Accuracy, %	*P* value	Accuracy, %	*P* value
pat[a] (baseline)	71.0	.04	71.0	.04	71.0	.04	71.0	.04
pat-gaze-cm-conf (full)	71.0	.05	74.2	.03	54.8	.24	61.3	.17
cm[b]	67.7	.08	67.7	.09	6.5	.998	35.5	.81
conf[c]	64.5	.16	58.1	.23	48.4	.63	41.9	.72
cm-conf	64.5	.13	61.3	.17	48.4	.56	41.9	.68
gaze[d]	83.9	.002	83.9	.00	38.7	.80	61.3	.14
gaze-conf	83.9	.004	80.6	.01	51.6	.34	61.3	.17
gaze-cm	71.0	.05	64.5	.10	61.3	.14	58.1	.21
gaze-cm-conf	71.0	.05	74.2	.04	45.2	.63	51.6	.36
pat-gaze	77.4	.02	71.0	.04	67.7	.09	74.2	.03
pat-gaze-cm	71.0	.05	64.5	.11	67.7	.08	64.5	.11
pat-gaze-conf	74.2	.04	67.7	.11	35.5	.85	64.5	.14
pat-cm	58.1	.21	54.8	.25	67.7	.09	61.3	.16
pat-conf	64.5	.14	61.3	.19	64.5	.13	51.6	.35
pat-cm-conf	61.3	.17	48.4	.45	64.5	.13	32.3	.88

[a]pat: participant metadata.

[b]cm: emotion confusion matrices.

[c]conf: emotion confusion details.

[d]gaze: gaze patterns.

Discussion

Principal Findings

In this study, we compared gaze and emotion recognition pattern data of 16 children with ASD with 17 NCs. Participants completed an in-lab computer-based emotion recognition task, where they wore an early prototype of the Superpower Glass system [9,31,32], a Google Glass–based emotion recognition learning aid, as well as a custom-built eye tracker attached to the glass that followed children's gaze looking at an emotional or distractor stimulus. In this limited data sample, we were unable to construct a machine learning classifier that reliably exploits these differences to predict ASD severity, to an accuracy significantly beyond that of the use of the age and gender data baseline. Although some models modestly outperformed a metadata-only baseline, prediction margins and significance did not hold up consistently across various ablations. Considering the large variance in gaze distraction in the ASD cohort, more data, perhaps more balanced for age and gender across cohorts, are likely required to develop a reliable model of greater use. A larger, more balanced corpus of data would also enable the use of more complex statistical models, such as artificial neural networks, which have the capacity to capture more complex structure within the data.

As an increasing amount of research work points to differences in facial attention in ASD [10-14], this study adds evidence that children with autism, when presented with facial stimuli relevant to a task involving social judgement, along with other, distracting stimuli, are more likely to attend to the distracting stimuli more. However, it also cautions that, just like with the literature on emotion recognition [16-20], this difference is subtle. Large and well-structured datasets are required to develop a diagnostic and phenotyping marker from this effect. Classical eye tracking studies have explored the subtlety of this effect in greater detail [30,47-52] and showed that it can vary depending on the emotional stimulus and setup. Some studies also observed that children with ASD pay more attention to the mouth than the eye region during an emotion recognition task [30,49-51]. The gaze data collected by our system was too noisy to test for these subtler effects. The prevailing belief in the literature was that facial attention differences are very clear in ASD, so we hypothesized that distraction ratios across an emotion recognition task alone were enough to distinguish ASD vs NC reliably. We now believe that integrating further subtlety through robust gaze-tracking hardware will be required to design a phenotyping marker from this effect.

Limitations

Our study had a number of limitations including:

- *Limited age range*: The age range of participants was limited to 6 to 17 years in this study. Unfortunately, this age range is not representative of children who are in critical periods of development for cognition and speech, and therefore further feasibility testing on younger children is necessary.

- *Imbalanced gender ratio*: In this study, we observed a gender ratio of 13 males to 3 females in our ASD cohort. Though males are substantially more likely to be diagnosed with autism than females by an average ratio of 4 males:1 female, the reported imbalance still presents a gender bias between our sample population of children with ASD to our NC children. Furthermore, there are gender differences in both neurotypical and ASD children on emotion recognition and gaze tasks [15], which may have been exacerbated by our imbalanced cohorts.
- *Limited coverage of the autism spectrum*: Only 3 of our 16 ASD participants had an ABIQ lower than 80 (between 55 and 79), suggesting that 13 of our 16 recruited ASD participants can be classified as children with high-functioning ASD. This limits our findings, as their performance may not be reflective of children from across the autism spectrum.
- *Chinrest added after study started*: Due to too much head movement by the first 4 participants, we added a chinrest to provide additional stability for the remaining subjects. This was an issue especially because more ASD subjects took part in the study later on. This confound may be exploited by the classifiers.
- *Nonstandard eye tracking:* The eye tracking system used in this study was custom-built and has not been evaluated on a standard dataset or compared with standard eye trackers. The mount and camera shifted during the study, requiring recalibration. A series of manual checks were employed to correct for these issues and ultimately manually annotate much of the pupil and world-gaze coordinates, as described in Multimedia Appendix 1. The mount occluded roughly 15% of the field of view in the lower region for most participants, which may have had a greater distracting effect on participants in the ASD group with tactile sensitivities.

Furthermore, more and better-structured data of this form may still leave our understanding incomplete for fundamental ethological reasons. There remain many uncertainties about the motivational factors, neurocognitive processes, and temporal requirements of combined facial engagement and emotion recognition. These uncertainties make it a challenge to design an experiment with “just right” parameters, even now, when technology seems increasingly up to the task. For the work described in this study, to measure both emotion recognition and attention to faces simultaneously, we used video-presented static photos of children and had the children look at these photos for 6 seconds to be sure they had the maximum opportunity to perform at their best. Although the photos in the CAFE dataset are naturalistic and representative of diverse children, this experimental condition is quite divergent from real-life social interactions with moving children and fleeting displays of facial emotion. Furthermore, there are obvious differences between looking at the face of a child on a video screen and naturalistic facial engagement between two children. Given the differences between these experimental conditions and “real life,” the field still does not have enough information to know how divergent our experimental data (or another with different parameters attempted) might be from the ecologically valid situations.

Future Outlook

To build a robust dataset toward at-home continuous phenotyping, the best course of action will be to capture the data at home as well. Key ingredients for building a robust dataset toward at-home continuous phenotyping are as follows: (1) large-scale in-home data acquisition, (2) augmented reality platforms integrating robust eye tracking, (3) game design for at-home analysis platforms, and (4) the use of clearly-defined and measurable experimental parameters [24]. We hope that in future studies, analyses on gaze and expression response in ASD can be made on data gathered in the home, in less-controlled settings, performed by families without the express need for specialist input. However, given that follow-up studies on the Superpower Glass system show heterogeneity in less-restricted data collection, this will likely be a challenge for effective human-computer interaction design: producing appropriate user interfaces and games that homogenize use in ways so as to lower the natural variability of measurement [28]. We expect data acquisition to be bolstered significantly by robust gaze-tracking hardware that can be developed at a low cost . Much of the low-cost gaze-tracking system envisioned for this study showed limitations once put on the children and had to be met with relatively expensive manual data recovery efforts. Further generations of the Superpower Glass system have not used the eye tracker in at-home studies because of its bulkiness and calibration issues [34,53]. As newer augmented reality devices entering the market begin to implement native gaze tracking, it is likely that these issues can be overcome.

Acknowledgments

The authors would like to sincerely thank their participants, without whom this research would not be possible. The work was supported in part by funds to DW from National Institutes of Health (1R01EB025025-01 and 1R21HD091500-01), The Hartwell Foundation, Bill and Melinda Gates Foundation, Coulter Foundation, Lucile Packard Foundation, and program grants from Stanford’s Human Centered Artificial Intelligence Program, Precision Health and Integrated Diagnostics Center, Beckman Center, Bio-X Center, Predictives and Diagnostics Accelerator Spectrum, Spark Program in Translational Research, and from the Wu Tsai Neurosciences Institute Neuroscience:Translate Program. They also acknowledge generous support from David Orr, Imma Calvo, Bobby Dekesyer, and Peter Sullivan.

Conflicts of Interest

DW is founder of Cognoa, a company that builds digital solutions for child health. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary Gaze Tracking Implementation Details.
[DOCX File , 363 KB - jmir_v22i4e13810_app1.docx]

References

1. Christensen DL, Bilder DA, Zahorodny W, Pettygrove S, Durkin MS, Fitzgerald RT, et al. Prevalence and characteristics of autism spectrum disorder among 4-year-old children in the autism and developmental disabilities monitoring network. J Dev Behav Pediatr 2016 Jan;37(1):1-8. [doi: 10.1097/DBP.0000000000000235] [Medline: 26651088]
2. Christensen DL, Braun KV, Baio J, Bilder D, Charles J, Constantino JN, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2012. MMWR Surveill Summ 2018 Nov 16;65(13):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6513a1] [Medline: 30439868]
3. Buescher AV, Cidav Z, Knapp M, Mandell DS. 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]
4. West MJ, Copland DA, Arnott WL, Nelson NL, Angwin AJ. Effects of prosodic and semantic cues on facial emotion recognition in relation to autism-like traits. J Autism Dev Disord 2018 Aug;48(8):2611-2618. [doi: 10.1007/s10803-018-3522-0] [Medline: 29492733]
5. Trevisan DA, Birmingham E. Are emotion recognition abilities related to everyday social functioning in ASD? A meta-analysis. Res Autism Spectr Disord 2016 Dec;32:24-42. [doi: 10.1016/j.rasd.2016.08.004]
6. Serret S, Hun S, Iakimova G, Lozada J, Anastassova M, Santos A, et al. Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new serious game: a pilot study. Mol Autism 2014;5:37 [FREE Full text] [doi: 10.1186/2040-2392-5-37] [Medline: 25018866]
7. Harms MB, Martin A, Wallace GL. Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychol Rev 2010 Sep;20(3):290-322. [doi: 10.1007/s11065-010-9138-6] [Medline: 20809200]
8. Fridenson-Hayo S, Berggren S, Lassalle A, Tal S, Pigat D, Bölte S, et al. Basic and complex emotion recognition in children with autism: cross-cultural findings. Mol Autism 2016;7:52 [FREE Full text] [doi: 10.1186/s13229-016-0113-9] [Medline: 28018573]
9. Daniels J, Schwartz JN, 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 Digit Med 2018;1:32 [FREE Full text] [doi: 10.1038/s41746-018-0035-3] [Medline: 31304314]
10. Schreibman L, Dawson G, Stahmer AC, Landa R, Rogers SJ, McGee GG, et al. Naturalistic developmental behavioral interventions: empirically validated treatments for autism spectrum disorder. J Autism Dev Disord 2015 Aug;45(8):2411-2428 [FREE Full text] [doi: 10.1007/s10803-015-2407-8] [Medline: 25737021]
11. Maglione MA, Gans D, Das L, Timbie J, Kasari C. Nonmedical interventions for children with ASD: recommended guidelines and further research needs. Pediatrics 2012 Nov 1;130(Supplement 2):S169-S178. [doi: 10.1542/peds.2012-0900o]
12. Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: the early start Denver model. Pediatrics 2010 Jan;125(1):e17-e23 [FREE Full text] [doi: 10.1542/peds.2009-0958] [Medline: 19948568]
13. Baron-Cohen S, Golan O, Ashwin E. Can emotion recognition be taught to children with autism spectrum conditions? Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3567-3574 [FREE Full text] [doi: 10.1098/rstb.2009.0191] [Medline: 19884151]
14. Sasson NJ, Turner-Brown LM, Holtzclaw TN, Lam KS, Bodfish JW. Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. Autism Res 2008 Feb;1(1):31-42 [FREE Full text] [doi: 10.1002/aur.4] [Medline: 19360648]
15. Kuusikko S, Haapsamo H, Jansson-Verkasalo E, Hurtig T, Mattila M, Ebeling H, et al. Emotion recognition in children and adolescents with autism spectrum disorders. J Autism Dev Disord 2009 Jun;39(6):938-945. [doi: 10.1007/s10803-009-0700-0] [Medline: 19205857]
16. Berggren S, Engström AC, Bölte S. Facial affect recognition in autism, ADHD and typical development. Cogn Neuropsychiatry 2016 May;21(3):213-227. [doi: 10.1080/13546805.2016.1171205] [Medline: 27099953]
17. Corden B, Chilvers R, Skuse D. Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome. Neuropsychologia 2008 Jan 15;46(1):137-147. [doi: 10.1016/j.neuropsychologia.2007.08.005] [Medline: 17920642]

18. Sasson NJ, Pinkham AE, Weittenhiller LP, Faso DJ, Simpson C. Context effects on facial affect recognition in schizophrenia and autism: behavioral and eye-tracking evidence. Schizophr Bull 2016 May;42(3):675-683 [FREE Full text] [doi: 10.1093/schbul/sbv176] [Medline: 26645375]
19. Xavier J, Vignaud V, Ruggiero R, Bodeau N, Cohen D, Chaby L. A multidimensional approach to the study of emotion recognition in autism spectrum disorders. Front Psychol 2015;6:1954 [FREE Full text] [doi: 10.3389/fpsyg.2015.01954] [Medline: 26733928]
20. Castelli F. Understanding emotions from standardized facial expressions in autism and normal development. Autism 2005 Oct;9(4):428-449. [doi: 10.1177/1362361305056082] [Medline: 16155058]
21. Calvo RA, D'Mello S. Affect detection: an interdisciplinary review of models, methods, and their applications. IEEE Trans Affective Comput 2010 Jan;1(1):18-37. [doi: 10.1109/t-affc.2010.1]
22. Brooks BL, Sherman EM, Strauss E. NEPSY-II: a developmental neuropsychological assessment, second edition. Child Neuropsychol 2009 Aug 10;16(1):80-101. [doi: 10.1080/09297040903146966]
23. Washington P, Voss C, Haber N, Tanaka S, Daniels J, Feinstein C, et al. A Wearable Social Interaction Aid for Children with Autism. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA'16; May 7 - 12, 2016; California, San Jose, USA p. 2348-2354. [doi: 10.1145/2851581.2892282]
24. Washington P, Park N, Srivastava P, Voss C, Kline A, Varma M, et al. Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging 2019 Dec 13:- [Online ahead of print]. [doi: 10.1016/j.bpsc.2019.11.015] [Medline: 32085921]
25. Tariq Q, Daniels J, Schwartz JN, 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 Nov;15(11):e1002705 [FREE Full text] [doi: 10.1371/journal.pmed.1002705] [Medline: 30481180]
26. Tariq Q, Fleming SL, Schwartz JN, Dunlap K, Corbin C, Washington P, et al. Detecting developmental delay and autism through machine learning models using home videos of Bangladeshi children: development and validation study. J Med Internet Res 2019 Apr 24;21(4):e13822 [FREE Full text] [doi: 10.2196/13822] [Medline: 31017583]
27. Kalantarian H, Jedoui K, Washington P, Tariq Q, Dunlap K, Schwartz J, et al. Labeling images with facial emotion and the potential for pediatric healthcare. Artif Intell Med 2019 Jul;98:77-86 [FREE Full text] [doi: 10.1016/j.artmed.2019.06.004] [Medline: 31521254]
28. Kalantarian H, Jedoui K, Dunlap K, Schwartz J, Washington P, Husic A, et al. The performance of emotion classifiers for children with parent-reported autism: a quantitative feasibility study. J Med Internet Res 2020:- preprint. [doi: 10.2196/13174]
29. Dawson G, Webb SJ, McPartland J. Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. Dev Neuropsychol 2005;27(3):403-424. [doi: 10.1207/s15326942dn2703_6] [Medline: 15843104]
30. Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J. Visual scanning of faces in autism. J Autism Dev Disord 2002 Aug;32(4):249-261. [doi: 10.1023/a:1016374617369] [Medline: 12199131]
31. 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 Jan;9(1):129-140 [FREE Full text] [doi: 10.1055/s-0038-1626727] [Medline: 29466819]
32. 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. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016 Presented at: UbiComp'16; September 12-16, 2016; New York, USA p. 1218-1226. [doi: 10.1145/2968219.2968310]
33. Washington P, Voss C, Kline A, Haber N, Daniels J, Fazel A, et al. SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2017 Presented at: ACM'17; September, 2017; Mountain View, CA, USA. [doi: 10.1145/3130977]
34. 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 Pediatr 2019 May 1;173(5):446-454 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.0285] [Medline: 30907929]
35. Keshav NU, Salisbury JP, Vahabzadeh A, Sahin NT. 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]
36. Sahin NT, Keshav NU, Salisbury JP, 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 4;5(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.8785] [Medline: 29301738]
37. Vahabzadeh A, Keshav NU, Salisbury JP, Sahin NT. 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]

38. Bölte S, Holtmann M, Poustka F. The social communication questionnaire (SCQ) as a screener for autism spectrum disorders: additional evidence and cross-cultural validity. J Am Acad Child Adolesc Psychiatry 2008 Jun;47(6):719-720. [doi: 10.1097/CHI.0b013e31816c42bd] [Medline: 18496329]
39. Roid GH, Barram RA. Essentials of Stanford-Binet Intelligence Scales (SB5) Assessment. New York, USA: John Wiley & Sons; 2004.
40. Constantino J. N. In: C.P. Gruber, Social responsiveness scale, second edition (SRS-2): Western Psychological Services Los Angeles, CA. USA: Western Psychological Services (WPS); 2012.
41. LoBue V, Thrasher C. The child affective facial expression (CAFE) set: validity and reliability from untrained adults. Front Psychol 2014;5:1532 [FREE Full text] [doi: 10.3389/fpsyg.2014.01532] [Medline: 25610415]
42. Dichter GS, Felder JN, Green SR, Rittenberg AM, Sasson NJ, Bodfish JW. Reward circuitry function in autism spectrum disorders. Soc Cogn Affect Neurosci 2012 Feb;7(2):160-172 [FREE Full text] [doi: 10.1093/scan/nsq095] [Medline: 21148176]
43. Pupil Labs. URL: https://pupil-labs.com [accessed 2019-01-14] [WebCite Cache ID 75PMr8BVb]
44. Donoser M, Bischof H. Efficient Maximally Stable Extremal Region (MSER) Tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006 Presented at: CVPR'06; June 17-22, 2006; New York, USA p. 553-560. [doi: 10.1109/CVPR.2006.107]
45. Timm F, Barth E. Accurate Eye Centre Localisation by Means of Gradients. In: Proceedings of the Sixth International Conference on Computer Vision Theory and Applications. 2011 Presented at: VISAPP'11; March 5-7, 2011; Vilamoura, Algarve, Portugal p. 125-130. [doi: 10.5220/0003326101250130]
46. Brox T, Bruhn A, Papenberg N, Weickert J. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Proceeding of the European Conference on Computer Vision. 2004 Presented at: ECCV'04; May 11-14, 2004; Prague, Czech Republic p. 25-36. [doi: 10.1007/978-3-540-24673-2_3]
47. Bal E, Harden E, Lamb D, van Hecke AV, Denver JW, Porges SW. Emotion recognition in children with autism spectrum disorders: relations to eye gaze and autonomic state. J Autism Dev Disord 2010 Mar;40(3):358-370. [doi: 10.1007/s10803-009-0884-3] [Medline: 19885725]
48. Dadds MR, El Masry Y, Wimalaweera S, Guastella AJ. Reduced eye gaze explains 'fear blindness' in childhood psychopathic traits. J Am Acad Child Adolesc Psychiatry 2008 Apr;47(4):455-463. [doi: 10.1097/CHI.0b013e31816407f1] [Medline: 18388767]
49. Spezio ML, Adolphs R, Hurley RS, Piven J. Abnormal use of facial information in high-functioning autism. J Autism Dev Disord 2007 May;37(5):929-939. [doi: 10.1007/s10803-006-0232-9] [Medline: 17006775]
50. van der Geest J, Kemner C, Verbaten M, van Engeland H. Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study. J Child Psychol Psychiatry 2002 Jul;43(5):669-678. [doi: 10.1111/1469-7610.00055] [Medline: 12120862]
51. Kliemann D, Dziobek I, Hatri A, Steimke R, Heekeren HR. Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. J Neurosci 2010 Sep 15;30(37):12281-12287 [FREE Full text] [doi: 10.1523/JNEUROSCI.0688-10.2010] [Medline: 20844124]
52. Frazier TW, Strauss M, Klingemier EW, Zetzer EE, Hardan AY, Eng C, et al. A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. J Am Acad Child Adolesc Psychiatry 2017 Jul;56(7):546-555 [FREE Full text] [doi: 10.1016/j.jaac.2017.05.005] [Medline: 28647006]
53. Haber N, Voss C, Wall D. IEEE Spectrum. 2020 Mar 26. Upgraded Google Glass Helps Autistic Kids 'See' Emotions URL: https://spectrum.ieee.org/biomedical/bionics/upgraded-google-glass-helps-autistic-kids-see-emotions [accessed 2020-03-31]

Abbreviations

ABIQ: Abbreviated Battery Intelligence Quotient
ASD: autism spectrum disorder
CAFE: Child Affective Facial Expression
IRB: institutional review board
NC: neurotypical control
SCQ: Social Communication Questionnaire
SRS: Social Responsiveness Scale

Edited by G Eysenbach; submitted 21.03.19; peer-reviewed by E Linstead, N Sahin, E Chong; comments to author 06.05.19; revised version received 28.06.19; accepted 09.02.20; published 22.04.20.

Please cite as:
Nag A, Haber N, Voss C, Tamura S, Daniels J, Ma J, Chiang B, Ramachandran S, Schwartz J, Winograd T, Feinstein C, Wall DP
Toward Continuous Social Phenotyping: Analyzing Gaze Patterns in an Emotion Recognition Task for Children With Autism Through Wearable Smart Glasses
J Med Internet Res 2020;22(4):e13810
URL: http://www.jmir.org/2020/4/e13810/
doi:10.2196/13810
PMID:32319961

Original Paper

Development of an Instrument to Assess Parents' Excessive Web-Based Searches for Information Pertaining to Their Children's Health: The "Children's Health Internet Research, Parental Inventory" (CHIRPI)

Antonia Barke[1], DiplPsych, DPhil, MA; Bettina K Doering[1], DiplPsych, PhD

Clinical and Biological Psychology, Catholic University Eichstaett-Ingolstadt, Eichstaett, Germany

Corresponding Author:
Antonia Barke, DiplPsych, DPhil, MA
Clinical and Biological Psychology
Catholic University Eichstaett-Ingolstadt
Ostenstrasse 25
Eichstaett, 85072
Germany
Phone: 49 842193 ext 21842
Email: abarke@gwdg.de

Abstract

Background: People often search the internet to obtain health-related information not only for themselves but also for family members and, in particular, their children. However, for a minority of parents, such searches may become excessive and distressing. Little is known about excessive web-based searching by parents for information regarding their children's health.

Objective: This study aimed to develop and validate an instrument designed to assess parents' web-based health information searching behavior, the Children's Health Internet Research, Parental Inventory (CHIRPI).

Methods: A pilot survey was used to establish the instrument (21 items). CHIRPI was validated online in a second sample (372/384, 96.9% mothers; mean age 32.7 years, SD 5.8). Item analyses, an exploratory factor analysis (EFA), and correlations with parents' perception of their children's health-related vulnerability (Child Vulnerability Scale, CVS), parental health anxiety (modified short Health Anxiety Inventory, mSHAI), and parental cyberchondria (Cyberchondria Severity Scale, CSS-15) were calculated. A subset of participants (n=73) provided retest data after 4 weeks. CHIRPI scores (total scores and subscale scores) of parents with a chronically ill child and parents who perceived their child to be vulnerable (CVS+; CVS>10) were compared with 2×2 analyses of variances (ANOVAs) with the factors Child's Health Status (chronically ill vs healthy) and perceived vulnerability (CVS+ vs CVS−).

Results: CHIRPI's internal consistency was standardized alpha=.89. The EFA identified three subscales: Symptom Focus (standardized alpha=.87), Implementing Advice (standardized alpha=.74) and Distress (standardized alpha=.89). The retest reliability of CHIRPI was measured as r_{tt}=0.78. CHIRPI correlated strongly with CSS-15 (r=0.66) and mSHAI (r=0.39). The ANOVAs comparing the CHIRPI total score and the subscale scores for parents having a chronically ill child and parents perceiving their child as vulnerable revealed the main effects for perceiving one's child as vulnerable but not for having a chronically ill child. No interactions were found. This pattern was observed for the CHIRPI total score (η^2=0.053) and each subscale (Symptom Focus η^2=0.012; Distress η^2=0.113; and Implementing Advice η^2=0.018).

Conclusions: The psychometric properties of CHIRPI are excellent. Correlations with mSHAI and CSS-15 indicate its validity. CHIRPI appears to be differentially sensitive to excessive searches owing to parents perceiving their child's health to be vulnerable rather than to higher informational needs of parents with chronically ill children. Therefore, it may help to identify parents who search excessively for web-based health information. CHIRPI (and, in particular, the Distress subscale) seems to capture a pattern of factors related to anxious health-related cognitions, emotions, and behaviors of parents, which is also applied to their children.

(J Med Internet Res 2020;22(4):e16148) doi:10.2196/16148

KEYWORDS

hypochondriasis; internet; health behavior; health knowledge, attitudes, practice; questionnaire; children; parents

Introduction

Background

The internet is a ubiquitous source of information. In 2017, 3.6 billion people worldwide have used the internet [1]. More than half of them (55%-85%) used the internet to search for health-related information [2-5]. Seeking health information on the internet has many benefits, such as convenience, greater perceived anonymity regarding sensitive issues [6,7], and health service user empowerment [2,8-10]. However, a variety of downsides have also been described. The quality of many health-related websites has come under scrutiny [9,11-13]. The dangers of using search engines to find health information, eg, escalating searches that lead the user from researching harmless symptoms to self-diagnosing with life-threatening diseases, have been pointed out [12-16]. The users' ability to appraise the material and filter irrelevant and unreliable information has also been questioned [11,12,16-19]. Some negative consequences of the relationship between health providers and patients have been reported [20,21]. Examples for such consequences comprise patients requesting inappropriate treatments on the basis of inaccurate, or inaccurately understood, Web-based information and frustration on the health professional's side (due to challenges to their expert opinion and time used to explain) and the patient's side (due to perceived unresponsiveness of the health professional to the Web-based researches).

People do not just search for information regarding their own health, but they also search for information relevant to the health of others in their family, particularly their children [3,4]. The reported prevalence of such searches varies from 30% to 35% [10,22], over 42% [16], 50% [3,4], and 56% [23] to 75% [24]. Searching for information related to one's children's health can be an expression of increased informational needs in the face of the new responsibilities of parenting [25].

Data about parents' web searches for information related to their children's health mirror data about Web searches for health information in general. Studies have reported that the majority of people (>60%) searching online for health information used search engines to find the websites [10,18] and that 75% of parents looking for information about their child's condition used Google [26]. Concerns about the quality of Web-based health information in general are mirrored by concerns about websites regarding children's health issues [14,27-30] including vaccination [31,32], other procedures [33], specific conditions such as prematurity [34], illnesses such as eczema [35], and more general health advice, such as recommendations for children's toothbrushing [36].

The reported benefits of searching online for information related to one's children's health include being better informed about health and health care provision [37], the availability of information outside doctors' hours [25], decreased use of services for minor, nonurgent issues [38], and improved decision making in urgent cases [38]. Sometimes parents are even able to arrive at the correct diagnosis when the attending pediatrician has failed to do so [39]. The vast majority (89%-90%) of parents who have used the internet to find information related to their children's health found it useful [26,40,41] and many (67%) said it had influenced the medical decisions made on behalf of their child [40]. Interestingly, only a minority of parents chose to discuss the results of their Web search with their child's doctor [26].

However, despite the clear benefits of searching the internet for health information, 18% [26] to 30% [42] of parents reported finding the websites confusing and distressing or felt more anxious after their search. In other words, in a small number of families, parents' search for Web-based health information can have detrimental effects. Apart from reports of cases where attempts to implement advice obtained online have gone wrong [27], parents' searches can negatively affect communication with health professionals [22,42-45] and lead to distress and anxiety for the parents [22,26,42]. In the case of searches for information regarding one's own health, a similar pattern of behavior, termed "cyberchondria," has been described [46-51]. Cyberchondria is characterized by Web searches for information about perceived symptoms and possible diagnoses. The searches are undertaken with the aim of assuaging worries and relieving anxiety, yet they increase them and potentially lead to further searches in a quest for reassurance. The increased worry may be a consequence of the search itself, the material encountered online, or the accompanying attentional and cognitive processes. Cyberchondria is a multidimensional construct with five main facets: compulsion, ie, Web searches interrupt activities of daily living; distress, ie, search behavior has negative emotional consequences; excessiveness, ie, repeated searching and, sometimes, escalating search behavior; need for reassurance, ie, desire to consult a health care professional after the Web search; mistrust of medical professionals, ie, putting more faith in the information one has researched by oneself than in trusting expert opinion. Cyberchondria is related to health anxiety [52,53]. The vicious circle that occurs when searching for health information for oneself [46] may also apply when searching for information related to one's children's health. Some authors call this pattern of behavior "cyberchondria by proxy" [22]. We do not recommend usage of this term as it suggests a resemblance to the Munchausen syndrome by proxy and thus appears unnecessarily stigmatizing, especially given that hardly anything is known about the consequences of such parental searches. As a neutral phrase, we suggest using "excessive search for Web-based information related to one's children's health."

Objectives

The aim of the study was to improve knowledge regarding parents' excessive Web-based searching for information related to one's children's health by developing an instrument to capture such search behavior. We also aimed at investigating whether such behavior is associated with personal health anxiety and with searching for Web-based health information related to one's own health, the perceived health status of one's child, and

other relevant health information, such as a chronic illness of the child or of the parents.

Methods

Ethics

The study was conducted according to the Declaration of Helsinki and approved by the Internal Review Board of Göttingen University (number 2017-162). Data were collected anonymously, and participants were not able to access the survey until they had viewed information about the study and provided consent.

Questionnaire Development

The first step was the generation of an item pool through brainstorming by the researchers and advanced students familiar with the construct of cyberchondria. We reviewed the existing research literature as referenced in the Introduction section. As no instruments exist to assess excessive searching for Web-based information related to one's children's health, we especially considered research in the field of adult cyberchondria [47-52] and previously established measurement instruments for cyberchondria such as the Cyberchondria Severity Scale (CSS) [48] as the basis for item development. Items of the CSS were assessed with regard to their suitability for parents' Web searches, and select item content was adapted for the item pool. In addition, informal interviews were conducted with parents of children aged 0 to 10 years, clinical psychologists specialized in treating children and adolescents, nursery school teachers, and pediatricians. This resulted in a pool of 84 items that were screened to remove items regarded as too linguistically complicated (eg, double negations, uncommon words) or redundant, leaving a final pool of 63 items. The 63 items were presented, in a randomized order, to an online sample of parents of children between the ages of 0 and 10 years. The parents were recruited online through parenting forums and advertisements on notice boards. A total of 486 parents, almost all being mothers (471/486, 96.9%), completed the questionnaire. Their mean age was 31.9 years (SD 5.3). After an analysis of the distribution of missing answers, item difficulty, and participant comments, exploratory factor analyses (EFAs) and reliability analyses were conducted. Items were selected in an iterative process with items excluded if they showed unsatisfactory item characteristics.

The final instrument, *Children's Health Internet Research Parental Inventory* (CHIRPI), consisted of 21 items organized into three 7-item subscales. The Distress subscale captures the aversive emotional consequences of parents' Web searches for information related to their children's health, such as anxiety or irritation, captures physiological arousal, such as difficulty falling asleep or finding it hard to relax, and captures the interference of these consequences with everyday activities. The Symptom Focus subscale captures parents' search behavior that is prompted by a perceived physical change in their children and is aimed at finding the causes for the symptoms or a cure for a presumed disease. The Implementing Advice subscale captures parents' tendency to implement Web-based advice and encompasses various behaviors, such as asking the child's doctor to prescribe specific medication or carry out particular diagnostic procedures.

Procedure Questionnaire Validation

Procedure

For the validation study, a Web-based questionnaire was generated using Lime Survey (Limesurvey GmbH, Hamburg, Germany) [54]. For a complete description according to the Checklist for the Reporting Results of Internet E-Surveys (CHERRIES [55]), see Multimedia Appendix 1. After viewing information about the study and providing consent, participants were linked to a questionnaire, which asked them to provide demographic information about themselves and their children. It included items on thoughts, emotions, and behaviors regarding their children's health. Then participants completed CHIRPI. This section was followed by questionnaires assessing the perceived health-related vulnerability of their children and parents' health anxiety and cyberchondria (see the following sections). At the end, participants were invited to take part in a brief retest in about three weeks' time. If they chose to do so, they were asked to generate a code so that their retest results could be linked with their original results while preserving their anonymity.

Material

Demographic Information About Parents and Children

The participants answered questions about their own age, sex, education, mother tongue, current employment, and whether they were a single parent. They also reported how many children they had and the age and sex of each child.

Health-Related Information

The participants were asked whether they worked in a health-related profession (yes/no), whether they, the other parent, or any of their children suffered from a chronic illness (yes/no). They also used a visual analogue scale (VAS;implemented as a horizontal slider) to respond to the following questions: (1) How anxious/worried are you about your child's health? (0=not at all anxious/worried to 100=very anxious/worried); (2) How do you rate your own medical knowledge/competence? (0=not at all competent to 100=very competent); (3) How would you rate your child's health? (0=very poor to 100=very good); (4) How has your experience with your child's doctors been? (0=very poor to 100=very good).

In addition, the participants reported how much time (in minutes) they spent searching the Web for information related to their children's health in a typical week in which their child (1) displays symptoms of some kind and (2) does not display any symptoms.

Parents' Searches for Web-Based Information Related to Their Children's Health

The validation version of CHIRPI consists of 21 items describing the cognitive, emotional, and behavioral aspects of parents' Web searches for information about their children's health (see the previous sections). Participants used a Likert scale (1=never to 5=always) to indicate how often they exhibited the behavior or experienced the emotional or cognitive state

described. See Multimedia Appendix 2 for the questionnaire in German and English.

Perceived Health-Related Vulnerability of the Child or Children

The parents' perception of the child's health-related vulnerability was measured with the *Child Vulnerability Scale* (CVS) [56]. As no German version existed, we translated the CVS according to established guidelines for such translations [57]. The scale consists of eight items describing parental perceptions of the child's health (eg, "In general my child seems less healthy than other children" and "When something is going around, my child usually catches it."). Parents responded using a 4-point scale (0=strongly disagree to 3=strongly agree). Scores can range from 0 to 24, with higher scores representing greater perceived vulnerability. Scores of 10 or more indicate that the parent perceives his or her child's health to be vulnerable [56]. The CVS had an internal consistency of Cronbach alpha=.74 in the original study [56] and alpha=.80 in our sample.

Anxiety Regarding the Parents' Own Health

Parents' anxiety about their own health was assessed with the modified version of the *Short Health Anxiety Inventory* (mSHAI) [58]. The mSHAI uses simpler answer formats than the original [59]; it consists of 14 items on health anxiety (eg, "I spend much of my time worrying about my health"), to which responses are given using a 5-point scale (1=strongly disagree to 5=strongly agree). Higher scores indicate greater health anxiety [60]. In our sample, the mSHAI had an internal consistency of standardized alpha=.94.

Cyberchondria With Regard to One's Own Health

Parents' tendency to conduct Web searches for information relating to their own health and the negative consequences of such searches were measured with the short version of the *Cyberchondria Severity Scale-15* (CSS-15) [46,48]. The CSS-15 consists of 15 items describing possible cognitions, emotions, or behaviors related to Web searches for health-related information (eg, "I feel more anxious or distressed after researching symptoms or perceived medical conditions online.") to which responses are given using a 5-point scale (1=never to 5=always). The short version showed good validity and good internal consistency in the original study (standardized alpha=.82) [46] and in our sample (standardized alpha=.84).

Retest

Participants who agreed to take part in a retest received an email invitation with a link to a short version of the Web-based survey 3 weeks after they had first completed it. They repeated basic demographic information (described in the Demographic Information About Parents and Children section), health-related information (described in the Health-Related Information section), and CHIRPI.

Data Analysis

We conducted a standard item analysis for CHIRPI, first inspecting the number of missing data points for each item and calculating item means, standard deviations, and difficulties. Interitem correlations and item-total correlations were also computed.

We investigated the factorial structure of CHIRPI using an EFA. The suitability of the data for a factor analysis was ascertained using the Kaiser-Meyer-Olkin (KMO) criterion and Bartlett's test. The number of factors to be extracted was determined using a Velicer minimum average partial (MAP) test [61]. Factors were extracted using a maximum likelihood estimation, and the extracted factors were rotated obliquely (Promax with a Kaiser normalization).

Reliability analyses were carried out by computing the internal consistency (standardized alpha) of the whole scale and the subscales and examining whether the scale and subscales would have benefitted from the removal of individual items. The retest reliabilities of the whole scale and the subscales were calculated as Pearson correlations between the first and second measurements. Parents who completed the retest too late (more than 5 weeks after first completing CHIRPI) were excluded from the retest. We also used independent sample *t* tests (2-tailed) to determine whether the subsample that had agreed to take part in the retest differed from the subsample that chose not to participate with regard to age, number of children, time spent searching for Web-based information related to their children's health, and scores in CHIRPI, CSS-15, mSHAI, and CVS.

The correlations were calculated between the CHIRPI scale and its subscales and the demographic information, global (VAS) ratings regarding worry about the child's health, perceived parental medical competence, perceived general health of the child, and previous experience with the child's doctors. In addition, we calculated the correlations between CHIRPI scores and scores in CVS, mSHAI, and CSS-15.

We used a 2×2 analysis of variance (ANOVA) with the factors child's health status (healthy/chronically ill) and the perceived health vulnerability of the child (CVS−: CVS below cutoff/CVS+: CVS above cutoff), and the CHIRPI total score as a dependent variable to investigate the influence of the child's objective health status and the parent's perceptions of the child's vulnerability on the CHIRPI score. These analysis was repeated for the subscales. As a measure of effect size, we reported eta-squared [62]. The data were analyzed using Statistica (TIBCO Software Inc, 2018, version 13, Palo Alto, CA), except for the EFA, which was conducted using SPSS (IBM, version 25, Armonk, New York).

Participants

The Web-based survey was widely publicized on German webpages and social media related to parenthood and children. Participation was open to all interested parents, and the resulting sample constituted a convenience sample. Participants were included in the analysis if they reported German as their native language and had at least one child under 10 years of age. A total of 515 participants gave informed consent, and 491 of the participants provided the basic data required to assess eligibility. The inclusion criteria were fulfilled by 480 respondents (nine reported another native language and in two cases the youngest child was older than 10 years). A total of 92 of the 480 respondents (92/480, 19.2%) stopped responding before they had completed CHIRPI. Furthermore, two participants provided implausible data and had to be excluded.

Data from a final set of 384 participants (372/384, 96.9% mothers and 12/384, 3.1% fathers) were analyzed. The mean age of the sample was 32.7 years (SD 5.8) and they had an average of 1.8 children (186 participants had 1 child, 137 participants had 2 children, 40 participants had 3 children, 13 participants had 4 children and 8 participants had five or more children). The mean age of the participants' youngest child was 3.4 years (SD 3.0). Approximately one in 10 (37/384, 9.6%) of the participating parents reported being single parents and a quarter (94/384, 25.4%) worked in a health-related profession. In total, 11.7% (45/384) reported that at least one of their children had a chronic illness, 22.9% (88/384) said they personally had a chronic illness, and 15.9% (61/384) said their partner suffered from a chronic illness; 37.8% (145/384) reported that at least one member of their immediate family had some form of a chronic illness.

Results

Item Analysis

There were very few missing responses in CHIRPI. The frequency of missing responses ranged from 0 (most items) to 3 (item 20; see Table 1). Item 16 was the most difficult, ie, least endorsed (p_i=0.07) and item 2, the easiest, ie, most endorsed (p_i=0.65). The mean item difficulty was p_i=0.29, and the interitem correlation for the whole scale was r=0.29. Item-total correlations ranged from r_{itc}=0.37 (item 20) to r_{itc}=0.65 (item 10), with a mean of r_{itc}=0.50.

Table 1. Item means and standard deviations, missing responses per item, item difficulties, item-total correlations, and the internal consistency (standardized alpha) of the scale with single item exclusions (n=384).

Item	Value, mean (SD)	Missing responses, n	Difficulty (p_i)	Item-total correlation (r_{itc})	Standardized alpha (if item is excluded)
1	3.2 (1.01)	0	0.56	0.52	.89
2	3.6 (1.08)	0	0.65	0.51	.89
3	1.8 (0.89)	0	0.20	0.38	.89
4	2.3 (0.92)	1	0.33	0.56	.89
5	1.5 (0.80)	0	0.11	0.38	.89
6	2.3 (0.96)	0	0.32	0.56	.89
7	1.7 (0.94)	0	0.17	0.55	.89
8	2.4 (0.85)	0	0.34	0.52	.89
9	2.0 (0.95)	0	0.25	0.58	.89
10	3.0 (1.03)	2	0.51	0.65	.89
11	1.7 (0.90)	0	0.17	0.40	.89
12	1.6 (0.84)	0	0.14	0.50	.89
13	1.7 (0.85)	1	0.16	0.43	.89
14	3.1 (1.11)	1	0.52	0.55	.89
15	1.6 (0.85)	1	0.15	0.53	.89
16	1.3 (0.60)	1	0.07	0.43	.89
17	2.8 (1.17)	0	0.45	0.56	.89
18	1.4 (0.71)	0	0.09	0.48	.89
19	3.3 (1.17)	0	0.57	0.58	.89
20	1.8 (0.89)	3	0.21	0.37	.89
21	1.7 (0.91)	0	0.19	0.58	.89

Factor Structure

The KMO criterion value of 0.88 and the Bartlett test chi-square value of 3421.3 (P<.001) indicated that the data were suitable for an EFA. The Velicer MAP test suggested the extraction of three factors. The maximum likelihood factor estimation converged after six iterations and three factors were extracted, which together explained 45.16% of the variance in the CHIRPI scores. The factor loadings ranged from 0.42 to 0.86, and no item showed double loadings of >0.30 with another factor (Table 2). Factor 1 aggregated items that were indicative of distress as a consequence of searching and was called "Distress." Factor 2 was related to a symptom-focused search strategy and was named "Symptom Focus." Factor 3 encompassed various behaviors aimed at implementing advice found online and was called "Implementing Advice." As the factors showed good internal consistency, we treated them as subscales. The subscales were moderately correlated, Distress×Symptom Focus: r=0.36 (P<.001); Distress×Implementing Advice: r=0.44 (P<.001); Symptom Focus×Implementing Advice: r=0.50 (P<.001).

Table 2. Results of the exploratory factor analysis: factor loadings, eigenvalues, and percentage of variance explained.

Item[a]	Factor 1: Distress[b]	Factor 2: Symptom focus[c]	Factor 3: Implementing advice[d]
15. I have difficulty falling asleep after searching online for information about illnesses or symptoms concerning my child.	0.87	−0.08	−0.02
21. I find it difficult to stop worrying about my child's health after searching online for information about illnesses or symptoms concerning my child.	0.86	−0.02	−0.02
9. After I have searched online for information about illnesses or symptoms that concern my child, I feel more anxious and distressed than before.	0.80	0.06	−0.07
7. I panic when I read online that one of my child's symptoms occurs in a rare or serious disease.	0.67	0.16	−0.10
4. I find it hard to relax after searching online for health information concerning my child.	0.67	−0.10	0.15
12. I become angry and irritated more easily after reading about my child's symptoms or illnesses online.	0.62	0.13	0.00
18. Searching online for health information relevant to my child interferes with my everyday activities, such as chores, hobbies or spending time with family and friends.	0.50	−0.09	0.27
10. When my child displays any symptom, I search on the Internet for information about it.	−0.03	0.80	0.05
17. When searching online for health information relevant to my child, I am mainly interested in what illness the observed symptoms point to.	0.17	0.76	−0.19
1. When I notice any change in my child's body I search for information about it on the Internet.	−0.06	0.73	−0.03
2. When I search online for health information relevant to my child, I am particularly interested in whether the problems or symptoms need treatment.	−0.07	0.68	0.02
6. Based on my online searches for health information relevant to my child's symptoms, I suspect he or she has a particular disease.	0.06	0.66	−0.03
19. When searching online for health information relevant to my child I am hoping to find practical advice.	−0.03	0.59	0.20
14. When searching online for health information relevant to my child, I am particularly interested in possible causes for the illness.	0.06	0.45	0.21
5. After I have searched online for health information I ask teachers or nursery staff for help in implementing the recommendations.	−0.01	−0.14	0.68
3. After I have searched online for health information I have prescribed certain activities to my child (e.g. exercise for muscle strengthening)	−0.08	0.08	0.52
16. I ask my child's doctor to prescribe particular drugs that I have read about online.	0.12	−0.04	0.50
8. When I find advice or recommendations during my online searches for health information (e.g. a particular diet) I apply them to my child.	0.11	−0.06	0.49
11. After searching online for health information I ask teachers or nursery staff for their help in observing my child's symptoms.	−0.04	0.25	0.48

Item[a]	Factor 1: Distress[b]	Factor 2: Symptom focus[c]	Factor 3: Implementing advice[d]
20. If I see freely available medication recommended during my online search for health information, I will give that medication to my child.	−0.07	0.16	0.43
13. I suggest specific diagnostic investigations that I have read about online to my child's doctor.	0.08	0.09	0.42

[a]Factor extraction method: maximum likelihood estimation; rotation: Promax with a Kaiser normalization.

[b]Eigenvalue=6.24; explained variance=29.71%; cumulative explained variance=29.71%.

[c]Eigenvalue=2.14; explained variance=10.21%; cumulative explained variance=39.92%.

[d]Eigenvalue=1.10; explained variance=5.24%; cumulative explained variance=45.16%.

Reliability

Internal Consistency

The whole scale had an internal consistency of standardized alpha=.89 and would not have been improved by the exclusion of any item. The internal consistencies of the subscales were as follows: Symptom Focus alpha=.87, Implementing Advice alpha=.74, and Distress alpha=.89 (Table 3). None of the subscales would have benefitted from excluding an item.

Table 3. Means, standard deviations, interitem correlations, mean item-total correlations, standardized alpha, and test-retest reliability of the Children's Health Internet Research Parental Inventory (CHIRPI) and its subscales.

Scale	Value, mean (SD)	Mean interitem correlation	Mean item-total correlation, r_{itc}	Standardized alpha	Test-retest reliability, r_{tt}
Distress	12.3 (4.77)	0.54	0.68	.89	0.84
Symptom focus	21.4 (5.59)	0.48	0.64	.87	0.70
Implementing advice	12.0 (3.57)	0.29	0.45	.74	0.61
CHIRPI (total scale)	45.7 (11.00)	0.29	0.50	.89	0.78

Retest Reliability

We calculated the 4-week-retest reliability of the subscales and the whole scale. A total of 96 participants agreed to take part in the retest and provided pseudonyms that could be matched with those provided in the original survey; 23 of them were excluded because they were late to fill in the retest. In the end, 73 datasets were included in the test-retest analysis, with a mean test-retest interval of 27.9 days (SD 3.1; range: 21-38). The retest reliabilities were as follows: r_{tt}=0.78; P<.001 (whole scale); r_{tt}=0.67; P<.001 (Symptom Focus), r_{tt}=0.59; P<.001 (Implementing Advice), and r_{tt}=0.84; P<.001 (Distress). The retest subsample did not differ from participants not included in the retest with regard to age, sex, or any questionnaire score but, on average, they had fewer children (for the full results, see Multimedia Appendix 3).

Indicators of Validity

Correlations With Health-Related Measures and Other Variables

The CHIRPI score was uncorrelated with parents' age and the level of education. Weak negative correlations indicated that parents with older children or more children tended to score lower on CHIRPI. There were moderate positive correlations between the CHIRPI score and time spent searching for Web-based health information in weeks when the child displayed symptoms (r=0.26) and a small correlation in symptom-free weeks (r=0.14). There was also a moderate correlation (r=0.38) between the CHIRPI score and the VAS rating of worry and anxiety about one's child's health. This correlation was clearly related to the greater distress that these parents felt regarding their search. In addition, there was a small (r=−0.20) negative association showing that the poorer the participants rated their experience with the child's doctors, the greater were their scores on CHIRPI. By far the strongest correlation was observed between the CHIRPI score and parental cyberchondria (r=0.66) followed by personal health anxiety of the parents (r=0.39) and parents' perception of the vulnerability of their child's health (r=0.29; see Table 4).

Turning to the subscales, the highest correlations involved Distress, which was strongly correlated with the CSS-15 (r=0.61), the mSHAI (r=0.50), the VAS rating of worry and anxiety regarding the child's health (r=0.50), and the perceived vulnerability regarding one's child (r=0.54). The other subscales were also most highly correlated with the CSS-15, r=0.51 (Symptom Focus) and r=0.46 (Implementing Advice). Details of all correlations are given in Table 4.

Table 4. Correlations between the full scale or subscales and demographic variables, time spent searching the internet for information related to one's children's health, health-related questionnaires, and the visual analog scales.

Variables[a]	Value, mean (SD)	CHIRPI[b] score		Distress		Symptom focus		Implementing advice	
		r	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
Age	32.7 (5.8)	−0.01	.88	−0.04	.47	0.02	.69	−0.03	.61
Years of education	16.2 (4.1)	0.00	.93	−0.07	.17	0.10	.046	−0.12	.02
Number of children	1.8 (1.1)	−0.12	.02	−0.09	.07	−0.15	.002	0.00	.96
Mean age of children (years)	3.4 (3.0)	−0.17[a]	<.001	−0.06	.24	−0.27[a]	<.001	0.03	.56
Search time, week without symptoms (mins)	7.9 (24.0)	0.14	.007	0.04	.49	0.11	.03	0.19[a]	<.001
Search time, week with symptoms (mins)	39.8 (52.5)	0.26[a]	<.001	0.11	.04	0.27[a]	<.001	0.20[a]	<.001
CVS[c]	5.9 (4.3)	0.29[a]	<.001	0.54[a]	<.001	0.15[a]	.005	0.27[a]	<.001
mSHAI[d]	27.3 (11.1)	0.39[a]	<.001	0.50[a]	<.001	0.21[a]	<.001	0.22[a]	<.001
CSS-15[e]	29.4 (8.0)	0.66[a]	<.001	0.61[a]	<.001	0.51[a]	<.001	0.46[a]	<.001
VAS[f] worry/anxiety regarding child's health	49.3 (26.4)	0.38[a]	<.001	0.50[a]	<.001	0.22[a]	<.001	0.21[a]	<.001
VAS child's health	86.5 (15.4)	−0.09	.08	−0.11	.03	0.00	.97	−0.09	.095
VAS participant's medical competence	61.9 (22.3)	0.03	.64	−0.08	.18	0.02	.68	0.14	.015
VAS experience with child's doctors	75.6 (23.6)	−0.20[c]	<.001	−0.09	.11	−0.19	<.001	−0.19[a]	<.001

[a]Significant after Bonferroni correction.

[b]CHIRPI: Children's Health Internet Research Parental Inventory.

[c]CVS: Child Vulnerability Scale.

[d]mSHAI: Modified Health Anxiety Inventory.

[e]CSS-15: Cyberchondria Severity Scale (15-item version).

[f]VAS: visual analog scale (0-100; 0 signifying no worries about the child's health, child's health is poor, parent has no medical competence, parent's experience with child's doctors has been poor).

Particular Groups of Parents

We identified groups of parents that may have different attitudes toward Web-based searches for health information regarding their children. The first group comprised parents with a chronically ill child (as opposed to parents of healthy children). The second group consisted of parents who perceived their child as being vulnerable with regard to health (CVS+ as opposed to CVS−). Within both groups, we compared the parents with their counterparts regarding the search behavior, health anxiety, and CHIRPI scores.

Parents With Chronically Ill Children

In our sample, 11.7% (45/384) of parents reported that they had a child with a chronic illness. They spent about twice as much time searching for Web-based health information in weeks with symptoms (mean 62.4 min, SD 95.6) compared with parents with healthy children (mean 36.9 min, SD 43.5) and worried more about their child's health (mean 60.9 min, SD 22.8) than their counterparts (mean 48.1 min, SD 26.5). Despite this, the only CHIRPI subscale on which they scored higher than parents of healthy children was Implementing Advice (mean 13.5, SD 4.4 vs mean 11.9, SD 3.6; see Table 5).

Table 5. Comparisons of parents with chronically ill children and parents with healthy children by means of independent samples t tests.

Variable	Chronically ill child (n=45), mean (SD)	Healthy child (n=337), mean (SD)	*t* value (df)	*P* value	Cohen *d*
CHIRPI[a]	48.9 (13.7)	45.2 (11.0)	2.04 (380)	.04	0.30
Distress	13.6 (5.8)	12.2 (4.8)	1.74 (380)	.08	0.26
Symptom focus	21.9 (5.6)	21.3 (5.7)	0.67 (380)	.50	0.11
Implementing advice	13.5 (4.4)	11.9 (3.6)	2.62 (380)	.009	0.38
CVS[b]	7.7 (4.5)	5.7 (4.2)	2.93 (351)	.004	0.47
mSHAI[c]	27.7 (9.5)	27.2 (11.4)	0.25 (345)	.81	0.04
CSS-15[d]	30.2 (8.0)	29.3 (8.0)	0.67 (335)	.50	0.11
Search time, week without symptoms (minutes)	14.7 (30.5)	7.1 (23.0)	1.93 (373)	.05	0.28
Search time, week with symptoms (minutes)	64.2 (95.6)	36.9 (43.5)	3.24 (378)	.001	0.37
VAS[e] worry/anxiety regarding child's health	60.9 (22.8)	48.1 (26.5)	2.78 (335)	.006	0.52
VAS child's health	73.8 (20.2)	88.2 (13.9)	−5.83 (361)	<.001	−0.83
VAS participant's medical competence	72.7 (19.1)	60.6 (22.3)	3.08 (307)	.002	0.59
VAS experience with child's doctors	63.8 (33.4)	77.1 (21.7)	−3.23 (321)	.001	−0.47

[a]CHIRPI: Children's Health Internet Research Parental Inventory.

[b]CVS: Child Vulnerability Scale.

[c]mSHAI: Modified Health Anxiety Inventory.

[d]CSS-15: Cyberchondria Severity Scale (15-item version).

[e]VAS: visual analog scale (0-100 with 0 designating no worry regarding child's health, poor child's health, no medical competence, poor experiences with child's doctors).

Perceived Health-Related Vulnerability of the Child

We compared parents who perceived their child's health to be vulnerable with parents who perceived their child's health to be less vulnerable. CVS+ parents reported markedly higher scores in CHIRPI (mean 52.0, SD 13.6) than their CVS− counterparts (mean 44.0, SD 10.1) and scored particularly high on the Distress subscale (CVS+: mean 16.4, SD 6.3; CVS−: mean 11.3, SD 4.0; Table 6). They also reported higher personal health anxiety (CVS+: mean 35.7, SD 11.9; CVS−: mean 25.3, SD 10.0), higher cyberchondria related to their own health (CVS+: mean 34.1, SD 9.6; CVS−: mean 28.3, SD 7.1), and greater worry and anxiety about their child's health (CVS+: mean 65.9, SD 22.0; CVS−: mean 44.9, SD 25.7) but did not report spending more time on Web searches. See Table 6 for the full results.

Table 6. Comparison of parents above (CVS+) and below (CVS−) the cutoff for the Child Vulnerability Scale using independent samples t tests.

Variable	CVS−[a] (n=283), mean (SD)	CVS+ (n=72), mean (SD)	*t* test (df)	*P* value	Cohen *d*
CHIRPI[b]	52.0 (13.6)	44.0 (10.1)	5.52 (353)	<.001	0.66
Distress	16.4 (6.3)	11.3 (4.0)	8.55 (353)	<.001	0.98
Symptom Focus	22.9 (5.8)	20.9 (5.5)	2.69 (353)	<.001	0.35
Implementing Advice	13.5 (4.3)	11.8 (3.4)	3.57 (353)	<.001	0.44
mSHAI[c]	35.7 (11.9)	25.3 (10.0)	7.42 (347)	<.001	0.95
CSS-15[d]	34.1 (9.6)	28.3 (7.1)	5.56 (337)	<.001	0.69
Search time, week without symptoms (minutes)	6.9 (12.6)	8.7 (27.0)	−0.52 (347)	.60	−0.08
Search time, week with symptoms (minutes)	44.4 (43.0)	40.4 (56.1)	0.56 (351)	.58	0.08
VAS[e] worry/anxiety regarding child's health	65.9 (22.0)	44.9 (25.7)	−6.1 (311)	<.001	0.88
VAS child's health	78.1 (20.5)	88.7 (12.8)	−5.29 (336)	<.001	−0.62
VAS participant's medical competence	60.0 (24.9)	63.1 (21.3)	−0.95 (291)	.34	-0.13
VAS experience with child's doctors	69.2 (26.5)	78.1 (21.8)	−2.77 (302)	<.001	−0.37

[a]CVS: Child Vulnerability Scale.

[b]CHIRPI: Children's Health Internet Research Parental Inventory.

[c]mSHAI: Modified Health Anxiety Inventory.

[d]CSS-15: Cyberchondria Severity Scale (15-item version).

[e]VAS: visual analog scale (0-100; 0 signifying no worries about child's health, child's health is poor, parent has no medical competence, parent's experience of child's doctors has been poor).

Influence of the Child's Health Status and Parents' Perception of Their Child's Health Vulnerability on the CHIRPI Scores

We investigated the relative influence of having a chronically ill child or perceiving the child as vulnerable with regard to their health on CHIRPI and its subscales by calculating four 2×2 ANOVAs with the factors child's health status (healthy/chronically ill) and perceiving the child as vulnerable (CVS−/CVS+). All analyses (CHIRPI total score and each subscale individually) revealed a main effect for the perceived vulnerability but no main effect for the child's health status or interactions of health status and perceived vulnerability of the child. See Multimedia Appendix 4 and Table 7.

Table 7. Results of the 2×2 analyses of variance with the factors child's health status and perceived vulnerability of the child and the dependent variables, Children's Health Internet Research, Parental Inventory (CHIRPI) total score, and its subscales.

Analyses of variance (2x2)	Sum of squares	*F* value (df)	*P* value	η^2
CHIRPI total score				
Intercept	292187.2	2444.256 (1)	N/A[a]	N/A
Health status	353.0	2.953 (1)	.09	0.008
Perceived vulnerability	2350.6	19.663 (1)	<.001	0.053
Health status×perceived vulnerability	77.7	0.650 (1)	.42	0.002
Error	41719.6	349 (N/A)	N/A	N/A
CHIRPI Symptom Focus				
Intercept	58975.60	1866.339 (1)	N/A	N/A
Health Status	7.76	0.246 (1)	.62	0.001
Perceived vulnerability	136.25	4.312 (1)	.04	0.012
Health status×perceived vulnerability	2.33	0.074 (1)	.79	0.000
Error	11028.27	349 (N/A)	N/A	N/A
CHIRPI Distress				
Intercept	24622.81	1190.417 (1)	N/A	N/A
Health status	44.88	2.170 (1)	.14	0.006
Perceived vulnerability	923.63	44.654 (1)	<.001	0.113
Health status×perceived vulnerability	15.40	0.745 (1)	.39	0.002
Error	7218.79	349 (N/A)	N/A	N/A
CHIRPI Implementing Advice				
Intercept	20614.99	1573.060 (1)	N/A	N/A
Health status	42.28	3.227 (1)	.07	0.009
Perceived vulnerability	85.44	6.519 (1)	.011	0.018
Health status×perceived vulnerability	0.29	0.022 (1)	.88	0.000

[a]N/A: not applicable.

Discussion

Principal Findings

CHIRPI is the first instrument to assess parents' excessive Web-based searching for information related to their children's health. The development of the inventory was guided by previous research on excessive Web-based research for health information [48-52], participative interviews with parents and children's health professionals, and psychometric standards. The inventory was validated in a good-sized online sample consisting mostly of mothers. Acceptance of the scale was high, and its psychometric characteristics are good to excellent.

Item Quality

Regarding the quality of the individual items, the general feedback in the validation and pilot tests indicated that the items were understandable and well accepted. The extremely low number of missing responses to individual items also suggested that participants did not find CHIRPI hard to complete. The statistical investigation showed that the item difficulties and item-total correlations were good to very good. In the context of the measurement of attitudes and behaviors, item difficulty is a measure of how frequently the statement made in an item is endorsed. Endorsement rates below 0.20 indicate comparatively low agreement, whereas endorsements above 0.80 point to widespread agreement. The endorsement rates for CHIRPI items ranged from 0.07 ("I ask my child's doctor to prescribe particular drugs that I have read about online.") to 0.65 ("When I search online for health information relevant to my child, I am particularly interested in whether the problems or symptoms need treatment."), with a mean of 0.29. For maximum diagnostic value, item difficulties should range from 0.20 to 0.80. However, considering that we are aiming to assess a behavioral excess, ie, a behavior that is relatively uncommon, the scale should differentiate better in the area of low endorsement. Therefore, the range of difficulties in CHIRPI seems to be appropriate. Item-total correlations capture the relationship between an individual item and the scale as a whole and those of the CHIRPI ranged from 0.37 to 0.65, with a mean of 0.50. These values are medium to high; in general, item-total correlations of 0.30 to 0.50 are regarded as medium and correlations of 0.50 and above, as high.

Factor Structure

The factor analysis revealed a clear factor structure without any double loadings, consisting of 3 moderately correlated factors, which together explained 45% of the variance. The first factor, Distress, captures how distressed a parent feels as a result of Web searches for health information. It captures cognitive (worry; item 21), emotional (feeling anxious, panicky, angry, or irritated; items 7, 9, and 12), and behavioral (finding it hard to relax or sleep or interference of the search with daily activities; items 4, 15, and 18) aspects of distress. The second factor, Symptom Focus, relates to the characteristics of the search behavior. It captures search behavior focused on changes the parent has noticed in his or her child's body and on attempts to determine the assumed causes of the symptoms or the presumed underlying illness. The third subscale, Implementing Advice, aggregates behaviors such as following general health advice for one's child that one has read about online. Examples for such use of internet advice may be the implementation of diets or exercise regimes, administering nonprescription medicine, asking the doctor for prescription drugs or diagnostic procedures, and engaging teachers or the nursery staff's assistance with observing symptoms or implementing health regimens.

It is important to note both similarities and differences of CHIRPI to the best-established instrument for assessing cyberchondria in adults searching for Web-based health information for themselves [48], especially as CSS-related research informed the generation of CHIRPI. Similar to the CSS, CHIRPI encompasses a Distress Scale, which consists of adapted item content of the respective CSS for parents' Web searches. It seems that this factor, reflecting the distress after searches, plays an important role in both constructs. In contrast to the CSS, however, CHIRPI hardly contains references to the compulsive nature of the Web searches or the mistrust toward medical professionals. Although items pertaining to this content were part of the original item pool for CHIRPI, they did not perform well during the questionnaire development phase and were thus eliminated by the iterative EFA process—perhaps because these facets are of less importance to parents' health-related Web searches. Excessiveness of the searches and reassurance by these searches were subsumed under the CHIRPI subscales, Symptom Focus, and Implementing Advice.

Reliability

We investigated two aspects of reliability: internal consistency and the 4-week retest reliability. The internal consistency of the total scale was high, with standardized alpha=.89, especially given its modest length (21 items). The subscales also showed good (Distress alpha=.89; Symptom Focus alpha=.87) or satisfactory (Implementing Advice alpha=.74) internal consistency. The comparatively low internal consistency of the Implementing Advice subscale might reflect the greater heterogeneity of behaviors described, which is indicated by the lower interitem correlations. Whereas the items of the Distress subscale capture aspects of distress that often co-occur (eg, increased worry, finding it harder to relax, problems with sleep), the items that form the Implementing Advice subscale describe behaviors that—albeit all indicative of the tendency to act on Web-based advice—need not co-occur. Participants who ask nursery staff or teachers for help with observing symptoms may not also ask their child's doctor to prescribe medications they learned about online.

Test-retest reliability is a measure of the degree of correspondence between questionnaire responses given at different points in time and is influenced by the quality of the questionnaire on the one hand and the stability of the measured construct on the other. At present, little is known about the temporal stability of health-related Web-based search behavior regarding one's children, in general, or excessive searching for such information, in particular. The 4-week retest reliability of r_{tt}=0.78 of CHIRPI suggests the phenomenon is stable across a medium time-frame of weeks and indicates that the instrument has good test-retest reliability. The test-retest reliability was highest for the Distress subscale (r_{tt}=0.84), which suggests that the tendency to experience worry and distress as a consequence of the searching may be an enduring trait. However, although the good test-retest reliability suggests the construct is reasonably stable and that the instrument used to measure it has adequate psychometric characteristics, further research in both these areas is warranted. In particular, it would be instructive to investigate the sensitivity of CHIRPI to change—eg, after an intervention aimed at reducing excessive Web searches by parents. In studies such as ours, where the subsample for the retest is self-selecting, the retest sample may be biased (eg, toward respondents particularly affected by the phenomenon under investigation), so we compared the data of the participants that formed our retest subsample against those who did not provide data for the retest. The samples did not differ with regard to the CHIRPI scores or scores on any of the other questionnaires, but parents with more children were less likely to participate again (small effect), possibly reflecting greater competition for their time.

Indicators of Validity

We investigated relationships between CHIRPI and related constructs, such as parental cyberchondria and worry about one's own health and the health of one's children to provide insights into the instrument's validity. In addition, we investigated, how the inventory performed in particular groups of parents, namely parents who reported that one of their children had a chronic illness and parents who perceived their child's health to be vulnerable (based on the CVS score).

Correlations With Health-Related Measures and Other Variables

Of all the subscales, the Distress subscale had the highest internal consistency, the highest retest reliability, and the highest correlation with the perceived vulnerability of the child's health. The Distress score was also highly correlated (r=0.46) with the parent's global rating of his or her worry and anxiety regarding the child's health. Interestingly, the Distress subscale was the only subscale to show this correlation, indicating that this subscale does assess the specific, distressing aspects of such health concerns. The Distress subscale was most highly correlated with general health anxiety and cyberchondria, ie, excessive Web searches related to one's own health. This points

to the existence of a general pattern of health anxiety-related cognitions, emotions, and behaviors that applies to the health of one's children as well as to one's own health

The Symptom Focus subscale was most strongly correlated with parents' search behavior regarding their own health, as measured by the CSS-15, and with time spent searching in a typical week in which the parent perceives symptoms of some sort in the child. This subscale is the only one that was correlated to the age and number of children. Having more children or older children was associated with a lower Symptom Focus score, although the effects were small to medium. The number of children one has and their mean age are not independent, so the observed effect may reflect a greater perceived need for information on a range of matters (including health) among first-time parents, who are adapting to the new task of parenting [24]. However, given the correlational nature of the data, questions of causality cannot be decided.

The Implementing Advice subscale was moderately correlated with parental cyberchondria and, interestingly, also with negative experiences with the child's doctors. Regarding the latter association, it is as easy to imagine that dissatisfaction with their child's doctor would lead parents to look online for advice and to follow such advice as it is to imagine that the parent-physician relationship might be disrupted by parental demands arising from Web-based research.

Differentiating the Influence of the Child's Health Status and Parents' Perception on the CHIRPI Scores

It seems highly likely that parents whose child has a chronic illness have a heightened need for medical information and make greater use of the internet as a source of such information. In our sample, 11.7% of participants reported that one of their children had a chronic illness, which accords well with the data in a study by Hölling et al [63] who estimated that every eighth child in Germany suffers from some form of chronic illness. This suggests that parents with chronically ill children were neither over- nor under-represented in our sample. Analyzing the differences in health-related Web searches, it appears that, compared with the parents of healthy children, parents with chronically ill children spend about twice as much time searching the internet for information related to their child's health in weeks in which the child displays symptoms (d=0.37). In weeks in which the child does not show any symptoms, no differences between these two groups were found. Parents with chronically ill children also had higher total CHIRPI scores, which could be traced back to the scores on the Implementing Advice subscale (d=0.38). However, these two groups of parents did not differ with regard to personal health anxiety and cyberchondria.

The above comparison pattern contrasts with the pattern of comparisons between parents who perceive their child's health to be vulnerable and those who do not. The former had much higher scores on all CHIRPI scales, as well as on the measures of personal health anxiety and cyberchondria.

To disentangle the contributions to excessive parental search behavior, we calculated an ANOVA that directly compared the influence of the two factors (having a chronically ill child and perceiving one's child as vulnerable) on CHIRPI scores. We found that CHIRPI is selectively sensitive to the concerns of parents who perceive their children to be vulnerable with regard to health and who show elevated scores of general health anxiety. This may reflect an overarching tendency to worry about health that extends to one's children. One could speculate that the behavior of parents with a chronically ill child represents a strategy for coping following receipt of a doctor's diagnosis of a chronic illness in their child. Whereas among parents who perceive their child's health to be vulnerable, checking them for symptoms, searching on the internet for symptom-related information, and experiencing distress could stem from the same health anxiety that causes these parents to worry about their own health and search the Web for related information. The differences warrant further investigation as such behaviors may relate to the transmission of health anxiety from parents to children. CHIRPI is intended to be sensitive to health anxiety and cyberchondria rather than to an increased need for health information due to chronic medical conditions. Thus, it speaks for the differential validity of CHIRPI and its sensitivity to health anxiety and cyberchondria that it does not show elevated scores for parents with chronically ill children simply in reaction to their legitimate need for health information.

Limitations

When interpreting the results, the following limitations must be borne in mind. First, our sample consisted predominantly of mothers; however, reflecting offline behaviors [64,65], the majority of those searching online for health information are women [2,10,18,66-68]. Second, our data are cross-sectional, so we cannot draw inferences about causality. Third, all our data are self-reported and may have been affected by a self-selection bias.

Conclusions

Most parents search on the internet for information related to their children's health and use it to their advantage. However, for a minority of parents, searching may escalate and have detrimental consequences similar to self-directed cyberchondria. Health professionals working with children worry that parents' Web searches for health information may affect their own role and their relationships with parents and children [69]. CHIRPI allows us to investigate such issues as it captures various forms of parents' excessive search behavior (eg, symptom-focused search) and their consequences, such as distress after the search or the implementation of newly acquired advice. CHIRPI has the potential to help identify parents who are at risk of child-directed cyberchondriac behavior. More research is needed to determine the potentially negative consequences of excessive searching for Web-based information related to one's children's health, eg, increased use of health services, children being subjected to unnecessary diagnostic procedures, and the transmission of unsubstantiated health concerns. The Distress subscale of CHIRPI, in particular, seems to be a useful marker of dysfunctional search influences, such as parents' anxiety about their own health or self-directed cyberchondria. It may indicate general patterns of factors related to anxious health-related cognitions, emotions, and behaviors, which the parents also apply to the health of their children. Hopefully, the

availability of CHIRPI will improve our understanding of the consequences of excessive Web-based searching for health information.

Practical Implications

Primarily, CHIRPI can be used to gain greater knowledge of health-related Web searches by parents and the possible consequences of such search behavior. The authors do not wish to suggest that searching for such information is problematic in itself. However, if it is coupled with health anxiety and cyberchondria regarding the parents themselves, it may become excessive and a source of distress for the parents. We would therefore suggest using this instrument to assess whether a person being treated for health anxiety also searches for their children and, if appropriate, discuss this behavior in the course of the treatment. Future research should address whether parents' excessive Web searches for health information regarding their children also leads to problematic consequences for the children themselves.

Acknowledgments

The authors gratefully acknowledge the help of Diana Festor, Philippa Gebhardt, and Sonja Kleemann with item development and data collection for the pilot study, and they acknowledge Nora Idris for her help with data collection for the validation study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys checklist.
[PDF File (Adobe PDF File), 380 KB - jmir_v22i4e16148_app1.pdf]

Multimedia Appendix 2
CHIRPI scale in German and English.
[DOCX File , 19 KB - jmir_v22i4e16148_app2.docx]

Multimedia Appendix 3
Independent samples t-tests comparing participants who opted to take part in the retest and those who did not.
[DOCX File , 17 KB - jmir_v22i4e16148_app3.docx]

Multimedia Appendix 4
Scores for the Children's Health Internet Research, Parental Inventory (CHIRPI) total score (A), and its subscales (B-D) as a function of the child's health status and the perceived vulnerability regarding the child's health assessed using the Child Vulnerability Scale (CVS).
[PNG File , 123 KB - jmir_v22i4e16148_app4.png]

References

1. Statista. 2018. Number of Internet Users Worldwide URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ [accessed 2018-11-17]
2. 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]
3. Fox S. Pew Research Center. 2011 May 12. The Social Life of Health Information, 2011 URL: https://www.pewresearch.org/internet/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2020-02-13]
4. Fox S, Rainie L. Pew Research Center. 2002 May 22. Vital Decisions: A Pew Internet Health Report URL: http://www.pewinternet.org/2002/05/22/vital-decisions-a-pew-internet-health-report/ [accessed 2020-02-13]
5. Van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct 2;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
6. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
7. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
8. Borzekowski DL, Rickert VI. Adolescent cybersurfing for health information: a new resource that crosses barriers. Arch Pediatr Adolesc Med 2001 Jul;155(7):813-817. [doi: 10.1001/archpedi.155.7.813] [Medline: 11434849]

9. 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. Br Med J 2002 Mar 9;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
10. Sillence E, Briggs P, Harris P, Fishwick L. Going online for health advice: changes in usage and trust practices over the last five years. Interact Comput 2007 May;19(3):397-406 [FREE Full text] [doi: 10.1016/j.intcom.2006.10.002]
11. Benigeri M, Pluye P. Shortcomings of health information on the internet. Health Promot Int 2003 Dec;18(4):381-386. [doi: 10.1093/heapro/dag409] [Medline: 14695369]
12. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality, and readability in English and Spanish. J Am Med Assoc 2001;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
13. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
14. Scullard P, Peacock C, Davies P. Googling children's health: reliability of medical advice on the internet. Arch Dis Child 2010 Aug;95(8):580-582. [doi: 10.1136/adc.2009.168856] [Medline: 20371593]
15. Wacogne I, Scott-Jupp R. The role of Google in children's health. Arch Dis Child 2010 Aug;95(8):576-577. [doi: 10.1136/adc.2009.176917] [Medline: 20573737]
16. White RW, Horvitz E. Cyberchondria. ACM Trans Inf Syst 2009;27(4):1-37. [doi: 10.1145/1629096.1629101]
17. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
18. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [doi: 10.1089/cpb.2004.7.497] [Medline: 15667044]
19. Norr AM, Allan NP, Boffa JW, Raines AM, Schmidt NB. Validation of the Cyberchondria Severity Scale (CSS): replication and extension with bifactor modeling. J Anxiety Disord 2015 Apr;31:58-64. [doi: 10.1016/j.janxdis.2015.02.001] [Medline: 25734759]
20. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
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;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
22. Aiken M, Kirwan G. The psychology of cyberchondria and cyberchondria by proxy. In: Power A, Kirwan G, editors. Cyberpsychology and New Media: A Thematic Reader. London: Psychology Press; 2013:158-169.
23. Goldman RD, Macpherson A. Internet health information use and e-mail access by parents attending a paediatric emergency department. Emerg Med J 2006 May;23(5):345-348 [FREE Full text] [doi: 10.1136/emj.2005.026872] [Medline: 16627833]
24. Ybarra M, Suman M. Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. Health Educ Res 2008 Jun;23(3):512-521 [FREE Full text] [doi: 10.1093/her/cyl062] [Medline: 16880222]
25. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 1;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
26. Sim NZ, Kitteringham L, Spitz L, Pierro A, Kiely E, Drake D, et al. Information on the world wide web--how useful is it for parents? J Pediatr Surg 2007 Feb;42(2):305-312. [doi: 10.1016/j.jpedsurg.2006.10.003] [Medline: 17270540]
27. Crocco AG, Villasis-Keever M, Jadad AR. Two wrongs don't make a right: harm aggravated by inaccurate information on the internet. Pediatrics 2002 Mar;109(3):522-523. [doi: 10.1542/peds.109.3.522] [Medline: 11875153]
28. Hargrave DR, Hargrave UA, Bouffet E. Quality of health information on the internet in pediatric neuro-oncology. Neuro Oncol 2006 Apr;8(2):175-182 [FREE Full text] [doi: 10.1215/15228517-2005-008] [Medline: 16533758]
29. Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the world wide web: systematic survey of advice on managing fever in children at home. Br Med J 1997 Jun 28;314(7098):1875-1879 [FREE Full text] [doi: 10.1136/bmj.314.7098.1875] [Medline: 9224132]
30. Pandolfini C, Impicciatore P, Bonati M. Parents on the web: risks for quality management of cough in children. Pediatrics 2000 Jan;105(1):e1. [doi: 10.1542/peds.105.1.e1] [Medline: 10617738]
31. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]
32. Ołpiński M. Anti-vaccination movement and parental refusals of immunization of children in USA. Pediatria Polska 2012 Jul;87(4):381-385. [doi: 10.1016/j.pepo.2012.05.003]
33. Acar B, Acar M, Ocak E, Kocaöz D, Koksal AO, Karasen RM. Accuracy of internet guidance on pediatric otolaryngology procedures. Int J Pediatr Otorhinolaryngol 2014 Dec;78(12):2190-2192. [doi: 10.1016/j.ijporl.2014.10.009] [Medline: 25455527]
34. Alderdice F, Gargan P, McCall E, Franck L. Online information for parents caring for their premature baby at home: a focus group study and systematic web search. Health Expect 2018 Aug;21(4):741-751 [FREE Full text] [doi: 10.1111/hex.12670] [Medline: 29380916]

35. Teasdale EJ, Muller I, Santer M. Carers' views of topical corticosteroid use in childhood eczema: a qualitative study of online discussion forums. Br J Dermatol 2017 Jun;176(6):1500-1507. [doi: 10.1111/bjd.15130] [Medline: 27753076]
36. de Oliveira BH, Grisolia BM, Dos Santos APP. Children's toothbrushing practices recommended on the internet by pediatric dentistry associations. Pediatr Dent 2016 Nov 15;38(7):484-488. [Medline: 28281953]
37. D'Alessandro DM, Dosa NP. Empowering children and families with information technology. Arch Pediatr Adolesc Med 2001 Oct;155(10):1131-1136. [doi: 10.1001/archpedi.155.10.1131] [Medline: 11576008]
38. Giesen M, Keizer E, van de Pol J, Knoben J, Wensing M, Giesen P. The impact of demand management strategies on parents' decision-making for out-of-hours primary care: findings from a survey in The Netherlands. BMJ Open 2017 May 9;7(5):e014605 [FREE Full text] [doi: 10.1136/bmjopen-2016-014605] [Medline: 28487458]
39. Bouwman MG, Teunissen QG, Wijburg FA, Linthorst GE. 'Doctor Google' ending the diagnostic odyssey in lysosomal storage disorders: parents using internet search engines as an efficient diagnostic strategy in rare diseases. Arch Dis Child 2010 Aug;95(8):642-644. [doi: 10.1136/adc.2009.171827] [Medline: 20418338]
40. Boston M, Ruwe E, Duggins A, Willging JP. Internet use by parents of children undergoing outpatient otolaryngology procedures. Arch Otolaryngol Head Neck Surg 2005 Aug;131(8):719-722. [doi: 10.1001/archotol.131.8.719] [Medline: 16103305]
41. Tuffrey C, Finlay F. Use of the internet by parents of paediatric outpatients. Arch Dis Child 2002 Dec;87(6):534-536 [FREE Full text] [doi: 10.1136/adc.87.6.534] [Medline: 12456558]
42. Baker JF, Devitt BM, Lynch S, Green CJ, Byrne DP, Kiely PJ. Internet use by parents of children attending a dedicated scoliosis outpatient clinic. Eur Spine J 2012 Oct;21(10):1972-1977 [FREE Full text] [doi: 10.1007/s00586-012-2429-2] [Medline: 22820914]
43. 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 US survey among 1.050 US physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
44. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
45. Wald HS, Dube CE, Anthony DC. Untangling the web--the impact of internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
46. Barke A, Bleichhardt G, Rief W, Doering BK. The Cyberchondria Severity Scale (CSS): German validation and development of a short form. Int J Behav Med 2016 Oct;23(5):595-605. [doi: 10.1007/s12529-016-9549-8] [Medline: 26931780]
47. Fergus TA. Cyberchondria and intolerance of uncertainty: examining when individuals experience health anxiety in response to Internet searches for medical information. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):735-739. [doi: 10.1089/cyber.2012.0671] [Medline: 23992476]
48. McElroy E, Shevlin M. The development and initial validation of the cyberchondria severity scale (CSS). J Anxiety Disord 2014 Mar;28(2):259-265. [doi: 10.1016/j.janxdis.2013.12.007] [Medline: 24508033]
49. Starcevic V. Cyberchondria: challenges of problematic online searches for health-related information. Psychother Psychosom 2017;86(3):129-133 [FREE Full text] [doi: 10.1159/000465525] [Medline: 28490037]
50. Starcevic V, Aboujaoude E. Cyberchondria, cyberbullying, cybersuicide, cybersex: 'new' psychopathologies for the 21st century? World Psychiatry 2015 Feb;14(1):97-100 [FREE Full text] [doi: 10.1002/wps.20195] [Medline: 25655165]
51. Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related internet use. Expert Rev Neurother 2013 Feb;13(2):205-213. [doi: 10.1586/ern.12.162] [Medline: 23368807]
52. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
53. Te Poel F, Baumgartner SE, Hartmann T, Tanis M. The curious case of cyberchondria: a longitudinal study on the reciprocal relationship between health anxiety and online health information seeking. J Anxiety Disord 2016 Oct;43:32-40. [doi: 10.1016/j.janxdis.2016.07.009] [Medline: 27497667]
54. LimeSurvey: An Open Source survey tool. Hamburg, Germany URL: https://www.limesurvey.org/ [accessed 2020-02-07]
55. 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]
56. Forsyth BW, Horwitz SM, Leventhal JM, Burger J, Leaf PJ. The child vulnerability scale: an instrument to measure parental perceptions of child vulnerability. J Pediatr Psychol 1996 Feb;21(1):89-101. [doi: 10.1093/jpepsy/21.1.89] [Medline: 8820075]
57. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
58. Bailer J, Rist F, Müller T, Mier D, Diener C, Ofer J. Erfassung von Krankheitsangst mit dem short health anxiety inventory (SHAI). Verhaltenstherapie Verhaltensmedizin 2013;34(4):378-398 [FREE Full text]

59. Salkovskis PM, Rimes KA, Warwick HM, Clark DM. The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. Psychol Med 2002 Jul;32(5):843-853. [doi: 10.1017/s0033291702005822] [Medline: 12171378]
60. Alberts NM, Hadjistavropoulos HD, Jones SL, Sharpe D. The Short Health Anxiety Inventory: a systematic review and meta-analysis. J Anxiety Disord 2013 Jan;27(1):68-78. [doi: 10.1016/j.janxdis.2012.10.009] [Medline: 23247202]
61. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. Behav Res Methods Instrum Comput 2000 Aug;32(3):396-402. [doi: 10.3758/bf03200807] [Medline: 11029811]
62. Levine TR, Hullett CR. Eta Squared, Partial Eta Squared, and misreporting of effect size in communication research. Human Commun Res 2002;28(4):612-625. [doi: 10.1111/j.1468-2958.2002.tb00828.x]
63. Hölling H, Schlack R, Dippelhofer A, Kurth B. [Personal, familial and social resources and health-related quality of life in children and adolescents with chronic conditions]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2008 Jun;51(6):606-620. [doi: 10.1007/s00103-008-0537-2] [Medline: 18465100]
64. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. Health Promot Int 2015 Sep;30(3):736-745. [doi: 10.1093/heapro/dat063] [Medline: 23985248]
65. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. Soc Sci Med 2015 Mar;128:151-158. [doi: 10.1016/j.socscimed.2015.01.028] [Medline: 25618604]
66. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb 20;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
67. Percheski C, Hargittai E. Health information-seeking in the digital age. J Am Coll Health 2011;59(5):379-386. [doi: 10.1080/07448481.2010.513406] [Medline: 21500056]
68. 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]
69. Holyoake D, Searle K. Cyberchondria: emerging themes for children's nurses in the internet age. Nurs Child Young People 2015 Jun;27(5):34-38. [doi: 10.7748/ncyp.27.5.34.e600] [Medline: 26059589]

Abbreviations

ANOVA: analysis of variance
CHIRPI: Children's Health Internet Research, Parental Inventory
CSS: Cyberchondria Severity Scale
CVS: Child Vulnerability Scale
EFA: exploratory factor analysis
KMO: Kaiser-Meyer-Olkin criterion
MAP: minimum average partial
mSHAI: modified short Health Anxiety Inventory
VAS: visual analog scale

Edited by G Eysenbach; submitted 05.09.19; peer-reviewed by Z Niu, E McElroy; comments to author 02.12.19; revised version received 29.12.19; accepted 27.01.20; published 15.04.20.

Please cite as:
Barke A, Doering BK
Development of an Instrument to Assess Parents' Excessive Web-Based Searches for Information Pertaining to Their Children's Health: The "Children's Health Internet Research, Parental Inventory" (CHIRPI)
J Med Internet Res 2020;22(4):e16148
URL: https://www.jmir.org/2020/4/e16148
doi:10.2196/16148
PMID:32293571

Original Paper

The French eHealth Acceptability Scale Using the Unified Theory of Acceptance and Use of Technology 2 Model: Instrument Validation Study

Meggy Hayotte[1], MSc; Pierre Thérouanne[2], PhD; Laura Gray[1], PhD; Karine Corrion[1], PhD; Fabienne d'Arripe-Longueville[1], PhD

[1]Laboratoire Motricité Humaine Expertise Sport Santé, Université Côte d'Azur, Nice, France

[2]Laboratoire d'Anthropologie et de Psychologie Cliniques, Cognitives et Sociales, Université Côte d'Azur, Nice, France

Corresponding Author:
Meggy Hayotte, MSc
Laboratoire Motricité Humaine Expertise Sport Santé
Université Côte d'Azur
261 Boulevard du Mercantour
Nice, 06200
France
Phone: 33 4 89 15 39 52
Email: meggy.hayotte@etu.univ-cotedazur.fr

Abstract

Background: Technology-based physical activity suggests new opportunities for public health initiatives. Yet only 45% of technology interventions are theoretically based, and the acceptability mechanisms have been insufficiently studied. Acceptability and acceptance theories have provided interesting insights, particularly the unified theory of acceptance and use of technology 2 (UTAUT2). In several studies, the psychometric qualities of acceptability scales have not been well demonstrated.

Objective: The aim of this study was to adapt the UTAUT2 to the electronic health (eHealth) context and provide a preliminary validation of the eHealth acceptability scale in a French sample.

Methods: In line with the reference validation methodologies, we carried out the following stages of validating the scale with a total of 576 volunteers: translation and adaptation, dimensionality tests, reliability tests, and construct validity tests. We used confirmatory factor analysis to validate a 22-item instrument with 7 subscales: Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation, Price Value, and Habit.

Results: The dimensionality tests showed that the bifactor confirmatory model presented the best fit indexes: χ^2_{173}=434.86 (P<.001), χ^2/df=2.51, comparative fit index=.97, Tucker-Lewis index=.95, and root mean square error of approximation=.053 (90% CI .047-.059). The invariance tests of the eHealth acceptability factor structure by sex demonstrated no significant differences between models, except for the strict model. The partial strict model demonstrated no difference from the strong model. Cronbach alphas ranged from .77 to .95 for the 7 factors. We measured the internal reliability with a 4-week interval. The intraclass correlation coefficients for each subscale ranged from .62 to .88, and there were no significant differences in the t tests from time 1 to time 2. Assessments for convergent validity demonstrated that the eHealth acceptability constructs were significantly and positively related to behavioral intention, usage, and constructs from the technology acceptance model and the theory of planned behavior.

Conclusions: The 22-item French-language eHealth acceptability scale, divided into 7 subscales, showed good psychometric qualities. This scale is thus a valid and reliable tool to assess the acceptability of eHealth technology in French-speaking samples and offers promising avenues in research, clinical practice, and marketing.

(J Med Internet Res 2020;22(4):e16520) doi:10.2196/16520

KEYWORDS

telemedicine; validation study; factor analysis, statistical; surveys and questionnaires; acceptability

Introduction

Background

Technology-based interventions to promote healthy behavior have been an emerging field of research for the past 10 to 20 years [1,2]. Among the healthy behaviors that are promoted, technology-based physical activity has brought to light new opportunities for public health interventions [3]. Several studies have evaluated the prospects of technologies such as exergames and active videogames [4], virtual reality [5], wearable physical activity trackers [6], website-delivered physical activity interventions [7], mobile phone apps [8,9], and video conferencing [10]. Electronic health (eHealth) physical activity promotion technologies have been designed not only for healthy adults [8], but also for vulnerable people in health care contexts, including cancer survivors [11,12], those in need of treatment for overweight and obesity [13,14] or cardiac rehabilitation [15], and older people [16,17]. All these technologies are popular (ie, positive assessment by many), with promising and reported positive outcomes [11-17]. However, the phenomena of usage cessation and losses to follow-up (ie, the law of attrition) are common problems [18]. Moreover, only 45% of the technology interventions are theoretically based, and the acceptability mechanisms have been insufficiently studied [1,19].

Acceptability and acceptance theories have provided interesting insights [20,21] into why some tools are chosen, accepted, and used more than others. The literature on acceptability and acceptance has emerged in different fields (eg, ergonomics, social psychology, management science) [22]. However, the acceptability and acceptance concepts have not been formally defined [23], and the distinction between the two has been based on the temporality of usage [21]. Acceptability refers to the a priori perceived use, whereas acceptance refers to the actual use [22]. Based on the proposed definitions [23], we define acceptability in this paper as the psychological antecedents of the behavioral intention to use technology without experience of the system. In the field of social psychology, the theory of reasoned action and the theory of planned behavior (TPB) [24,25] hold that attitudes and representations determine behavioral intention and real behavior [22]. These theories are the foundation for the technology acceptance model (TAM) [26], the most frequently used model in health informatics [27]. Nevertheless, several extensions have been proposed—TAM2 [28] and TAM3 [29]—revealing that the original TAM was not optimal in eHealth [27]. The unified theory of acceptance and use of technology (UTAUT), particularly its extension, UTAUT2, is today the most complete model, as it combines theory of reasoned action, TAM, a motivational model, TPB, a combined TPB and TAM, a model of personal computer use, diffusion of innovations theory, and social cognitive theory [29-31]. The UTAUT2 comprises 26 items divided into 8 constructs: Performance Expectancy (PE, 3 items), Effort Expectancy (EE, 4 items), Social Influence (SI, 3 items), Facilitating Conditions (FC, 4 items), Hedonic Motivation (HM, 3 items), Price Value (PV, 3 items), Habit (HT, 3 items), and Behavioral Intention (BI, 3 items).

Acceptability assessments in several studies in eHealth contexts have been based on tools without or with only partially demonstrated psychometric qualities [32,33]. However, to ensure the quality of future research, it is necessary to have scales with validated psychometric qualities [34]. The UTAUT model can be considered as a relevant framework for assessing the acceptability of eHealth, particularly for patient-centered assessment [20]. Yet, for theoretically based technologies in the health and wellness field, only 2 studies have been based on the UTAUT model [1], and 2 were conducted in France [27]. The scarcity French studies [1,27] may be due to the absence of validated scales in French to evaluate acceptability. To our knowledge, validated scales in the French language have been based on the TAM model [35] or on other definitions of acceptability in which the concept of acceptability is merged with the definition of usability [22,36]. The UTAUT2 [31] has already been translated into other languages (eg, German [37], Turkish [38], and Portuguese [39]) and has proven its validity; however, the psychometric qualities of the scales have been only partially demonstrated.

Objective

The aims of this study were to adapt the UTAUT2 [31] to the eHealth context and to validate this version, which we called the eHealth acceptability scale, in French-speaking samples. This validated tool would allow for the development of further studies in this field.

Methods

Study Design

In line with the guidelines for scale validation from Vallerand et al [40] and Boateng et al [41], we conducted successive stages: translation and adaptation, dimensionality tests, reliability tests, and construct validity tests.

We managed the administration of the scale using LimeSurvey CE, version 2.06+ (LimeSurvey CE). We distributed the scale link by email or face-to-face at the end of students' courses. We also distributed the link by email to health professionals and adults registered for adapted physical activity. In addition, we posted the link online via social media networks.

Study Population

We recruited participants in various categories of the general population: students (studying sports, psychology, management, and computer science at a university in the South of France), health professionals (in the field of obesity), and adults with health conditions (ie, diabetes, cardiovascular disease, and obesity) registered for adapted physical activity sessions. To conduct the successive stages of validation, we divided the participants into 5 sample groups.

This study was approved by the French National Commission for Information Technology and Civil Liberties (authorization no: UCA-E18-00), and all participants gave their electronic consent before participation.

Measures

Sociodemographics

The sociodemographic information, provided by all participants after they had completed the scale items, included their sex, year of birth, education level, and professional status.

eHealth Acceptability Scale

The UTAUT2 [31], originally developed in English in the field of mobile internet use, has 2 sections, 1 for the UTAUT2 scale comprising 26 items divided into 8 constructs, and the other for assessing the usage frequency of various apps for mobile internet. According to the definition we chose, acceptability corresponds to the psychological antecedents of the behavioral intention to use technology without experience of the system. Based on this definition, we excluded BI from the eHealth acceptability scale.

We produced this French adaptation of the UTAUT2 scale using the back-translation method [42]. Original items were translated individually by 4 researchers in the field of psychology and compiled to obtain a single French version. This French version was back-translated by 4 researchers unaware of the original version. The back-translators were subsequently asked to compare their own translation with the original to specify the differences. Differences were noted for 2 items, which were adjusted with the same procedure until all back-translators concluded that there was no difference. We then used a committee approach to replace mobile internet with a global expression that would include all the eHealth apps. We chose information and communication technologies for health, abbreviated as ICT for health, in reference to the wording used in a similar French questionnaire [36].

The preliminary version of the eHealth acceptability scale comprised 23 items divided into 7 subscales: PE (3 items), EE (4 items), SI (3 items), FC (4 items), HM (3 items), PV (3 items), and HT (3 items). Participants answered on a 7-point scale with labeled anchors ranging from 1, “strongly disagree,” to 7, “strongly agree.” We chose this 7-point scale because the participants were not familiar with the study context [43]. We administered this preliminary version of the eHealth acceptability scale to samples 1 to 3 and its adjusted form after the first confirmatory factor analysis (CFA) to samples 4 and 5. Sample 5 participants completed the scale a second time after 4 weeks for the test-retest reliability assessments.

Behavioral Intention

BI comes from the original UTAUT2 [31]. The 3 items were translated following the same procedure described above. The participants in all samples answered on a 7-point scale with labeled anchors ranging from 1, “strongly disagree,” to 7, “strongly agree.” BI was theoretically positively related to the constructs of the eHealth acceptability scale.

Usage

We measured usage in all samples as the frequency of eHealth technology use on a 7-point scale ranging from 1, “never,” to 7, “many times per day,” for 5 technologies: mobile health apps, forums or social networks for health, videos for health management, exergames or active video games, and health trackers. Usage was theoretically positively related to the constructs of the eHealth acceptability scale, especially FC and HT [31].

Technology Acceptance Model Constructs

Perceived Ease of Use (PEOU, 5 items), extracted from the TAM [26,35], was theoretically positively related to the constructs of the eHealth acceptability scale, especially EE [30]. This subscale was measured on a 7-point scale with labeled anchors ranging from 1, “strongly disagree,” to 7, “strongly agree.” Only the sample 4 participants completed this subscale to test for convergent validity.

Theory of Planned Behavior Constructs

Subjective Norms (SN, 3 items) and Perceived Behavioral Control (PBC, 5 items) extracted from the TPB [44,45] were theoretically positively related to the constructs of the eHealth acceptability scale, especially SI and FC [30]. These subscales were measured on 7-point scales with labeled anchors ranging from 1, “strongly disagree,” to 7, “strongly agree.” Only the sample 4 participants completed these subscales to test for convergent validity.

Statistical Analyses

We performed all statistical analyses with IBM SPSS version 23 (IBM Corporation) and IBM SPSS Amos version 23 (IBM Corporation). We examined the missing data trends. The cutoff for an acceptable percentage of missing data has not been well established in the literature [46]. However, 5% is considered inconsequential [47], and the risk of statistical bias is considered when the rate is higher than 10% [48]. In our global sample, the missing rate was under 10%. For structural equation modeling, the maximum likelihood estimation and the multiple imputation for handling missing data presented close to equivalent good properties [49]. We applied the maximum likelihood estimations (considered the standard for structural equation models [46]) to be used in Amos v23.

Tests of Dimensionality

We ran tests of dimensionality using maximum likelihood estimation CFA in structural equation modeling according to several models [50]. We used the following indicators to assess competence of the model fit [51-54]: chi-square (significant values $P \leq .05$), chi-square over degrees of freedom (significant values ≤3.00), comparative fit index (CFI; value >.90), Tucker-Lewis index (TLI; value >.90), root mean square error of approximation (RMSEA; value <.08), and the 90% confidence interval of RMSEA (ranging from .00 to .08).

We computed invariance of the eHealth acceptability scale between the sexes according to Gregorich’s methodology [55]. In the CFA framework, we tested a hierarchy of hypotheses to increasingly constrain the model. These hypotheses included configural (ie, no constraint), metric (ie, equal loads), strong (ie, equal covariances), and strict (ie, equal residuals) factorial invariance multigroup comparisons [55]. In addition to the previous indicators, we used the Akaike information criterion, expected cross-validation index, delta χ^2/df ($\Delta\chi^2/df$), delta CFI (ΔCFI), and delta RMSEA (ΔRMSEA). Nonsignificant $\Delta\chi^2/df$,

CFI differences <.01, and RMSEA differences <.015 indicated that the invariance hypothesis was not rejected [51,56].

Tests of Reliability

We calculated Cronbach alpha coefficients [57] to assess the internal consistency of each subscale; a value >.70 is considered satisfactory and a value >.60 is considered marginally acceptable [58]. We measured the test-retest reliability twice with an acceptable interval of 4 weeks [59] and a minimum sample size of 50 as recommended [60]. Data analyses involved the calculation of intraclass correlation coefficients (ICCs), the 95% confidence interval of the ICCs, and paired-sample *t* tests. We expected ICCs >.60 and the absence of significant differences in the *t* tests [40].

Tests of Construct Validity

We used Pearson correlation coefficients to measure the association between variables for the analysis of convergent validity. A significant correlation of .30 between the scale and each of the other theoretically appropriate measures was required [61].

Results

Study Population

To conduct the successive stages of validation, we divided the participants into 5 samples. Samples 1 (n=20), 2 (n=10), 3 (n=227), and 4 (n=319) were independent groups, and sample 5 (n=61) was a subgroup of sample 4. The global sample included 576 volunteers, mainly students (n=349, 60.6%), with 53.5% men (n=303) and a mean age of 26.8 (SD 10.9) years. We excluded 18 volunteers because they had never used eHealth technology. Table 1 presents detailed participant characteristics for each sample.

Table 1. Sociodemographic characteristics in each sample (N=576).

Characteristics	Sample 1 (n=20), n (%)	Sample 2 (n=10), n (%)	Sample 3 (n=227), n (%)	Sample 4 (n=319), n (%)	Sample 5[a] (n=61), n (%)
Age group, years					
18-24	3 (15.0)	2 (20.0)	128 (56.4)	238 (74.6)	25 (41.0)
25-34	8 (40.0)	5 (50.0)	46 (20.3)	41 (12.9)	13 (21.3)
≥35	9 (45.0)	3 (30.0)	43 (18.9)	38 (11.9)	23 (37.7)
Missing data	0	0	10 (4.4)	2 (0.6)	0
Sex					
Female	8 (40.0)	5 (50.0)	117 (51.5)	132 (41.4)	40 (65.6)
Male	12 (60.0)	5 (50.0)	100 (44.1)	186 (58.3)	21 (34.4)
Missing data	0	0	10 (4.4)	1 (0.3)	0
Education, years					
<12	7 (35.0)	2 (20.0)	1 (0.4)	0	0
12	5 (25.0)	4 (40.0)	125 (55.1)	143 (44.8)	9 (14.8)
15	6 (30.0)	1 (10.0)	30 (13.2)	116 (36.4)	19 (31.1)
≥17	2 (10.0)	3 (30.0)	61 (26.9)	60 (18.8)	33 (54.1)
Missing data	0	0	10 (4.4)	0	0
Professional status					
Unemployed	0	0	5 (2.2)	7 (2.2)	0
Student	3 (15.0)	8 (80.0)	103 (45.4)	235 (73.7)	25 (41.0)
Employed	15 (75.0)	0	103 (45.4)	74 (23.2)	36 (59.0)
Retired	2 (10.0)	2 (20.0)	6 (2.6)	3 (0.9)	0
Missing data	0	0	10 (4.4)	0	0

[a]Sample 5 was a subsample of sample 4.

Translation and Adaptation

We performed the first content clarity analysis on sample 1 (n=20), which revealed an acceptable clarity score (mean range from 4.40 to 7.00; mean 6.22, SD 0.71). Only 3 items (ie, EE2, SI2, and SI3) obtained a score of less than 5, which we rephrased according to participants' suggestions. We performed a second content clarity analysis on sample 2 (n=10) regarding the 3 rephrased items. The clarity score increased for 2 items (SI2: mean range 4.40 to 6.20; SI3: mean range 4.90 to 6.20) but decreased for the third (EE2: mean range 4.65 to 3.30). We retained the 2 items with increased clarity scores in their rephrased form and the item with a decreased clarity score in

its original translated wording. Multimedia Appendix 1 shows the preliminary pool of 23 items with their mean clarity scores.

Tests of Dimensionality

We conducted a first maximum likelihood CFA on sample 3 (n=227) with the 23-item and 7-factor model. Standardized factor loadings were all higher than the recommended value of .50 [62], except for item FC4, for which the factor loading was .27. As a result, we removed item FC4. We conducted a second CFA using sample 4 (n=319) with the 22-item (ie, without FC4 item) and 7-factor correlated model (χ^2_{188}=471.80, *P*<.001). Fit indexes were as follows: χ^2/df=2.51, CFI=.94, TLI=.91, and RMSEA=.069 (90% CI 0.061-0.077), revealing an acceptable model fit, with good standardized factor loadings for all items (ie, ≥0.63).

Based on the recommendations of Myers et al [50], we examined several models to assess the dimensionality of the scale, using samples 3 and 4 merged (n=546). Table 2 presents model fit indexes for each model. First, the unidimensional model did not present good fit indexes. Second, the first-order all-factor correlated model presented good fit indexes, as previously demonstrated. Third, the hierarchical second-order model presented acceptable fit indexes. Fourth, the bifactor confirmatory model presented the best fit indexes: χ^2_{173}=434.86 (*P*<.001), χ^2/df=2.51, CFI=.97, TLI=.95, and RMSEA=.053 (90% CI .047-.059). These results sustained the possibility of extracting a global acceptability score from the scale.

Table 2. Fit indexes of the structural equation models (n=546).

Models	χ^2	χ^2 *df*	*P* value	RMSEA[a] (90% CI)	TLI[b]	CFI[c]	$\Delta\chi^2$	$\Delta\chi^2$ *df*	ΔP
Unidimensional	4721.73	209	<.001	.199 (.194-.204)	.27	.39	N/A[d]	N/A	N/A
First-order all-factor correlated	532.29	188	<.001	.058 (.052-.064)	.94	.95	4189.44	21	<.001
Hierarchical second-order	825.98	202	<.001	.075 (.070-.081)	.90	.92	293.69	14	<.001
Bifactor confirmatory	434.86	173	<.001	.053 (.047-.059)	.95	.97	391.12	29	<.001

[a]RMSEA: root mean square error of approximation.

[b]TLI: Tucker-Lewis index.

[c]CFI: comparative fit index.

[d]N/A: not applicable.

We tested the invariance of the scale factorial structure following Gregorich's recommendations [55], with samples 3 and 4 merged (n=535; 11 without sex information). The invariance tests were based on multigroup comparisons: female group (n=249) and male group (n=286). Each group presented good fit indexes for the CFA model (Table 3). We tested invariance in the 22-item 7-factor correlated model. Dimensional, metric, strong, and strict models presented good fit indexes (ie, CFI, TLI, and RMSEA) with significant chi-square *P* values (ie, *P*<.001). No significant differences between models were reported, except for the strict model (Table 3). A partial strict model, unconstrained for error of measurement for items EE2 and HM1, showed good fit indexes with no significant difference from the strong model.

Table 3. Fit indexes of structural modeling to assess sex invariance (n=535).

Models	χ^2	χ^2 *df*	*P* value	RMSEA[a]	TLI[b]	CFI[c]	ECVI[d]	AIC[e]	$\Delta\chi^2$	$\Delta\chi^2$ *df*	Δ*P*	ΔCFI	ΔRMSEA
Male (n=286)	427.42	188	<.001	.067	.93	.94	2.11	601.42	N/A[f]	N/A	N/A	N/A	N/A
Female (n=249)	330.71	188	<.001	.055	.96	.96	2.04	504.71	N/A	N/A	N/A	N/A	N/A
Dimensional[g]	758,12	376	<.001	.044	.94	.95	1.91	1018.12	N/A	N/A	N/A	N/A	N/A
Metric[h]	770.21	389	<.001	.043	.94	.95	1.88	1004.21	12.09	13	.520	0	0.001
Strong[i]	801.51	417	<.001	.042	.95	.95	1.84	979.51	31.30	28	.304	0	0.001
Strict[j]	915.77	439	<.001	.045	.94	.94	1.97	1049.77	114.26	22	<.001	0.012	0.002
Partial strict[k]	908.84	438	<.001	.044	.94	.94	1.91	1015.86	43.40	41	.370	0.011	0.002

[a]RMSEA: root mean square error of approximation.
[b]TLI: Tucker-Lewis index.
[c]CFI: comparative fit index.
[d]ECVI: expected cross-validation index.
[e]AIC: Akaike information criterion.
[f]N/A: not applicable.
[g]No invariance.
[h]Equal loads.
[i]Equal covariances.
[j]Equal residuals.
[k]Equal residuals except for items EE2 and HM1.

Tests of Reliability

Cronbach alphas ranged from .77 to .95 in samples 3 and 4 (n=546) for the 7 eHealth acceptability factors (ie, α_{PE}=.84; α_{EE}=.88; α_{SI}=.95; α_{FC}=.78; α_{HM}=.92; α_{PV}=.86; α_{HT}=.77) and were .93 for BI and .60 for usage.

We measured test-retest reliability in sample 5 (n=61) twice with an acceptable interval of 4 weeks [59]. Table 4 presents the results of the ICC and *t* tests. The ICCs for each construct ranged from .62 to .88. Thus, there were no significant differences in the *t* tests from time 1 to time 2.

Table 4. Descriptive statistics for the test-retest reliability in sample 5 (n=61).

Items	Score, mean (SD)		*t* test[a]	*P* value	ICC[b] (95% CI)	*P* value
	Time 1	Time 2				
Performance Expectancy	4.67 (1.45)	4.46 (1.40)	t_{60}=1.74	.09	.88 (.80-.93)	<.001
Effort Expectancy	5.43 (1.14)	5.60 (1.16)	t_{60}=–1.22	.23	.74 (.57-.84)	<.001
Social Influence	3.67 (1.39)	3.52 (1.66)	t_{60}=0.87	.39	.77 (.62-.86)	<.001
Facilitating Conditions	5.82 (0.92)	5.83 (1.09)	t_{60}=–0.04	.97	.62 (.38-.78)	<.001
Hedonic Motivation	5.16 (1.29)	5.10 (1.21)	t_{60}=0.46	.65	.80 (.67-.88)	<.001
Price Value	4.42 (1.14)	4.45 (1.11)	t_{60}=–0.22	.83	.62 (.36-.77)	<.001
Habit	3.27 (1.35)	3.28 (1.40)	t_{60}=–0.07	.94	.77 (.61-.86)	<.001

[a]Paired-sample *t* test.
[b]ICC: intraclass correlation coefficient.

Tests of Construct Validity

We assessed convergent validity using Pearson correlation coefficients in sample 4 (n=319). BI was related to the eHealth acceptability subscales in the expected directions, even though the effect sizes were small for EE, FC, and PV. Usage was related to HT as expected, but not with FC. PEOU, SN, and PBC were related to the eHealth acceptability subscales in the expected directions. We observed additional significant correlation coefficients between constructs. Table 5 presents the complete matrix.

Table 5. Matrix of Pearson correlations in sample 4 (n=319)[a,b].

Items	PE[c]	EE[d]	SI[e]	FC[f]	HM[g]	PV[h]	HT[i]	BI[j]	Usage	PEOU[k]	PBC[l]	SN[m]
PE	N/A[n]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
EE	*.33*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SI	*.54*	.14	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
FC	.12	*.62*	.12	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HM	*.40*	*.47*	.21	*.40*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
PV	.24	.23	.19	.28	.28	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HT	*.57*	.21	*.54*	N/A	.30	.22	N/A	N/A	N/A	N/A	N/A	N/A
BI	*.58*	.27	*.52*	.22	*.45*	.22	*.65*	N/A	N/A	N/A	N/A	N/A
Usage	*.43*	.24	.30	.12	.28	N/A	*.50*	*.46*	N/A	N/A	N/A	N/A
PEOU	.21	*.64*	.15	*.59*	*.36*	*.31*	.19	.26	.17	N/A	N/A	N/A
PBC	*.31*	*.43*	.22	*.49*	.30	.30	*.31*	*.41*	*.32*	*.51*	N/A	N/A
SN	*.47*	.13	*.71*	.13	.24	.25	*.52*	*.55*	*.36*	.24	*.31*	N/A

[a]Significant correlations between subscales (ie, >.30) [61] are shown in italics.

[b]Shows only correlations with $P<.05$.

[c]PE: Performance Expectancy.

[d]EE: Effort Expectancy.

[e]SI: Social Influence.

[f]FC: Facilitating Conditions.

[g]HM: Hedonic Motivation.

[h]PV: Price Value.

[i]HT: Habit.

[j]BI: Behavioral Intention.

[k]PEOU: Perceived Ease of Use.

[l]PBC: Perceived Behavioral Control.

[m]SN: Subjective Norms.

[n]N/A: not applicable.

Discussion

Principal Findings

This study aimed to fill a gap in the acceptability literature by providing a validated scale based on the UTAUT2 model [31] that would be suitable for eHealth contexts in French-speaking samples. The eHealth acceptability scale comprised 22 items divided into 7 subscales: PE (3 items), EE (4 items), SI (3 items), FC (3 items), HM (3 items), PV (3 items), and HT (3 items).

The dimensionality tests showed that the first-order all-factor correlated model and the bifactor confirmatory model had good fit indexes. The results confirmed the possibility of both using the subscales individually and extracting a global score of acceptability. The internal consistency evaluated by Cronbach alphas was considered satisfactory [57] and thus was confirmed. The ICCs for each subscale were above .60 and there were no significant differences in the *t* test over a 4-week period. These results demonstrated the temporal stability of the eHealth acceptability scale. Although it might seem important to attain strict factorial invariance, practical experience suggests that this is almost unachievable [55]. The partial strict factorial invariance pointed to the sex invariance in our analysis. This conclusion was one of the major findings, as it confirms that the eHealth acceptability scale can be used in male and female French-speaking samples.

Convergent validity assessments showed that subscales of the eHealth acceptability scale were significantly positively related to BI, usage, and the PEOU construct from the TAM [26], and significantly positively related to the SN and PCB constructs from the TPB [44]. These preliminary results need to be confirmed in future studies.

The strength of this scale validation was that it followed all the steps recommended by Boateng [41].

Limitations

Some limitations must nevertheless be acknowledged. One of these limitations, as in all rating scales, is the self-reported nature of the responses, which can be biased based on social desirability [63]. Another limitation is the homogeneity of the samples we used. Most of the participants were young and students. Few participants with low socioeconomic status were included, which limited generalizability. In populations that are not familiar with eHealth tools, it may be necessary to deliver specific education, notably by providing a description of the terms used. In addition, given the age distribution of our sample

(ie, centered on ages 18-34 years), we could not test the age invariance. Furthermore, the simultaneous modification of the language (ie, into French) and the context (ie, adaptation to eHealth) may have led to potential interactions and is a limitation. The study would probably have been stronger if we had validated a French-language instrument before changing the context.

Comparison With Prior Work

Compared with the English-language UTAUT2 model [31], the French eHealth acceptability scale comprised 22 items divided into 7 subscales: PE (3 items), EE (4 items), SI (3 items), FC (3 items), HM (3 items), PV (3 items), and HT (3 items), according to our analyses. We removed item FC4 for its inconsistency; the low loading was also observed to a lesser extent in the German translation [37], although not removed. The sex invariance demonstrated in our analysis was not provided in the original version [31], nor in the other translations [37,38].

Future Directions

In future studies, it will be necessary to test the constructs of the eHealth acceptability scale, which was based on the UTAUT2 model, in French samples and to estimate the explained variance in BI and usage. In addition, evaluation of age invariance will be necessary. The suggested adaptation to the eHealth context could also be replicated in other languages. Specifically, an English validation of the eHealth acceptability scale would be of interest in order to provide a common tool across French- and English-speaking samples. This scale could be used in future research to identify acceptability correlates in different contexts. It could also be used in clinical practice before implementing a new technology in health care or in the field of marketing as new technologies are developed.

Conclusions

We designed a 22-item French-language eHealth acceptability scale, divided into 7 subscales. The scale demonstrated good psychometric qualities (ie, reliability, dimensionality, validity). With this preliminary validation, the scale can be used with men and women to assess the acceptability of eHealth technology in French-speaking samples and offers promising avenues in research, clinical practice, and marketing.

Acknowledgments

The authors sincerely thank Marie Barralis, Lise Arena, Valentin Montmirail, and Dirk Steiner for their contributions to the translations; the Centre Spécialisé Obésité PACA Est, Faculty of Physical Education and Sports Studies of Toulon University; and Valentine Filleul and Raphaëlle Ladune for their valuable assistance in collecting the data. We also thank all the participants.

This study was supported by the French government, managed by the National Agency of Research as part of the UCA^{JEDI} Future Investments project, reference number ANR-15-IDEX-01, as well as a PhD grant to MH from the Région Sud Provence-Alpes Côte d'Azur, France, and cosupported by the association Azur Sport Santé.

The institutions funding this work played no role in the study design; the collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preliminary version of the eHealth acceptability scale and adapted items of the UTAUT2.
[DOCX File , 34 KB - jmir_v22i4e16520_app1.docx]

References

1. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91. [doi: 10.1177/1460458216650979] [Medline: 27245673]
2. 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]
3. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. J Behav Med 2017 Feb;40(1):112-126. [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
4. Sween J, Wallington SF, Sheppard V, Taylor T, Llanos AA, Adams-Campbell LL. The role of exergaming in improving physical activity: a review. J Phys Act Health 2014 May;11(4):864-870 [FREE Full text] [doi: 10.1123/jpah.2011-0425] [Medline: 25078529]
5. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Hum Behav 2017 May;70:317-327. [doi: 10.1016/j.chb.2017.01.013]

6. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current state of commercial wearable technology in physical activity monitoring 2015-2017. Int J Exerc Sci 2018;11(7):503-515 [FREE Full text] [Medline: 29541338]
7. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
8. Jee H. Review of researches on smartphone applications for physical activity promotion in healthy adults. J Exerc Rehabil 2017 Feb;13(1):3-11 [FREE Full text] [doi: 10.12965/jer.1732928.464] [Medline: 28349027]
9. Mollee JS, Middelweerd A, Kurvers RL, Klein MCA. What technological features are used in smartphone apps that promote physical activity? A review and content analysis. Pers Ubiquit Comput 2017 Jul 5;21(4):633-643. [doi: 10.1007/s00779-017-1023-3]
10. Hakala S, Rintala A, Immonen J, Karvanen J, Heinonen A, Sjögren T. Effectiveness of physical activity promoting technology-based distance interventions compared to usual care. Systematic review, meta-analysis and meta-regression. Eur J Phys Rehabil Med 2017 Dec;53(6):953-967. [doi: 10.23736/S1973-9087.17.04585-3] [Medline: 28466628]
11. Gell NM, Grover KW, Humble M, Sexton M, Dittus K. Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. Support Care Cancer 2017 Apr;25(4):1291-1300. [doi: 10.1007/s00520-016-3523-5] [Medline: 27957621]
12. Rabin C, Dunsiger S, Ness KK, Marcus BH. Internet-based physical activity intervention targeting young adult cancer survivors. J Adolesc Young Adult Oncol 2011 Dec;1(4):188-194 [FREE Full text] [doi: 10.1089/jayao.2011.0040] [Medline: 23610737]
13. Raaijmakers LCH, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: a systematic review. Appetite 2015 Dec;95:138-151. [doi: 10.1016/j.appet.2015.07.008] [Medline: 26165415]
14. Chaplais E, Naughton G, Thivel D, Courteix D, Greene D. Smartphone interventions for weight treatment and behavioral change in pediatric obesity: a systematic review. Telemed J E Health 2015 Oct;21(10):822-830. [doi: 10.1089/tmj.2014.0197] [Medline: 26290954]
15. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013;2(6):e000568 [FREE Full text] [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
16. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
17. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
18. 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]
19. Petit A, Cambon L. Exploratory study of the implications of research on the use of smart connected devices for prevention: a scoping review. BMC Public Health 2016 Dec 11;16:552. [doi: 10.1186/s12889-016-3225-4] [Medline: 27401769]
20. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
21. Bobillier Chaumon ME. [The situated acceptance of ICT in/for the activity: towards a clinical use]. Psychol Travail Organ 2016 Mar;22(1):4-21. [doi: 10.1016/j.pto.2016.01.001]
22. Alexandre B, Reynaud E, Osiurak F, Navarro J. Acceptance and acceptability criteria: a literature review. Cogn Technol Work 2018 Jan 24;20(2):165-177. [doi: 10.1007/s10111-018-0459-1]
23. Adell E, Varhalyi A, Nilsson L. The definition of acceptance and acceptability. In: Horberry T, Regan MA, Stevens A, editors. Driver Acceptance of New Technology: Theory, Measurement and Optimisation. London, UK: CRC Press; 2018.
24. Fishbein M, Ajzen I. Belief, Attitude, Intention, And Behavior: An Introduction To Theory And Research. Reading, MA: Addison-Wesley; 1975.
25. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.
26. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
27. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Dec;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
28. Venkatesh V, Davis FD. 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]
29. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
30. Chang A. UTAUT and UTAUT 2: a review and agenda for future research. Winners 2012 Sep 28;13(2):10. [doi: 10.21512/tw.v13i2.656]
31. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]

32. Direito A, Tooley M, Hinbarji M, Albatal R, Jiang Y, Whittaker R, et al. Tailored daily activity: an adaptive physical activity smartphone intervention. Telemed J E Health 2019 May 07. [doi: 10.1089/tmj.2019.0034] [Medline: 31063038]
33. Schneider KL, Ferrara J, Lance B, Karetas A, Druker S, Panza E, et al. Acceptability of an online health videogame to improve diet and physical activity in elementary school students: “Fitter Critters”. Games Health J 2012 Aug;1(4):262-268 [FREE Full text] [doi: 10.1089/g4h.2012.0009] [Medline: 24761317]
34. Howard MC, Jayne BS. An analysis of more than 1,400 articles, 900 scales, and 17 years of research: the state of scales in cyberpsychology, behavior, and social networking. Cyberpsychol Behav Soc Netw 2015 Mar;18(3):181-187. [doi: 10.1089/cyber.2014.0418] [Medline: 25751050]
35. Février F, Jamet E, Rouxel G. Quel outil d'évaluation de l'acceptabilité des nouvelles technologies pour des études francophones? New York, NY: ACM Press; 2008 Presented at: 20th Conference on l'Interaction Homme-Machine; September 2-5, 2008; Metz, France p. 199-204. [doi: 10.1145/1512714.1512753]
36. Brangier É, Hammes S. [How to measure the human-technology-organization relationship?]. Perspect Interdisc Travail Santé 2007;9(2):71-78. [doi: 10.4000/pistes.1466]
37. Harborth D, Pape S. German translation of the unified theory of acceptance and use of technology 2 (UTAUT2) questionnaire. SSRN J 2018 Mar 28. [doi: 10.2139/ssrn.3147708]
38. Yilmaz B, Kavanoz S. The validity and reliability of Turkish version of unified theory of acceptance and use of technology. J Turk Stud 2017 Jan 01;12(32):127-146. [doi: 10.7827/TurkishStudies.12064]
39. Macedo IM. Predicting the acceptance and use of information and communication technology by older adults: an empirical examination of the revised UTAUT2. Comput Hum Behav 2017 Oct;75:935-948. [doi: 10.1016/j.chb.2017.06.013]
40. Vallerand RJ. Toward a methodology for the transcultural validation of psychological questionnaires: implications for research in the French language. Can Psychol 1989;30(4):662-680. [doi: 10.1037/h0079856]
41. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149 [FREE Full text] [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
42. Brislin RW. The wording and translation of research instruments. In: Lonner WJ, Berry JW, editors. Field Methods in Cross-Cultural Research. Thousand Oaks, CA: Sage Publications, Inc; 1986:137-164.
43. Chyung SYY, Roberts K, Swanson I, Hankinson A. Evidence-based survey design: the use of a midpoint on the Likert scale. Perform Improv 2017 Nov 20;56(10):15-23. [doi: 10.1002/pfi.21727]
44. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
45. Gagné C, Godin G. Les Théories Sociales Cognitives: Guide pour la Mesure des Variables et le Développement de Questionnaire. Sainte-Foy, QC: Groupe de recherche sur les aspects psychosociaux de la santé, Université Laval; 1999.
46. Dong Y, Peng CJ. Principled missing data methods for researchers. Springerplus 2013 Dec;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]
47. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999 Mar;8(1):3-15. [Medline: 10347857]
48. Bennett DA. How can I deal with missing data in my study? Aust N Z J Public Health 2001 Oct;25(5):464-469. [Medline: 11688629]
49. Allison PD. Missing data techniques for structural equation modeling. J Abnorm Psychol 2003 Nov;112(4):545-557. [doi: 10.1037/0021-843X.112.4.545] [Medline: 14674868]
50. Myers ND, Martin JJ, Ntoumanis N, Celimli S, Bartholomew KJ. Exploratory bifactor analysis in sport, exercise, and performance psychology: a substantive-methodological synergy. Sport Exerc Perform Psychol 2014;3(4):258-272. [doi: 10.1037/spy0000015]
51. Byrne BM. Factor analytic models: viewing the structure of an assessment instrument from three perspectives. J Pers Assess 2005 Aug;85(1):17-32. [doi: 10.1207/s15327752jpa8501_02] [Medline: 16083381]
52. Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P. An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Sociol Methods Res 2008 Jan 01;36(4):462-494 [FREE Full text] [doi: 10.1177/0049124108314720] [Medline: 19756246]
53. Hoyle RH, Panter AT. Writing about structural equation models. In: Hoyle RH, editor. Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA: Sage Publications, Inc; 1995:158-176.
54. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equation Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
55. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. Med Care 2006 Nov;44(11 Suppl 3):S78-S94 [FREE Full text] [doi: 10.1097/01.mlr.0000245454.12228.8f] [Medline: 17060839]
56. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Res Methods 2016 Jun 29;3(1):4-70. [doi: 10.1177/109442810031002]
57. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]

58. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
59. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol 2003 Aug;56(8):730-735. [doi: 10.1016/s0895-4356(03)00084-2] [Medline: 12954464]
60. Altman DG. Practical Statistics For Medical Research (Chapman & Hall/CRC Texts In Statistical Science). London, UK: Chapman & Hall/CRC; 1990.
61. Machin D, Campbell MJ, Tan SB, Tan SH. Sample Size Tables for Clinical Studies. Third edition. Oxford, UK: Wiley-Blackwell; 2008.
62. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis: Pearson New International Edition. Harlow, UK: Pearson Education Limited; 2014.
63. Phillips DL, Clancy KJ. Some effects of "social desirability" in survey studies. Am J Sociol 1972 Mar;77(5):921-940. [doi: 10.1086/225231]

Abbreviations

BI: Behavioral Intention
CFA: confirmatory factor analysis
CFI: comparative fit index
EE: Effort Expectancy
eHealth: electronic health
FC: Facilitating Conditions
HM: Hedonic Motivation
HT: Habit
ICC: intraclass correlation coefficient
PBC: Perceived Behavioral Control
PE: Performance Expectancy
PEOU: Perceived Ease of Use
PV: Price Value
RMSEA: root mean square error of approximation
SI: Social Influence
SN: Subjective Norms
TAM: technology acceptance model
TLI: Tucker-Lewis index
TPB: theory of planned behavior
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 06.10.19; peer-reviewed by A Saptono, U Öberg; comments to author 10.11.19; revised version received 22.11.19; accepted 15.12.19; published 15.04.20.

Please cite as:
Hayotte M, Thérouanne P, Gray L, Corrion K, d'Arripe-Longueville F
The French eHealth Acceptability Scale Using the Unified Theory of Acceptance and Use of Technology 2 Model: Instrument Validation Study
J Med Internet Res 2020;22(4):e16520
URL: https://www.jmir.org/2020/4/e16520
doi:10.2196/16520
PMID:32293569

Original Paper

Exploring the Hierarchical Influence of Cognitive Functions for Alzheimer Disease: The Framingham Heart Study

Huitong Ding[1,2,3], BSc; Ning An[1,2], PhD; Rhoda Au[3,4,5,6], PhD; Sherral Devine[3,5], PhD; Sanford H Auerbach[5,6], MD; Joseph Massaro[5,7], PhD; Prajakta Joshi[3,5], MPH; Xue Liu[3,5], MPH, MSc; Yulin Liu[3,5], MSc; Elizabeth Mahon[3,5], BA; Ting FA Ang[3,4,5*], MD, MPH; Honghuang Lin[5,8*], PhD

[1]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China

[2]Key Laboratory of Knowledge Engineering with Big Data of Ministry of Education, Hefei University of Technology, Hefei, China

[3]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, United States

[4]Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

[5]The Framingham Heart Study, Framingham, MA, United States

[6]Department of Neurology, Boston University School of Medicine, Boston, MA, United States

[7]Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States

[8]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Honghuang Lin, PhD
Section of Computational Biomedicine
Department of Medicine
Boston University School of Medicine
72 E Concord St, E632
Boston, MA, 02118
United States
Phone: 1 6173580091
Email: hhlin@bu.edu

Abstract

Background: Although some neuropsychological (NP) tests are considered more central for the diagnosis of Alzheimer disease (AD), there is a lack of understanding about the interaction between different cognitive tests.

Objective: This study aimed to demonstrate a global view of hierarchical probabilistic dependencies between NP tests and the likelihood of cognitive impairment to assist physicians in recognizing AD precursors.

Methods: Our study included 2091 participants from the Framingham Heart Study. These participants had undergone a variety of NP tests, including Wechsler Memory Scale, Wechsler Adult Intelligence Scale, and Boston Naming Test. Heterogeneous cognitive Bayesian networks were developed to understand the relationship between NP tests and the cognitive status. The performance of probabilistic inference was evaluated by the 10-fold cross validation.

Results: A total of 4512 NP tests were used to build the Bayesian network for the dementia diagnosis. The network demonstrated conditional dependency between different cognitive functions that precede the development of dementia. The prediction model reached an accuracy of 82.24%, with sensitivity of 63.98% and specificity of 92.74%. This probabilistic diagnostic system can also be applied to participants that exhibit more heterogeneous profiles or with missing responses for some NP tests.

Conclusions: We developed a probabilistic dependency network for AD diagnosis from 11 NP tests. Our study revealed important psychological functional segregations and precursor evidence of AD development and heterogeneity.

(J Med Internet Res 2020;22(4):e15376) doi:10.2196/15376

KEYWORDS

Alzheimer disease; neuropsychological test; stratification; Bayesian network; clustering

Introduction

Background

Alzheimer disease (AD) is a chronic neurodegenerative disease characterized by cognitive decline [1]. Neuropsychological (NP) tests—a key measure of phenotypic expression of one's cognition state—are commonly used by practitioners to assess cognitive dysfunction, especially in the memory, attention, and executive domains [2,3]. However, given the extensive variability in performance patterns across a standard comprehensive protocol of NP tests, physicians often find themselves making clinical decisions with certain degrees of uncertainty, and the situation is compounded when patients are unable to complete the tests because of a multitude of reasons. Given the data heterogeneity within and across NP tests, conventional qualitative classification is unable to accurately portray the clinical manifestation of a spectrum disorder such as AD. To date, many studies examined various cognitive domains individually [4], as separate entities, when in fact different regions of the brain work simultaneously and not in silos [5]. Therefore, to better characterize the complexity of AD, we need to identify and describe the hierarchical interaction pattern among NP tests and their symbiotic relationship with each other, to help illuminate the indices of neurodegenerative processes. Some researchers have proposed to focus on AD precursors of cognitive decline to reduce AD clinical trial failures [1]. We contend that the relationship is bidirectional. Different patterns of symptoms are indicative of cognitive impairment, whereas the presence of cognitive impairment impacts the symptoms associated with subsequent risk. Furthermore, to enhance clinical utility, a full global use of available observations will aid physicians with AD diagnosis, particularly for those patients who exhibit more heterogeneous NP profiles.

Many risk factors of AD have been identified in past decades [6,7]. Apolipoprotein E4 (ApoE4) status has been demonstrated to be a significant genetic risk factor for AD [8]. Although the factors underlying the sex differences have generally been weakly investigated, the difference indeed exhibits influence on the development and progression of AD [9]. ApoE4 tended to have different effects on AD between men and women [10]. In addition, education has been recognized as another potential risk factor, where people with different levels of education tended to show different risks of AD [11]. However, the hierarchical interplay between different risk factors and their effects on cognitive status are yet to be investigated.

Objective

The objective of this study was to represent the intricate interplay of various NP tests with probabilistic graphical models and provide a top-down theoretical view to demonstrate the relationship between NP tests and cognitive status.

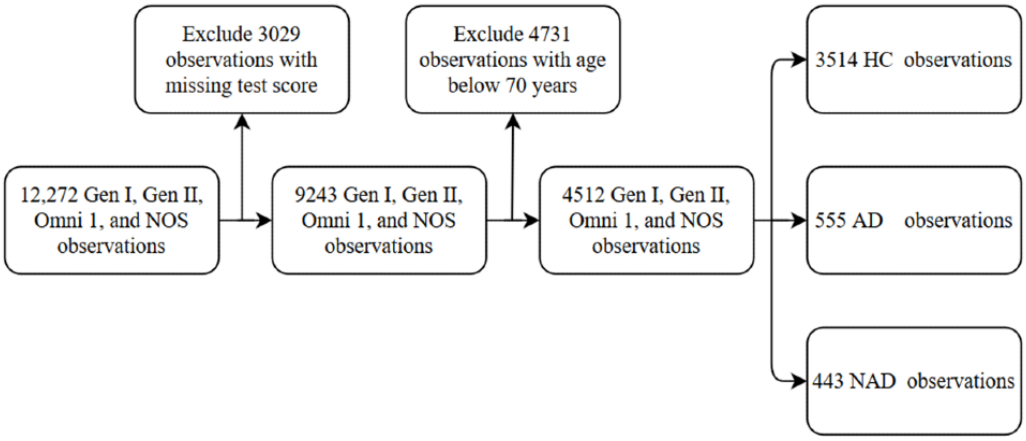
Methods

Study Population

The Framingham Heart Study (FHS) is a community-based longitudinal observational study that began in 1948. Details of FHS cohorts have been previously described [12]. Briefly, three generations of participants have been enrolled since 1948. To reflect the increasing ethnic diversity in Framingham, two additional cohorts, Omni Study 1 and Omni Study, were enrolled in 1994 and 2002, respectively. Every 2 to 8 years, each participant is given a comprehensive physical examination and queried for various lifestyles. NP tests have been administered through ancillary studies using standardized testing protocols and scoring procedures since 1981 [13]. Routine quality assurance processes were performed to keep consistency of these tests over time [14]. This study included all participants with valid NP tests from the original cohort (Gen I), offspring cohort (Gen II), multiethnic Omni 1 cohort, and new offspring spouse cohort [15]. Given the fact that AD primarily affects participants of advanced age, and the average age of dementia onset among FHS participants is around 85 years, our study was restricted to participants who were 70 years or older [16,17].

The dementia diagnosis was based on the Diagnostic and Statistical Manual of Mental Disorders, fourth edition, whereas AD diagnosis was based on the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer disease and Related Disorders Association [18]. All dementia diagnoses were adjudicated by an expert panel consisting of at least one neurologist and one neuropsychologist, using information from various sources such as NP assessments, neurology examinations, family interviews, FHS health exams, and external medical records [19]. According to their cognitive status, participants were grouped as healthy control (HC), AD, and non-Alzheimer dementia (NAD). Details of the dementia surveillance have been published [20-22]. The process of sample selection is shown in Figure 1.

Figure 1. The process of sample selection. AD: Alzheimer disease; HC: healthy control; NAD: non-Alzheimer dementia; NOS: new offspring spouse.



Identification of Cognitive Function Clusters

We performed correlation analysis to explore the dependency between NP tests. A cognitive function cluster is a set of NP tests that have stronger correlation than those outside of the cluster. The correlation between NP tests was assessed by Pearson chi-square test. Given the limited number of NP tests in our study, we used correlation coefficient 0.6 as the correlation cut-off, similar to previous studies [23,24]. Correlations between NP tests and cognitive status were evaluated by one-way analysis of variance [25].

Bayesian Network for Modeling Hierarchical Probabilistic Dependencies

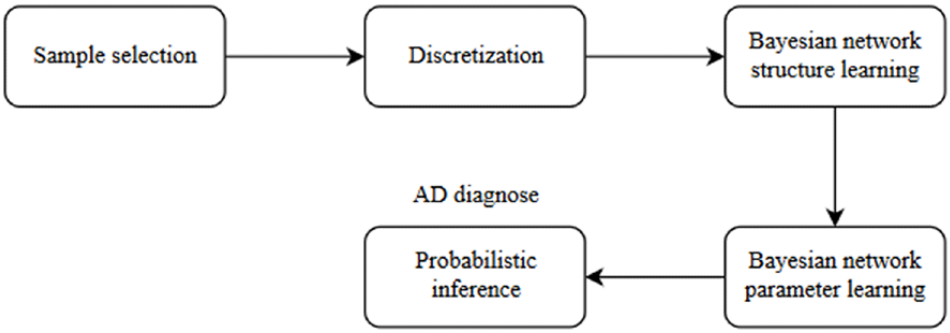
Bayesian network is a representative probabilistic machine learning method [26], which explores the information contained in experimental data to evaluate the probability of specific hypotheses. It can summarize a complex system into a simplified representation to capture the hierarchical interplay among components and provide insights on how each component influences others [27]. A Bayesian network is represented by a directed acyclic graph composed of nodes and edges. In this study, we used nodes to represent NP tests and cognitive status, and edges to represent the influence between nodes. For example, Test A–>Cognitive Status means that Test A is the parent node of Cognitive Status, and Cognitive Status is the child node of Test A. The edge direction suggests that Test A has an influence on Cognitive Status, which is formulated as the conditional probability of how the Cognitive Status depends on Test A. In contrast, given that Test A has no incoming edges, its probability does not depend on other factors. These dependency relationships could propagate through the network and influence downstream tests. It is worth to note that the conditional probability depends only on parent nodes but not grandparent nodes. For example, if we also observe Cognitive Status–>Test B, it means that Test B is only directly dependent on Cognitive Status but not grandparent node Test A, although Cognitive Status is dependent on Test A.

Figure 2 shows the flowchart of building a Bayesian network. Each observation includes 11 NP test scores and the cognitive status. Each continuous variable was discretized by partitioning around medoids method, which was used to find the intrinsic structures in NP tests and assign observations into homogeneous clusters [28,29]. The optimal number of clusters was determined by the silhouette width [30].

A search-and-score strategy was then used to build Bayesian networks from NP tests. The algorithm first assigned a likelihood score to each candidate structure. The score represented how well that structure fits the NP tests, which was evaluated by the Bayesian Information Criterion [31]. Unnecessary complex structures could fit existing data well but lack the generalizability to new data. Therefore, to recover the underlying Bayesian network structures, we included a penalty term equal to the Minimum Description Length score [32]. The method was previously shown to outperform other scoring functions such as Bayesian Dirichlet equivalence score, Akaike information criterion, and factorized normalized maximum likelihood [31,33]. Two searching methods were then used to find the optimal structure. One was Heuristic Hill-Climbing greedy search, which aims to optimize the local score [34] but cannot apply any prior knowledge about the expected structure of Bayesian network. The other one was Tabu search, which was used for validation and could search the space of directed graphs while escaping local optimum [35]. Bootstrap was adopted to minimize the uncertainty of the model [36].

In Bayesian network parameter learning, two parametric estimation methods were used, including maximum likelihood estimation and Bayesian parameter estimation [37]. To further validate the Bayesian network, logic sampling method [38] was used to generate simulated data based on learned Bayesian network and check whether it was consistent with prior information about the correlation of NP tests. Details of the learning process are provided in Multimedia Appendix 1, Methods.

Figure 2. The flowchart of building a Bayesian network. AD: Alzheimer disease.



Markov Blanket to Select Neuropsychological Tests

Following the principle of the filter-based feature selection, an optimal subset of NP tests was derived from the data itself but not the performance matrix [39]. One key feature is conditional independence, which defines a sufficient subset S as follows [40]: S∈G is a sufficient subset of NP tests if and only if P(Status|G)=P(Status|S), where G is a set consisting of all 11 NP tests. Cognitive status is conditionally independent of other NP tests given S. The set of locally affecting variables is called the Markov blanket [41]. In Bayesian networks, the Markov blanket of cognitive status is a set of NP tests that consists of parent nodes, child nodes, and spouse nodes of cognitive status [41]. NP test A is the spouse of NP test B because they have common children. NP tests in Markov blanket are directly connected to cognitive status, which, therefore, determined the probability distribution of cognitive status. It provides a direction for detecting the potential causal cognitive functions for AD [42,43]. Pearson chi-square test was used to demonstrate the conditional independence between NP tests and cognitive status [44].

Probabilistic Inference

Once Bayesian networks were built, each participant's cognitive status was derived using an averaging likelihood weighting simulation method, which is an approximating inference method [34]. It calculated the posterior probabilities of cognitive status from observed NP tests. The details are provided in Multimedia Appendix 1, Methods.

To demonstrate heterogeneity of hierarchical influence of cognitive functions and cognitive status, the analysis was conducted for the full observations and also stratified by sex (male or female), ApoE4 status (OMIM 107741), and education level (beyond high school/high school graduate and below). Participants with missing education information were excluded from the education-stratified analyses (11 observations). Similarly, for the ApoE-stratified analyses, participants who did not consent to genetic analyses or without ApoE4 information were excluded (200 observations).

All participants had provided written informed consent. This study was approved by the Institutional Review Board of Boston University Medical Campus. The data collection was monitored by a National Heart, Lung, and Blood Institute Observational Study Monitoring Board and complied with the Strengthening the Reporting of Observational Studies in Epidemiology reporting guideline [45].

Results

Sample Characteristics

Our study included 4512 sets of NP tests from 2091 participants, primarily of European ancestry (1166 females, mean age 79 [SD 6] years). On average, each participant underwent 2.2 NP examinations. One examination of the participant is considered as a study sample. Table 1 shows the clinical characteristics of study samples.

Although 32 NP tests have been administered at the FHS (Multimedia Appendix 1), this study was focused on 11 NP tests that were administered to more than 85% of participants between 1999 and 2016. These tests included the first version of Wechsler Memory Scale Logical Memory Immediate Recall (LMi) [46], Logical Memory Delayed Recall (LMd) [46], and Logical Memory Recognition (LMr) [46]; Visual Reproductions Immediate Recall (VRi) [46], Visual Reproductions Delayed Recall (VRd) [46], and Visual Reproductions Recognition (VRr) [46]; and Paired Associate Learning Immediate Recall (PASi) [46], and the first version of Wechsler Adult Intelligence Scale similarities test (SIM) [47]. Given the importance in the measurement of confrontational word retrieval and verbal memory, our study also included Boston Naming Test 30 item Even Version (BNT30) and hard-pair scores from PASi and PASd [46,48].

Table 1. Clinical characteristics of study samples. A total of 4512 sets of neuropsychological tests from 2091 participants were included.

Characteristics	Healthy control (n=3514)	Alzheimer disease (n=555)	Non-Alzheimer dementia (n=443)
Age at neuropsychological exam (years)			
Mean (SD)	79 (6)	85 (6)	84 (6)
Range	70-101	70-103	70-97
Male, n (%)	1521 (43.3)	179 (32.3)	220 (49.7)
Highest level of education attained[a]			
High school and below, n (%)[b]	1491 (42.5)	358 (65.2)	241 (54.5)
Beyond high school, n (%)[b]	2019 (57.5)	191 (34.8)	201 (45.5)
ApoE4[c] **allele**			
ApoE4(−), n (%)[b]	2794 (82.9)	346 (65.3)	327 (79.2)
ApoE4(+), n (%)[b]	575 (17.1)	184 (34.7)	86 (20.8)
Neuropsychological test scores, mean (SD)			
Verbal memory			
Logical Memory Immediate Recall	11.2 (3.7)	4.8 (3.8)	7.9 (3.9)
Logical Memory Delayed Recall	10.2 (3.9)	3.0 (4.0)	6.5 (4.1)
Logical Memory Recognition	9.4 (1.4)	7.1 (2.3)	8.5 (1.7)
Visual memory			
Visual Reproductions Immediate Recall	7.1 (3.0)	3.1 (2.3)	4.0 (2.5)
Visual Reproductions Delayed Recall	6.1 (3.1)	1.6 (1.9)	2.7 (2.4)
Visual Reproductions Recognition	2.6 (1.1)	1.3 (1.1)	1.7 (1.1)
New learning			
Paired Associate Learning Immediate Recall	12.8 (3.3)	8.4 (2.9)	9.9 (2.8)
Hard score of Paired Associate Learning Delayed Recall	2.0 (1.3)	0.5 (0.9)	1.0 (1.1)
Hard Score of Paired Associate Learning Immediate Recall	4.4 (3.0)	1.1 (1.7)	2.0 (2.0)
Abstract reasoning			
Similarities Test	15.5 (3.9)	9.8 (5.0)	11.6 (4.7)
Language and naming			
Boston Naming Test, 30-item Even Version	26.1 (3.4)	19.4 (5.9)	22.3 (5.4)

[a]Valid education data (n): Healthy control (3510); Alzheimer disease (549); Non-Alzheimer dementia (442).

[b]Values were calculated based on the subset with valid data.

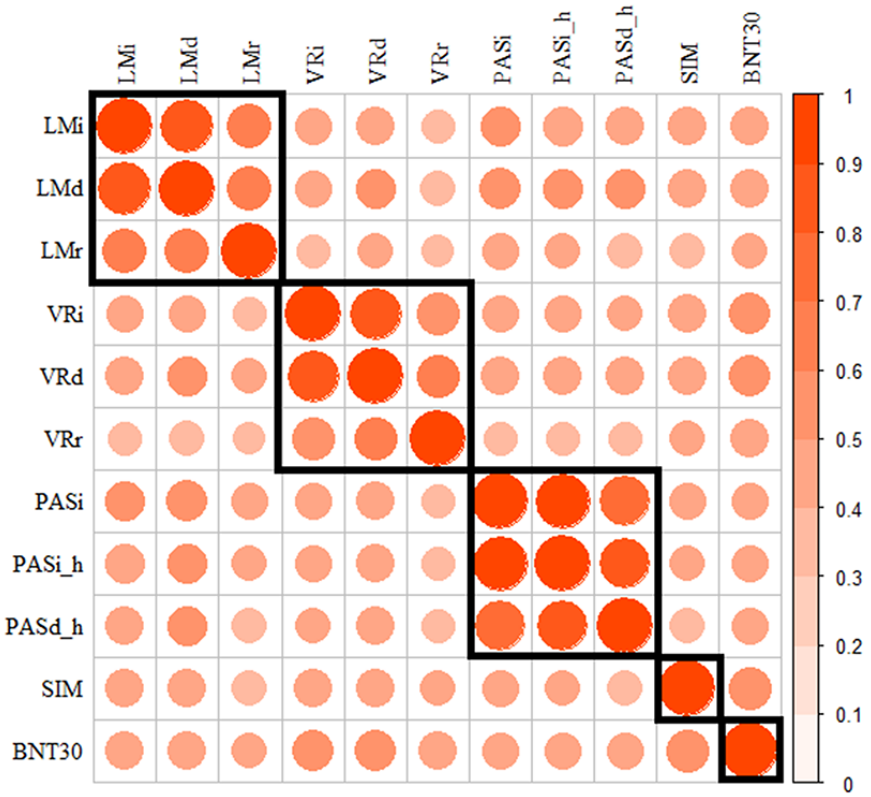
[c]ApoE4: Apolipoprotein E4. Participants who did not consent to genetic analyses or with no ApoE4 information were excluded. Valid genetic data (n): Healthy control (3369); Alzheimer disease (530); Non-Alzheimer dementia (413).

Correlation Clusters of Neuropsychological Tests

We performed unsupervised clustering to investigate the correlation between NP tests. As shown in Figure 3, these NP tests could be divided into five clusters, each representing a distinct cognitive function. The intracluster correlation coefficients between NP test pairs were all higher than 0.60, which formed a clear cluster boundary to distinguish different cognitive functions without overlapping. In contrast, the intercluster correlation coefficients between NP test pairs were mostly lower than 0.50. The correlation of NP tests in subpopulations is shown in Multimedia Appendix 1.

As expected, three cognitive outcome groups had quite different mean NP test scores (Tukey-Kramer test, $P<.001$), suggesting a strong correlation between NP test and cognitive status (Multimedia Appendix 1). The association remained significant after Bonferroni correction for multiple testing.

Figure 3. Correlation clusters between different neuropsychological tests. The red rectangles represent different clusters of tests. Bigger and redder nodes represent higher correlation, whereas whiter and smaller nodes represent lower correlation. BNT30: Boston Naming Test 30 item Even Version; LMd: Logical Memory Delayed Recall; LMi: Logical Memory Immediate Recall; LMr: Logical Memory Recognition; PASd_h: Hard score of Paired Associate Learning Delayed Recall; PASi: Paired Associate Learning Immediate Recall; PASi_h: Hard Score of Paired Associate Learning Immediate Recall; SIM: similarities test; VRd: Visual Reproductions Delayed Recall; VRi: Visual Reproductions Immediate Recall; VRr: Visual Reproductions Recognition.



Bayesian Networks

Figure 4 shows the Bayesian network of hierarchical influence between NP tests and cognitive status. The network consists of nodes and edges, which represent the conditional dependence between NP tests and cognitive status. The parent node has an influence on the predictability of child nodes [49]. The first precursor of dementia is SIM, following the sequence of SIM–>BNT30–>VRi–>VRd, and eventually leading to dementia. On the other hand, the cognitive status could also influence visual memory indirectly via logical memory. Figure 5 shows Bayesian networks in subpopulations stratified by sex, ApoE, and education. For males, LMd (eg, verbal memory) directly influences the cognitive status, which then influences other NP tests. In other words, changes in verbal memory function are a precursor of dementia, which is consistent with the focus on memory as the key cognitive symptom of dementia [50]. The cognitive functions of visuospatial processing, visual memory, language, and verbal reasoning could also influence cognitive status. For females, the first precursor of dementia is SIM, following the sequence of SIM–>BNT30–>VRi–>VRd until the influencing cognitive status. Cognitive status also influences verbal memory and visual memory. Cognitive status influenced PASi indirectly via LMi. For participants carrying ApoE4 alleles, VRd influences cognitive status, followed by other NP tests. For participants without ApoE4 alleles, the relationship of NP tests and cognitive status is similar to the relationship among female only participants. For participants with low degrees of education, LMd influences cognitive status, which would then influence other NP tests. For participants with advanced degrees of education, VRd influences cognitive status and then LMi indirectly via PASi. VRr influences BNT30, which then influences SIM. It is worth noting that similar precursors of dementia were observed in females and participants without ApoE4 alleles. Gender-stratified ApoE4+ models can be found in Multimedia Appendix 1.

Taking the Bayesian network of females as an example (Figure 5), information is transmitted in a sequence of SIM–>BNT30–>VRi–>VRd–>Cognitive Status. If we have the VRd score, cognitive status becomes independent of SIM, BNT30, and VRi. Similarly, the block also exists in a diverging connection (LMi<–Cognitive Status–>PASi). Two child nodes, LMi and PASi, are related to each other by the Cognitive Status. However, if the participant is diagnosed with AD, LMi and PASi become conditionally independent. In other words, the decline of verbal memory and visual memory function does not influence each other. In the correlation network, all NP tests are related to cognitive status, but the hierarchical influence among NP tests and cognitive function cannot be distinguished. Some NP tests' predictability of cognitive status is influenced by other NP tests.

The Markov blanket of cognitive status is the parent and child of the cognitive status node in Figure 4. Seven NP tests were included into Markov blankets of cognitive status for all Bayesian networks (Figure 5). The most frequent ones were

LMi and VRd, which were parents of the cognitive status node in 4 and 5 Bayesian networks, respectively. The NP tests in Markov blankets have a direct first-level influence on cognitive status. As shown in Table 2, the degree of association between NP tests and cognitive status is declining when conditioned on the subset of Markov blanket. This reveals that the relationship between specific NP test and cognitive status is influenced by other tests. It provides a global and hierarchical view to understanding the relationship between NP tests and cognitive status. Future functional analyses can determine the specific role of these cognitive functions in AD pathogenesis.

The scores of NP tests were parameterized using probability tables as shown in Multimedia Appendix 1. The marginal probability and the probability dependency between NP tests were determined using the maximum likelihood estimation. Once the network was constructed and the probability was specified, Bayes theorem was used to propagate probability through the network to infer cognitive status. The performance of the model was evaluated using 10-fold cross validation. As shown in Multimedia Appendix 1, the overall accuracy was 82.2%, with sensitivity of 64.0% and specificity of 92.7%. The models stratified by sex, ApoE status, and education have similar performances.

Figure 4. Bayesian network shows the hierarchical influence between neuropsychological tests and cognitive status. BNT30: Boston Naming Test 30 item Even Version; LMd: Logical Memory Delayed Recall; LMi: Logical Memory Immediate Recall; LMr: Logical Memory Recognition; PASd_h: Hard score of Paired Associate Learning Delayed Recall; PASi: Paired Associate Learning Immediate Recall; PASi_h: Hard Score of Paired Associate Learning Immediate Recall; SIM: similarities test; VRd: Visual Reproductions Delayed Recall; VRi: Visual Reproductions Immediate Recall; VRr: Visual Reproductions Recognition. Nodes with the same color represent NP tests measuring the same cognitive function.

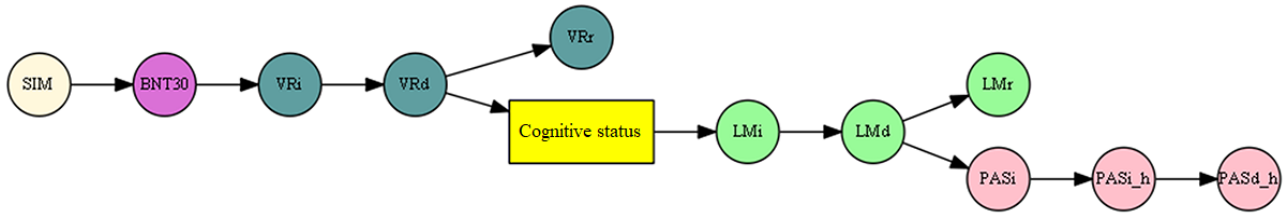


Figure 5. Hierarchical influence of neuropsychological tests and cognitive status in subpopulations. BNT30: Boston Naming Test 30 item Even Version; LMd: Logical Memory Delayed Recall; LMi: Logical Memory Immediate Recall; LMr: Logical Memory Recognition; PASd_h: Hard score of Paired Associate Learning Delayed Recall; PASi: Paired Associate Learning Immediate Recall; PASi_h: Hard Score of Paired Associate Learning Immediate Recall; SIM: similarities test; VRd: Visual Reproductions Delayed Recall; VRi: Visual Reproductions Immediate Recall; VRr: Visual Reproductions Recognition. Nodes with the same color represent NP tests measuring the same cognitive function.

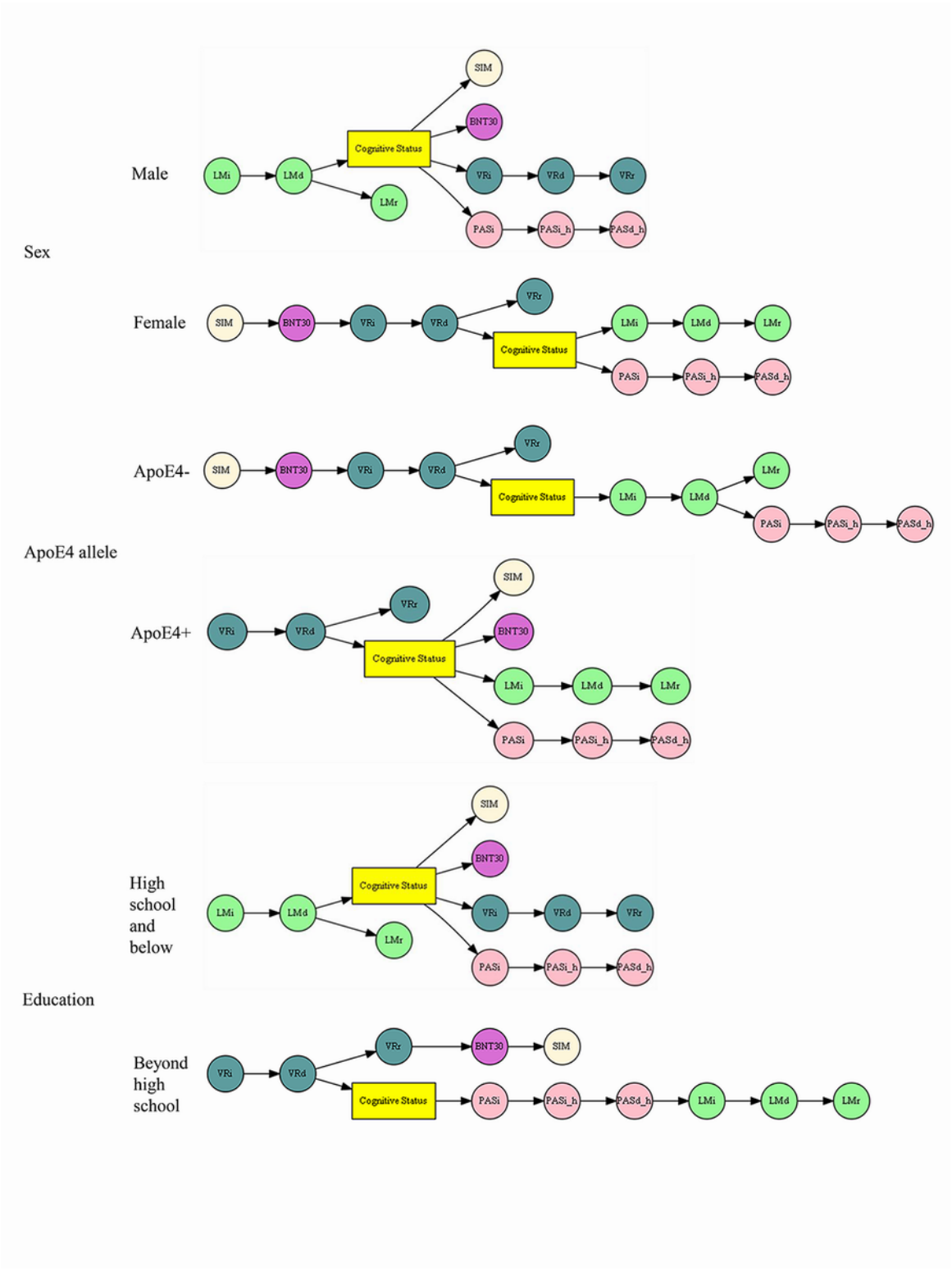


Table 2. The *P* values of association between neuropsychological (NP) tests and cognitive status when conditioning on NP tests in Markov blanket. The significant degree was computed using Pearson chi-square test with adjusted degrees of freedom.

Neuropsychological tests	ϕ[a], *P* value	VRd[b], *P* value	LMi[c], *P* value	VRd and LMi, *P* value
Logical Memory Delayed Recall	<.001	<.001	<.001	>.99
Logical Memory Recognition	<.001	<.001	.62	>.99
Visual Reproductions Immediate Recall	<.001	.18	<.001	>.99
Visual Reproductions Recognition	<.001	<.001	<.001	>.99
Boston Naming Test 30 item Even Version	<.001	<.001	<.001	>.99
Similarities test	<.001	<.001	<.001	>.99
Paired Associate Learning Immediate Recall	<.001	<.001	<.001	>.99
Hard Score of Paired Associate Learning Immediate Recall	<.001	<.001	<.001	>.99
Hard score of Paired Associate Learning Delayed Recall	<.001	<.001	<.001	>.99

[a]ϕ represents empty condition set.

[b]VRd: Visual Reproductions Delayed Recall.

[c]LMi: Logical Memory Immediate Recall.

Alzheimer Disease Probabilistic Inference

We used Bayesian networks of total population to illustrate AD inferences, in which test scores were modeled as discrete intervals via clustering. Cognitive status was modeled by HC, AD, and NAD. SIM had no parent nodes, and their probabilities did not depend on other tests. Therefore, the probability for SIM was characterized by its marginal probabilities. For the tests influenced by others, conditional probabilities were used to reflect that relationship. The probabilities of a test can only be expressed by its immediate parent node. Although SIM influences cognitive status, only VRd was used in cognitive status's conditional probability table (CPT). On the basis of the CPT, the prior probability of each NP test can be calculated. Influence between tests can also back-propagate along an edge. Using the Bayes theorem, we can compute the posteriors to infer cognitive status when NP tests are only partially available. Missing data from NP tests occur for various reasons, some of which are independent of cognitive status. This approach allows us to use the available information to make the best probabilistic inference, which does not depend on the same review sequence of NP tests. To reduce sampling effects, the inference process was repeated 100 times, and the mean of inference probability was used as the final clinical decision (Multimedia Appendix 1). To further validate whether the network structure can reflect the true relationship between NP tests and AD diagnosis, we simulated 3000 participants with 11 NP tests and correlated them with the cognitive status. As shown in Multimedia Appendix 1, all the relationships of simulated NP tests data were highly consistent with those derived from the original data. The discretized intervals derived from different populations are shown in Table 3 and Multimedia Appendix 1.

Table 3. Discretization of neuropsychological test scores.

Neuropsychological test	Score interval[a]
Logical Memory Immediate Recall	(0,3), (4,6), (7,8), (9,9), (10,10), (11,11), (12,12), (13,13), (14,16), (17,23)
Logical Memory Delayed Recall	(0,0), (1,3), (4,6), (7,8), (9,9), (10,10), (11,11), (12,12), (13,15), (16,24)
Logical Memory Recognition	(0,1), (2,3), (4,4), (5,5), (6,6), (7,7), (8,8), (9,9), (10,10), (11,11)
Visual Reproductions Immediate Recall	(0,2), (3,3), (4,4), (5,5), (6,6), (7,7), (8,8), (9,9), (10,10), (11,14)
Visual Reproductions Delayed Recall	(0,0), (1,1), (2,2), (3,3), (4,4), (5,5), (6,6), (7,7), (8,9), (10,14)
Visual Reproductions Recognition	(0,0), (1,1), (2,2), (3,3), (4,4)
Paired Associate Learning Immediate Recall	(0,7), (7.5,8.5), (9,9.5), (10,11), (11.5,12.5), (13,13.5), (14,14), (14.5,15.5), (16,17.5), (18,21)
Hard Score of Paired Associate Learning Immediate Recall	(0,0), (1,1), (2,2), (3,3), (4,4), (5,5), (6,6), (7,7), (8,8), (9,12)
Hard score of Paired Associate Learning Delayed Recall	(0,0), (1,1), (2,2), (3,3), (4,4)
Similarities Test	(0,6), (7,10), (11,12), (13,13), (14,14), (15,15), (16,16), (17,17), (18,19), (20,26)
Boston Naming Test 30 Item Even Version	(0,12), (13,17), (18,20), (21,23), (24,25), (26,26), (27,27), (28,28), (29,29), (30,30)

[a]Parentheses are used to indicate closed intervals, in case the square brackets are mistaken for references.

Discussion

Principal Findings

Various analytic models have been used to identify informative NP tests for AD prediction [4]. However, it is challenging to understand the hierarchical influence between NP tests. Our study explores the hierarchical probabilistic dependency of 11 NP tests and adjudicated cognitive outcomes—HC, AD, and NAD—based on the participants' cognition at the time of their NP tests. These tests were incorporated into our Bayesian network to establish a probabilistic-based framework for cognitive outcomes. Within this theoretical hierarchical influence structure, upstream NP tests affect downstream NP tests, allowing us to probabilistically inference downstream NP tests. We observed interdependence between NP tests and overall cognition of an individual, where higher-level cognitive functions—represented by individual NP tests that preceded cognitive outcome in the hierarchy—are considered as precursors of dementia, whereas the lower-level cognitive functions are impacted as a result of dementia.

Machine learning methods have been used to predict cognitive decline [51,52]. The identification of preclinical patterns of cognitive decline is essential for the prevention and early treatment of AD [3,53,54]. NP tests have been used for AD diagnosis for a long time. Not surprisingly, all NP tests were found to be related to AD. However, the hierarchical relationship between NP tests was not previously fully investigated. Our data-driven findings depict the interplay of cognitive functions and further identify tests that influence, or are influenced by, cognitive status. Current use of cognitive tests largely focuses on declines that are in later disease states and thus cannot be used to detect preclinical disease states. We advocate that the decline of some cognitive functions is only the later manifestations of AD, not the precursor (ie, pre-AD performance metric). From a preventive medicine perspective, tests that are identified as precursors can be taken as reference for physicians to be considered as potential targets for AD intervention and perhaps prevention. The value of using patterns of NP tests is that they can be used as early screening tools to identify at-risk patients and provide interventions, including nonpharmacological therapies, which may delay or perhaps stop disease progression altogether.

Our study represents a significant step forward in how to better characterize preclinical AD heterogeneity. As cognitive functions tend to interrelate in complex ways, we explored conditional probability dependencies to present a main relationship between cognitive functions. For males, decline of memory function influences cognitive status, which also influences the functions of language, verbal reasoning, and visuospatial simultaneously. In comparison, for females, the relationships are more complex; verbal reasoning function influences language function, which influences visuospatial function and, subsequently, cognitive status. Cognitive status in females influences the function of logical and visual memory. It is noteworthy that although the interplay of cognitive functions has a different sequence for males vs females, logical memory and visual reproduction functions are typically the first precursors in AD. Our results suggest that gender-specific evaluations need to be considered by clinicians in AD clinics similar to other diseases such as heart attacks [55].

We also quantified functional connectivity through statistical correlations and coherence [56]. Functional connectivity is defined as a function of probability distributions over observed multivariate responses. The hierarchical influence among cognitive functions under various AD risk factors was assessed to understand psychological functional segregation for heterogeneous AD beyond just those associated with sex differences.

Our study has several advantages. In terms of method, the proposed structure of Bayesian network combined with inference offers several advantages in the inference of disease status. First, it provides a likelihood of the diagnosis, which is more intuitive and meaningful in a clinical setting [57,58]. In addition, we were able to impute the missing data by the probabilistic inference strategy, which was based on the network structure and training samples to capture the global assessment between NP tests. The long follow-up period, beginning in 1976, and the minimal loss to follow-up at FHS make it an ideal study population for AD research [59]. The AD diagnosis was adjudicated and verified by a panel consisting of at least one neurologist and one neuropsychologist; hence, outcome bias was minimized [3]. The NP tests within the FHS NP battery are well known and widely administered by many clinicians and researchers. Given these strengths, the results of this study can be readily translated into real-world application.

We also acknowledge several limitations of our study. The study participants are primarily of European ancestry, they have higher levels of education than the general population in the United States, and the majority of NP tests were carried out in English. In addition, the average age of dementia onset among FHS participants is around 85 years, which is higher than the expected average. Therefore, findings of our study might not be generalizable to other populations, such as these with lower educational attainment, other ancestries, or non-English–speaking groups. Moreover, we did not further distinguish participants with mild cognitive impairment from HCs. The models presented in this study were solely based on 11 NP tests from five categories of cognitive function. Introduction of other NP tests might reveal additional interactions between cognitive functions and help to strengthen the overall model.

Conclusions

We developed a probabilistic dependency network for AD diagnosis from 11 NP tests. Our study revealed important psychological functional segregations and precursor evidence of AD development. Future validations with additional samples and NP tests would provide a more comprehensive picture of cognitive function and identify potential NP biomarkers for AD surveillance.

Acknowledgments

The authors would like to thank the FHS study staff for their many years of hard work in the examination of subjects and acquisition of data. This work was supported by the National Heart, Lung, and Blood Institute contract (N01-HC-25195) and by grants from the National Institute on Aging AG-008122, AG-16495, AG-062109, AG049810, AG054156, and from the National Institute of Neurological Disorders and Stroke, NS017950, and from the National "111 Project", No. B14025. It was also supported by Pfizer, Inc; the Boston University Digital Health Initiative; Boston University Alzheimer Disease Center Pilot Grant; and the National Center for Advancing Translational Sciences, National Institutes of Health, through Boston University Clinical & Translational Science Institute Grant Number 1UL1TR001430. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institutes of Health or the US Department of Health and Human Services.

Conflicts of Interest

RA has received grant funding support from Evidation Health and Biogen. She has been on the scientific advisory board of Optum Labs and serves on the scientific advisory board of Signant Health and Cognition Therapeutics; none of which have any conflict of interest with the contents of this project.

Multimedia Appendix 1
Details of methods and supplement results.
[PDF File (Adobe PDF File), 716 KB - jmir_v22i4e15376_app1.pdf]

References

1. Au R, Piers RJ, Lancashire L. Back to the future: Alzheimer's disease heterogeneity revisited. Alzheimers Dement (Amst) 2015 Sep;1(3):368-370 [FREE Full text] [doi: 10.1016/j.dadm.2015.05.006] [Medline: 27077132]
2. Elias MF, D'Agostino RB, Elias PK, Wolf PA. Neuropsychological test performance, cognitive functioning, blood pressure, and age: the Framingham Heart Study. Exp Aging Res 1995;21(4):369-391. [doi: 10.1080/03610739508253991] [Medline: 8595803]
3. Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. Arch Neurol 2000 Jun;57(6):808-813. [doi: 10.1001/archneur.57.6.808] [Medline: 10867777]
4. Ang TF, An N, Ding H, Devine S, Auerbach SH, Massaro J, et al. Using data science to diagnose and characterize heterogeneity of Alzheimer's disease. Alzheimers Dement (N Y) 2019;5:264-271 [FREE Full text] [doi: 10.1016/j.trci.2019.05.002] [Medline: 31304232]
5. Riddle D, Lichtenwalner R. Neurogenesis in the adult and aging brain. In: Riddle DR, editor. Brain Aging: Models, Methods, and Mechanisms. Boca Raton, Florida: CRC Press; 2007.
6. Pike CJ. Sex and the development of Alzheimer's disease. J Neurosci Res 2017 Jan 2;95(1-2):671-680 [FREE Full text] [doi: 10.1002/jnr.23827] [Medline: 27870425]
7. Bangen KJ, Beiser A, Delano-Wood L, Nation DA, Lamar M, Libon DJ, et al. APOE genotype modifies the relationship between midlife vascular risk factors and later cognitive decline. J Stroke Cerebrovasc Dis 2013 Nov;22(8):1361-1369 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2013.03.013] [Medline: 23601373]
8. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. J Am Med Assoc 1997;278(16):1349-1356. [doi: 10.1001/jama.1997.03550160069041] [Medline: 9343467]
9. Cahill L. Why sex matters for neuroscience. Nat Rev Neurosci 2006 Jun;7(6):477-484. [doi: 10.1038/nrn1909] [Medline: 16688123]
10. Altmann A, Tian L, Henderson VW, Greicius MD, Alzheimer's Disease Neuroimaging Initiative Investigators. Sex modifies the APOE-related risk of developing Alzheimer disease. Ann Neurol 2014 Apr;75(4):563-573 [FREE Full text] [doi: 10.1002/ana.24135] [Medline: 24623176]
11. Demir EY, Özcan T. Evaluating the relationship between education level and cognitive impairment with the Montreal Cognitive Assessment Test. Psychogeriatrics 2015 Sep;15(3):186-190 [FREE Full text] [doi: 10.1111/psyg.12093] [Medline: 25560276]
12. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 1979 Sep;110(3):281-290. [doi: 10.1093/oxfordjournals.aje.a112813] [Medline: 474565]
13. Au R, Seshadri S, Wolf PA, Elias MF, Elias PK, Sullivan L, et al. New norms for a new generation: cognitive performance in the framingham offspring cohort. Exp Aging Res 2004;30(4):333-358. [doi: 10.1080/03610730490484380] [Medline: 15371099]

14. Hankee LD, Preis SR, Beiser AS, Devine SA, Liu Y, Seshadri S, et al. Qualitative neuropsychological measures: normative data on executive functioning tests from the Framingham offspring study. Exp Aging Res 2013;39(5):515-535 [FREE Full text] [doi: 10.1080/0361073X.2013.839029] [Medline: 24151914]
15. Dawber TR, Meadors GF, Moore Jr FE. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 1951 Mar;41(3):279-286. [doi: 10.2105/ajph.41.3.279] [Medline: 14819398]
16. Gustafson D, Rothenberg E, Blennow K, Steen B, Skoog I. An 18-year follow-up of overweight and risk of Alzheimer disease. Arch Intern Med 2003 Jul 14;163(13):1524-1528. [doi: 10.1001/archinte.163.13.1524] [Medline: 12860573]
17. Skoog I, Lernfelt B, Landahl S, Palmertz B, Andreasson LA, Nilsson L, et al. 15-year longitudinal study of blood pressure and dementia. Lancet 1996 Apr 27;347(9009):1141-1145. [doi: 10.1016/s0140-6736(96)90608-x] [Medline: 8609748]
18. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
19. Seshadri S, Beiser A, Au R, Wolf PA, Evans DA, Wilson RS, et al. Operationalizing diagnostic criteria for Alzheimer's disease and other age-related cognitive impairment-Part 2. Alzheimers Dement 2011 Jan;7(1):35-52 [FREE Full text] [doi: 10.1016/j.jalz.2010.12.002] [Medline: 21255742]
20. Satizabal CL, Beiser AS, Chouraki V, Chêne G, Dufouil C, Seshadri S. Incidence of dementia over three decades in the Framingham Heart Study. N Engl J Med 2016;374(6):523-532. [doi: 10.1056/nejmoa1504327]
21. Farmer ME, White LR, Kittner SJ, Kaplan E, Moes E, McNamara P, et al. Neuropsychological test performance in Framingham: a descriptive study. Psychol Rep 1987 Jun;60(3 Pt 2):1023-1040. [doi: 10.2466/pr0.1987.60.3c.1023] [Medline: 3628637]
22. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
23. Tomasi D, Volkow ND. Aging and functional brain networks. Mol Psychiatry 2012 May;17(5):471, 549-471, 558 [FREE Full text] [doi: 10.1038/mp.2011.81] [Medline: 21727896]
24. Hu X, Le TH, Parrish T, Erhard P. Retrospective estimation and correction of physiological fluctuation in functional MRI. Magn Reson Med 1995 Aug;34(2):201-212. [doi: 10.1002/mrm.1910340211] [Medline: 7476079]
25. Shi Y, Yamada K, Liddelow SA, Smith ST, Zhao L, Luo W, Alzheimer's Disease Neuroimaging Initiative, et al. ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. Nature 2017 Sep 28;549(7673):523-527 [FREE Full text] [doi: 10.1038/nature24016] [Medline: 28959956]
26. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015 May 28;521(7553):452-459. [doi: 10.1038/nature14541] [Medline: 26017444]
27. Pearl J. Causal inference in statistics: An overview. Statist Surv 2009;3:96-146. [doi: 10.1214/09-ss057]
28. Park HS, Jun CH. A simple and fast algorithm for K-medoids clustering. Expert Syst Appl 2009;36(2):3336-3341. [doi: 10.1016/j.eswa.2008.01.039]
29. Kaufman L, Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis. Marblehead, MA: John Wiley & Sons; 2005.
30. Arunajadai SG. RNA unwinding by NS3 helicase: a statistical approach. PLoS One 2009 Sep 22;4(9):e6937 [FREE Full text] [doi: 10.1371/journal.pone.0006937] [Medline: 19771177]
31. Liu ZF, Malone B, Yuan CH. Empirical evaluation of scoring functions for Bayesian network model selection. BMC Bioinform 2012;13(Suppl 15):S14. [doi: 10.1186/1471-2105-13-s15-s14]
32. Rissanen J. Stochastic complexity. J Roy Stat Soc B Met 1987;49(3):223-239. [doi: 10.1111/j.2517-6161.1987.tb01694.x]
33. Vázquez-Castellanos JF, Serrano-Villar S, Latorre A, Artacho A, Ferrús ML, Madrid N, et al. Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. Mucosal Immunol 2015 Jul;8(4):760-772. [doi: 10.1038/mi.2014.107] [Medline: 25407519]
34. Russell SJ, Norvig P. Artificial Intelligence: A Modern Approach. Third Edition. Malaysia: Pearson Education Limited; 2016.
35. Beretta S, Castelli M, Gonçalves I, Henriques R, Ramazzotti D. Learning the structure of Bayesian networks: a quantitative assessment of the effect of different algorithmic schemes. Complexity 2018;2018:1-12. [doi: 10.1155/2018/1591878]
36. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. J Comput Biol 2000;7(3-4):601-620. [doi: 10.1089/106652700750050961] [Medline: 11108481]
37. Silvey SD. Statistical Inference. Boca Raton, Florida: Chapman and Hall/CRC; 2017.
38. Korb KB, Nicholson AE. Bayesian Artificial Intelligence. Boca Raton, FL: CRC Press; 2010.
39. Tsamardinos I, Aliferis C. Towards Principled Feature Selection: Relevancy, Filters and Wrappers. In: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics.: Association for Computing Machinery; 2003 Presented at: AISTATS'03; January 3-6, 2003; Key West, Florida, USA URL: https://pdfs.semanticscholar.org/6ff6/b42a16a2e70813f6728cf4d0b672939235ec.pdf
40. Bell DA, Wang H. A formalism for relevance and its application in feature subset selection. Mach Learn 2000;41(2):175-195. [doi: 10.1023/A:1007612503587]

41. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Berlin, Germany: Elsevier; 1997.
42. Han B, Chen X, Talebizadeh Z. FEPI-MB: identifying SNPs-disease association using a Markov Blanket-based approach. BMC Bioinformatics 2011 Nov 24;12(Suppl 12):S3 [FREE Full text] [doi: 10.1186/1471-2105-12-S12-S3] [Medline: 22168374]
43. Han B, Chen XW, Talebizadeh Z, Xu H. Genetic studies of complex human diseases: characterizing SNP-disease associations using Bayesian networks. BMC Syst Biol 2012;6(Suppl 3):S14 [FREE Full text] [doi: 10.1186/1752-0509-6-S3-S14] [Medline: 23281790]
44. Plackett RL. Karl Pearson and the Chi-Squared Test. Int Stat Rev 1983;51(1):59-72. [doi: 10.2307/1402731]
45. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
46. Wechsler D, Stone C. Wechsler Memory Scale (WMS). New York: The Psychological Corporation; 1948.
47. Wechsler D. Wechsler Adult Intelligence Scale (WAIS). New York: The Psychological Corporation; 1955.
48. Kaplan E, Goodglass H, Weintraub S. Boston Naming Test. Austin,TX: Pro-ed; 2001.
49. Scutari M, Auconi P, Caldarelli G, Franchi L. Bayesian Networks Analysis of Malocclusion Data. Sci Rep 2017 Nov 10;7(1):15236 [FREE Full text] [doi: 10.1038/s41598-017-15293-w] [Medline: 29127377]
50. Sundermann EE, Biegon A, Rubin LH, Lipton RB, Landau S, Maki PM, Alzheimer's Disease Neuroimaging Initiative. Does the female advantage in verbal memory contribute to underestimating Alzheimer's disease pathology in women versus men? J Alzheimers Dis 2017;56(3):947-957. [doi: 10.3233/JAD-160716] [Medline: 28106548]
51. Johnson P, Vandewater L, Wilson W, Maruff P, Savage G, Graham P, et al. Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease. BMC Bioinformatics 2014;15(Suppl 16):S11 [FREE Full text] [doi: 10.1186/1471-2105-15-S16-S11] [Medline: 25521394]
52. van der Linde I, Bright P. A genetic algorithm to find optimal reading test word subsets for estimating full-scale IQ. PLoS One 2018;13(10):e0205754. [doi: 10.1371/journal.pone.0205754] [Medline: 30335801]
53. Small BJ, Fratiglioni L, Viitanen M, Winblad B, Bäckman L. The course of cognitive impairment in preclinical Alzheimer disease: three- and 6-year follow-up of a population-based sample. Arch Neurol 2000 Jun;57(6):839-844. [doi: 10.1001/archneur.57.6.839] [Medline: 10867781]
54. Rubin EH, Storandt M, Miller JP, Kinscherf DA, Grant EA, Morris JC, et al. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. Arch Neurol 1998 Mar;55(3):395-401. [doi: 10.1001/archneur.55.3.395] [Medline: 9520014]
55. Khamis RY, Ammari T, Mikhail GW. Gender differences in coronary heart disease. Heart 2016 Jul 15;102(14):1142-1149. [doi: 10.1136/heartjnl-2014-306463] [Medline: 27126397]
56. Friston KJ. Functional and effective connectivity: a review. Brain Connect 2011;1(1):13-36. [doi: 10.1089/brain.2011.0008] [Medline: 22432952]
57. A´lvarez I, Go´rriz J, Rami´rez J, Salas-Gonzalez D, Lo´pez M, Puntonet C, et al. Alzheimer's diagnosis using eigenbrains and support vector machines. Electron Lett 2009;45(7):342. [doi: 10.1049/el.2009.3415]
58. Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, et al. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 2008 Feb 1;39(3):1186-1197 [FREE Full text] [doi: 10.1016/j.neuroimage.2007.09.073] [Medline: 18054253]
59. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol 2015 Dec;44(6):1800-1813 [FREE Full text] [doi: 10.1093/ije/dyv337] [Medline: 26705418]

Abbreviations

AD: Alzheimer disease
ApoE4: apolipoprotein E4
BNT30: Boston Naming Test 30 item Even Version
CPT: conditional probability table
FHS: Framingham Heart Study
HC: healthy control
LMi: Logical Memory Immediate Recall
LMd: Logical Memory Delayed Recall
NAD: non-Alzheimer dementia
NP: neuropsychological
PASi: Paired Associate Learning Immediate Recall
SIM: similarities test
VRi: Visual Reproductions Immediate Recall
VRd: Visual Reproductions Delayed Recall

VRr: Visual Reproductions Recognition

Edited by G Eysenbach; submitted 05.07.19; peer-reviewed by XT Zeng, I van der Linde; comments to author 22.10.19; revised version received 13.01.20; accepted 24.01.20; published 23.04.20.

Please cite as:
Ding H, An N, Au R, Devine S, Auerbach SH, Massaro J, Joshi P, Liu X, Liu Y, Mahon E, Ang TFA, Lin H
Exploring the Hierarchical Influence of Cognitive Functions for Alzheimer Disease: The Framingham Heart Study
J Med Internet Res 2020;22(4):e15376
URL: http://www.jmir.org/2020/4/e15376/
doi:10.2196/15376
PMID:32324139

Original Paper

Development and Validation of a Comprehensive Well-Being Scale for People in the University Environment (Pitt Wellness Scale) Using a Crowdsourcing Approach: Cross-Sectional Study

Leming Zhou[1], PhD; Bambang Parmanto[1], PhD

Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Leming Zhou, PhD
Department of Health Information Management
University of Pittsburgh
6021 Forbes Tower
3600 Forbes Ave at Meyran Ave
Pittsburgh, PA, 15260
United States
Phone: 1 412 383 6653
Fax: 1 412 383 6655
Email: Leming.Zhou@pitt.edu

Abstract

Background: Well-being has multiple domains, and these domains are unique to the population being examined. Therefore, to precisely assess the well-being of a population, a scale specifically designed for that population is needed.

Objective: The goal of this study was to design and validate a comprehensive well-being scale for people in a university environment, including students, faculty, and staff.

Methods: A crowdsourcing approach was used to determine relevant domains for the comprehensive well-being scale in this population and identify specific questions to include in each domain. A web-based questionnaire (Q1) was used to collect opinions from a group of university students, faculty, and staff about the domains and subdomains of the scale. A draft of a new well-being scale (Q2) was created in response to the information collected via Q1, and a second group of study participants was invited to evaluate the relevance and clarity of each statement. A newly created well-being scale (Q3) was then used by a third group of university students, faculty, and staff. A psychometric analysis was performed on the data collected via Q3 to determine the validity and reliability of the well-being scale.

Results: In the first step, a group of 518 university community members (students, faculty, and staff) indicated the domains and subdomains that they desired to have in a comprehensive well-being scale. In the second step, a second group of 167 students, faculty, and staff evaluated the relevance and clarity of the proposed statements in each domain. In the third step, a third group of 546 students, faculty, and staff provided their responses to the new well-being scale (Pitt Wellness Scale). The psychometric analysis indicated that the reliability of the well-being scale was high.

Conclusions: Using a crowdsourcing approach, we successfully created a comprehensive and highly reliable well-being scale for people in the university environment. Our new Pitt Wellness Scale may be used to measure the well-being of people in the university environment.

(J Med Internet Res 2020;22(4):e15075) doi:10.2196/15075

KEYWORDS

crowdsourcing; questionnaire design; university

Introduction

Background

Well-being is "a good or satisfactory condition of existence; a state characterized by health, happiness, and prosperity" [1]. Well-being is commonly assessed using well-being scales.

Few well-being scales, such as the Patient-Reported Outcomes Measurement Information System (PROMIS), 36-Item Short-Form Health Survey (SF-36), and World Health Organization Quality of Life scale, have been designed for the general population [2-4]. However, while these generic well-being scales are useful for large-scale assessments and for obtaining an overall impression of a population, they may not be able to accurately reflect the well-being situation of a particular population. Therefore, a population-specific well-being scale is needed to precisely assess the well-being of the target population.

Many well-being and quality of life scales have been created for different purposes and for different target populations [5-8]. Some well-being and quality of life scales were specifically created for people with particular diseases, such as depression, stroke, and cancer [9-12]. Others were created for particular populations, such as children and adults [13-15]. These specific scales are very useful for assessing the well-being of the target population, but they are not appropriate for other populations.

People in the university environment (students, faculty, and staff) can be considered a specific population. The activities conducted by people in the university environment and the relationships among them are different from those of people in government offices, companies, hospitals, and even elementary and secondary education schools. People in the university environment are focused on higher education, research, and career development. At the same time, universities are not ivory towers. People in the university environment (students, faculty, and staff) are not isolated from the world, and they have a life outside of teaching, research, service, and learning. Like other people, they experience various problems in real life, such as physical disease, mental problems, financial pressure, and problems related to handling relationships with difficult people around them.

In recent years, several important discussions have arisen regarding college students' health issues, faculty and student relationships, and university employees' job satisfaction [16-22]. These discussions often only focused on a specific issue, such as physical health [23,24], harmful lifestyle [25,26], and mental health [16,27,28]. In many cases, however, these issues are intertwined; for instance, mental illness or abnormal behavior may be triggered by heavy academic workload, severe financial pressure, and poor relationships with others [17,29,30]. Therefore, it is necessary to use tools to perform a *comprehensive* well-being assessment in order to provide a foundation for well-being improvement interventions. At present, there is no well-being scale specifically designed for people in the university environment (students, faculty, and staff).

Well-being is a higher order construct, and thus, it includes multiple lower order constructs or domains [6,31,32]. The commonly covered domains in well-being scales are physical, emotional (or mental), social (or relational), spiritual, and financial (or socioeconomic) [6,32]. Some well-being scales also cover occupational, environmental, and intellectual domains [6,32]. These domains may have one or multiple subdomains. For instance, the physical domain may include subdomains, such as physical health, daily living activities, pain, and sleep; the social domain may include participation, friends, and other relationships; and the mental domain may include happiness, depression, stress coping skills, and communication.

To conduct a comprehensive well-being assessment for people in the university environment, we need a scale with multiple domains [6,31,32]. There are a number of domain-specific scales, such as the social interaction anxiety scale (SIAS) for social interaction anxiety, and a number of PROMIS scales for pain [33], smoking [34], and depression [35]. However, we cannot simply use a combination of multiple existing domain-specific scales to build our well-being scale for two major reasons. First, the comprehensive well-being of different populations needs to be measured using different sets of domains. The combination of these domains can only be determined by the target population. Second, the wording of some statements and subdomains for existing domain-specific scales may not be applicable for our target population, as many scales were created with certain populations in mind, such as elderly people, healthy young professionals, people with cancer, and people who play a particular role (eg, caregiver). Hence, in this project, it was necessary to first identify the domains and related subdomains of a comprehensive well-being scale relevant for people in the university environment and then create statements that use language appropriate to this population for each subdomain.

The typical scale development approach used by researchers involves conducting a literature review, drafting a scale for a small group of experts to review, and then releasing the new scale to a group of recruited study participants to collect responses. In this typical approach, a sample of the target population is only involved in the last stage of scale development. This is a shortcoming in that study participants are simply asked to provide responses to the statements in the scale, and thus, any domains or subdomains that researchers may have missed in the draft of the scale will not be brought to the researchers' attention. Crowdsourcing is one way to overcome this issue.

In recent years, crowdsourcing has been used to collect ideas from a crowd [36-38]. The benefit of crowdsourcing is that the collected wisdom of the crowd can be identified by using feedback obtained from a large pool of the target population [39-45]. On the other hand, most people in the crowd do not have formal training in research or scale development; therefore, the information from the crowd cannot be solely depended on to create a new scale. We have adopted what we believe is a better strategy. It involves combining these two approaches in the development of our new comprehensive well-being scale in order to retain the advantages of these two methods while avoiding their limitations.

More specifically, in this combined approach, information from the literature was used to guide the development of the new scale, and the crowd participated in all stages of the scale development and evaluation (domain and subdomain determination, statement relevance and clarity evaluation, and response to statements) to fully reflect their ideas in the new scale.

Before describing the objective of this study, we present the definitions of several commonly referred well-being domains below. They have been adopted from a previous study [6].

Physical wellness refers to "the quality and performance of bodily functioning." *Emotional wellness* reflects "the psychological, cognitive, and emotional quality of a person's life." *Social wellness* is about "how well an individual is connected to others in their local and wider social community." *Spiritual wellness* is about "meaning, a connection to something greater than oneself, and in some cases, faith in a higher power." *Financial wellness* refers to "an individual's financial management skills and financial security." *Occupational wellness* indicates "an individual's career development opportunities and job satisfaction." *Intellectual wellness* refers to "an individual's ability to handle tasks in daily life and on the job, and their self-assessment of their performance."

Objective

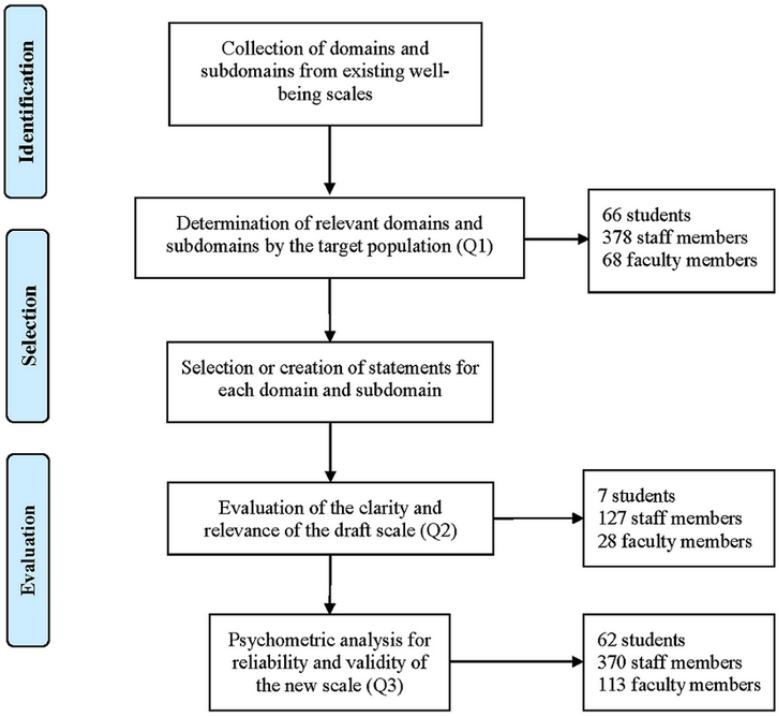
The goal of this study was to create a comprehensive well-being scale for people in the university environment, using a combined approach (traditional scale development method and crowdsourcing). The study also sought to achieve acceptable reliability of the new scale to demonstrate the benefits of using this combined scale development approach.

Methods

User-Centered Approach for Scale Development and Evaluation

User-centered design is the process of developing a tool from the perspective of how it will be understood and used by users [46]. Therefore, in a user-centered design, the target users of a tool are actively involved in all stages of the product development. In scale development, this includes domain and subdomain identification, statement selection in terms of validity, and scale evaluation in terms of validity and reliability. These are the general steps we took in this study (Figure 1). The details of each step are provided in the following sections.

Figure 1. Flowchart of the design and evaluation of the well-being questionnaire.



Study Procedure

In this study, well-being scales and their corresponding domains and subdomains were collected from several recent review studies on well-being and quality of life scales [5-8,31,47]. A web-based questionnaire (Q1) was created to collect opinions from people in the university environment (students, faculty, and staff) on these domains and subdomains for their own well-being assessment. The obtained results were used to guide the creation of the first draft of a new comprehensive well-being scale. This draft was provided to people in the university environment via another web-based questionnaire (Q2) to obtain their evaluation of the relevance and clarity of each statement. A revised well-being scale based on input from the draft was then released to people in the university environment in order to collect responses to its statements via a third web-based questionnaire (Q3). All of the study participants were encouraged to provide comments and suggestions on each statement in these three questionnaires and on the new scale. A psychometric analysis was performed to evaluate the reliability and validity of the new comprehensive well-being scale. This

study protocol was approved by the Human Research Protection Office at the University of Pittsburgh (Pitt). The details of each step are provided below.

Step 1: Collection of Domains and Subdomains From Existing Well-Being Scales

Well-being and quality of life scales from six recent review studies were collected [5-8,31,47]. The domains and subdomains of these scales were compiled. The statements of these scales were also compiled.

Step 2: Determination of Relevant Domains and Subdomains According to the Target Population

It is known that domains and their subdomains vary widely in different well-being scales [6]. In this study, to get an idea of which domains and subdomains are most valuable in assessing the well-being of a university population, study participants were asked to fill out a web-based questionnaire (Q1) with a list of domains and subdomains. Study participants were asked to provide their opinions on the relevance of these domains and subdomains for their own well-being assessment.

Step 3: Evaluation of the Clarity and Relevance of a Draft Scale

Statements were selected from existing well-being and quality of life scales for domains and subdomains identified as relevant to the university population in step 2. For domains having only few already existing statements, such as the intellectual domain, new statements were created. The collection of these statements in each domain and subdomain formed the first draft of the new well-being scale. We had multiple rounds of discussions on the clarity of each statement and relevance of each statement to the corresponding domain and subdomain. A final draft of 77 statements was provided via a web-based questionnaire (Q2) to people in the university environment for evaluation of the relevance and clarity of each statement. These study participants evaluated the relevance and clarity of each statement for use in a comprehensive well-being scale, using a scale from 1 to 4, where 1 meant no relevance or clarity and 4 meant high relevance or clarity. In response to these evaluations, if the average relevance of a statement was lower than 2.5, it was removed from the scale. If the clarity of a statement was rated 1 or 2, the wording of the statement was adjusted. We had multiple face-to-face meetings to discuss the rating and wording of statements for finalizing the draft scale.

Step 4: Questionnaire Study and Psychometric Analysis for Reliability and Validity

After we agreed on the content validity of the statements in the new well-being scale, the scale was released to people in the university environment via a web-based questionnaire (Q3) in order to collect study participants' answers to the statements in the questionnaire. The obtained data were used to evaluate the reliability and validity of the new scale. The details of the data analysis are presented in a later section.

Participant Recruitment

In this study, the study participants were current Pitt students, staff, and faculty who were randomly selected by a bulk email system. Former students, staff, and faculty were excluded from the study, because they might have been working in a different environment for a long time and hence their opinion might not reflect the actual well-being of someone currently in the university environment.

To recruit study participants, emails about the purpose of the study and links to the corresponding questionnaires (Q1, Q2, and Q3) were randomly distributed to approximately 2000 current students, staff, and faculty at Pitt via a bulk email system (Read Green) at different time points for each questionnaire study. This Pitt bulk email system has all the email addresses of current Pitt students, staff, and faculty. According to the Pitt Fact Book 2019, the total number of email addresses included in the bulk email system was close to 50,000 (one per person). When we requested to make an announcement via this bulk email system, we were required to indicate the number of people and the categories of the university members. The number of email addresses requested was directly linked to the charge of the email distribution service. The bulk email system randomly picked email addresses from each indicated category (students, staff, and faculty) among the 50,000 email addresses, for a total of 2000 email addresses, and sent out the announcement. Since the three announcements were made at three different time points (separated by approximately 1 month), the 2000 email addresses in Q1, Q2, and Q3 could be completely different or have very limited overlap. In other words, one Pitt student, staff member, or faculty member might have received one, two (unlikely), or three (very unlikely) email announcements because of the randomness of the email selection.

To participate in the study, students, faculty, and staff who received the email message could click on the link to the questionnaire given in the email and provide their responses on the web-based Qualtrics system. The purpose of the study was also described at the beginning of each questionnaire. Study participation was voluntary, and participants could stop participating in the study at any time. They could also request that their entered data be removed in the comments section of the questionnaires.

Participants were asked to provide some basic demographic information, such as age, gender, race, education, and role at Pitt, before they responded to any other statements in the questionnaires. Their responses were stored anonymously, since they were not required to provide their name, department, or job. The Internet Protocol (IP) addresses of their computers were hidden to the investigators.

Data Analysis

A descriptive analysis was performed on the collected data to understand the demographic characteristics of the study participants and the overall results from the data-collection questionnaires, such as the mean and SD values of individual statements. The comments and suggestions collected by open-ended questions in the three web-based questionnaires were summarized briefly.

Cronbach alpha was calculated for each domain of the scale and the entire well-being scale to evaluate the reliability of the scale. Cronbach alpha is a commonly used measurement of

internal consistency for questionnaires. For research and exploratory studies, Cronbach alpha values from .7 to .8 are considered acceptable, whereas a value around .9 is considered excellent [48].

Exploratory factor analysis and confirmatory factor analysis were performed to determine and verify the constructs of the new well-being scale. In the exploratory factor analysis, the extraction method was principal component analysis and the rotation method was Oblimin with Kaiser Normalization [49,50]. The factor loadings obtained in the exploratory factor analysis were used to determine whether each statement should be included in the well-being scale and in one specific domain. Here, 0.32 was used as the guiding value for the evaluation [51]. However, in certain cases, we overruled this value and chose to keep a statement in the scale, even if the factor loadings were smaller than 0.32 or multiple factor loadings were greater than 0.32, using judgement skills gained from our extensive experience in scale development and the opinions of the target population obtained in the first web-based questionnaire study (Q1). R package LAVAAN 0.5 (Yves Rosseel et al, Belgium) was used for the confirmatory factor analysis. The estimator was maximum likelihood. A two-layer multi-factor model was used in this analysis. The domains were latent variables, and their items were the observables. All the domains together were used to measure overall well-being. All the statistical analyses were performed using R 3.3 (The R Foundation, Vienna, Austria) and IBM SPSS version 24 (IBM Corp, Armonk, New York, USA).

Results

Identified Scales

In total, 165 well-being and quality of life scales were collected from previous review studies. The total number of statements in these scales was approximately 4700. We cannot provide an exact number for the total because some scales have multiple versions with different numbers of items. A few hundred domains were covered in these scales; however, most of them were only mentioned in one or a few scales. We chose the following seven frequently covered domains for well-being assessment: physical, emotional, social, spiritual, financial, occupational, and intellectual. Their subdomains, which were found in multiple scales, were identified as well. These domains and subdomains were listed in the web-based questionnaire (Q1) so that the study participants could make selections. The definitions of the seven domains were given in the questionnaire so that every participant would know the meaning of each domain. The subdomains were more specific, and thus, no definitions were provided for them.

Domains and Subdomains

After Q1 was distributed to approximately 2000 students, faculty, and staff via email, 518 of the recipients chose to answer the questionnaire. Their mean age was 41.6 years (SD 13.4). Further details on their demographics are summarized in Table 1.

The responses from these 518 study participants in the first questionnaire study (Q1) are summarized in Table 2. The responses were organized into categories. These categories were then broken down into domains and subdomains. In Table 2, the information is listed in the order of importance for inclusion in a comprehensive well-being assessment, as indicated by the participants' responses.

Table 1. Demographic information of the 518 study participants in the first questionnaire study.

Characteristic	Value, n (%)
Role	
Student	66 (12.7)
Staff	378 (73.0)
Faculty	68 (13.1)
Gender	
Male	107 (20.7)
Female	405 (78.2)
Undeclared	6 (1.1)
Race	
African American	21 (4.1)
White American	455 (87.8)
Asian American	24 (4.6)
Other (mixed race, Native American, or Hispanic)	13 (2.6)
Education	
High school or lower	12 (2.3)
Some college credits, no degree	40 (7.7)
Associate degree	20 (3.9)
Bachelor’s degree	194 (37.5)
Master’s degree	170 (32.8)
Professional degree	13 (2.5)
Doctoral degree	66 (12.7)

Table 2. Summary of answers from the study participants in the first questionnaire study (N=518).

Question	Value, n (%)
Which of the following domains are important for a comprehensive well-being assessment?	
Physical wellness	505 (97.5)
Emotional wellness	493 (95.2)
Financial wellness	374 (72.2)
Social wellness	338 (65.3)
Occupational wellness	329 (63.5)
Spiritual wellness	245 (47.3)
Intellectual wellness	236 (45.6)
Other (eg, environmental wellness)	21 (4.1)
Please indicate the subdomains for each domain, which you believe are important for your comprehensive well-being assessment.	
Physical wellness	
Physical activity	483 (93.2)
Nutrition	480 (92.7)
Sleep	468 (90.3)
Overall health	408 (78.8)
Chronic disease	216 (41.7)
Medication dependence	178 (34.4)
Appetite	114 (22.0)
Other	25 (4.8)
Emotional wellness	
Stress	416 (80.3)
Positive attitude	386 (74.5)
Anxiety	369 (71.2)
Resilience	306 (59.1)
Depression or bipolar disorder	300 (57.9)
Traumatic events	225 (43.4)
Posttraumatic stress disorder	200 (38.6)
Negative attitude	152 (29.3)
Other	36 (6.9)
Social wellness	
Relationship with family, friends, and colleagues	488 (94.2)
Connection with others	425 (82.0)
Social participation	341 (65.8)
Smoking, alcohol, and drug use	117 (22.6)
Other	16 (3.1)
Financial wellness	
Preparedness for short-term and long-term financial emergency	434 (83.8)
Skills for financial management	427 (82.4)
Income level	274 (52.9)
Other	35 (6.8)
Spiritual wellness	
Purpose of life	311 (60.0)

Question	Value, n (%)
Satisfaction with the current belief system	266 (51.4)
View of the world	259 (50.0)
Meaning of life	207 (40.0)
Meditation	181 (34.9)
Spiritual activities	160 (30.9)
Religion	153 (29.5)
Other	26 (5.0)
Occupational wellness	
Job satisfaction	486 (93.8)
Job security	405 (78.2)
Career development opportunities	362 (69.9)
Job stress	326 (62.9)
Job performance	234 (45.2)
Career ambition	189 (36.5)
Workaholic (job and life balance)	85 (16.4)
Other	34 (6.6)
Intellectual wellness	
Capacity for thinking and acquiring knowledge	455 (87.8)
View on life-long learning (burden, part of life, or enjoy)	386 (74.5)
Informal education experience	254 (49.0)
Formal education experience	230 (44.4)
Other	19 (3.7)
How many questions are reasonable for assessing each domain of your well-being?	
10	231 (44.6)
5	111 (21.4)
15	90 (17.4)
20	71 (13.7)

One additional category, environmental wellness, was added by 21 (4.05%) participants as an aspect of wellness that is important, indicating that these individuals live a lifestyle that is mindful of their surroundings. Study participants also identified many more subdomains in each domain than were in the original questionnaire. Multimedia Appendix 1 provides a list of additional subdomains mentioned by some study participants. Some study participants also made general comments on the well-being scale creation activity itself.

> *Great idea. You cannot improve something if you don't track it first.* [Participant #341 in Q1]
>
> *Make sure it is available to all doctors that a person will visit so they can use it as a baseline.* [Participant #355 in Q1]

Relevance and Clarity of the Proposed Statements

A draft of a well-being scale with 77 statements, which included those domains and subdomains designated as important by the participants in the Q1 study, was created. Seven domains were included, and on average, there were 10 statements in each domain (also according to the responses of many participants in the Q1 study). Most of these statements were selected or modified from existing well-being and quality of life scales, except for those in the financial and intellectual domains, which were mainly written by us, as existing scales did not include many such statements.

This draft was randomly distributed to another 2000 university members to obtain their feedback (as members of the target population) on the relevance and clarity of each statement via the second web-based questionnaire (Q2). In total, 167 participants responded, and of these, 143 (85.6%) provided their ratings on the relevance and clarity of all 77 statements and the other 24 provided their ratings for at least one-third of the statements in the draft scale. The mean age of this group of participants was 44.0 years (SD 12.99). Among the 167 participants, there were 127 staff members (76.0%), 28 faculty members (16.8%), and 7 students (4.2%). Most participants were female (132/167, 79.0%). There were 30 (18.0%) male participants and 5 (3.0%) who did not indicate gender. As indicated in the Methods section, statements with an average

rating of relevance lower than 2.5 were directly removed from the draft scale. The wording of statements was adjusted if participants were confused by the statements (clarity rating was 1 or 2). A few statements were removed because they were highly personal and study participants expressed a strong objection to them (eg, a statement about sexual activity). At the end of this step, the updated well-being scale had 47 statements in total.

Evaluation of the New Well-Being Scale

The updated well-being scale (named the Pitt Wellness Scale), several demographic questions, and few open-ended questions for comments and suggestions were combined to create the third web-based questionnaire (Q3). The link to Q3 was again randomly distributed to approximately 2000 university community members. In total, 671 individuals clicked on the link to this questionnaire, and 546 of them provided responses to all of the statements in the new well-being scale. This new scale was evaluated using the responses from these 546 participants. The mean duration of response to all the statements was 535.32 seconds (minimum 117, maximum 14,794, SD 1187.33; less than 10 minutes), which is an acceptable length of time for most people. The mean age of the participants in the Q3 study was 43.7 years (SD 13.54). Further details on their demographics are provided in Table 3.

Descriptive statistics of responses were calculated, and a reliability test of the scale was performed. For most statements, response options ranged on a scale from 1 (strongly agree) to 7 (strongly disagree). Eight statements (self-assessed level of wellness for each domain and overall wellness) had options ranging on a scale from 1 (excellent) to 5 (terrible). The options for the level of pain statement ranged from 0 (no pain) to 10 (most severe pain ever). After the reliability analysis, three statements were removed to improve the reliability of the well-being scale. Therefore, the final version of the new well-being scale included 44 statements. The overall Cronbach alpha of the 44-item scale was .933. Table 4 presents the descriptive statistics, Cronbach alpha value of each domain, and number of items in each domain.

An exploratory factor analysis was performed on the responses from the 546 study participants, assuming there were seven factors in this scale. The obtained pattern matrix is shown in Table 5. Here, rotated factor loadings greater than 0.32 are shown. Two statements (WO and P6) with factor loadings less than 0.32 are also shown. WO is for overall well-being and therefore does not belong to any domain. P6 is about appetite, which is highly relevant to both physical health and mental health. Therefore, although the highest factor loading for P6 was 0.301, we still chose to keep this statement in the scale. The reliability evaluation indicated that both physical and mental domains had higher reliability when P6 was in the physical domain (Cronbach alpha, .705 [without P6] vs .714 [with P6] for the physical domain and .857 [with P6] vs .860 [without P6] for the mental domain). Therefore, we kept P6 in the physical domain.

Table 3. Demographic information of the 546 study participants who provided responses to all the statements in the new well-being scale.

Characteristic	Value, n (%)
Role	
Student	62 (11.4)
Faculty	113 (20.7)
Staff	370 (67.8)
Undeclared	1 (0.2)
Gender	
Male	128 (23.4)
Female	411 (75.3)
Undeclared	7 (1.3)
Race	
African American	16 (2.9)
White American	496 (90.8)
Asian American	22 (4.0)
Other	12 (2.2)
Education	
High school or lower	4 (0.7)
Some college credits, no degree	24 (4.4)
Some technical training, no degree	11 (2.0)
Associate degree	18 (3.3)
Bachelor’s degree	194 (35.5)
Master’s degree	163 (29.9)
Professional degree	22 (4.0)
Doctoral degree	110 (20.1)
Marital status	
Single	137 (25.1)
Married or long-term committed relationship	369 (67.6)
Divorced or separated	34 (6.2)
Widowed	6 (1.1)
Household income	
≤US $10,000	7 (1.3)
US $10,001-25,000	14 (2.6)
US $25,001-50,000	114 (20.9)
US $50,001-75,000	91 (16.7)
US $75,001-100,000	84 (15.4)
US $100,001-125,000	74 (13.6)
>US $125,000	143 (26.2)
Declined to answer	19 (3.5)

Table 4. Descriptive statistics of study participants' responses to 44 statements in the seven domains of the new scale and the reliability of each domain (N=546).

Statements	Value, mean (SD)
Physical domain, seven items (Cronbach alpha=.714)	
P1. I feel rested when I wake up in the morning.	3.60 (1.54)
P2. Each week, I exercise moderately for at least 30 minutes (for instance, walking briskly, bicycling slower than 10 miles per hour, playing tennis, and ballroom dancing).	3.14 (1.13)
P3. Because of my health status, I am physically able to exercise as much as I would like to.	2.60 (1.63)
P4. I usually have enough energy for everyday activities.	2.52 (1.32)
P5. My chronic pain level is (0=no pain, 10=most severe pain ever).	1.23 (1.46)
P6. My appetite has been good recently.	2.15 (1.05)
PO. My overall physical health is (1=excellent, 5=terrible).	2.09 (0.69)
Mental domain, seven items (Cronbach alpha=.860)	
M1. I am generally satisfied with my quality of life.	2.44 (1.21)
M2. I am generally self-accepting.	2.57 (1.29)
M3. I feel hopeful about the future.	2.46 (1.25)
M4. I feel that I have control over my emotions.	2.40 (1.05)
M5. I believe that life is what you make it.	2.31 (1.21)
M6. I am open to new opportunities if my first plan does not work out.	2.03 (0.79)
MO. My overall mental health is (1=excellent, 5=terrible).	2.05 (0.78)
Social domain, six items (Cronbach alpha=.781)	
S1. I am living in a safe community.	1.82 (0.88)
S2. When something good happens to me, I share the experience with my family and/or friends.	1.90 (0.92)
S3. I am satisfied with my ability to meet the needs of people who depend on me.	2.12 (0.98)
S4. I am satisfied with my current level of social activities.	2.84 (1.38)
S5. I have people in my life who care about me.	1.56 (0.78)
SO. My overall social wellness is (1=excellent, 5=terrible).	2.02 (0.77)
Financial domain, five items (Cronbach alpha=.856)	
F1. If I incur an unexpected above average expense, I would still be stable financially.	3.03 (1.84)
F2. I have someone to help with my financial affairs, if needed.	2.95 (1.77)
F3. I am saving for retirement and for emergencies.	2.28 (1.46)
F4. My income is adequate for my current needs.	3.21 (1.79)
FO. My overall financial wellness is (1=excellent, 5=terrible).	2.39 (0.93)
Spiritual domain, six items (Cronbach alpha=.892)	
SP1. I feel that my life is meaningful.	2.21 (1.11)
SP2. I feel inner and/or spiritual strength in difficult times.	2.69 (1.38)
SP3. I have a sense of direction for my life.	2.45 (1.19)
SP4. I know what is really important in my life.	2.01 (0.98)
SP5. My personal beliefs (religious or not) help me to cope with difficulties in life.	2.32 (1.11)
SPO. My overall spiritual wellness is (1=excellent, 5=terrible).	2.14 (0.79)
Occupational domain, seven items (Cronbach alpha=.844)	
O1. I feel I have input on deciding how my job gets done.	2.52 (1.37)
O2. I am satisfied with the amount of time required by my job duties.	2.86 (1.49)
O3. My employer provides me many career development opportunities.	3.06 (1.63)
O4. I feel comfortable working with my colleagues.	2.21 (1.22)

Statements	Value, mean (SD)
O5. My work and life are well-balanced.	2.97 (1.50)
O6. My job security is high.	2.90 (1.43)
OO. My overall occupational wellness is (1=excellent, 5=terrible).	2.22 (0.80)
Intellectual domain, five items (Cronbach alpha=.828)	
I1. I am satisfied with the quality of my work.	2.15 (0.99)
I2. I am aware of my intellectual strengths.	1.92 (0.80)
I3. I can rely upon my talents and skills to handle unexpected situations.	1.89 (0.75)
I4. I am satisfied with my ability to make decisions.	2.07 (0.90)
IO. My overall intellectual wellness is (1=excellent, 5=terrible).	1.78 (0.60)
WO. My overall well-being is (1=excellent, 5=terrible).[a]	1.99 (0.64)

[a]The statement is for self-assessment of overall wellness. It is in the scale but does not belong to any domain.

Table 5. Rotated factor loadings of the exploratory factor analysis with 44 statements from 546 study participants.

Statement ID	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
P1	0.485[a]	0.351	—	—	—	—	—
P2	0.534[a]	—	—	—	—	—	—
P3	0.758[a]	—	—	—	—	—	—
P4	0.609[a]	—	—	—	—	—	—
P5	0.610[a]	—	—	—	—	—	—
P6	0.170	0.301	—	—	—	—	—
PO	0.621[a]	—	—	—	—	—	—
M1	—[b]	0.447[a]	—	—	—	—	—
M2	—	0.655[a]	—	—	—	—	—
M3	—	0.376[a]	—	—	0.351	—	—
M4	—	0.632[a]	—	—	—	—	—
M5	—	0.488[a]	—	—	—	—	—
M6	—	0.413[a]	—	—	—	—	0.321
MO	—	0.528[a]	—	—	—	—	—
S1	—	—	0.339[a]	—	—	—	—
S2	—	—	0.612[a]	—	—	—	—
S3	—	—	0.574[a]	—	—	—	—
S4	—	0.363	0.590[a]	—	—	—	—
S5	—	—	0.555[a]	—	—	—	—
SO	—	—	0.625[a]	—	—	—	—
F1	—	—	—	0.887[a]	—	—	—
F2	—	—	—	0.683[a]	—	—	—
F3	—	—	—	0.708[a]	—	—	—
F4	—	—	—	0.829[a]	—	—	—
FO	—	—	—	0.891[a]	—	—	—
SP1	—	—	—	—	0.563[a]	—	—
SP2	—	—	—	—	0.820[a]	—	—
SP3	—	—	—	—	0.640[a]	—	—
SP4	—	—	—	—	0.604[a]	—	—
SP5	—	—	—	—	0.809[a]	—	—
SPO	—	—	—	—	0.796[a]	—	—
O1	—	—	—	—	—	−0.709[a]	—
O2	—	—	—	—	—	−0.759[a]	—
O3	—	—	—	—	—	−0.730[a]	—
O4	—	—	—	—	—	−0.676[a]	—
O5	—	—	—	—	—	−0.686[a]	—

Statement ID	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
O6	—	—	—	—	—	−0.578[a]	—
OO	—	—	—	—	—	−0.784[a]	—
I1	—	—	—	—	—	−0.345	0.408[a]
I2	—	—	—	—	—	—	0.860[a]
I3	—	—	—	—	—	—	0.869[a]
I4	—	—	—	—	—	—	0.768[a]
IO	—	—	—	—	—	—	0.839[a]
WO	0.156	0.288	0.170	0.085	0.173	−0.205	0.225

[a]Selected domain.

[b]Factor loadings <0.32.

There are several statements with high loading factors in more than one domain. For instance, P1 is about sleep quality, which is related to both physical and mental wellness. In these cases, both factor loading and Cronbach alpha were used to determine which domain is more appropriate for the statements. Typically, the domain with the higher factor loading and higher Cronbach alpha was chosen.

The confirmatory factor analysis assessed the fit of the seven-factor structure using the responses from the 546 study participants. For this seven-factor two-layer model, the comparative fit index (CFI) was 0.866, Tucker-Lewis index (TLI) was 0.859, and the root mean square error of approximation (RMSEA) was 0.058, suggesting adequate model fit.

Comments

Some study participants provided brief comments after providing responses in the well-being scale. Some of these comments were specific to the university and are not shown here. Others were more generic and may be applicable to other places as well. These comments are presented below.

> *I also have a second job which affects the answers given here.* [Participant #42 in Q3]
>
> *Occupational distress mostly from direct supervisor.* [Participant #394 in Q3]
>
> *Currently pregnant, so my physical well-being is not what it should be.* [Participant #547 in Q3]
>
> *Great survey; easy to understand and take!* [Participant #646 in Q3]
>
> *Not a big fan of spiritual wellness. I know this survey says this could be religious or not, but isn't there another word without a religious connotation?* [Participant #648 in Q3]

Discussion

Principal Findings

The goal of this study was to develop a new comprehensive well-being scale for people in the university environment and evaluate its reliability and validity. We used a combined method (traditional survey design method and crowdsourcing) to create a new well-being scale for people in the university environment. This is a user-centered approach since the target population is involved in all the stages of scale development. The benefits of this combined method and user-centered approach were that findings were incorporated from previous well-being scale development studies and ideas and opinions were gathered from a large number of people in the target population. The obtained scale was shown to be highly reliable (Cronbach alpha of the scale was .933). A summary of the uniqueness of this study is given below.

First, the candidates for domains, subdomains, and scale statements were collected from a large number of existing well-being and quality of life scales identified by several recent review studies [6-8]. This provided a solid foundation for the validity and reliability of the new scale. We also had extensive experience in creating and evaluating scales in previous studies [7,52-61]. Second, in the three steps involving users, a total of 1,231 study participants from the target population contributed their ideas to the development of this scale at different stages. This made it possible to fully incorporate their needs and ideas into the new scale. This is typically not done in the traditional scale development approach. Third, the combined method generated a highly reliable new scale, and this result demonstrates that our approach is feasible for scale development.

In the past, scales were typically created by experts according to their experience and the literature. The target populations were only involved in the last step for the final evaluation. The quality of the scale was strongly determined by the knowledge of the experts in the field and their understanding of the target population. The application of crowdsourcing in this project reduced this dependence.

Crowdsourcing has been used in many previous studies [37-45,62-65]; however, the role of study participants was mainly limited to providing responses to already existing questionnaires, instead of being involved in all stages of the questionnaire development and evaluation. In this study, guided by the user-centered approach, samples of the target population were involved in all stages of questionnaire development and

evaluation. They provided invaluable ideas for building this highly reliable well-being scale.

Comparison With Prior Work

There are many other well-being scales. However, none of them were specifically designed for people in the university environment. Additionally, because this environment includes people who have different ages and different roles, the typical well-being scales for the workplace do not apply well [66-69], especially for students. This study used the findings from other well-being scale development studies and adopted a new approach to building a scale for people in the university environment. This new scale is considered better than other generic or employee well-being scales for more precise well-being assessment of this particular population.

Limitations

A scoring system for this scale has not been established yet, and thus, it is not feasible to compare the well-being outcomes from this scale with those from other existing scales. In the next step, we will develop a scoring system for the scale and compare the obtained scores with the results from other well-established scales domain by domain. We will perform another study to evaluate the relationship among the domains and determine the weight these domains should have in the overall well-being measure. For this purpose, we are currently creating a website that allows people in the university environment to complete multiple well-being scales online, including this new scale and several other scales. The obtained data will be used for comparison and further analysis.

This study included a large number of staff members (n=867) but a relatively smaller number of students (n=135). The number of participants in each category was sufficient to obtain study results. However, since the number of study participants in the three categories was not well balanced, the results may be biased to some extent. It may be necessary to increase focus on these populations by designing well-being scales for people in each category (ie, one scale for staff, one for faculty, and one for students).

Conclusions

By using a combined approach (a traditional scale development method and crowdsourcing for idea collection at multiple stages of scale development), a highly reliable and comprehensive well-being scale was created for people in the university environment. This scale may be used for reliable well-being assessment in the population of this environment. The results of the well-being assessment may be used to guide the design of well-being improvement interventions.

Acknowledgments

This work was supported in part by funding from Pitt's HealthyU Initiative to the Department of Health Information Management and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant numbers 90RE5018). We gratefully acknowledge three students (Hsin-Chieh Yang, Anne Chen, and Gina Boito) for their work during the well-being scale review and the preparation of statements in the first draft. We would like to thank Pitt students, staff, and faculty for their answers to and valuable comments on the well-being scale.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Other subdomains mentioned by study participants in Q1.
[DOCX File , 13 KB - jmir_v22i4e15075_app1.docx]

References

1. Dictionary.com. 2020. Definition of Well-being URL: https://www.dictionary.com/browse/well-being [accessed 2020-02-15]
2. Edelen MO, Stucky BD, Hansen M, Tucker JS, Shadel WG, Cai L. The PROMIS Smoking Initiative: initial validity evidence for six new smoking item banks. Nicotine Tob Res 2014 Sep;16 Suppl 3:S250-S260 [FREE Full text] [doi: 10.1093/ntr/ntu065] [Medline: 25118232]
3. Patel AA, Donegan D, Albert T. The 36-Item Short Form. J Am Acad Orthop Surg 2007;15(2):126-134. [doi: 10.5435/00124635-200702000-00007]
4. The WHOQOL Group. The World Health Organization quality of life assessment (WHOQOL): Development and general psychometric properties. Social Science & Medicine 1998 Jun;46(12):1569-1585. [doi: 10.1016/s0277-9536(98)00009-4]
5. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics 2000 Jan;17(1):13-35. [doi: 10.2165/00019053-200017010-00002] [Medline: 10747763]
6. Linton M, Dieppe P, Medina-Lara A. Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. BMJ Open 2016 Dec 07;6(7):e010641 [FREE Full text] [doi: 10.1136/bmjopen-2015-010641] [Medline: 27388349]
7. Peterson KA, Zhou L, Watzlaf VJM. A Comprehensive Review of Quality of Life Surveys for Trauma-Affected Communities. Perspect Health Inf Manag 2019;16(Winter):1e [FREE Full text] [Medline: 30766456]

8. Cooke PJ, Melchert TP, Connor K. Measuring Well-Being: A Review of Instruments. The Counseling Psychologist 2016 Aug 05;44(5):730-757. [doi: 10.1177/0011000016633507]
9. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Oct;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
10. Loe BS, Stillwell D, Gibbons C. Computerized Adaptive Testing Provides Reliable and Efficient Depression Measurement Using the CES-D Scale. J Med Internet Res 2017 Dec 20;19(9):e302 [FREE Full text] [doi: 10.2196/jmir.7453] [Medline: 28931496]
11. Love MF, Sharrief A, Chaoul A, Savitz S, Beauchamp JES. Mind-Body Interventions, Psychological Stressors, and Quality of Life in Stroke Survivors: A systematic review. Stroke 2019 Feb;50(2):434-440. [doi: 10.1161/STROKEAHA.118.021150] [Medline: 30612536]
12. Gershon R, Cella D, Dineen K, Rosenbloom S, Peterman A, Lai J. Item response theory and health-related quality of life in cancer. Expert Rev Pharmacoecon Outcomes Res 2003 Dec;3(6):783-791. [doi: 10.1586/14737167.3.6.783] [Medline: 19807355]
13. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults. American Journal of Preventive Medicine 1998 May;14(4):245-258. [doi: 10.1016/s0749-3797(98)00017-8]
14. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH roadmap cooperative group during its first two years. Med Care 2007;45(5 Suppl 1):S3-S11. [doi: 10.1097/01.mlr.0000258615.42478.55]
15. Forrest CB, Bevans KB, Tucker C, Riley AW, Ravens-Sieberer U, Gardner W, et al. Commentary: the patient-reported outcome measurement information system (PROMIS®) for children and youth: application to pediatric psychology. J Pediatr Psychol 2012 Jul;37(6):614-621 [FREE Full text] [doi: 10.1093/jpepsy/jss038] [Medline: 22362923]
16. Evans TM, Bira L, Gastelum JB, Weiss LT, Vanderford NL. Evidence for a mental health crisis in graduate education. Nat Biotechnol 2018 Mar 06;36(3):282-284. [doi: 10.1038/nbt.4089] [Medline: 29509732]
17. The Graduate Assembly. University of California, Berkeley. 2014. Graduate Student Happiness & Well-Being Report URL: http://ga.berkeley.edu/wp-content/uploads/2015/04/wellbeingreport_2014.pdf [accessed 2020-02-15]
18. Nature Editor. What to do to improve postgraduate mental health. Nature 2018 May;557(7704):140. [doi: 10.1038/d41586-018-05074-4] [Medline: 29743703]
19. Reay D. I'd whisper to my student self: you are not alone. Nature 2018 May 8;557(7704):160-161. [doi: 10.1038/d41586-018-05080-6] [Medline: 29740141]
20. Powell K. Work–life balance: Break or burn out. Nature 2017 May 18;545(7654):375-377. [doi: 10.1038/nj7654-375a]
21. Kenney WL. A mentor's acid test. Nature 2017 May 18;545(7654):377-377. [doi: 10.1038/nj7654-377a]
22. Woolston C. Graduate survey: A love–hurt relationship. Nature 2017 Oct 26;550(7677):549-552. [doi: 10.1038/nj7677-549a]
23. Deliens T, Deforche B, De Bourdeaudhuij I, Clarys P. Determinants of physical activity and sedentary behaviour in university students: a qualitative study using focus group discussions. BMC Public Health 2015 Feb 28;15:201 [FREE Full text] [doi: 10.1186/s12889-015-1553-4] [Medline: 25881120]
24. Salzwedel A, Rabe S, Zahn T, Neuwirth J, Eichler S, Haubold K, et al. User Interest in Digital Health Technologies to Encourage Physical Activity: Results of a Survey in Students and Staff of a German University. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e51 [FREE Full text] [doi: 10.2196/mhealth.7192] [Medline: 28428156]
25. Quinn PD, Fromme K. Alcohol use and related problems among college students and their noncollege peers: the competing roles of personality and peer influence. J Stud Alcohol Drugs 2011 Jul;72(4):622-632 [FREE Full text] [doi: 10.15288/jsad.2011.72.622] [Medline: 21683044]
26. Bendtsen M, Bendtsen P. Feasibility and user perception of a fully automated push-based multiple-session alcohol intervention for university students: randomized controlled trial. JMIR Mhealth Uhealth 2014 Jun;2(2):e30 [FREE Full text] [doi: 10.2196/mhealth.3233] [Medline: 25098296]
27. McCall HC, Richardson CG, Helgadottir FD, Chen FS. Evaluating a Web-Based Social Anxiety Intervention Among University Students: Randomized Controlled Trial. J Med Internet Res 2018 Dec 21;20(3):e91 [FREE Full text] [doi: 10.2196/jmir.8630] [Medline: 29563078]
28. Palacios JE, Richards D, Palmer R, Coudray C, Hofmann SG, Palmieri PA, et al. Supported Internet-Delivered Cognitive Behavioral Therapy Programs for Depression, Anxiety, and Stress in University Students: Open, Non-Randomised Trial of Acceptability, Effectiveness, and Satisfaction. JMIR Ment Health 2018 Dec 14;5(4):e11467 [FREE Full text] [doi: 10.2196/11467] [Medline: 30552076]
29. Uhlorn BL. Science. 2019 Oct 10. How I made peace with my imperfections in grad school URL: https://www.sciencemag.org/careers/2019/10/how-i-made-peace-my-imperfections-grad-school [accessed 2020-02-15]
30. Sohn E. How to handle the dark days of depression. Nature 2018 May 8;557(7704):267-269. [doi: 10.1038/d41586-018-05088-y] [Medline: 29740142]
31. Lindert J, Bain PA, Kubzansky LD, Stein C. Well-being measurement and the WHO health policy Health 2010: systematic review of measurement scales. Eur J Public Health 2015 Jan 22;25(4):731-740. [doi: 10.1093/eurpub/cku193] [Medline: 25616594]

32. Rath T, Harter J. Wellbeing: The Five Essential Elements. New York, NY: Gallup Press; May 2010.
33. Revicki DA, Cook KF, Amtmann D, Harnam N, Chen W, Keefe FJ. Exploratory and confirmatory factor analysis of the PROMIS pain quality item bank. Qual Life Res 2014 Feb;23(1):245-255. [doi: 10.1007/s11136-013-0467-9] [Medline: 23836435]
34. Edelen MO, Tucker JS, Shadel WG, Stucky BD, Cerully J, Li Z, et al. Development of the PROMIS Health Expectancies of Smoking Item Banks. Nicotine Tob Res 2014 Aug 12;16(Suppl 3):S223-S231. [doi: 10.1093/ntr/ntu053] [Medline: 25118229]
35. Pilkonis PA, Yu L, Dodds NE, Johnston KL, Maihoefer CC, Lawrence SM. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. J Psychiatr Res 2014 Sep;56:112-119 [FREE Full text] [doi: 10.1016/j.jpsychires.2014.05.010] [Medline: 24931848]
36. Créquit P, Mansouri G, Benchoufi M, Vivot A, Ravaud P. Mapping of Crowdsourcing in Health: Systematic Review. J Med Internet Res 2018 May 15;20(5):e187 [FREE Full text] [doi: 10.2196/jmir.9330] [Medline: 29764795]
37. Gottlieb A, Hoehndorf R, Dumontier M, Altman RB. Ranking adverse drug reactions with crowdsourcing. J Med Internet Res 2015 Mar 23;17(3):e80 [FREE Full text] [doi: 10.2196/jmir.3962] [Medline: 25800813]
38. Washington P, Kalantarian H, Tariq Q, Schwartz J, Dunlap K, Chrisman B, et al. Validity of Online Screening for Autism: Crowdsourcing Study Comparing Paid and Unpaid Diagnostic Tasks. J Med Internet Res 2019 May 23;21(5):e13668 [FREE Full text] [doi: 10.2196/13668] [Medline: 31124463]
39. Zide M, Caswell K, Peterson E, Aberle DR, Bui AA, Arnold CW. Consumers' Patient Portal Preferences and Health Literacy: A Survey Using Crowdsourcing. JMIR Res Protoc 2016 Jun 08;5(2):e104 [FREE Full text] [doi: 10.2196/resprot.5122] [Medline: 27278634]
40. Bardos J, Friedenthal J, Spiegelman J, Williams Z. Cloud Based Surveys to Assess Patient Perceptions of Health Care: 1000 Respondents in 3 days for US $300. JMIR Res Protoc 2016 Aug 23;5(3):e166 [FREE Full text] [doi: 10.2196/resprot.5772] [Medline: 27554915]
41. Dainty KN, Vaid H, Brooks SC. North American Public Opinion Survey on the Acceptability of Crowdsourcing Basic Life Support for Out-of-Hospital Cardiac Arrest With the PulsePoint Mobile Phone App. JMIR Mhealth Uhealth 2017 May 17;5(5):e63 [FREE Full text] [doi: 10.2196/mhealth.6926] [Medline: 28526668]
42. DePalma MT, Rizzotti MC, Branneman M. Assessing Diabetes-Relevant Data Provided by Undergraduate and Crowdsourced Web-Based Survey Participants for Honesty and Accuracy. JMIR Diabetes 2017 Jul 12;2(2):e11 [FREE Full text] [doi: 10.2196/diabetes.7473] [Medline: 30291072]
43. Zhou J, Bell D, Nusrat S, Hingle M, Surdeanu M, Kobourov S. Calorie Estimation From Pictures of Food: Crowdsourcing Study. Interact J Med Res 2018 Nov 05;7(2):e17 [FREE Full text] [doi: 10.2196/ijmr.9359] [Medline: 30401671]
44. Wong AY, Lauridsen HH, Samartzis D, Macedo L, Ferreira PH, Ferreira ML. Global Consensus From Clinicians Regarding Low Back Pain Outcome Indicators for Older Adults: Pairwise Wiki Survey Using Crowdsourcing. JMIR Rehabil Assist Technol 2019 Jan 15;6(1):e11127 [FREE Full text] [doi: 10.2196/11127] [Medline: 30664493]
45. Nguyen VT, Young B, Ravaud P, Naidoo N, Benchoufi M, Boutron I. Overcoming Barriers to Mobilizing Collective Intelligence in Research: Qualitative Study of Researchers With Experience of Collective Intelligence. J Med Internet Res 2019 Jul 02;21(7):e13792 [FREE Full text] [doi: 10.2196/13792] [Medline: 31267977]
46. International Organization for Standardization. ISO 9241-210: Ergonomics of human-system interaction. 2019. Part 210: Human-centred design for interactive systems URL: https://www.iso.org/standard/77520.html [accessed 2019-01-15]
47. Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. Health Qual Life Outcomes 2011 Feb 04;9:8 [FREE Full text] [doi: 10.1186/1477-7525-9-8] [Medline: 21294858]
48. Landauer TK. Behavioral research methods in human-computer interaction. In: Helander MG, Landauer TK, Prabhu PV, editors. Handbook of Human-Computer Interaction. Amsterdam, Netherlands: Elsevier BV; 1997:203-227.
49. Brown JD. Choosing the right type of rotation in PCA and EFA. Shiken: JALT Testing & Evaluation SIG Newsletter 2009;13(3):20-25.
50. Costello A, Osborn J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation 2005;10(7):1-9. [doi: 10.4135/9781412995627.d8]
51. Tabachnick B, Fidell L. Using Multivariate Statistics. Boston: Allyn and Bacon; 2001.
52. Zhou L, DeAlmeida D, Parmanto B. Applying a User-Centered Approach to Building a Mobile Personal Health Record App: Development and Usability Study. JMIR Mhealth Uhealth 2019 Jul 05;7(7):e13194 [FREE Full text] [doi: 10.2196/13194] [Medline: 31278732]
53. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and Facilitators of the Use of Mobile Health Apps From a Security Perspective: Mixed-Methods Study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
54. Wang E, Zhou L, Chen SK, Hill K, Parmanto B. Development and evaluation of a mobile AAC: a virtual therapist and speech assistant for people with communication disabilities. Disabil Rehabil Assist Technol 2018 Nov;13(8):731-739. [doi: 10.1080/17483107.2017.1369592] [Medline: 28949264]

55. Zhou L, Parmanto B, Joshi J. Development and Evaluation of a New Security and Privacy Track in a Health Informatics Graduate Program: Multidisciplinary Collaboration in Education. JMIR Med Educ 2018 Dec 21;4(2):e19 [FREE Full text] [doi: 10.2196/mededu.9081] [Medline: 30578227]
56. Zhou L, Saptono A, Setiawan IM, Parmanto B. Making Self-Management Mobile Health Apps Accessible to People With Disabilities: Qualitative Single-Subject Study. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e15060 [FREE Full text] [doi: 10.2196/15060] [Medline: 31899453]
57. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): Development and Validation Study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
58. Wang E, Zhou L, Chen SK, Hill K, Parmanto B. An mHealth Platform for Supporting Clinical Data Integration into Augmentative and Alternative Communication Service Delivery: User-Centered Design and Usability Evaluation. JMIR Rehabil Assist Technol 2018 Jul 24;5(2):e14 [FREE Full text] [doi: 10.2196/rehab.9009] [Medline: 30042092]
59. Zhou L, Parmanto B, Alfikri Z, Bao J. A Mobile App for Assisting Users to Make Informed Selections in Security Settings for Protecting Personal Health Data: Development and Feasibility Study. JMIR Mhealth Uhealth 2018 Dec 11;6(12):e11210 [FREE Full text] [doi: 10.2196/11210] [Medline: 30538088]
60. Zhou L, Thieret R, Watzlaf V, Dealmeida D, Parmanto B. A Telehealth Privacy and Security Self-Assessment Questionnaire for Telehealth Providers: Development and Validation. Int J Telerehabil 2019 Jun 12;11(1):3-14 [FREE Full text] [doi: 10.5195/ijt.2019.6276] [Medline: 31341542]
61. Parmanto B, Lewis AN, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016 Jul 01;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
62. Mastboom MJ, Planje R, van de Sande MA. The Patient Perspective on the Impact of Tenosynovial Giant Cell Tumors on Daily Living: Crowdsourcing Study on Physical Function and Quality of Life. Interact J Med Res 2018 Feb 23;7(1):e4 [FREE Full text] [doi: 10.2196/ijmr.9325] [Medline: 29475829]
63. Lalor JP, Woolf B, Yu H. Improving Electronic Health Record Note Comprehension With NoteAid: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Crowdsourced Workers. J Med Internet Res 2019 Jan 16;21(1):e10793 [FREE Full text] [doi: 10.2196/10793] [Medline: 30664453]
64. Nama N, Sampson M, Barrowman N, Sandarage R, Menon K, Macartney G, et al. Crowdsourcing the Citation Screening Process for Systematic Reviews: Validation Study. J Med Internet Res 2019 Apr 29;21(4):e12953 [FREE Full text] [doi: 10.2196/12953] [Medline: 31033444]
65. Wu D, Huang W, Zhao P, Li C, Cao B, Wang Y, et al. A Crowdsourced Physician Finder Prototype Platform for Men Who Have Sex with Men in China: Qualitative Study of Acceptability and Feasibility. JMIR Public Health Surveill 2019 Oct 08;5(4):e13027 [FREE Full text] [doi: 10.2196/13027] [Medline: 31596245]
66. Mäkikangas A, Kinnunen U, Feldt T, Schaufeli W. The longitudinal development of employee well-being: a systematic review. Work & Stress 2016 Jan 12;30(1):46-70. [doi: 10.1080/02678373.2015.1126870]
67. Danna K, Griffin RW. Health and Well-Being in the Workplace: A Review and Synthesis of the Literature. Journal of Management 1999 Jun;25(3):357-384. [doi: 10.1177/014920639902500305] [Medline: 22973420]
68. Jarden RJ, Sandham M, Siegert RJ, Koziol-McLain J. Quality appraisal of workers' wellbeing measures: a systematic review protocol. Syst Rev 2018 Dec 20;7(1):240 [FREE Full text] [doi: 10.1186/s13643-018-0905-4] [Medline: 30572952]
69. Di Castro VC, Hernandes JC, Mendonça ME, Porto CC. Life satisfaction and positive and negative feelings of workers: a systematic review protocol. Syst Rev 2018 Dec 23;7(1):243 [FREE Full text] [doi: 10.1186/s13643-018-0903-6] [Medline: 30579364]

Abbreviations

Pitt: University of Pittsburgh
PROMIS: Patient Reported Outcomes Measurement Information System

Edited by G Eysenbach; submitted 17.06.19; peer-reviewed by K Gordon, J Koepp, M Oroojeni Mohammad Javad; comments to author 30.09.19; revised version received 23.02.20; accepted 19.03.20; published 29.04.20.

Please cite as:
Zhou L, Parmanto B
Development and Validation of a Comprehensive Well-Being Scale for People in the University Environment (Pitt Wellness Scale) Using a Crowdsourcing Approach: Cross-Sectional Study
J Med Internet Res 2020;22(4):e15075
URL: http://www.jmir.org/2020/4/e15075/
doi:10.2196/15075
PMID:32347801

Original Paper

Developing a Competency-Based Learning and Assessment System for Residency Training: Analysis Study of User Requirements and Acceptance

Cheng-Ting Hsiao[1,2,3], MD; Fremen ChihChen Chou[4,5,6], MD; Chih-Cheng Hsieh[7], MD; Li Chun Chang[8], MD; Chih-Ming Hsu[3,9], MD

[1]Department of Emergency Medicine, Chang Gung Memorial Hospital, Puzi City, Chiayi County, Taiwan
[2]School of Medicine, Chang Gung University, Taoyuan, Taiwan
[3]Chang Gung Medical Education Research Centre, Taoyuan, Taiwan
[4]Center for Faculty Development, Department of Emergency Medicine, China Medical University Hospital, Taichung, Taiwan
[5]School of Medicine, China Medical University, Taichung, Taiwan
[6]Department of Emergency Medicine, China Medical University Hospital, Taichung, Taiwan
[7]Management, Chang Gung Memorial Hospital, Yunlin, Taiwan
[8]Nursing, Chang Gung University of Science and Technology, Taoyuan, Taiwan
[9]Medical Education Department, Chang Gung Memorial Hospital, Puzi City, Chiayi County, Taiwan

Corresponding Author:
Chih-Ming Hsu, MD
Medical Education Department
Chang Gung Memorial Hospital
No 6, Sec West
Jiapu Road
Puzi City, Chiayi County, 61363
Taiwan
Phone: 886 5 3621000 ext 3128
Email: kan200068@cgmh.org.tw

Abstract

Background: The increasingly complex medical environment highlights the importance of milestones and entrustable professional activities (EPAs) to realize the ideals of competency-based medical education (CBME). However, if enormous amounts of assessment results need to be compiled, the development of a digital system to manage, integrate, and synthesize learning and assessment data will be necessary. Furthermore, this system should be able to facilitate real-time assessment with feedback and therefore enhance users' learning through coaching in the moment in the clinical workplace.

Objective: The main purpose of this study was to develop a competency-based electronic platform system to provide resident physicians with clinical assessments and learning in order to enhance the learning of trainees and reduce the burden of assessments.

Methods: A competency-based learning and assessment system (CBLAS) for residency training was designed, developed, and evaluated in this study. Opinion interviews and a focus group consensus meeting of key users, including trainees, clinical teachers, and administrative staff, were conducted as needs assessments. The structure of the CBLAS was designed according to the thematic analysis of needs assessments. Clinical teachers' acceptance of using CBME assessments, according to the constructs of attitude, perceived usefulness, and perceived ease of use, was surveyed in the beginning and half a year after implementation of the CBLAS. Additionally, the satisfaction of using the CBLAS, according to information, system, and service qualities, was surveyed after implementation.

Results: The main functions of the CBLAS, including milestones, EPAs, learning portfolios, teacher/student feedback, e-books, learning materials, assessment progress tracking, and statistical analysis of assessment results, were designed and developed for responding to nine themes, which emerged from the needs assessments of the three user groups. Twenty clinical teachers responded to the CBME assessment acceptance surveys before and after CBLAS implementation, which revealed a significant improvement in the factor of "attitude" (P=.02) but no significant differences in the two factors of "usefulness" (P=.09) and "ease of use" (P=.58) for CBME assessments. Furthermore, satisfaction surveys were performed in 117 users, and 87.2% (102/117) were

satisfied with the CBLAS in terms of information, system, and service qualities. There was no significant difference in satisfaction among different user groups.

Conclusions: The CBLAS is a user-centered platform that supports clinical teachers' assessment exercises and residents' learning, as well as administrative work for staff according to users' needs assessments and operationalized features of CBME assessments. With the system, clinical teachers had a more positive attitude to conduct the assessment activities of milestones and EPAs and learners could arrange their study schedules to enhance their learning effectiveness. The CBLAS sheds light on how to effectively design and develop a digital system to execute milestone- and EPA-based assessments for enhancing competency-based education among residents, according to our experiences in Taiwan.

(J Med Internet Res 2020;22(4):e15655) doi:10.2196/15655

KEYWORDS

competency-based learning; milestones; entrustable professional activities; assessment; learning platform

Introduction

Background

Medical education systems require continual improvement if they are to react effectively to the challenges of rapid changes in technology and the sociocultural environment of the 21st century [1]. To address the medicine-related challenges at hand, equipping doctors with essential competencies has been strongly emphasized in competency-based medical education (CBME) [2]. CBME focuses on practice-based competencies, which have been analyzed and developed according to the needs of the society and patients. CBME does not emphasize time-based training. Instead, it promises greater accountability, flexibility, and learner-centeredness [3]. However, to enable implementation of CBME in the usual clinical teaching and training assessment routines, the most influential and persuasive specific practices are those of "entrustable professional activities (EPAs)" proposed by the Dutch scholar ten Cate [4] and "milestones," which were planned by the Accreditation Council for Graduate Medical Education (ACGME) [5].

Competency-Based Medical Education

Medical education in the United States and Canada, as reported by Flexner in 1910, spurred a reform movement, which laid the foundation of a structure- and procedure-oriented medical education model [6]. In this previous report, Flexner mentioned that medical students should have a foundation in elementary science before admission to medical schools. In addition, medical schools should provide students with clinical training, laboratory training, full-time teachers, and a teaching hospital. It was not until 2010 that the Carnegie Foundation for the Advancement of Teaching published "Educating Physicians," which was like the second version of the report by Flexner. This report clearly outlined the following five major plans for medical education reform in the 21st century: (1) standardization of learning outcomes and individualization of the learning process; (2) integration of formal knowledge and clinical experience; (3) establishment of a holistic medicine and health care team; (4) self-learning and growth; and (5) professionalism and recognition. These reform plans resonate with the CBME model.

CBME, proposed by the World Health Organization in the 1970s, aims to educate "health professionals who can practice medicine at a defined level of proficiency, in accord with local conditions, to meet local needs [7]." After the Institute of Medicine published its report "Crossing the Quality Chasm" in 2001, this important educational issue re-emerged and was widely discussed. CBME is seen as a different teaching method, as it predefines the competencies medical students should acquire before graduation [8] and suggests different methods of evaluating educational outcomes [9]. To identify desirable training results, the concept of "progression of competence" was created [10]. As CBME emphasizes learner-centeredness in students, learners enjoy greater autonomy and flexibility in scheduling their curricula. Moreover, the clear learning goals and the transparent process differ from traditional medical training. Thus, this educational model can better meet the needs of both the society and patients [2,11-13].

Entrustable Professional Activities

An EPA, proposed by ten Cate in 2005 [4], is defined as unsupervised task execution once the learner has attained a sufficient level of competency. During the assessment activity, the clinical teacher should observe and assess the trainee's performance and give feedback accordingly [14]. The term "entrustable" was coined to emphasize the importance of "trust" and to reinforce an "entrustment decision" as well [15]. Entrustable competencies are an important aspect of the doctor-patient relationship. Patients visit doctors for treatment because they trust them. In addition, when teachers teach students to attain a certain proficiency level, they also "trust" them to be able to practice their skills in a professional task independently [5]. EPAs embody the core competencies that residents must learn. Through assessment activities when performing a task, residents improve their proficiency of relevant competencies in a professional context, which ensures the realization of CBME and advances medical education.

Milestones

The concept of residency milestones originated with Nasca, the chief executive officer of the ACGME in 2008, according to the guidelines of which the following two points were mentioned: (1) development of milestones for each specialty and (2) development of assessment tools for milestones [16]. Therefore, a milestone implies an achievement or behavior presented by a physician who has the competency to carry out EPAs [17]. Compared with the current residency training program, the competencies emphasized with milestones are closer to the competencies required by a specific specialty. Milestones specify the core knowledge, skills, attitudes, and

beliefs a medical specialist should acquire; moreover, it could also be applied to map out the progression of a medical student from admission to graduation.

The concept of milestones to describe a learner's progress and performance originated from the skill acquisition model proposed by Dreyfus and Dreyfus [18]. Moreover, clinical teachers use milestones as elements of a standardized assessment tool to indicate a learner's training outcome. The Milestone Project can help achieve the goals set out in the Outcome Project by serving as indicators of the quality of a training project and the progress of a learner. From that, we know that developing core competency–based milestones is a key step in realizing the spirit of CBME [19].

As patients' conditions vary, assessments based on diverse content, several types of provisions, and multiple dimensions can be helpful to teachers in keeping them well informed of their students' learning conditions. Furthermore, teachers can use the digital assessment platform to conduct assessments as part of their duty, according to which they can provide useful feedback. They can also simultaneously encourage students to self-reflect and direct them toward revising their learning plans. As a result, students' competencies can be developed [20].

Berz et al pointed out in their study that many teachers recognized that compared with milestone-based assessment tools, traditional tools are too narrow in scope, while also being inconvenient to use in the continuity of a clinical setting [21]. The development of a new assessment tool could enhance the quality and quantity of the assessment, as well as the satisfaction of users [21].

However, the data produced from assessments required for training programs based on milestones and EPAs are enormous, and compiling such a large amount of data would consume much time and resources. As a result, the implementations of milestones and EPAs have been hindered owing to the restricted medical and administrative resources. Many studies on related literature point out that the digitization of CBME will be critical for its future development. Lockyer et al noted that effective assessment is essential for the implementation of CBME [22]. Further, effective assessment needs effective information management and documentation, as well as continuous improvements to the assessment system [22]. Additionally, Schumacher et al claimed that as effective assessment requires huge amounts of data, highly effective data management is critically important [23]. Therefore, this work needs to be supported by digital assessments and mobile apps.

The collection and management of enormous assessments in the clinical environment need technology support, and the synthesis of data to inform the "progression of competence" of a trainee relies much on information technology. In the United States, according to the requirement of the ACGME, the Clinical Competency Committee is responsible for collecting data on the assessment results of residents every 6 months and holding discussions to decide each resident's proficiency level and final training outcome [24]. Analysis and discussion to synthesize the data without the assistance of a digital system would be a formidable challenge in many medical environments, such as those in Taiwan, where medical and administrative manpower resources are limited.

For learning whether trainees' proficiency levels have met the training goals, a digital system is required to achieve the following goals: (1) comprehensively digitize the activities of the residency assessments; (2) allow teachers and students access to the system for online learning and feedback, without time and space restrictions; (3) reduce trainees' cognitive load and enhance their learning effectiveness; (4) provide trainees clearer learning goals and resources by integrating the assessment activities and learning into one system; (5) support the operation of a huge amount of statistical analyses and diagrams; (6) conduct assessments on the online system; and (7) create a feedback exchange platform where teachers and trainees can freely share their experiences.

Thus far, in Chinese-speaking areas, only few units have conducted a literature review and begun development of CBME digitization. The development of our Competency-Based Learning and Assessment System (CBLAS) is expected to be the pioneer of CBME digitization. Moreover, as the system was developed in Mandarin Chinese, it could serve as an example of system development for other Chinese-speaking countries.

Considering the aforementioned situations and facts, this study set the following three goals: (1) plan a competency-based and learner-centered training program, objective assessment model, and assessment scale for later integration into the system; (2) design the structure, build a web-based system for residency assessments and learning according to user needs analyses (trainees, teachers, and administrative staff), and implement the system in two medical centers and two regional hospitals in Taiwan; and (3) investigate clinical teachers' acceptance toward CBME assessments and the assessment time before and after implementation of the system and evaluate system satisfaction by different users after implementation.

Methods

Study Methods

The methods of this study included four parts. First, the needs assessments of key users, including trainees, clinical teachers, and administrative staff, for designing the CBLAS were conducted through several opinion interviews and a focus group consensus meeting. Second, the structure of the CBLAS was designed and developed according to needs assessment analyses and consensus. Third, clinical teachers' acceptance (based on the constructs of attitude, perceived usefulness, and perceived ease of use) of using CBME assessments was surveyed in the beginning of the implementation of the CBLAS and the acceptance was followed up with a survey half a year later. Moreover, we sampled five clinical teachers to observe the average time of performing an assessment in the clinical workplace. Fourth, a total of 117 users (three groups including teachers, trainees, and administrative staff) were surveyed for their satisfaction of using the CBLAS, according to information quality, system quality, and service quality.

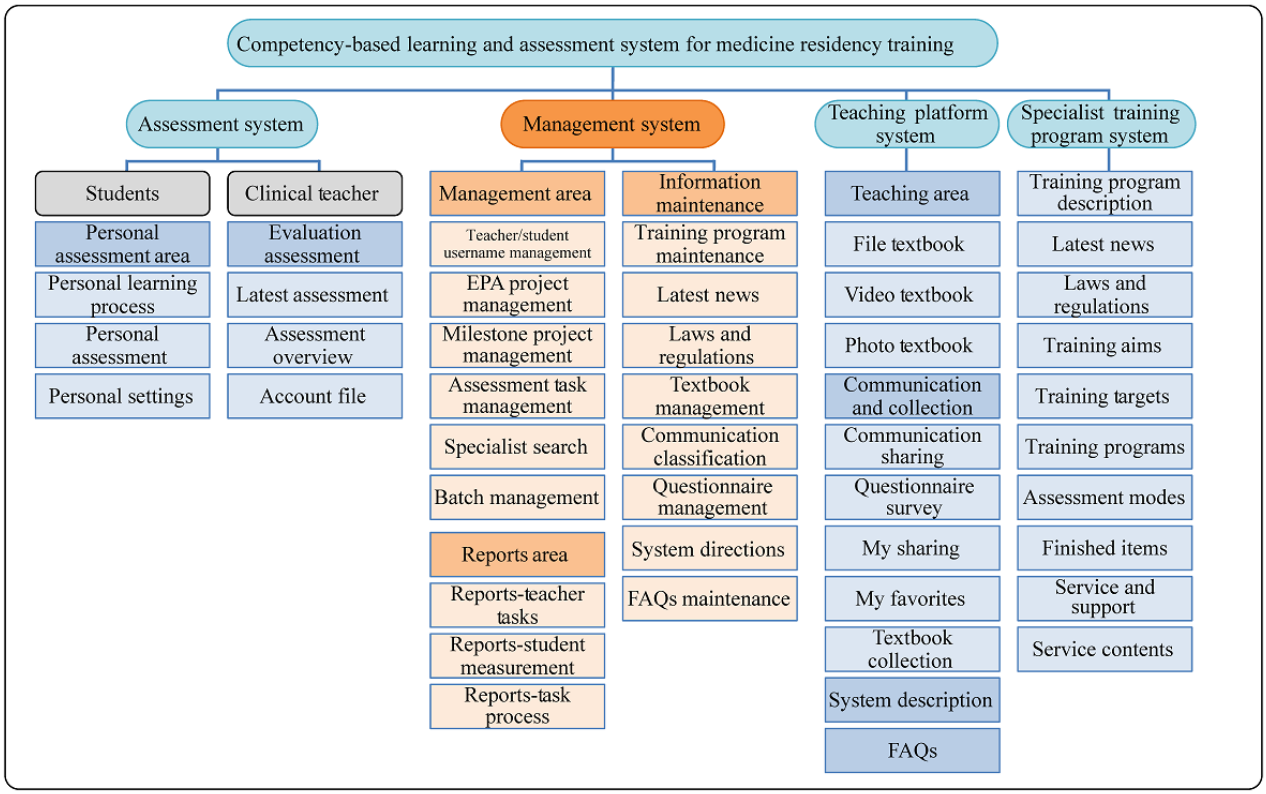
Needs Assessments of Key Users for the Competency-Based Learning and Assessment System

The development of the CBLAS was centered on the users' needs. We collected opinions from clinical teachers, trainees, and administrative staff regarding the difficulties encountered when performing assessments. The challenges included heavy workload with repetitive tasks, assessment items that were often missed out, waste of time and paper resources to print assessments, and difficulties in recovering backend data. Additionally, we invited five clinical teachers, three trainees, and three administrative staff to participate in a user focus group consensus meeting to discuss the possible solutions for CBME assessment challenges with the web-based system. The records of the consensus meeting and previous opinion interviews were further analyzed with a thematic analysis technique, in which the common themes of the users' needs were identified by grouping ideas with commonality. We aspired to build a fast and easy system to meet teachers' and trainees' needs. Through continuous improvements, modifications, and updates, the functions of competency assessment and teaching were expected to be enhanced. As a result, the learning outcomes could eventually be optimized.

Development of the System Structure

Based on the thematic analysis of the users' needs and the literature review, we divided the CBLAS into the following four major functions: (1) training program, where the training plan, training goals, and program plans of a specific specialty can be viewed and questions regarding the training program can be answered; (2) assessment, where milestones and EPAs are used by teachers and the results can be checked by trainees; (3) learning, where learning materials can be provided and teachers and students can communicate; and (4) management, where the control center of the abovementioned three subsystems is located and is only available to administrators and department secretaries (Figure 1). The details of the design have been reported in the Results with system functions and interface design.

Figure 1. System structure.



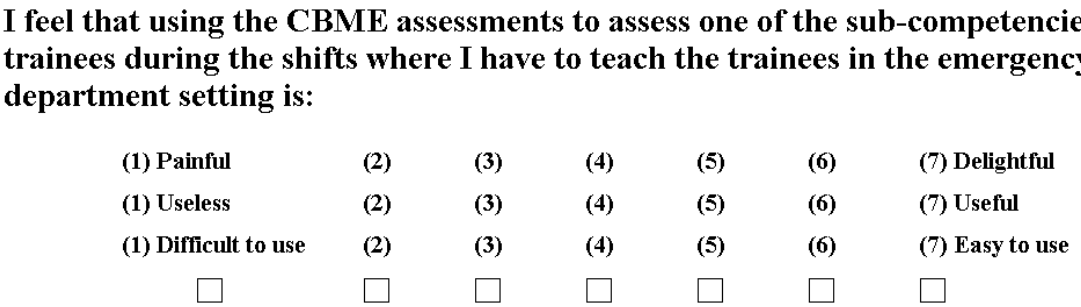
Acceptance Survey for Competency-Based Medical Education Assessments

The success of CBME relies much on frequent assessments with clinical teachers' commitment on purposeful observation and meaningful feedback. Clinical teachers' acceptance of performing CBME assessments would be the key to success. This study applied the constructs in the Technology Acceptance Model, which was proposed by Davis in 1989 [25], to design a survey to investigate clinical teachers' acceptance of CBME assessments. When users are introduced to a new tool, three main factors influence their acceptance decision [26]. The three factors are perceived usefulness, perceived ease of use, and attitude. Perceived usefulness indicates that users believe that they will improve work performance when using a certain tool. Perceived ease of use indicates that users think that using a particular tool will not need too much effort. The construct of attitude in acceptance represents the users' emotional perceptions toward using the tool.

First, we designed the survey with Google Forms to investigate teachers' acceptance of CBME assessments. The survey questions were designed according to three factors, including attitude toward using CBME assessments (three items),

perceived usefulness of CBME assessments (two items), and perceived ease of use of CBME assessments (two items). A 7-point semantic differential scale was used in the questionnaire to rate the variables. An example item of each of the three factors (attitude, usefulness, and ease of use) is shown in Figure 2.

Figure 2. Example item of the questionnaire. CBME: competency-based medical education.



We invited clinical teachers by convenience sampling to respond to the acceptance survey in the beginning of implementation, and they were followed up after 6 months. The paired *t*-test was applied to assess the significance of the difference between the two survey times. We also analyzed the acceptance rate of clinical teachers in terms of each factor, where the percentage of users who rated all items in the factor ≥5 points was considered as the acceptance rate for the factor. Beside clinical teachers' acceptance, we further sampled five clinical teachers to observe how much time they needed to record their feedback after a clinical assessment with a traditional paper tool and with the CBLAS. The average time for inputting a word (a Chinese character) was counted to descriptively provide a sense of the difference in recording qualitative feedback in the clinical workplace between a traditional paper tool and the CBLAS.

Satisfaction Survey for the Competency-Based Learning and Assessment System

We assessed users' satisfaction with the CBLAS after it was implemented for more than half a year. The subjects included clinical teachers, learners, and administrative staff. A questionnaire with a Likert 7-point scale was used in the survey for users to rate their perceived satisfaction in terms of information quality, system quality, and service quality. In the scale, 1 point indicated strongly disagree, 4 points indicated undecided, and 7 points indicated strongly agree with the statement of the question item. Example items of the three factors are as follows: (1) information quality, I feel that the information provided by the CBLAS fits my needs; (2) system quality, I feel that the CBLAS is stable and does not easily crash; (3) service quality, I feel satisfied that CBLAS maintenance personnel can respond to my needs in real time when I encounter problems with the CBLAS.

One-way analysis of variance was applied to analyze the difference in satisfaction between different user groups. When analyzing the satisfaction percentage of the users for a quality factor, we considered scores of all items in the factor >5 points to indicate that users were satisfied with the CBLAS for that factor.

Results

Summary of the Needs of Different Key User Groups

After thematic analysis of the records from opinion interviews and a focus group consensus meeting involving five clinical teachers, three trainees, and three administrative staff, we identified the needs that different user groups specified for the CBLAS (Textbox 1).

Textbox 1. Themes of the needs of different key user groups.

Residents (trainees)

1. Learner-centeredness
   - Being able to adjust his/her own learning progress without time and space restrictions
   - Being able to understand and organize future goals through milestones
2. Being able to use different types of devices with the support of responsive web design

Clinical teachers

1. Alleviating time- and space-related problems to improve the delivery of training programs and assessment activities
2. Being fully informed of each trainee's learning progress and needs
3. Being able to provide feedback in real time, communicate, give appropriate assistance, and share knowledge according to the trainee's competency level

Medical Education Department (administrative staff)

1. Reducing the workload in terms of the administrative procedure and large volume of statistical data
2. Reducing paper use and carbon emission
3. Reducing barriers in promoting competency-based training
4. Systematizing records and assessment results

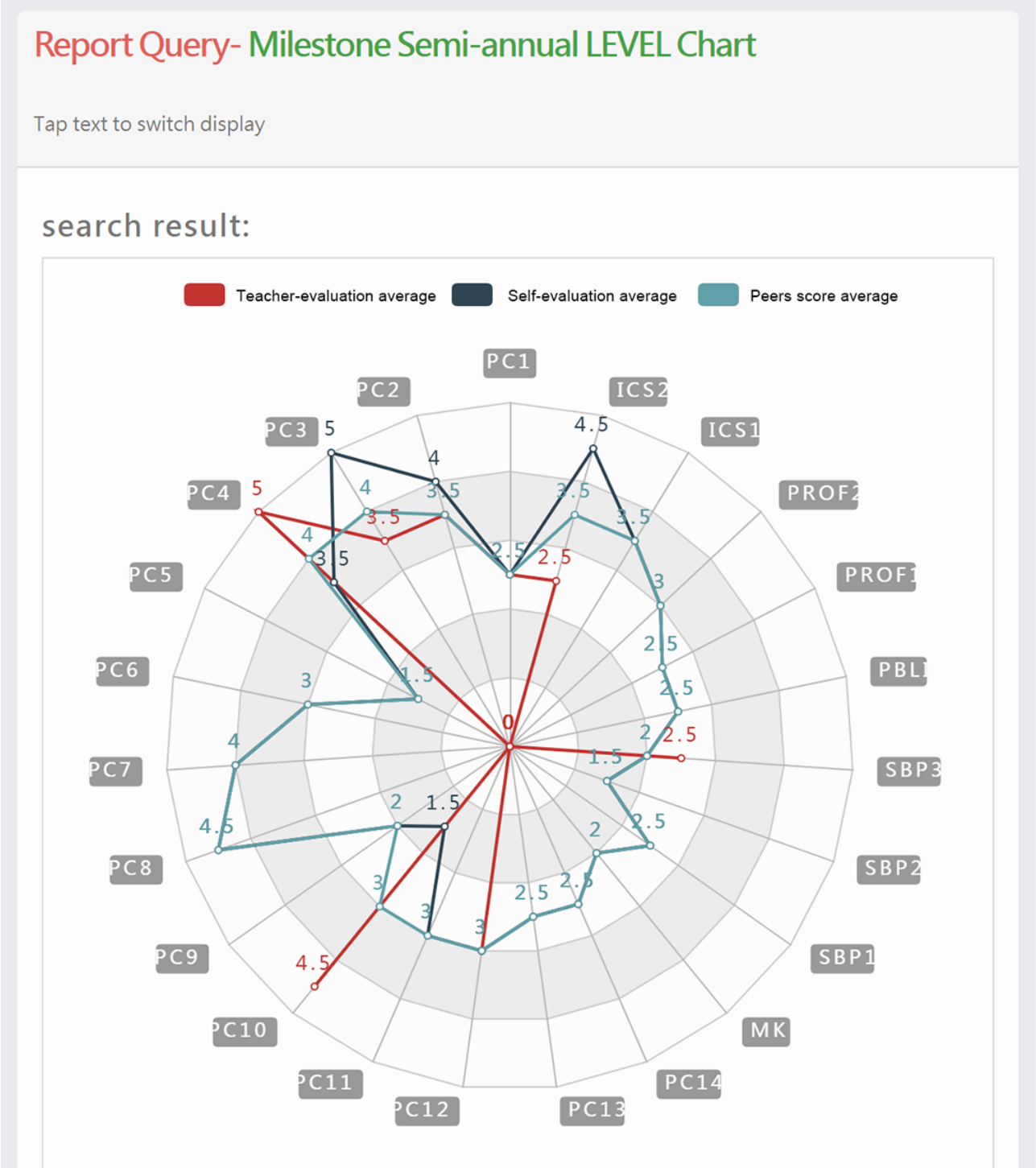
System Functions and Interface Design

Assessment Procedure

The administrator is responsible for creating new personal accounts and managing each batch's trainees. The department secretaries assign clinical teachers to trainees each month for assessment activities, which include shift-based milestones and EPAs. The CBLAS ensures that a certain amount of assessment results for each trainee comes from different teachers' assessments so that the reliability and validity of the assessment results can be ensured.

After data collection, the CBLAS further organizes and analyzes the assessment results. By presenting charts through data visualization, users will become aware of the gap between the expected training goal and the trainee's actual competency level, which will help with his/her further training. In addition, radar charts are produced to show a trainee's proficiency level in each subcompetency for indicating whether the trainee is ahead or behind his/her peers (Figure 3). Furthermore, the radar charts serve as a reference for further training required by the learner if the trainee falls behind, so that the goal of teaching each student according to his/her aptitude can be realized.

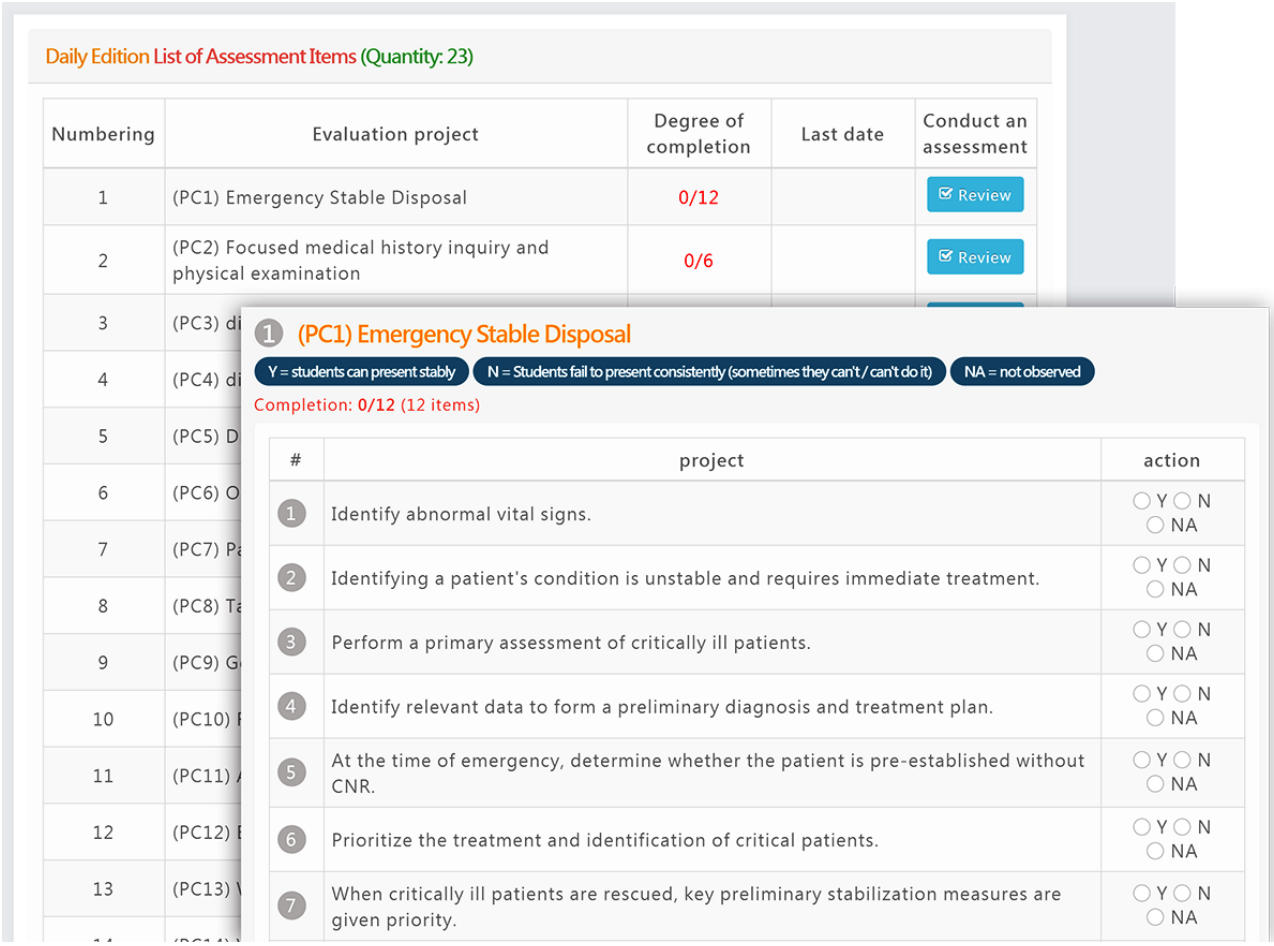
Figure 3. Radar chart.



Shift-Based Milestone Interface

The shift-based milestone interface is trainee-centered. Teachers assess an assigned trainee's performance by clicking on the name of the trainee and then on the subcompetency (Figure 4, left) to execute the milestone assessment and feedback activity (Figure 4, right).

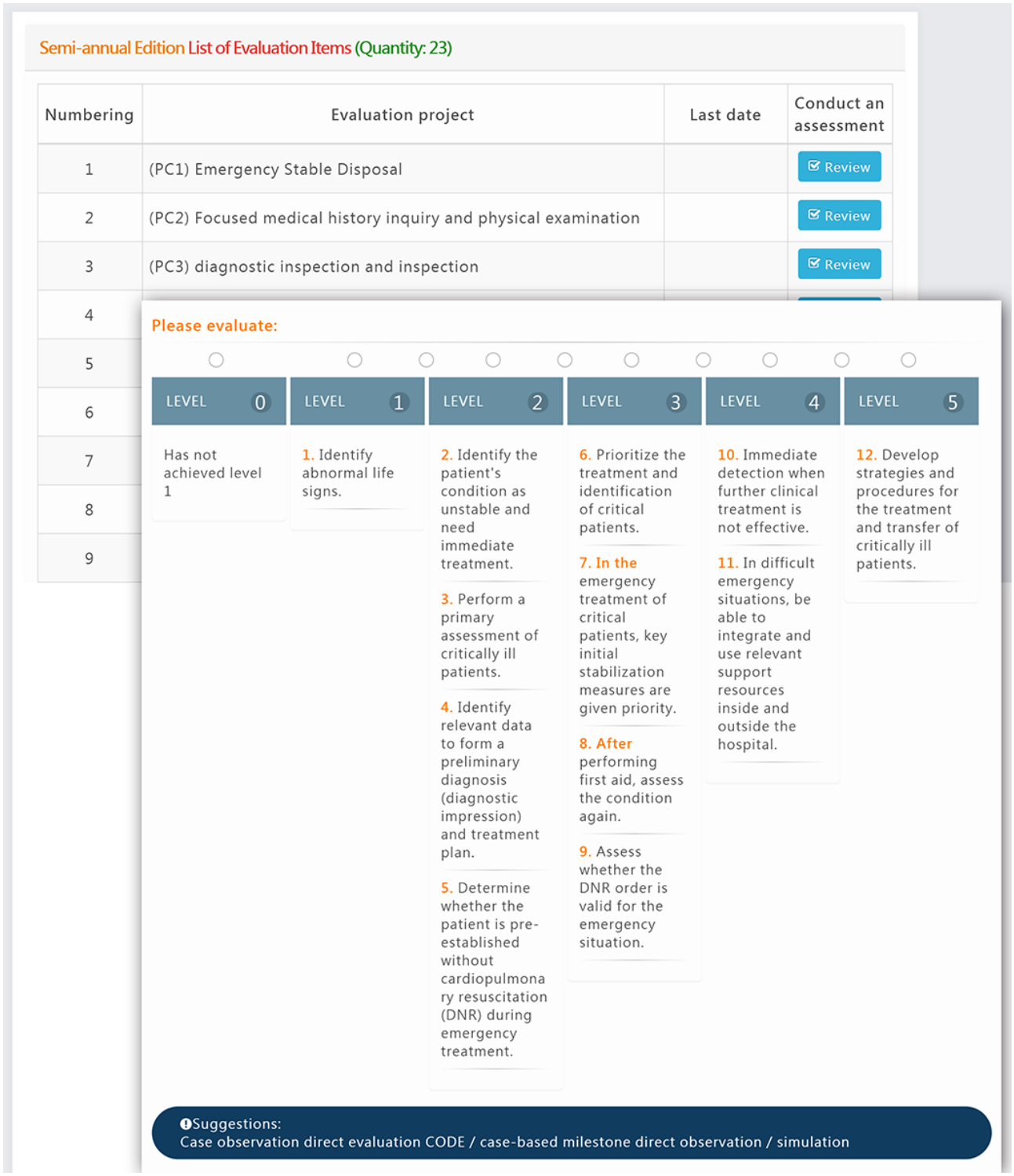
Figure 4. Daily milestone interface.



Half-Year Milestone Interface

The half-year milestone interface is for trainees who have received training for at least 6 months. It is where the assigned teacher provides the assessment to the trainee and the trainee inserts his/her own self-assessment as well. After the teacher clicks on the name of the trainee, the CBLAS shows all 23 subcompetencies to be assessed (Figure 5). The assessment can be saved before the result submission. In other words, the teacher can freely allocate his/her time to complete the assessment activity.

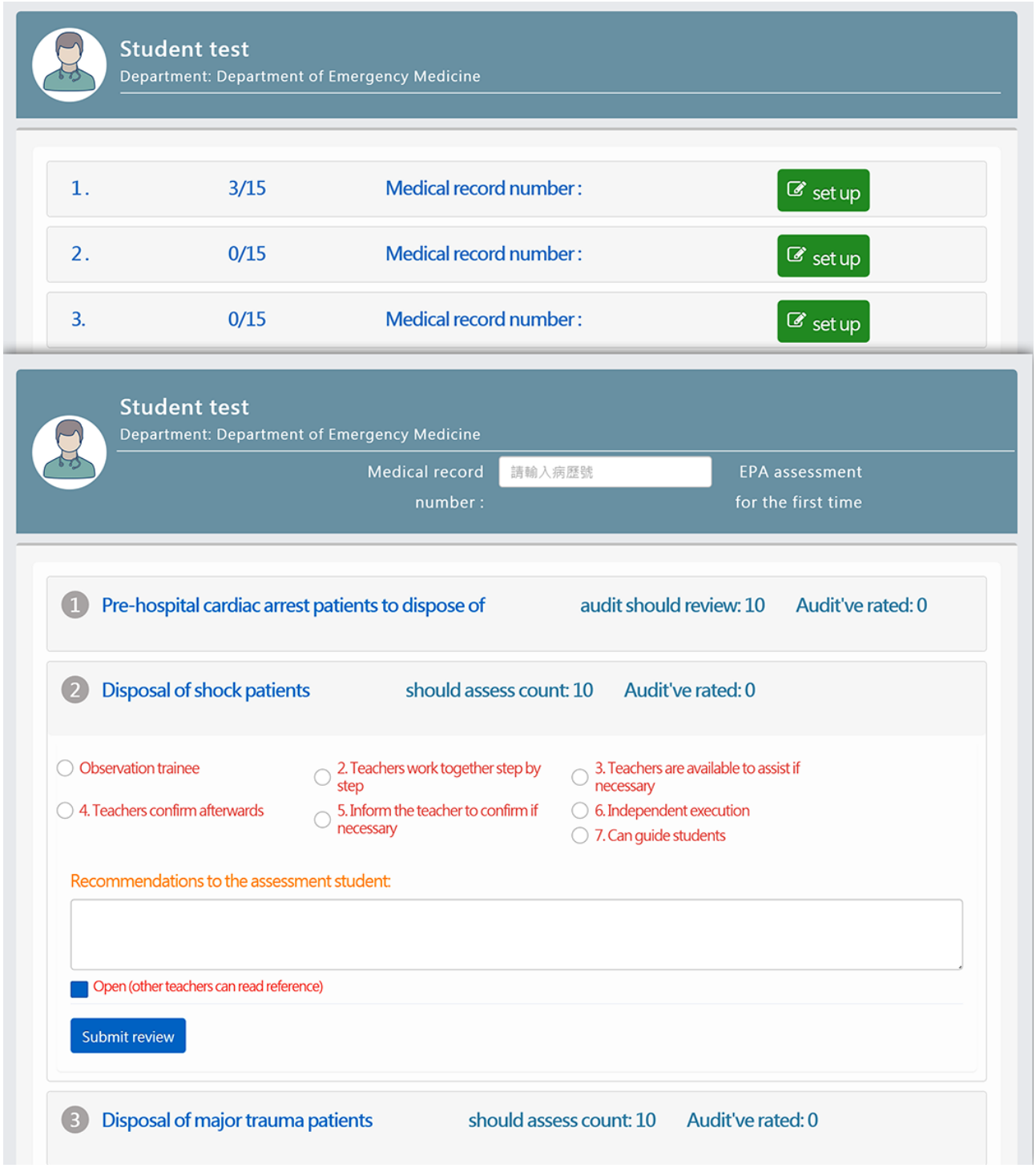
Figure 5. Half-year milestone interface.



Entrustable Professional Activity Assessment Interface

The EPA assessment interface centers on the trainees and serves as a place where they can communicate with their clinical teachers about the EPAs that need to be assessed. After the clinical teachers click on the name of a trainee and enter the medical number to continue to the next page, they can select the EPAs to provide assessment and feedback (Figure 6).

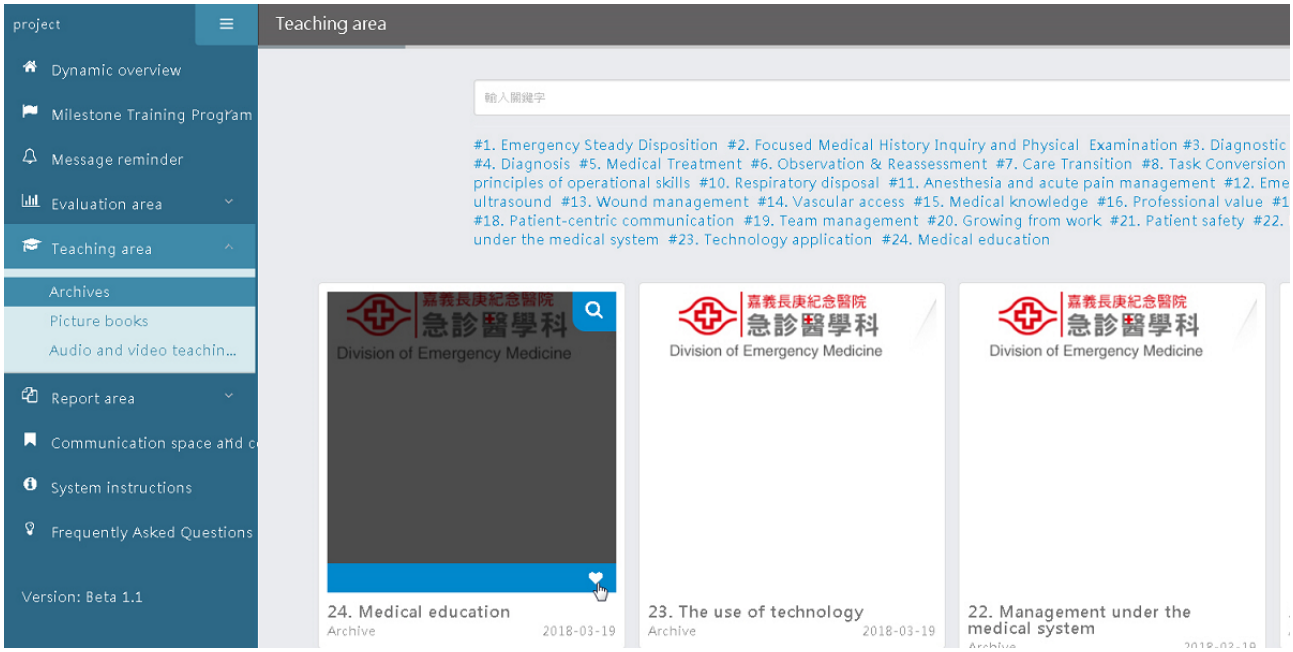
Figure 6. Entrustable professional activity assessment interface.



E-Learning Platform Interface

The e-learning platform interface comprises different specialties, and the learning materials of a specific specialty are shown depending on the trainee's identity, so that his/her cognitive load can be reduced. Learning materials can be divided into the following three major types: documents, pictures, and videos. Trainees can search for materials based on their needs and can add certain materials into "My Favorites," so that they can check them at any later time (Figure 7).

Figure 7. E-learning platform interface.



Instant Notifications and Reminders

Users automatically receive reminders from the CBLAS when they need to provide an assessment. They can simultaneously also receive updates from the CBLAS. Compared with the paper assessments used in the past, clinical teachers have more flexibility in terms of when they need to provide and edit assessments of a trainee's subcompetencies. No assessment sheets need to be prepared in advance, which can enhance the work efficiency of residency competency assessments. Moreover, the results are shown in the CBLAS in real time, which helps users complete direct observations and clinical assessments at any time and ensures that teachers are fully informed of trainees' learning progress. As a result, teaching quality and teaching-related work efficiency can be enhanced.

Clinical Teachers' Perceived Acceptance of Competency-Based Medical Education Assessments Before and After System Implementation

A total of 20 clinical teachers responded to the CBME acceptance survey before and after the implementation of the CBLAS. We found that their attitudes toward the CBLAS significantly improved after implementation (mean score change from 5.15 [SD 0.72] to 5.70 [SD 0.61], P=.02); however, there was not much difference before and after implementation in the perceived usefulness and perceived ease of use for CBME assessments. When we counted the acceptance percentage of the clinical teachers according to their attitude, the acceptance percentage increased from 60% to 90%. The change in the attitude percentage indicated that 30% of the participants changed their attitude from negative or undecided for at least one item in the attitude factor to positive for all items in the attitude factor. The results are presented in Table 1. Beside acceptance, we also randomly sampled five clinical teachers regarding convenience before and after system implementation to observe the time needed for inputting their feedback after an assessment in the emergency department clinical environment. Five different teachers were sampled for the before and after tests. The average time for inputting a word (a Chinese character) reduced from around 2.2 seconds a word by hand writing to 0.5 seconds a word by speech input.

Table 1. Acceptance of competency-based medical education assessments before and after system implementation (n=20).

Assessment	Before implementation		After implementation		*t* value
	Score, mean (SD)	Acceptance percentage[a]	Score, mean (SD)	Acceptance percentage[a]	
Attitude	5.15 (0.72)	60%	5.70 (0.61)	90%	−2.55[b]
Usefulness	6.10 (0.91)	90%	6.05 (0.81)	95%	0.191
Ease of use	6.05 (0.89)	90%	5.90 (0.89)	90%	0.557
Overall acceptance	5.77 (0.73)	—[c]	5.88 (0.66)	—	−0.54

[a]Percentage of users who rated all items in the factor with a score ≥5.

[b]P<.05.

[c]Three factors were designed to evaluate the acceptance of CBME assessments; therefore, the overall acceptence percentage was not counted in this study.

Satisfaction Survey for the Competency-Based Learning and Assessment System

A total of 117 users responded to the satisfaction survey, with 52 responses from teachers, 46 from learners, and 19 from administrative staff. The detailed results are shown in Table 2. All three groups of key users showed high scores for the perceived quality of three factors, with the mean score ranging from 5.90 to 6.47 on the 7-point Likert scale. The three factors included information quality, system quality, and service quality. There was no difference in satisfaction among the three groups of users. The satisfaction percentage for all users regarding the three factors ranged from 87% to 90%.

Table 2. Perceived quality of the competency-based learning and assessment system (n=117).

Perceived quality	User score, mean (SD)			F (ANOVA[a])	*P* value	Percentage of satisfaction[b]
	Clinical teachers	Trainees	Administrative staff			
Information quality	5.90 (1.01)	5.98 (1.16)	6.03 (0.62)	1.05	.36	87%
System quality	6.06 (0.75)	6.07 (0.87)	6.15 (0.58)	0.10	.90	90%
Service quality	6.07 (0.83)	6.05 (1.21)	6.47 (0.47)	1.51	.23	89%
Overall	6.01 (0.82)	6.03 (1.02)	6.30 (0.51)	—[c]	—	—

[a]ANOVA: analysis of variance.

[b]Percentage of users who rated all items in the factor with a score ≥5.

[c]Three factors were designed to evaluate the acceptance of CBME assessments; therefore, the overall acceptence percentage was not counted in this study.

Discussion

Principal Findings

In the rapidly changing 21st century, the life cycle of knowledge is getting shorter. Lifelong learning will be a normal approach. The CBLAS mentioned in this article is an assessment system for clinical skills learning. It is designed to connect learning with the real workplace to facilitate competency-based performance to a defined proficiency with frequent assessment and feedback to ensure progression of competence [10,14].

Medical education and training have no resistance to rapidly changing technology. Many pedagogical strategies were developed to facilitate an active learner-centered teaching and learning approach [25]. Information technology platforms and assessment tools can provide unique, timely, cost-effective, and valuable opportunities to improve the outcomes of education and training [27]. Heller et al mentioned that the Competence-Based Knowledge Structure for Personalized Learning system focuses on knowledge learning through multifaceted learning to improve learners' level of knowledge. The assessment takes place when learners finish learning in order to find out whether they have acquired the required knowledge. This e-learning platform emphasizes personalized learning paths and efficient assessment of knowledge and competencies [28]. Zehry et al also mentioned in a 2011 study that these e-learning platforms have several distinct advantages over traditional teaching and learning models, including the ability to update materials in a timely manner for ensuring the ability to provide students with the latest evidence-based content [29]. The CBLAS was designed according to the concept of CBME and the needs of key user groups. This system allows updating of teaching resources, and users can share, search, and discuss on the platform in real time. Thus, it is a valuable resource for updating knowledge. Beyond knowledge and cognition, multiple effective workplace-based assessments and feedback are essential for the implementation of CBME. The CBLAS provides authentic workplace-based assessments and immediate feedback to connect knowledge and clinical practice. We found that clinical teachers' attitudes had greatly improved after the implementation of the CBLAS. The possible reasons are that the design of this system framework meets the needs of clinical teachers and practical experience promotes the understanding of CBME.

Our study results showed that all three groups of key users had high scores for the perceived quality of the CBLAS. CBME is a paradigm shift from traditional assessments to milestone-based assessments. Schumacher et al found that "Data and Assessment System Development" is an important resource needed to implement this new system [23]. Workplace-based assessments of milestones and EPAs will generate large amounts of data for all trainees across a program. These data will need to be collected accurately and effectively. Additionally, the data should be presented and synthesized in a meaningful manner to facilitate interpretation and review in clinical competency committees. The important point is data accuracy. Every assessment result is only stored in the CBLAS after being confirmed by the clinical teacher. If some assessment items are missed or the same assessment item is checked twice, a pop-up message will remind the teacher to make corrections before clicking on confirm and sending the result back to the CBLAS. The CBLAS is equipped with an automatic submission check function, which can prevent teachers from accidentally failing to assess trainees. Apart from this, the CBLAS matches up teachers with assigned trainees and sends timely reminders to the teachers to avoid repeated assessments and insufficient number of assessments.

In the past, the collection and organization of huge amounts of data were often the most troublesome tasks for administrative staff. The CBLAS can efficiently save time on the delivery of questionnaires because after the assessment is finished, it usually takes few days to submit and collect the paper questionnaire, as well as create a file for it. Further, some teachers even forget

to submit the assessment questionnaire. Moreover, the assessment data can be easily categorized, shared, integrated, and analyzed. The CBLAS not only meets the needs of teachers and learners, but also considers the needs of administrative staff.

Future Plan

Although the CBLAS currently focuses on Western medicine professionals, it has reserved some space to include other specialties in future use. Therefore, we developed an e-learning platform that integrates CBME, which would serve not only as a reference for major departments and countries, but also as a basis for further development. Some improvements were suggested for the CBLAS, and these are presented below.

Improve the Program and Learning Content by Simplifying the Content but Targeting the Milestones

To ensure the quality of online learning on different types of devices and to keep pace with the trend of online education, we are consistently making improvements to the CBLAS according to analysis data obtained from pre- and postsurvey user feedback. With a focus on clarifying the learning goals, we made changes to the learning material and teaching content, such as visualizing the content and simplifying the text, which will help users to read the interface on smartphones, tablets, or other devices with ease.

Increase the Interactivity to Enhance Community Cooperation and User Engagement

Currently, the teacher-student feedback system has been completed, and in the future, we hope to build a community system where teachers and students can form small learning groups with specific learning goals. Through such a system, teachers and students can have discussions, exchange opinions, and share their learning experiences. As a result, it can increase users' cooperation in learning and enhance their engagement through smoother group communication.

Guide Learners to Access More External Learning Materials and Curricula

Currently, the speed of developing new learning materials cannot keep pace with the speed of medical knowledge advancement. In the future, we hope to find a faster and more effective way of obtaining accurate and appropriate learning materials that teachers can provide to learners in order to ensure that students can easily access learning materials most suitable for them.

Conclusions

Flexibility and active learning methods take precedence in contemporary medical education [30]. Using an electronic platform for learning will help clinicians avoid geographical and time constraints and promote participation in medical education [31]. Additionally, this will be considered most valuable when using an electronic platform to learn and for its instant feedback, self-assessment, simple interface, extended completion time, and thematic associations [32]. An electronic platform must be taken seriously and adopted quickly, and it should be revised according to the different cultural needs of different countries to facilitate both cognitive knowledge learning and workplace competency-based learning, advocating both learner-centered learning and health care quality [3,7,23,27].

The CBLAS in this study was built with a focus on three key users' needs to realize competency-based medical education for residency training and to build a visionary learning platform. It not only enhances the quality and efficiency of clinical teachers' instructions and trainees' learning progress with milestone assessments at any time, but also addresses administrative heavy workload for educational staff. The CBLAS can show different interfaces according to the user's identity to fit the user's needs, which helps reduce the user's cognitive load and increase the ease and feasibility of CBLAS use. Consequently, the results showed that 87.2% (102/117) of users were satisfied with the CBLAS, without any differences between different user groups.

In conclusion, as expected from the literature [23], our study showed the potential of digital systems to facilitate CBME according to detailed user needs assessments and careful design and development of the system to respond to all the different key users' needs in the clinical workplace, with operational concepts for CBME, such as EPAs and milestones. More detailed qualitative interviews to explore how the system plays a facilitating role in clinical assessments and the learning process will be the next step to advance our knowledge about the design of competency-based education systems.

Acknowledgments

We deeply thank Chang Gung Medical Education Research Centre and the Centre Director Professor Lynn Monrouxe for her assistance with this manuscript. The authors also thank the Ministry of Science and Technology (MOST 105-2511-S-182-003-MY2 and MOST 107-2511-H-182A-003), China Medical University Hospital (DMR-107-196), and the Research Committee of Chang Gung Memorial Hospital, Taiwan (no CDRPG6F0011) for financial assistance.

Authors' Contributions

CMH conceived the study design and drafted the manuscript, FCC contributed to the design of the app and reviewed the manuscript, CCH and LCC performed the data and statistical analysis, and CTH conceived the study concept and design and wrote the manuscript. All the authors contributed to the development and testing of the system.

Conflicts of Interest

None declared.

References

1. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. The Lancet 2010 Dec;376(9756):1923-1958. [doi: 10.1016/s0140-6736(10)61854-5]
2. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. Acad Med 2002 May;77(5):361-367. [doi: 10.1097/00001888-200205000-00003] [Medline: 12010689]
3. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. Medical Teacher 2010 Jul 27;32(8):638-645. [doi: 10.3109/0142159x.2010.501190]
4. ten Cate O. Entrustability of professional activities and competency-based training. Med Educ 2005 Dec;39(12):1176-1177. [doi: 10.1111/j.1365-2929.2005.02341.x] [Medline: 16313574]
5. Chiu H. The Development of Competency-Based Medical Education. Formosan Journal of Medicine 2017;21(3):319-324. [doi: 10.6320/FJM.2017.21(3).11]
6. Flexner A. The Carnegie Foundation for the Advancement of Teaching. In: Medical Education In The United States And Canada, Bulletin Number Four, 1910. New York: Andesite Press; 2015.
7. McGaghie W, Miller G, Sajid AW, Telder TV. Competency-based curriculum development on medical education: an introduction. Public Health Pap 1978(68):11-91. [Medline: 664734]
8. Harden RM. AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education. Medical Teacher 2009 Jul 03;21(1):7-14. [doi: 10.1080/01421599979969]
9. Harden RM, Crosby JR, Davis MH, Friedman M. AMEE Guide No. 14: Outcome-based education: Part 5-From competency to meta-competency: a model for the specification of learning outcomes. Med Teach 1999 Jul 03;21(6):546-552. [doi: 10.1080/01421599978951] [Medline: 21281173]
10. Bandiera G, Lendrum D. Daily encounter cards facilitate competency-based feedback while leniency bias persists. CJEM 2008 Jan;10(1):44-50. [doi: 10.1017/s1481803500010009] [Medline: 18226318]
11. Brown TC, McCleary LE, Stenchever MA, Poulson AM. A competency-based educational approach to reproductive biology. American Journal of Obstetrics and Gynecology 1973 Aug;116(7):1036-1043. [doi: 10.1016/s0002-9378(16)33856-x]
12. Weinstein HM, Russell ML. Competency-based psychiatric education. Am J Psychiatry 1976 Aug;133(8):935-939. [doi: 10.1176/ajp.133.8.935] [Medline: 942007]
13. Botticelli M, Anderson A. MD-level competence in internal medicine. Objectives and the flexible clerkship. Arch Intern Med 1981 Feb;141(2):235-236. [Medline: 7458519]
14. Chi CS. EPA-based Assessment in Residency Training Program. Tungs' Medical Journal 2016;10:1-4. [doi: 10.29838/TMJ]
15. Chan CY. Entrustable Professional Activities -- Doubts and Understanding. The Changhua Journal of Medicine 2017;15:104-113. [doi: 10.6501/CJM.1502.007]
16. Beeson MS, Carter WA, Christopher TA, Heidt JW, Jones JH, Meyer LE, et al. Emergency Medicine Milestones. Journal of Graduate Medical Education 2013 Mar;5(1s1):5-13. [doi: 10.4300/jgme-05-01s1-02]
17. Carraccio C, Burke AE. Beyond Competencies and Milestones: Adding Meaning Through Context. Journal of Graduate Medical Education 2010 Sep;2(3):419-422. [doi: 10.4300/jgme-d-10-00127.1]
18. Dreyfus H, Dreyfus S. Five Steps from Novice to Experts. In: Mind Over Machine: The Power Of Human Intuition And Expertise In The Era Of The Computer. New York: Free Press; 1986:18-36.
19. Nasca TJ. ACGME Bulletin.: ACGME; 2008 May. The CEO's First Column - The Next Step in the Outcomes-Based Accreditation Project URL: https://www.acgme.org/Portals/0/PFAssets/bulletin/bulletin5_08.pdf [accessed 2018-04-03]
20. Tekian A, Watling CJ, Roberts TE, Steinert Y, Norcini J. Qualitative and quantitative feedback in the context of competency-based education. Med Teach 2017 Dec;39(12):1245-1249. [doi: 10.1080/0142159X.2017.1372564] [Medline: 28927332]
21. Berz JP, Cheng T, Quintiliani LM. Milestones-based direct observation tools in internal medicine resident continuity clinic. BMC Med Educ 2017 Dec 04;17(1):240 [FREE Full text] [doi: 10.1186/s12909-017-1077-y] [Medline: 29202742]
22. Lockyer J, Carraccio C, Chan M, Hart D, Smee S, Touchie C, ICBME Collaborators. Core principles of assessment in competency-based medical education. Med Teach 2017 Jun;39(6):609-616. [doi: 10.1080/0142159X.2017.1315082] [Medline: 28598746]
23. Schumacher DJ, Spector ND, Calaman S, West DC, Cruz M, Frohna JG, et al. Putting the pediatrics milestones into practice: a consensus roadmap and resource analysis. Pediatrics 2014 May;133(5):898-906. [doi: 10.1542/peds.2013-2917] [Medline: 24733873]
24. Andolsek K, Padmore J, Hauer K, Edgar L, Holmboe E. Accreditation Council for Graduate Medical Education. 2017. Clinical Competency Committees: A Guidebook for Programs URL: https://www.acgme.org/Portals/0/ACGMEClinicalCompetencyCommitteeGuidebook.pdf [accessed 2018-06-13]
25. 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]
26. Ajzen I, Madden TJ. Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. Journal of Experimental Social Psychology 1986 Sep;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]

27. Dawd S. The Promise of E-Platform Technology in Medical Education. Ethiop J Health Sci 2016 Mar;26(2):171-176 [FREE Full text] [doi: 10.4314/ejhs.v26i2.11] [Medline: 27222630]
28. Heller J, Steiner C, Hockemeyer C, Albert D. Competence-based knowledge structures for personalised learning. International Journal on E-learning 2006;5(1):75-88.
29. Zehry K, Halder N, Theodosiou L. E-Learning in medical education in the United Kingdom. Procedia - Social and Behavioral Sciences 2011 May;15(5):3163-3167. [doi: 10.1016/j.sbspro.2011.04.265] [Medline: 30872122]
30. Moran J, Briscoe G, Peglow S. Current Technology in Advancing Medical Education: Perspectives for Learning and Providing Care. Acad Psychiatry 2018 Dec 13;42(6):796-799. [doi: 10.1007/s40596-018-0946-y] [Medline: 29949053]
31. Cook DA, Steinert Y. Online learning for faculty development: a review of the literature. Med Teach 2013 Nov 05;35(11):930-937. [doi: 10.3109/0142159X.2013.827328] [Medline: 24006931]
32. Maertens H, Madani A, Landry T, Vermassen F, Van Herzeele I, Aggarwal R. Systematic review of e-learning for surgical training. Br J Surg 2016 Oct 18;103(11):1428-1437. [doi: 10.1002/bjs.10236] [Medline: 27537708]

Abbreviations

ACGME: Accreditation Council for Graduate Medical Education
CBLAS: competency-based learning and assessment system
CBME: competency-based medical education
EPA: entrustable professional activity

Edited by N Zary; submitted 28.07.19; peer-reviewed by V Della Mea, P Bertheau; comments to author 21.10.19; revised version received 17.12.19; accepted 10.02.20; published 14.04.20.

Please cite as:
Hsiao CT, Chou FC, Hsieh CC, Chang LC, Hsu CM
Developing a Competency-Based Learning and Assessment System for Residency Training: Analysis Study of User Requirements and Acceptance
J Med Internet Res 2020;22(4):e15655
URL: https://www.jmir.org/2020/4/e15655
doi:10.2196/15655
PMID:32286233

Original Paper

A Clinical Teaching Blended Learning Program to Enhance Registered Nurse Preceptors' Teaching Competencies: Pretest and Posttest Study

Xi Vivien Wu[1], MEd, PhD; Yuchen Chi[1], BSc; Umadevi Panneer Selvam[2], MSc; M Kamala Devi[1], PhD; Wenru Wang[1], PhD; Yah Shih Chan[1], MSc; Fong Chi Wee[2], MSc; Shengdong Zhao[3], PhD; Vibhor Sehgal[4], BTech; Neo Kim Emily Ang[1], PhD

[1]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]Department of Education and Practice, Nursing Service, Tan Tock Seng Hospital, Singapore, Singapore

[3]School of Computing, National University of Singapore, Singapore, Singapore

[4]University of California Berkeley, Berkeley, CA, United States

Corresponding Author:
Xi Vivien Wu, MEd, PhD
Alice Lee Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Level 2, Clinical Research Centre
Block MD11, 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66012756
Email: nurwux@nus.edu.sg

Abstract

Background: Clinical nursing education provides opportunities for students to learn in multiple patient care settings, receive appropriate guidance, and foster the development of clinical competence and professionalism. Nurse preceptors guide students to integrate theory into practice, teach clinical skills, assess clinical competencies, and enhance problem-solving and critical thinking skills. Previous research has indicated that the teaching competencies of nurse preceptors can be transferred to students' clinical learning to enhance their clinical competencies.

Objective: This study aimed to develop a clinical teaching blended learning (CTBL) program with the aid of web-based clinical pedagogy (WCP) and case-based learning for nurse preceptors and to examine the effectiveness of the CTBL program on nurse preceptors' clinical teaching competencies, self-efficacies, attitudes toward web-based learning, and blended learning outcomes.

Methods: A quasi-experimental single-group pretest and posttest design was adopted. A total of 150 nurse preceptors participated in the CTBL program, which was conducted from September 2019 to December 2019. A set of questionnaires, including the clinical teaching competence inventory, preceptor self-efficacy questionnaire, attitudes toward web-based continuing learning survey, and e-learning experience questionnaire, was used to assess the outcomes before and after the CTBL program.

Results: Compared with the baseline, the participants had significantly higher total mean scores and subdomain scores for clinical teaching competence (mean 129.95, SD 16.38; $P<.001$), self-efficacy (mean 70.40, SD 9.35; $P<.001$), attitudes toward web-based continuing learning (mean 84.68, SD 14.76; $P<.001$), and blended learning outcomes (mean 122.13, SD 14.86; $P<.001$) after the CTBL program.

Conclusions: The CTBL program provides a comprehensive coverage of clinical teaching pedagogy and assessment strategies. The combination of the WCP and case-based approach provides a variety of learning modes to fit into the diverse learning needs of the preceptors. The CTBL program allows the preceptors to receive direct feedback from the facilitators during face-to-face sessions. Preceptors also gave feedback that the web-based workload is manageable. This study provides evidence that the CTBL program increases the clinical teaching competencies and self-efficacies of the preceptors and promotes positive attitudes toward web-based learning and better blended learning outcomes. The health care organization can consider the integration of flexible learning and intellect platforms for preceptorship education.

(J Med Internet Res 2020;22(4):e18604) doi:10.2196/18604

KEYWORDS

blended learning; case-based learning; clinical pedagogy; clinical teaching competency; web-based program; nurse preceptor

Introduction

Background

Clinical nursing education provides opportunities for students to learn in multiple patient care settings, receive appropriate guidance, and foster the development of clinical competence and professionalism. In a preceptorship program, a nursing student is assigned to a registered nurse preceptor who is responsible for clinical guidance. Preceptors nurture the development of clinical knowledge, skills, and professional attitudes in nursing students through guiding, role modeling, and facilitating professional development [1]. The preceptors' levels of competence impact students' learning experiences and clinical competencies. However, preceptors often lack teaching knowledge and experience, which leads to role ambiguity and unfamiliarity with clinical education systems [2]. Hence, training in clinical teaching pedagogy is recommended to assist with preceptors' professional development. Innovative continuing education promotes the professional development of health care professionals and fits into their busy working schedules [3]. Technology alone is not able to meet the diverse learning needs of health care professionals in the clinical environment. Nevertheless, blended learning is recognized as an effective approach to provide an alternative learning strategy and promote the integration of theory and practice [4]. It is paramount that continuing education courses integrate technology, increase the flexibility and responsiveness of the workforce, and offer alternative means to attend the courses [5].

Literature Review

Registered Nurse Preceptors' Teaching Competencies and Self-Efficacies

The transition from a nursing student to a professional nurse is challenging and may lead to anxiety and frustration, which consequently result in job dissatisfaction, low productivity, and attrition [6]. It is critical to retain nurses at the workplace to address the global nursing shortage. Nurse preceptors play important roles in clinical education. Preceptors facilitate the clinical learning of students through the process of socialization by helping students to adjust to their working cultures and environment and become functional members of their teams. Positive preceptorship experience prepares nursing students for smooth transitions into the reality of practice [7]. In addition, preceptors help nursing students to acquire a sense of professional identity through caring and teaching [8]. What students learn from nurse preceptors will be reflected in their daily practices when they become independent nurses [8]. Hence, the competency levels of nurse preceptors directly affect the quality of future nurses.

Previous studies indicate that it is crucial to prepare preceptors to be pedagogically ready to embark on the preceptorship process [2,9-11]. Researchers have reported that preceptors who have greater knowledge of preceptor roles tend to display higher self-efficacies and confidence in their capabilities of clinical teaching [12]. Furthermore, the way nursing students are precepted shapes workplace cultures and environments, which may help with the issue of retaining nurses at the workplace to address the global nursing shortage [11,13]. Therefore, nurse preceptors' teaching competency is positively related to the quality of future nurses, which translates to improved patient safety, quality of care, and staff retention [14,15].

Web-Based Learning for Health Care Professionals

Nowadays, it is crucial for health care workers to access continuous learning opportunities to update their knowledge and skills. However, health care institutions always face challenges in supporting their staff for continuous professional developments, such as costs of training, staff absence from clinical areas, and limited time to attend courses [16]. Electronic learning (e-learning) using information technology as a medium provides an alternative to build virtual platforms for learners' interactions [17]. With the advancement of information technology, e-learning platforms are able to demonstrate outcomes comparable with those of face-to-face programs by using interactive simulation videos and discussion forums [18-21]. E-learning is flexible and allows nurses to learn at their own pace [22]. Thus, it is evident as an effective alternative means to train health care professionals within resource-limited settings [23].

Although e-learning has been widely used in hospital training programs, there are still emerging concerns. It requires a certain level of computer literacy and skills. Nurses who have not taken web-based courses require more support to adapt to e-learning platforms [22,24,25]. Some studies highlighted concerns regarding the limitations of learners' assessments and feedback when using e-learning programs, which can significantly impact learning outcomes [26,27]. In addition, learners may feel isolated and disengaged during e-learning, which may result in a poor completion rate [28].

Blended Learning and a Case-Based Approach for Health Care Professionals

Blended learning involves a systematic combination of face-to-face interactions and technology-mediated interactions among learners, facilitators, and resources [29]. Compared with e-learning sessions, the teaching and learning styles of face-to-face sessions should be redesigned to fit into the learning needs of health care professionals [30]. Blended learning can engage case studies, group discussions, and debates to emphasize social elements during the session [16,30]. This will increase human interactions and peer support to motivate learners in the learning process [28,31].

During face-to-face sessions, case-based learning is a highly advocated learning approach in health professional education because it encourages learners to apply knowledge to solve

problems [32,33]. It allows learners to either discuss in groups or to have a discussion guided by a facilitator after a case presentation [34]. Case-based learning promotes active and reflective learning and facilitates learners to improve their critical thinking and problem-solving skills [34,35]. Blended learning brings e-learning and face-to-face learning together and provides learners with exposure to various learning experiences. Studies have highlighted the advantages of blended learning, including more flexibility, less time restrictions, greater pedagogic richness, and more cost-effectiveness [36,37]. It can support diverse learners to meet their learning needs. In addition, blended learning helps learners improve their sense of autonomy and responsibility and allows facilitators to maintain learners' engagement and motivation [16,38,39].

A review of the literature shows that the number of blended learning studies has been limited, and the area of study focuses more on learners' and facilitators' satisfaction, perception, and experience [16,38,40]. As an innovative approach, blended learning has not been validated for nurse preceptor training. Hence, this study aimed to develop a clinical teaching blended learning (CTBL) program with the aid of web-based clinical pedagogy (WCP) and case-based learning for nurse preceptors and to examine the effectiveness of the CTBL program on nurse preceptors' clinical teaching competencies, self-efficacies, attitudes toward web-based learning, and blended learning outcomes.

Methods

Research Questions

The study was designed to address the following research questions:

1. What is the effect of the CTBL program on the clinical teaching competencies of nurse preceptors?
2. What is the effect of the CTBL program on nurse preceptors' self-efficacies in conducting clinical teaching and assessment?
3. What is the effect of the CTBL program on nurse preceptors' attitudes toward web-based continuing learning?
4. What are the effects of the CTBL program on nurse preceptors' blended learning outcomes?

Study Design and Setting

A single-group prospective pretest and posttest design was adopted for this study. The CTBL program with the aid of WCP and case-based learning for nurse preceptors was developed. The study was conducted at an acute tertiary hospital located in the central region of Singapore.

Participants

The target population was nurses who were assigned the roles of clinical teaching and assessments of the nursing students. The inclusion criteria included nurse preceptors who (1) were guiding nursing students for clinical teaching and assessments, (2) were aged 21 years and older, (3) had not completed the clinical teaching course provided by the hospital, and (4) used smart mobile devices (smartphones, tablets, or laptops) in their daily lives. The exclusion criteria included nurse preceptors who did not use smart mobile devices. A total of 150 registered nurse preceptors were eligible for participation in this study.

The sample size needs to be determined by the number of participants who are required to maintain statistical power for statistical tests used in data analysis. The sample size was calculated using a paired *t* test, with an adjusted alpha of .0167 (Bonferroni approach) and an estimated effect size and power of 0.8. With the use of the IBM SPSS Version 25 statistical tool G*power [41], the sample size was calculated to be 120. Factoring at a 20.0% attrition rate, the estimated sample size was 144. A total of 150 participants were recruited for this study.

Intervention

The CTBL program consists of a WCP program and a face-to-face case-based learning workshop. The CTBL program was carried out as a 1-day workshop (8 hours). In the morning session, the participants were given access and time to attend the WCP program. The WCP program was developed by a research team using a 3-step process: integrate the theoretical framework, evidence from systematic review, and feedback from content validity test by experts and the result of the pilot test with the nurse preceptors. The WCP program consists of 8 modules: (1) introduction of preceptorship, (2) planning care with preceptee, (3) conducting clinical assessment, (4) facilitating clinical learning, (5) creating a positive clinical learning environment, (6) providing constructive feedback, (7) handling challenging situations, and (8) managing underperforming preceptees. The details of the WCP program have been reported in another paper published by the team [21] (Figure 1).

In the afternoon session, face-to-face case-based learning was conducted for preceptors by a facilitator who was also a nurse educator. In total, 3 case studies were used in case-based learning. The team developed case studies based on common situations that preceptors might encounter in the process of clinical teaching. Preceptors were divided into small groups to discuss, present, and share how they planned to manage such situations. The group discussion and presentation were facilitated by the nurse educator. Learner-to-learner and learner-to-facilitator communication and interaction were maintained throughout the case discussion. Table 1 presents a sample of the case study. Case-based learning is frequently used in medical and nursing education. Studies have demonstrated that case-based learning promotes group processes, collaborative learning, critical thinking, and decision making. In addition, learners are encouraged to participate actively in the discussion and critique of each other's work [34,42].

Figure 1. Modules in the web-based clinical pedagogy program.

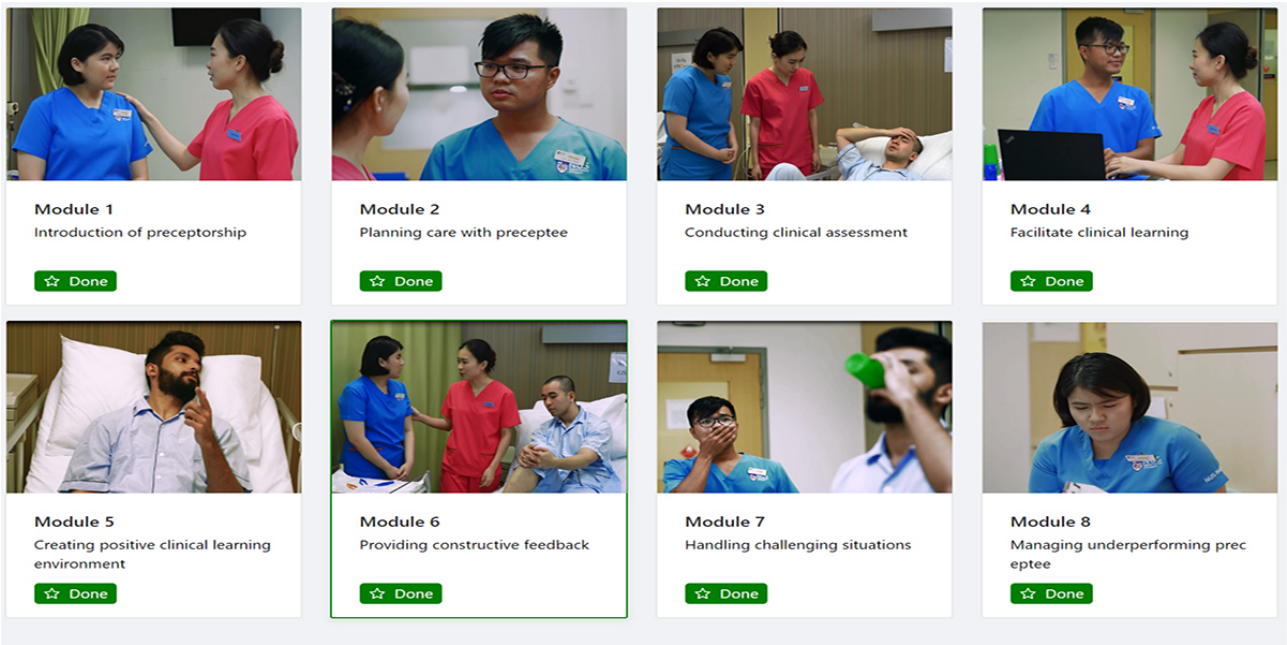


Table 1. Sample of the case study.

Case study	Case description	Guiding questions
Case study 1	You are the preceptor of a final-year nursing student. After completing the hospital orientation program, your nurse manager introduces your preceptee to you. After a discussion with your preceptee, you find out that she was previously trained at a different hospital and had never done any attachment in ABC hospital.	• Share how you will welcome your preceptee • Discuss some concerns that you may have about your preceptee • How would you assess her level of competency for her expected skills?
Case study 2	You have a preceptee who is a matured learner. You realize that he had 20 years of working experience in a financial sector as a manager. However, he was retrenched from the previous job and took up nursing as a midcareer switch job. You experience disagreements between you and your preceptee frequently. At the same time, you realize that he asks valid questions at work in an unacceptable tone. You have other colleagues giving feedback to you that your preceptee is not a good team player.	• Discuss the possible reasons for the preceptee's behavior • Share how you plan to work with your preceptee
Case study 3	You are the preceptor to a nursing student from a foreign country. She has just completed the hospital orientation program. You have observed that she appeared withdrawn and lost in the ward when caring for patients. She demonstrates an unenthusiastic attitude toward learning and you have never seen her ask questions.	• Discuss possible reasons for her behavior • Explore her learning needs • Discuss your plan in approaching the situation

Outcome Measures

Sociodemographic profiles, including gender, age, race, educational level, job title, clinical department, area of specialization, years of working experience as a registered nurse, and years of experience in clinical teaching, were obtained using structured questionnaires. In total, 4 instruments were used to evaluate the learning outcomes of the nurse preceptors. The face and content validity for the clinical teaching competence inventory (CTCI), preceptor self-efficacy questionnaire (PSEQ), attitudes toward a web-based continuing learning survey (AWCLS), and e-Learning experience questionnaire (LEQ) were assessed by the research team and the experts committee in this study.

Clinical Teaching Competence Inventory

The CTCI was developed in Taiwan [43]. It consists of 4 domains of teaching and assessment competencies and 31 items. The 4 domains are student evaluation (assessment), goal setting and individual teaching, teaching strategies, and demonstration of organized knowledge. The CTCI was psychometrically tested, and the results indicated that the instrument has adequate content validity (scale content validity index =0.75) and internal consistency of reliability (Cronbach alpha=.88) for assessing clinical teaching and assessment behavior in practice settings.

Preceptor Self-Efficacy Questionnaire

The PSEQ [44] consists of 21 items and uses a 4-point scale (Cronbach alpha=.93). The participants used the PSEQ to rate their confidence in teaching strategies, learning critical thinking,

challenging situations, providing feedback and evaluations, and overall confidence in precepting nursing students.

Attitudes Toward a Web-Based Continuing Learning Survey

The AWCLS was developed based on relevant studies and the technology acceptance model [45]. There are 4 dimensions and 18 items using a 7-point Likert scale in the survey. The 4 dimensions are (1) the perceived usefulness domain assesses the nurses' perceptions of the extent to which they perceive that the influence of web-based continuing learning is useful, (2) the perceived ease of use domain evaluates the nurses' perceptions of the extent to which web-based continuing learning is easy to use, (3) the behavior domain measures the nurses' willingness to take up web-based continuing learning, and (4) the affection domain assesses the nurses' perceptions regarding positive feelings about web-based continuing learning. The results of the alpha coefficients (Cronbach alpha=.96) for the scales indicated that the AWCLS is appropriate for evaluating nurses' attitudes toward web-based continuing learning [46].

e-Learning Experience Questionnaire

The LEQ is used to measure the outcomes of a blended learning environment that combines e-learning and face-to-face interactions [47]. The instrument consists of 32 items that are grouped into 9 domains: quality of teaching in an e-learning context, participants' interaction and engagement, clarity of goals and standards for online component, quality of online resources, appropriateness of the assessment in an e-learning context, appropriateness of workload related to online materials and activities, issues related to participants management, degree to which online materials and activities support face-to-face learning, and overall satisfaction with the quality of online materials and activities. The Cronbach alpha values range from .61 to .84 for different segments [47].

Ethical Considerations

The National Health Group Domain Specific Review Board (approval number: 2018/00138) approved the ethical aspect of the study. Permission and support to conduct the study were obtained from the senior management of the hospital. A participant information sheet was given to each potential participant with details on the aims and procedures of the study. The participants were reassured that participation in the study was voluntary and that their identities would remain anonymous. They were informed that they had the right to withdraw from the study at any time.

Data Collection Procedure

Recruitment emails were sent to each potential participant. The research team explained the research process and obtained consent from potential participants. Nurse preceptors were invited to complete the sociodemographic profile and questionnaires (CTCI, PSAI, AWCLS, and LEQ) before they started the CTBL program (baseline). Posttest questionnaires were administered at the end of the CTBL program. A total of 6 training workshops were conducted between September 2019 and December 2019. Each workshop consisted of 20 to 30 participants. Data were collected using *MySurvey*, a secured web-based platform subscribed by the university. *MySurvey* operated on Verint Enterprise Feedback Management 15.2, and the website was deployed on a campus on-premise server and maintained by the information technology center of the university.

Data Analysis

Data were retrieved from the *MySurvey* platform. A total of 2 research assistants checked the accuracy of the data. SPSS 25.0 (SPSS institute) was used to analyze the data. Descriptive statistics, such as frequency, mean, and SD, were used to analyze the sociodemographic data and study outcomes. The paired *t* test was used to compare the pre- and postintervention score changes at a significance level of .05. The final analysis was based on 150 participants who had completed all the data collection. The normality of outcome variables was examined using the Shapiro-Wilk test and quantile-quantile plot normal distribution graphics and by visually inspecting the histogram generated. Normality was also determined through skewness values divided by SE and kurtosis divided by SE. Normality was indicated by a value of less than −2.0 or more than 2.0 [48].

Results

The study examined the impact of the CTBL program on nurse preceptors' clinical teaching competencies, self-efficacies, attitudes toward web-based learning, and blended learning outcomes. The results demonstrated significant improvement in most of the outcome measures, as illustrated in Figure 2.

Sociodemographic Characteristics

Table 2 demonstrates the sociodemographic characteristics of the participants. More than 90.0% (140/150) of the participants were female, and the mean age was 31.6 years. The ethnicity distributions were Chinese (38/150, 25.3%), Malay (26/150, 17.3%), Indian (29/150, 19.3%), Filipino (52/150, 34.7%), and other ethnic groups (5/150, 3.3%). The majority of the participants (108/150, 72.0%) were graduates who had a Bachelor of Nursing degree, and 28.0% (42/150) of the participants were diploma or certificate holders. More than two-thirds of the participants (103/150, 68.7%) were senior/registered nurses, and one-third of them (47/150, 31.3%) were enrolled nurses. The average working experience was 8 years. The majority of the participants had relevant clinical teaching experiences, ranging from 1 to 3 years (98/150, 65.3%) to 3 to 10 years (32/150, 21.3%). However, 12.7% (19/150) of the participants did not have any clinical teaching experience.

Figure 2. Mean score pretest and posttest (N=150) for AWCLS. AWCLS: attitudes toward a web-based continuing learning survey; CTCI: clinical teaching competence inventory; LEQ: e-Learning experience questionnaire; PSEQ: preceptor self-efficacy questionnaire.

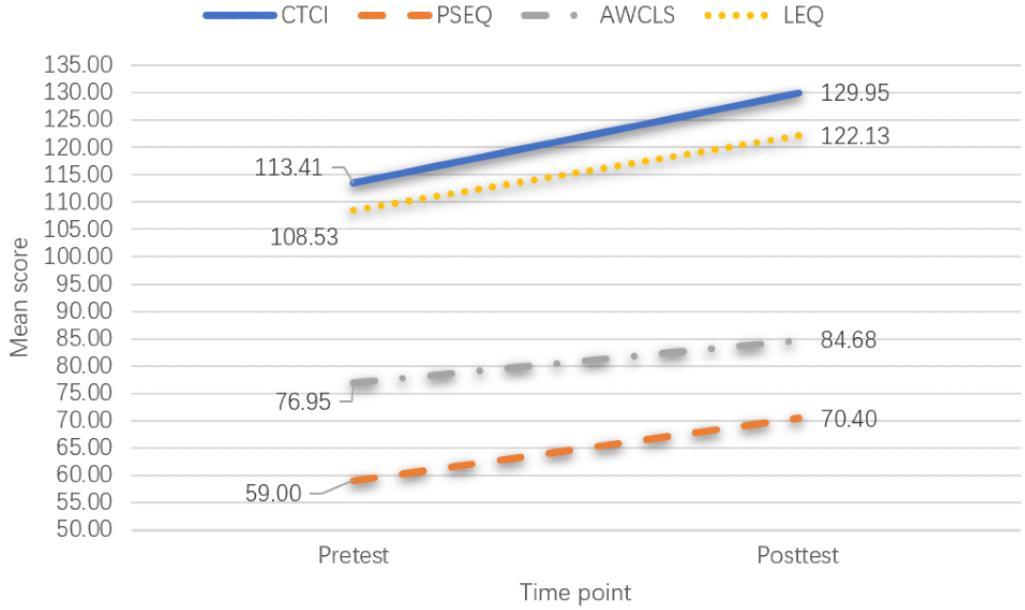


Table 2. Sociodemographic characteristics of the participants (N=150).

Characteristic	Value
Age (years), mean (SD)	31.6 (6.4)
Gender, n (%)	
Female	140 (93.3)
Male	10 (6.7)
Ethnicity, n (%)	
Chinese	38 (25.4)
Malay	26 (17.3)
Indian	29 (19.3)
Filipino	52 (34.7)
Others	5 (3.3)
Education level, n (%)	
Bachelor of nursing	108 (72.0)
Diploma of nursing	31 (20.7)
Certificate of nursing	11 (7.3)
Years of working, mean (SD)	8 (5.0)
Clinical teaching experience (years), n (%)	
0	19 (12.7)
1-3	98 (65.3)
3-6	20 (13.3)
7-10	12 (8.0)
>10	1 (0.7)
Current position, n (%)	
Enrolled nurse	47 (31.3)
Registered nurse	68 (45.4)
Senior registered nurse	35 (23.3)

Clinical Teaching Competencies

The participants had significant increases in their clinical teaching competency scores after completion of the CTBL program (mean 129.95, SD 16.38; *P*<.001) compared with the baseline (mean 113.41, SD 16.67; *P*<.001). Overall, the postintervention scores were significantly higher than the preintervention scores in all 4 subdomains: student evaluation (mean 37.57, SD 4.80; *P*<.001), goal setting and individual teaching (mean 30.18, SD 4.26; *P*<.001), teaching strategies (mean 38.22, SD 5.084; *P*<.001), and demonstration of organized knowledge (mean 16.63, SD 2.26, *P*<.001; Table 3).

Table 3. Mean scores for the clinical teaching competence inventory, preceptor self-efficacy questionnaire, and attitudes toward a web-based continuing learning survey, before and after the intervention (N=150).

Measure	Preintervention, mean (SD)	Postintervention, mean (SD)	Difference within group, mean (95% CI)	*t* test (*df*=149)	*P* value
Clinical teaching competency inventory					**<.001**
Student evaluation	31.98 (5.56)	37.57 (4.80)	5.59 (4.63-6.54)	11.53	
Goal setting and individual teaching	25.74 (5.18)	30.18 (4.26)	4.44 (3.58-5.30)	10.2	
Teaching strategies	34.79 (5.26)	38.22 (5.084)	3.43 (2.39-4.46)	6.53	
Demonstration of organized knowledge	14.45 (2.45)	16.63 (2.26)	2.19 (1.73-2.64)	9.54	
Total score	113.41 (16.67)	129.95 (16.38)	16.53 (13.37-19.70)	10.33	
Preceptor self-efficacy questionnaire					**<.001**
Total score	59.00 (11.05)	70.40 (9.35)	10.56 (8.70-12.42)	11.23	
Attitude toward web-based continuing learning survey					**<.001**
Perceived usefulness	25.74 (5.30)	28.25 (5.42)	2.51 (1.70-3.33)	6.07	
Perceived ease of use	20.79 (4.11)	22.85 (3.85)	2.05 (1.41-2.70)	6.27	
Behavior	14.96 (3.43)	16.73 (3.24)	1.77 (1.24-2.31)	6.53	
Affection	15.45 (3.70)	16.85 (3.47)	1.39 (0.85-1.94)	5.04	
Total score	76.95 (15.33)	84.68 (14.76)	7.73 (5.49-9.98)	6.81	

Preceptor Self-Efficacies

The mean score changes of the preceptors' self-efficacies from pre- to postintervention are shown in Table 3. The participants had a significantly higher total score for self-efficacy after the CTBL program (mean 70.40, SD 9.35; *P*<.001) than at the baseline (mean 59.00, SD 11.05; *P*<.001).

Attitudes Toward Web-Based Continuing Learning

The mean changes in attitudes toward web-based continuing learning are shown in Table 3. The total mean score for attitudes toward web-based continuing learning increased significantly postintervention (mean 84.68, SD 14.76; *P*<.001) compared with the baseline (mean 76.95, SD 15.33; *P*<.001). The postintervention scores were significantly higher than baseline in all 4 subdomains: perceived usefulness (mean 28.25, SD 5.42; *P*<.001), perceived ease of use (mean 22.85, SD 3.85; *P*<.001), behavior (mean 16.73, SD 3.24; *P*<.001), and affection (mean 16.85, SD 3.47; *P*<.001) toward web-based continuing learning.

Blended Learning Outcomes

Multimedia Appendix 1 indicates the mean score changes of the blended learning outcomes from pre- to postintervention. The instrument is a Likert scale that uses *strongly disagree*, *disagree*, *neutral*, *agree*, and *strongly agree*. To facilitate the display, *strongly disagree* and *disagree* were combined as *disagree* and *agree* and *strongly agree* were combined as agree. The participants had a significantly higher total mean score for blended learning outcomes after the CTBL program (mean 122.13, SD 14.86; *P*<.001) than at the baseline (mean 108.53, SD 14.07; *P*<.001). In general, the postintervention scores were significantly higher than the preintervention scores, in most of the subdomains:

- Subdomain 1: quality of teaching in e-learning context (mean 25.62, SD 2.80; *P*<.001),
- Subdomain 2: participants' interaction and engagement (mean 14.93, SD 2.24; *P*<.001),
- Subdomain 3: clarity of goals and standards for online component (mean 11.78, SD 1.63; *P*<.001),
- Subdomain 4: quality of online resources (mean 15.84, SD 2.19; *P*<.001),
- Subdomain 5: appropriateness of assessment in e-learning context (mean 10.85, SD 1.49; *P*<.001),
- Subdomain 7: issues related to participants management (mean 11.95, SD 1.60; *P*<.001),
- Subdomain 8: degree to which online materials and activities support face-to-face learning (mean 15.95, SD 2.13; *P*<.001), and
- Subdomain 9: overall satisfaction with the quality of online materials and activities (mean 4.09, SD 0.63; *P*<.001).

However, for subdomain 6: appropriateness of workload related to online materials and activities, although slightly improved,

there was no significant difference in the score before and after the intervention (mean 9.49, SD 1.80; *P*=.16).

Discussion

Principal Findings

The results of the study have provided empirical evidence for our hypotheses in that nurse preceptors who received the CTBL program scored higher in terms of clinical teaching competencies, self-efficacies in conducting clinical teaching and assessment, attitudes toward web-based learning, and blended learning outcomes. In a systematic review, it was identified that the younger generation was more likely to access web-based programs [49]. In this study, the mean age of the participants was 31.6 years. This is considered a relatively young group of nurses. Hence, they could be more positive and acceptable to new modes of learning, such as blended and web-based platforms.

Our study suggested that the blended learning approach provided pedagogical benefits to the participants to enhance their competencies and self-efficacies in clinical teaching. The importance of having a competent clinical preceptor is universally recognized, as it forms the basis of a good clinical mentorship to enhance the learning outcomes of a student in clinical practice [50]. Self-efficacy is important because job satisfaction and confidence not only impact preceptors but also impact the retention, turnover, and job satisfaction of newly recruited nurses. Good clinical teaching and mentorship often result in positive outcomes, such as increased job satisfaction, productivity, quality of care, and patient safety for students or newly recruited nurses [51,52]. Most preceptors perceived improvement of professional knowledge and teaching competencies as the most important benefit of preceptorship and motivation for precepting students [53,54]. However, in a study by Seo et al [55], the majority of the preceptors viewed monetary compensation as a valuable incentive, whereas the others preferred support for education and promotion in career. Hence, appropriate extrinsic compensation, for example, opportunities for promotion and monetary rewards could be considered to attract and encourage nurses to take on the role of preceptors.

The blended learning approach has shown evidence of its relevance and usefulness in medical and nursing training [56,57]. It has been shown to be an effective method of training in health care professionals with high levels of learner satisfaction [58-60]. On the basis of Tobin's [61] 3-dimensional framework to evaluate participants' perception of the web-based environment, we evaluated the CTBL program. First, the *emancipatory activities* dimension of the CTBL program allowed the participants to control the pace, place, and depth of learning by offering the WCP program, and the participants could have more flexibility in learning and reviewing the contents of the program. Second, the *coparticipatory activities* dimension (the case-based approach), provides a platform for the participants to explore and discuss various options in clinical teaching and assessment among the peer learners and with the instructors. Third, in the *qualia* dimension of the CTBL program, the participants enjoyed the program in general, and this is reflected in the overall improvement in competency and self-efficacy in clinical teaching and positive attitudes toward web-based learning and blended learning outcomes.

The CTBL program consists of WCP and case-based learning. This program replaced traditional lectures in the classroom using a web-based asynchronous learning mode. Research shows that health care professionals appreciate asynchronous, digital teaching and learning [62]. In fact, a challenge for educators is to meet the learning needs of 21st century learners in providing more flexible digital learning platforms [63]. However, case-based learning allows participants to interact with peers and instructors through face-to-face discussions. In fact, the results of the LEQ indicated that the participants valued this active learning part of the CTBL program. Research has shown that the higher the learner's engagement level, the greater the learners' learning potential [64]. In fact, a number of preceptors were actively engaged in discussions during the development process of the CTBL program. The early involvement of the preceptors contributed to the identification of the learning needs and preferences of the targeted participants, the verification of the contents of the program, and creative teaching pedagogies. Researchers have identified five important elements in the process of developing new technological learning methods: clarifications of learners' expectations, help to recognize the bigger picture, stimulation of interactions, creations of a structure, and context-specific contents [65].

In the LEQ, items 1, 3, 11, 15, and 22 were phrased negatively. Hence, these items were reverse coded in the data analysis. There were 3 items in the LEQ that did not demonstrate significant changes after the intervention. For item 1, "to do well in the online quizzes all you really need is a good memory" (*P*=.18), 40 participants (40/150, 26.7%) selected *agree* postintervention compared with 20 participants (20/150, 13.3%) at the baseline. In fact, good performance in quizzes does not require only good memory but should also reflect learners' mastery of the contents. The quizzes in the CTBL program were designed in such a way that learners should apply knowledge learned from the module to solve problems in the scenarios. Research has demonstrated that less experienced preceptors are more unwilling to fail incompetent students in clinical settings because of a lack of confidence in their own capabilities [66]. Junior preceptors often feel impeded in their decision-making abilities because of lack of clinical teaching knowledge and experiences. The scenarios in the WCP and case-based approach provided a platform for the nurse preceptors to have a real sense of clinical situations and how they could provide guidance and support to their preceptee and develop their judgment skills. Quizzes after each scenario further evaluated the preceptors' ability to apply learnt knowledge to solve issues that they may encounter in real clinical settings.

For item 11, "the workload for the online component of this unit of study is too heavy" (*P*=.47), the number of participants who selected *disagree* increased from 12.0% (18/150) at baseline to 37.3% (56/150) postintervention. After the CTBL program, 3 times the number of participants were inclined to reasonable workload for web-based learning. This could be an indicator that learners need to have real hands-on experience to change their preperceived idea of web-based learning. The workload

of web-based learning is a debatable hot topic. Faculties argued that web-based education appears to have acquired an unstoppable momentum [67]. Notably, limited studies have explored the workload of web-based learning.

Items 3 and 15 are related to the feedback provided to the learners. For item 3, "I received too much feedback online from my teacher/facilitator" (P=.02), although there was a significant change, it is interesting to note that 34.6% (52/150) of participants selected *disagree* postintervention compared with 22.0% (33/150) of participants at baseline. More participants concurred that the feedback from the facilitator was reasonable. For item 15, "I didn't receive enough helpful online feedback from my teacher/facilitator" (P=.29), 28.0% (42/150) of participants selected *disagree* postintervention compared with 14.0% (21/150) of participants at baseline. Similarly, the number of participants who experienced good quality feedback doubled. This shows that the CTBL program provided both sufficient and good quality feedback to the learners. It is evident that efficient interactions between the facilitator and learners contribute to the creation of a positive learning climate [68]. Learners value timely, quality feedback and frequent interaction in the blended learning environment.

Limitations

The study population was restricted to one health care institution in Singapore. Therefore, it may limit the generalizability of the study. However, as the participants were nurse preceptors from various clinical settings, such as medical, surgical, orthopedic, cardiac, and renal, this could be a good representation of nurses from a wide range of clinical settings. One of the limitations of the study is the lack of a control group. Hence, practitioners and management should exercise caution when applying the results of this study. It would be beneficial to use randomized controlled trials in a variety of health care settings. Thus, stronger evidence on the effectiveness of the CTBL program can be generated.

Implication for Practice

The CTBL program is a further enhancement of the WCP program [21], which provides comprehensive coverage on clinical teaching pedagogy and assessment strategies. In addition, case-based learning provides a platform for preceptors to interact and discuss clinical teaching experiences with their peers and facilitators. The combination of WCP and the face-to-face approach provides a variety of learning modes that fit into the diverse learning needs of the preceptors. The CTBL program allows preceptors to receive direct feedback from facilitators during face-to-face sessions. This study also sheds some light on web-based workload as the preceptors gave feedback that the web-based workload was manageable. Future research can explore the workload of web-based learning and effective ways of providing feedback on the web because these areas are understudied.

For course developers, it is important to consider learning styles, learning needs, approaches, and possible platforms in the initial design phase, and logistic issues, such as Wi-Fi connection, servers, and physical classrooms, should be taken into consideration. In addition, the comprehensive approach of the evaluation process enabled the provision of scientific evidence for the effectiveness of the program. More research can explore the cost-effectiveness of the program in terms of saving manpower, cost, time, academic standards, etc.

The CTBL program can be used in other acute hospitals, community hospitals, primary care settings, and higher education institutions to prepare nurse preceptors to be pedagogically ready. Nurse preceptors need resources and support to develop their teaching-coaching capabilities. Comprehensive clinical teaching preparation programs and a variety of learning platforms, such as web-based and face-to-face discussions, provide open communication lines and viable resources for active preceptors. Ongoing support should be offered to preceptors, such as support from nurse managers and recognition and compensatory systems, to encourage and reward them to take up preceptorship roles. Such resources promote preceptors in embracing clinical teaching pedagogy and the spirit of life-long learning. Hence, nursing students have more enriched clinical learning experiences through guidance from preceptors with competent clinical teaching skills.

Conclusions

Currently, with an increasing reliance of the clinical nursing education system on preceptors to facilitate students' learning, it is imperative to provide preceptors with essential pedagogical knowledge. Clinical teaching preparation will enhance role satisfaction and sustain the willingness to perform as a preceptor. The CTBL program is an innovative clinical teaching program for nurse preceptors. The program consists of a WCP and a face-to-face case-based learning workshop, which meet the diverse learning needs of the preceptors and provide a flexible platform that fits into the busy working schedules of preceptors. This study provides empirical evidence that the CTBL program increases the clinical teaching competencies and self-efficacies of preceptors and promotes positive attitudes toward web-based learning and better blended learning outcomes. Health care organizations can consider the integration of flexible learning and intellectual platforms for preceptorship education, such as web-based, case-based, and blended learning approaches.

Acknowledgments

The authors would like to thank Dr Mark Gan, who provided expert opinions on the education perspectives and evaluation of the program. The authors would like to thank the nurses who participated in the research. They would also like to thank Tan Tock Seng Hospital, Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, and the National University of Singapore for providing support for the study. The authors would like to thank the National University Health System Medical Publications Support Unit, Singapore, for assistance in professionally editing the language of this manuscript. The study was funded by the Ministry of Education, Singapore, and Tertiary Education Research grant (MOE 2017-TRF-009).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mean scores for the e-Learning experience questionnaire before and after the intervention (N=150).
[DOCX File , 32 KB - jmir_v22i4e18604_app1.docx]

References

1. Li A, Su Y. Exploring the relationship between personality features and teaching self-efficacy in clinical nursing preceptors. J Nurs Res 2014 Sep;22(3):176-182. [doi: 10.1097/jnr.0000000000000041] [Medline: 25111111]
2. Wu XV, Enskär K, Heng DG, Pua LH, Wang W. The perspectives of preceptors regarding clinical assessment for undergraduate nursing students. Int Nurs Rev 2016 Sep;63(3):473-481. [doi: 10.1111/inr.12272] [Medline: 27100137]
3. Singapore Ministry of Health. 2016. Skills Framework for Healthcare URL: https://www.moh.gov.sg/docs/librariesprovider4/default-document-library/skills-framework-for-healthcare.pdf [accessed 2020-04-16]
4. McCutcheon K, O'Halloran P, Lohan M. Online learning versus blended learning of clinical supervisee skills with pre-registration nursing students: a randomised controlled trial. Int J Nurs Stud 2018 Jun;82:30-39 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.02.005] [Medline: 29574394]
5. Fortune J. The virtual learning environment: an alternative, flexible and accessible method of neonatal nurse education. J Neonatal Nurs 2007 Dec;13(6):231-235. [doi: 10.1016/j.jnn.2007.09.005]
6. Aboshaiqah A, Qasim A. Nursing interns' perception of clinical competence upon completion of preceptorship experience in Saudi Arabia. Nurse Educ Today 2018 Sep;68:53-60. [doi: 10.1016/j.nedt.2018.05.021] [Medline: 29886285]
7. Eta V, Atanga M, Atashili J, D'Cruz G. Nurses and challenges faced as clinical educators: a survey of a group of nurses in Cameroon. Pan Afr Med J 2011;8(1):28 [FREE Full text] [doi: 10.4314/pamj.v8i1.71085] [Medline: 22121437]
8. Carlson E, Pilhammar E, Wann-Hansson C. 'This is nursing': nursing roles as mediated by precepting nurses during clinical practice. Nurse Educ Today 2010 Nov;30(8):763-767. [doi: 10.1016/j.nedt.2010.01.020] [Medline: 20378215]
9. Wu XV, Enskär K, Pua LH, Heng DG, Wang W. Development and psychometric testing of holistic clinical assessment tool (HCAT) for undergraduate nursing students. BMC Med Educ 2016 Sep 22;16(1):248 [FREE Full text] [doi: 10.1186/s12909-016-0768-0] [Medline: 27658587]
10. Wu XV, Wang W, Pua LH, Heng DG, Enskär K. Undergraduate nursing students' perspectives on clinical assessment at transition to practice. Contemp Nurse 2015;51(2-3):272-285. [doi: 10.1080/10376178.2016.1163232] [Medline: 26956057]
11. Wu XV, Enskär K, Pua LH, Heng DG, Wang W. Clinical nurse leaders' and academics' perspectives in clinical assessment of final-year nursing students: a qualitative study. Nurs Health Sci 2017 Sep;19(3):287-293. [doi: 10.1111/nhs.12342] [Medline: 28371304]
12. Larsen R, Zahner SJ. The impact of Web-delivered education on preceptor role self-efficacy and knowledge in public health nurses. Public Health Nurs 2011;28(4):349-356. [doi: 10.1111/j.1525-1446.2010.00933.x] [Medline: 21736613]
13. Trede F, Sutton K, Bernoth M. Conceptualisations and perceptions of the nurse preceptor's role: a scoping review. Nurse Educ Today 2016 Jan;36:268-274. [doi: 10.1016/j.nedt.2015.07.032] [Medline: 26276530]
14. Arbabi H, Johnson J, Forgrave D. The effect of preceptorship on nurses' training and preparation with implications for Qatar: a literature review. J Nurs Educ Prac 2018 Feb 1;8(7):44. [doi: 10.5430/jnep.v8n7p44]
15. Kim KH. Clinical competence among senior nursing students after their preceptorship experiences. J Prof Nurs 2007;23(6):369-375. [doi: 10.1016/j.profnurs.2007.01.019] [Medline: 18053963]
16. Glogowska M, Young P, Lockyer L, Moule P. How 'blended' is blended learning?: students' perceptions of issues around the integration of online and face-to-face learning in a continuing professional development (CPD) health care context. Nurse Educ Today 2011 Nov;31(8):887-891. [doi: 10.1016/j.nedt.2011.02.003] [Medline: 21388722]
17. Tudorache IC, Iordache AM, Iordache MT. The advantages and the disadvantages of e-learning in European union. J Inform Syst Manage 2012;6(2):1-7 [FREE Full text]
18. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. J Am Med Assoc 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
19. George R, Papachristou N, Belisario J, Wang W, Wark P, Cotic Z, et al. Online elearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
20. Wu XV, Chan YS, Tan KH, Wang W. A systematic review of online learning programs for nurse preceptors. Nurse Educ Today 2018 Jan;60:11-22. [doi: 10.1016/j.nedt.2017.09.010] [Medline: 28987893]
21. Wu XV, Chi Y, Chan YS, Wang W, Ang EN, Zhao S, et al. A web-based clinical pedagogy program to enhance registered nurse preceptors' teaching competencies-an innovative process of development and pilot program evaluation. Nurse Educ Today 2020 Jan;84:104215. [doi: 10.1016/j.nedt.2019.104215] [Medline: 31683138]
22. Xing W, Ao L, Xiao H, Liang L. Chinese nurses' preferences for and attitudes about e-learning in continuing education: a correlational study. J Contin Educ Nurs 2020 Feb 1;51(2):87-96. [doi: 10.3928/00220124-20200115-08] [Medline: 31978247]

23. Frehywot S, Vovides Y, Talib Z, Mikhail N, Ross H, Wohltjen H, et al. E-learning in medical education in resource constrained low- and middle-income countries. Hum Resour Health 2013 Feb 4;11:4 [FREE Full text] [doi: 10.1186/1478-4491-11-4] [Medline: 23379467]
24. Moule P, Ward R, Lockyer L. Issues with e-learning in nursing and health education in the UK: are new technologies being embraced in the teaching and learning environments? J Res Nurs 2010 May 27;16(1):77-90. [doi: 10.1177/1744987110370940]
25. Parker EB, Wassef ME. Flexible online learning options for graduate nursing students. Nurse Educ 2010;35(6):243-247. [doi: 10.1097/NNE.0b013e3181f7f177] [Medline: 20975408]
26. Cottrell S, Donaldson JH. Exploring the opinions of registered nurses working in a clinical transfusion environment on the contribution of e-learning to personal learning and clinical practice: results of a small scale educational research study. Nurse Educ Pract 2013 May;13(3):221-227. [doi: 10.1016/j.nepr.2013.01.014] [Medline: 23474272]
27. Talebian S, Mohammadi HM, Rezvanfar A. Information and communication technology (ICT) in higher education: advantages, disadvantages, conveniences and limitations of applying e-learning to agricultural students in Iran. Procedia Soc Behav Sci 2014 Oct;152:300-305. [doi: 10.1016/j.sbspro.2014.09.199]
28. Haigh J. Information technology in health professional education: why IT matters. Nurse Educ Today 2004 Oct;24(7):547-552. [doi: 10.1016/j.nedt.2004.07.008] [Medline: 15465170]
29. Bliuc A, Goodyear P, Ellis RA. Research focus and methodological choices in studies into students' experiences of blended learning in higher education. Internet High Educ 2007 Jan;10(4):231-244. [doi: 10.1016/j.iheduc.2007.08.001]
30. Garrison D, Kanuka H. Blended learning: uncovering its transformative potential in higher education. Internet High Educ 2004 Apr;7(2):95-105. [doi: 10.1016/j.iheduc.2004.02.001]
31. Sit JW, Chung JW, Chow MC, Wong TK. Experiences of online learning: students' perspective. Nurse Educ Today 2005 Feb;25(2):140-147. [doi: 10.1016/j.nedt.2004.11.004] [Medline: 15701540]
32. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet 2010 Dec 4;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
33. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME guide no 23. Med Teach 2012;34(6):e421-e444. [doi: 10.3109/0142159X.2012.680939] [Medline: 22578051]
34. Turk B, Ertl S, Wong G, Wadowski P, Löffler-Stastka H. Does case-based blended-learning expedite the transfer of declarative knowledge to procedural knowledge in practice? BMC Med Educ 2019 Dec 3;19(1):447 [FREE Full text] [doi: 10.1186/s12909-019-1884-4] [Medline: 31796049]
35. Bonney K. Case study teaching method improves student performance and perceptions of learning gains. J Microbiol Biol Educ 2015 May;16(1):21-28 [FREE Full text] [doi: 10.1128/jmbe.v16i1.846] [Medline: 25949753]
36. Graham RC. Blended learning systems: definition, current trends, and future directions. In: Bonk CJ, Graham CR, editors. The Handbook of Blended Learning: Global Perspectives, Local Designs. San Francisco, CA: Pfeiffer Publications; 2006:3-21.
37. Ocak MA. Blend or not to blend: a study investigating faculty members perceptions of blended teaching. World J Educ Technol 2010;2(3):196-210 [FREE Full text]
38. Jokinen P, Mikkonen I. Teachers' experiences of teaching in a blended learning environment. Nurse Educ Pract 2013 Nov;13(6):524-528. [doi: 10.1016/j.nepr.2013.03.014] [Medline: 23608218]
39. Smyth S, Houghton C, Cooney A, Casey D. Students' experiences of blended learning across a range of postgraduate programmes. Nurse Educ Today 2012 May;32(4):464-468. [doi: 10.1016/j.nedt.2011.05.014] [Medline: 21645947]
40. Posey L, Pintz C. Transitioning a bachelor of science in nursing program to blended learning: successes, challenges & outcomes. Nurse Educ Pract 2017 Sep;26:126-133. [doi: 10.1016/j.nepr.2016.10.006] [Medline: 28277260]
41. 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. [doi: 10.3758/bf03193146] [Medline: 17695343]
42. Wu XV, Enskär K, Lee CC, Wang W. A systematic review of clinical assessment for undergraduate nursing students. Nurse Educ Today 2015 Feb;35(2):347-359. [doi: 10.1016/j.nedt.2014.11.016] [Medline: 25497138]
43. Hsu L, Hsieh S, Chiu H, Chen Y. Clinical teaching competence inventory for nursing preceptors: instrument development and testing. Contemp Nurse 2014;46(2):214-224. [doi: 10.5172/conu.2014.46.2.214] [Medline: 24787255]
44. Parsons R. Improving preceptor self-efficacy using an online educational program. Int J Nurs Educ Scholarsh 2007;4(21):1-17. [doi: 10.2202/1548-923X.1339] [Medline: 18052919]
45. Liang J, Wu S, Tsai C. Nurses' internet self-efficacy and attitudes toward web-based continuing learning. Nurse Educ Today 2011 Nov;31(8):768-773. [doi: 10.1016/j.nedt.2010.11.021] [Medline: 21176859]
46. Chiu Y, Tsai C. The roles of social factor and internet self-efficacy in nurses' web-based continuing learning. Nurse Educ Today 2014 Mar;34(3):446-450. [doi: 10.1016/j.nedt.2013.04.013] [Medline: 23669599]
47. Ginns P, Ellis R. Quality in blended learning: exploring the relationships between on-line and face-to-face teaching and learning. Internet High Educ 2007 Jan;10(1):53-64. [doi: 10.1016/j.iheduc.2006.10.003]

48. George D, Mallery P. SPSS for Windows Step by Step: A Simple Study Guide and Reference, 17.0 Update. Tenth Edition. Boston, MA: Allyn & Bacon; 2010.
49. Micallef R, Kayyali R. A systematic review of models used and preferences for continuing education and continuing professional development of pharmacists. Pharmacy (Basel) 2019 Nov 16;7(4):154 [FREE Full text] [doi: 10.3390/pharmacy7040154] [Medline: 31744140]
50. World Health Organisation. 2017. Patient Safety: Making Health Care Safer URL: http://apps.who.int/iris/bitstream/10665/255507/1/WHO-HIS-SDS-2017.11-eng [accessed 2020-11-11]
51. Harrington S. Mentoring new nurse practitioners to accelerate their development as primary care providers: a literature review. J Am Acad Nurse Pract 2011 Apr;23(4):168-174. [doi: 10.1111/j.1745-7599.2011.00601.x] [Medline: 21489010]
52. Kacel B, Millar M, Norris D. Measurement of nurse practitioner job satisfaction in a midwestern state. J Am Acad Nurse Pract 2005 Jan;17(1):27-32. [doi: 10.1111/j.1041-2972.2005.00007.x] [Medline: 15679881]
53. Latessa R, Beaty N, Landis S, Colvin G, Janes C. The satisfaction, motivation, and future of community preceptors: the north Carolina experience. Acad Med 2007 Jul;82(7):698-703. [doi: 10.1097/ACM.0b013e318067483c] [Medline: 17595570]
54. Payakachat N, Ounpraseuth S, Ragland D, Murawski MM. Job and career satisfaction among pharmacy preceptors. Am J Pharm Educ 2011 Oct 10;75(8):153 [FREE Full text] [doi: 10.5688/ajpe758153] [Medline: 22102743]
55. Seo H, Ryu K, Lee S, Noh J, La H, Kim E. Stress, satisfaction, and competency of hospital pharmacy preceptors under the new pharmacy program in south Korea. Am J Pharm Educ 2018 Oct;82(8):6351 [FREE Full text] [doi: 10.5688/ajpe6351] [Medline: 30425399]
56. Knox S, Cullen W, Dunne C. Continuous professional competence (CPC) for emergency medical technicians in Ireland: educational needs assessment. BMC Emerg Med 2013 Dec 17;13:25 [FREE Full text] [doi: 10.1186/1471-227X-13-25] [Medline: 24345064]
57. Knox S, Cullen W, Dunne C. Continuous professional competence (CPC) for Irish paramedics and advanced paramedics: a national study. BMC Med Educ 2014 Mar 2;14:41 [FREE Full text] [doi: 10.1186/1472-6920-14-41] [Medline: 24580830]
58. Howlett D, Vincent T, Watson G, Owens E, Webb R, Gainsborough N, et al. Blending online techniques with traditional face to face teaching methods to deliver final year undergraduate radiology learning content. Eur J Radiol 2011 Jun;78(3):334-341. [doi: 10.1016/j.ejrad.2009.07.028] [Medline: 19729259]
59. Makhdoom N, Khoshhal KI, Algaidi S, Heissam K, Zolaly MA. 'Blended learning' as an effective teaching and learning strategy in clinical medicine: a comparative cross-sectional university-based study. J Taibah Univ Sci 2013 Apr;8(1):12-17. [doi: 10.1016/j.jtumed.2013.01.002]
60. Wu J, Tennyson RD, Hsia T. A study of student satisfaction in a blended e-learning system environment. Comput Educ 2010 Aug;55(1):155-164. [doi: 10.1016/j.compedu.2009.12.012]
61. Tobin K. Qualitative perceptions of learning environments on the world wide web. Learn Environ Res 1998;1(2):139. [doi: 10.1023/A:1009953715583]
62. Sinclair P, Kable A, Levett-Jones T. The effectiveness of internet-based e-learning on clinician behavior and patient outcomes: a systematic review protocol. JBI Database System Rev Implement Rep 2015 Jan;13(1):52-64. [doi: 10.11124/jbisrir-2015-1919] [Medline: 26447007]
63. Johannesen M, Øgrim L, Giæver T. Notion in motion: teachers' digital competence. Nordic J Digit Literacy 2014;4:300-312 [FREE Full text]
64. Dutra DK. Implementation of case studies in undergraduate didactic nursing courses: a qualitative study. BMC Nurs 2013 Jul 4;12(1):15-23 [FREE Full text] [doi: 10.1186/1472-6955-12-15] [Medline: 23826925]
65. Haraldseid C, Friberg F, Aase K. How can students contribute? A qualitative study of active student involvement in development of technological learning material for clinical skills training. BMC Nurs 2016;15(1):2 [FREE Full text] [doi: 10.1186/s12912-016-0125-y] [Medline: 26759529]
66. Timmins F, Cassidy S, Nugent O, Lydon C, Part S, Keane L, et al. Reluctance to fail nursing students in practice-implications for nurse managers. J Nurs Manag 2017 Oct;25(7):489-490. [doi: 10.1111/jonm.12508] [Medline: 29024202]
67. Allen IE, Seaman J. Changing Course: Ten Years of Tracking Online Education in the United States. London, UK: Babson Survey Research Group and Quahog Research Group; 2013.
68. Keržič D, Tomaževič N, Aristovnik A, Umek L. Exploring critical factors of the perceived usefulness of blended learning for higher education students. PLoS One 2019;14(11):e0223767 [FREE Full text] [doi: 10.1371/journal.pone.0223767] [Medline: 31751345]

Abbreviations

AWCLS: attitudes toward a web-based continuing learning survey
CTBL: clinical teaching blended learning
CTCI: clinical teaching competence inventory
e-learning: electronic learning
LEQ: e-Learning experience questionnaire
PSEQ: preceptor self-efficacy questionnaire

WCP: web-based clinical pedagogy

Edited by G Eysenbach; submitted 09.03.20; peer-reviewed by A Younas, P Sousa, M Grasl; comments to author 24.03.20; revised version received 26.03.20; accepted 26.03.20; published 24.04.20.

Please cite as:
Wu XV, Chi Y, Panneer Selvam U, Devi MK, Wang W, Chan YS, Wee FC, Zhao S, Sehgal V, Ang NKE
A Clinical Teaching Blended Learning Program to Enhance Registered Nurse Preceptors' Teaching Competencies: Pretest and Posttest Study
J Med Internet Res 2020;22(4):e18604
URL: http://www.jmir.org/2020/4/e18604/
doi:10.2196/18604
PMID:32329743

Original Paper

Global Reach of an Online COVID-19 Course in Multiple Languages on OpenWHO in the First Quarter of 2020: Analysis of Platform Use Data

Heini Utunen[1], MA, BSocSci; Ngouille Ndiaye[1], MPH; Corentin Piroux[1], MScM; Richelle George[1], BA; Melissa Attias[1], MSc, BA; Gaya Gamhewage[1], MD, MPH

Learning and Capacity Development Unit, World Health Organization Health Emergencies Programme, World Health Organization, Geneva, Switzerland

Corresponding Author:
Heini Utunen, MA, BSocSci
Learning and Capacity Development Unit
World Health Organization Health Emergencies Programme
World Health Organization
20 Avenue Appia
Geneva, 1211
Switzerland
Phone: 41 795933476
Email: utunenh@who.int

Abstract

Background: At the onset of the coronavirus outbreak, the World Health Organization's (WHO) Health Emergencies Learning and Capacity Development Unit, together with the WHO's health technical lead on coronaviruses, developed a massive open online course within 3 weeks as part of the global response to the emergency. The introductory coronavirus disease (COVID-19) course was launched on January 26, 2020, on the health emergencies learning platform OpenWHO.org.

Objective: The aim of this paper is to investigate the geographic reach of different language courses accessed by a worldwide audience seeking information on COVID-19. Users' professional identities and backgrounds were explored to inform course owners on the use case. The course was developed and delivered via the open-access learning platform OpenWHO.org. The self-paced resources are available in a total of 13 languages and were produced between January 26 and March 25, 2020.

Methods: Data were collected from the online courses' statistical data and metrics reporting system on the OpenWHO platform. User patterns and locations were analyzed based on Google Analytics and the platform's own statistics capabilities, and data sets were overlaid. This analysis was conducted based on user location, with the data disaggregated according to the six WHO regions, the top 10 countries, and the proportion of use for each language version. Data included affiliation, gender, age, and other parameters for 32.43% (52,214/161,007) of the users who indicated their background.

Results: As of March 25, 2020, the introductory COVID-19 course totaled 232,890 enrollments across all languages. The Spanish language course was comprised of more than half (n=118,754, 50.99%) of all course enrollments, and the English language course was comprised of 38.21% (n=88,988) of enrollments. The WHO's Region of the Americas accounted for most of the course enrollments, with more than 72.47% (138,503/191,130) enrollment across all languages. Other regions were more evenly distributed with less than 10% enrollment for each. A total of 32.43% (52,214/161,007) of users specified a professional affiliation by choosing from the 12 most common backgrounds in the OpenWHO user profiles. Before the COVID-19 pandemic, users were spread over the 11 distinct affiliations, with a small fraction of users identifying themselves as "Other." With the COVID-19 introductory course, the largest number of users selected "Other" (16,527/52,214, 31.65%), suggesting a large number of users who were not health professionals or academics. The top 10 countries with the most users across all languages were Argentina, Chile, Colombia, Ecuador, India, Mexico, Peru, Spain, the United Kingdom, and the United States.

Conclusions: The online course has addressed a worldwide learning need by providing WHO's technical guidance packaged in simple formats for access and use. The learning material development was expedited to meet the onset of the epidemic. Initial data suggest that the various language versions of the course, in particular Spanish, have reached new user groups, fulfilling the platform's aim of providing learning everywhere to anyone that is interested. User surveys will be carried out to measure the real impact.

(J Med Internet Res 2020;22(4):e19076) doi:10.2196/19076

KEYWORDS

online learning; OpenWHO; novel coronavirus; COVID-19; coronavirus; pandemic; WHO; e-learning; MOOC; public health

Introduction

Background

The focus of this study is the World Health Organization's (WHO) health emergencies platform OpenWHO.org, which hosts online learning resources for outbreaks and epidemics. OpenWHO is an open source online platform adjusted for low bandwidths with mobile and download capabilities. This free, web-based knowledge transfer platform was designed for massive, real time use during a pandemic and has been in a real test during the early part of 2020 in offering free online courses to improve the response and preparedness for coronavirus disease (COVID-19).

The COVID-19 resources are hosted on two learning channels on the platform: one for courses in official UN languages and a second for courses in additional national languages. The first course related to COVID-19, "Introduction to emerging respiratory viruses, including COVID-19: methods for detection, prevention, response and control," was launched on OpenWHO on January 26, 2020, following the first WHO Emergency Committee meeting on January 22 and 23, 2020.

The course has four modules:

1. Introduction to emerging respiratory viruses, including COVID-19
2. Detecting emerging respiratory viruses, including COVID-19: surveillance and laboratory investigation
3. Risk communication and community engagement
4. Preventing and responding to an emerging respiratory virus, including COVID-19

This first edition of the course includes introductory information on the novel coronavirus and other coronaviruses, and basic information for anyone wanting to understand the new epidemic. By the end of the course, learners should be able to describe the nature of emerging respiratory viruses, how to detect and assess an outbreak, and strategies for preventing and controlling outbreaks due to novel respiratory viruses, as well as what strategies should be used to communicate risk and engage communities. The course was packaged and presented through PowerPoint slide decks, and video recordings by WHO health experts in different areas of work were added during the following days and weeks. This simple packaging allows for material to be used on multiple devices, in low-bandwidth settings, and through an offline function in the mobile app. The packaging of the online courses with videos and slide decks also supports the ability to update the material to reflect the frequently changing and updated WHO technical guidance.

Due to the changing content during the first months of the Public Health Emergency of International Concern and the first weeks of the pandemic, the OpenWHO team did not make available any quizzes as is usual for other courses on the platform and, thus, was not providing a certificate of completion. These features will be added once the technical content can be considered more final and established.

The timeline in Figure 1 shows the launch of the course in different languages and the total enrollments as of March 27, 2020, the day the OpenWHO platform reached 1 million enrollments.

The course was published on January 26, 2020, in English and was gradually published in all other UN languages. Another 7 national languages were produced by the courtesy of dozens of volunteers providing spontaneous translation offers from WHO country and regional offices and headquarters, as well as volunteers from public health institutes and educational units. All volunteers were seeking to provide support to maximize the local-level uptake of courses for an effective response to the pandemic. Among other efforts, OpenWHO released an introductory video to COVID-19 in Indian sign language, the first sign language resource on the platform.

This study looks at the course use in all of the WHO regions. The WHO Member States are grouped into six WHO regions: African Region, Region of the Americas, Eastern Mediterranean Region, European Region, South-East Asia Region, and Western Pacific Region, as shown in Figure 2.

Figure 1. Introduction to coronavirus disease course languages and enrollment figures as of March 27, 2020.

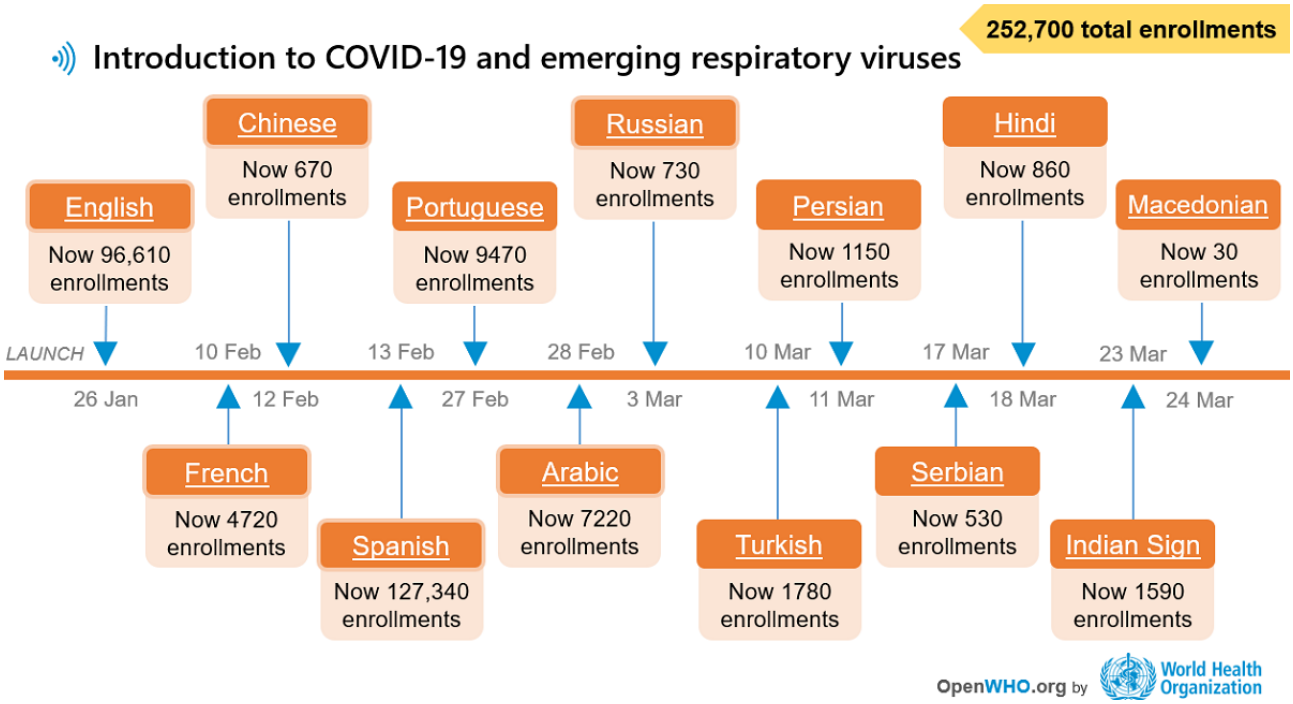
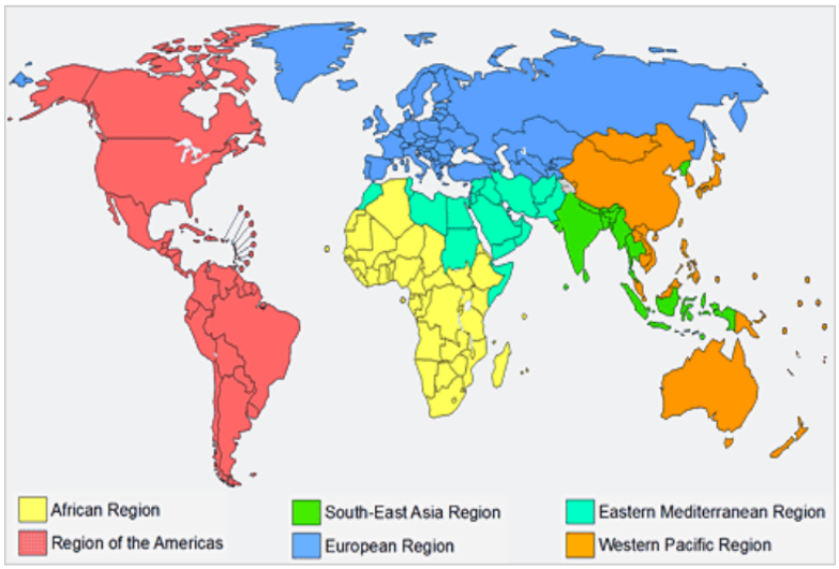


Figure 2. World Health Organization regions.



Relevant Literature

According to the WHO [1], "health literacy represents the cognitive and social skills that determine the motivation and ability of individuals to gain access to understand and use information in ways which promote and maintain good health." Research has found that the WHO is a respected source of accurate health information during epidemics, suggesting that the organization has a platform to shape health behavior, which less-trusted sources such as governments may lack [2]. The OpenWHO team has focused on designing knowledge transfer resources for health emergencies in formats and languages that are suitable for frontline responders and affected communities. This approach has prioritized multi-language production, recognizing that language can be a key obstacle to health literacy [3].

A 2015 study in Kenya by Translators Without Borders found that providing health information in Swahili—the lingua franca throughout the country—produced a significant increase in comprehension compared with providing the same information in English [4]. A 2019 study by the same organization, which partners with the WHO team on many translations, found that the local form of Swahili was the most effective language for risk communication and community engagement for the Ebola response in Goma, Democratic Republic of the Congo, compared with the French and standard Swahili languages [5]. Providing health information in individuals' native languages has also shown to improve knowledge about illnesses and medications in a patient population in Sri Lanka, as well as the

understanding of oral health information among Vietnamese-speaking mothers in Australia [6,7].

Materials produced for OpenWHO are designed with additional accessibility considerations in mind. The resources are offered as downloadable slides that combine images and short texts, which can be read on a mobile device. Video and audio formats are also being integrated for those with strong oral cultures. In addition, the open-access nature of the platform can empower individuals who are more health literate to strengthen the health literacy of their communities, particularly in group-oriented societies. In a cholera-endemic neighborhood in Ghana, researchers found that household units impacted individual health literacy; nearly three-quarters of households surveyed followed suggestions from household members on how to prevent cholera [8].

By making materials from a reputable source available in multiple languages and in easily accessible, portable formats, OpenWHO's COVID-19 resources aim to contribute to improved health literacy.

Methods

This article investigates the geographic reach of different language courses accessed by the worldwide audience population while seeking health information on COVID-19. The users' professional identities are explored to inform the course owners on the use case. The course was developed and delivered via the learning platform OpenWHO.org. The self-paced introductory course was provided in a total of 13 languages between January 26 and March 25, 2020. The 13 languages were Arabic, Chinese, English, French, Hindi, Indian sign language, Macedonian, Persian, Portuguese, Russian, Serbian, Spanish, and Turkish.

This study's preliminary objective was to demonstrate the rapid surge of learners accessing the digitized learning materials as the COVID-19 epidemic grew into a pandemic during the early part of 2020. The aim of this study was to obtain a better understanding of the origin and type of people who sought access to online learning related to the emerging health crisis.

Statistical data for the identical courses in 13 languages were generated. More in-depth analysis was carried out on the English and Spanish language courses given the large use case in these two languages—89.20% (207,742/232,890) of all learners.

The data was collected from the online courses' statistical data and metrics reporting system on the OpenWHO platform. User patterns and locations were analyzed based on Google Analytics and the OpenWHO platform's own statistical capabilities, and data sets were overlaid.

This snapshot analysis was conducted based on user location with the data disaggregated according to the six WHO regions, the top 10 countries with the most users, and the proportion of use for each language version. Data included affiliation, gender, primary language, age, and other parameters for approximately 30% of users who indicated their background.

Results

Introductory COVID-19 Course User Metrics

During the first 2 months of the course's availability (January 26 to March 25, 2020), all 13 languages combined gathered 232,890 enrollments. The use of materials intensified after the declaration of the COVID-19 pandemic on March 11, 2020 (Figure 3).

The two most popular languages were the Spanish language course, which comprised half (n118,754/232,890, 50.99%) of all introductory COVID-19 course enrollments as of March 25, 2020, and the English language course, which comprised 38.21% (n=88,988) of enrollments. Despite being launched 2 weeks later, the Spanish version rapidly surpassed the original English course on March 19, 2020, becoming the main language driving the increase in enrollment. The English and Spanish course users jointly amounted to 210,000 users, representing 90.17% of all users across all language versions.

As expected, the number of accumulated enrollments across all language versions of the course rose steadily with each new language version launched (Figure 4).

In terms of user locations, the top 10 countries with the most users across all languages, accounting for (40,893/57,763) 70.79% of the total course users, were Ecuador (n=9521), Mexico (n=8236), Colombia (n=6137), Chile (n=4067), the United States (n=3836), Argentina (n=3609), Spain (n=1766), the United Kingdom (n=1499), India (n=1117), and Peru (n=925).

This makeup represents a marked shift in comparison with other top courses on the platform. For example, prior to the current outbreak of COVID-19, the platform's most popular emergency-related course, eProtect Ebola (offered in English and French), consisted primarily of users from Africa, Europe, and North America, with no South or Central American countries appearing in the top 20 for either language version of the course. This pattern reflected what was a general trend across the platform in the months preceding the launch of the introductory COVID-19 course; the platform's top 5 countries most commonly consisted of, in descending order, the United States, India, the United Kingdom, Portugal, and Nigeria, with only the presence of India breaking the aforementioned trend. As such, the launch of the introductory COVID-19 course has brought with it a change in the demographic of users on the platform, with a marked increase in traction from Central and South America.

Figure 3. Introduction to coronavirus disease course use by language as of March 25, 2020.

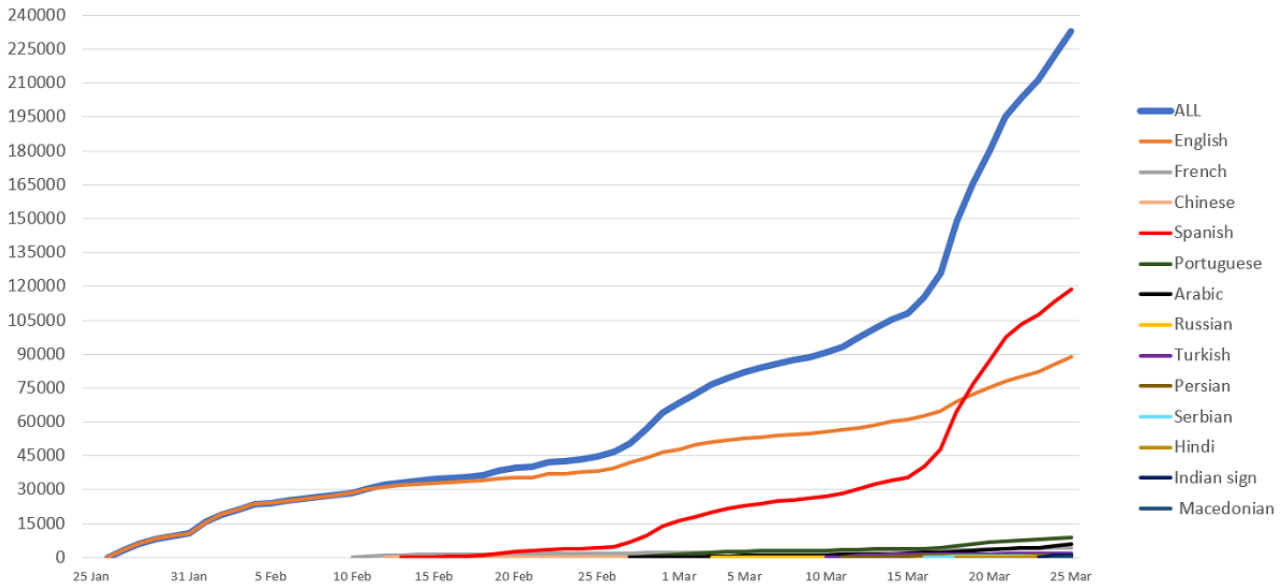
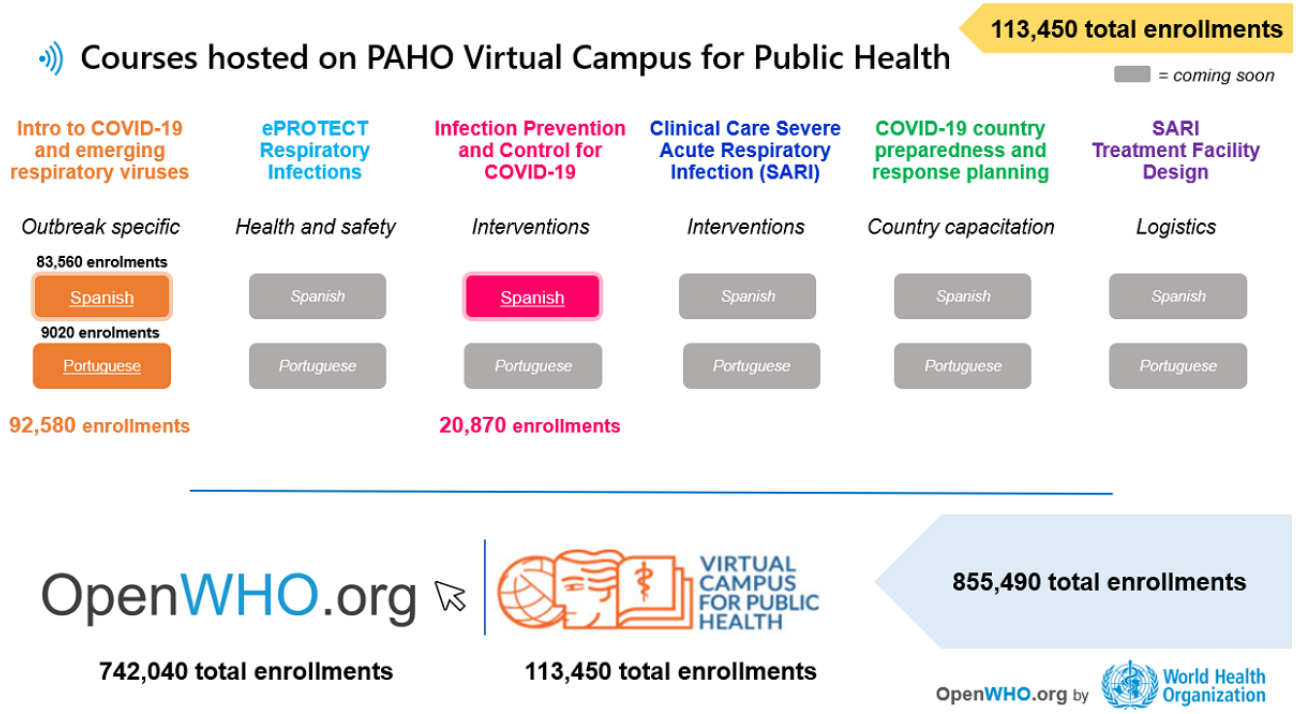


Figure 4. OpenWHO courses hosted on the Pan-American Health Organization virtual campus as of March 27, 2020. COVID-19: coronavirus disease.



Geographical Distribution of Users

The worldwide distribution of users is displayed in Figure 5 according to the WHO regions. The Region of the Americas accounted for most of the enrollments in the course, with more than (138,503/191,130) 72.47% of the total enrollments across all languages. Other regions were more evenly distributed: African Region (n=7643, 4.00%), Eastern Mediterranean Region (n=12,945, 6.77%), European Region (n=18,259, 9.55%), South-East Asia Region (n=7245, 3.79%), Western Pacific Region (n=5291, 2.77%), and territories (n=1244, 0.65%).

Users originating from the Region of the Americas accounted for over half of the enrollments in the Spanish language course. The second most popular language choice for users in the region was English, with a majority of English-course users enrolling from the United States of America, totaling more than a quarter (n=23,794/138,503, 28.75%) of the total English course users (Table 1).

Each language version provided interesting findings. The Serbian language course, for example, was used more in 4 other countries (Ecuador, Chile, Colombia, and Bosnia and Herzegovina) than it was in Serbia. The Indian sign language course attracted 1000 enrollments in the first 24 hours, with the largest use in the city of Baghdad, Iraq. Just after Portugal, the second top country for the Portuguese course was Mexico. For this version, almost as many enrollments came from Mexico City as from Lisbon. After Lisbon and Mexico City, 3 other

Spanish-speaking cities were in the top 5 for the Portuguese version (Bogotá and Medellín in Colombia, and Quito in Ecuador), exceeding the numbers of enrollments from any other Portuguese-speaking city. São Paulo and Rio de Janeiro (Brazil) were the sixth and seventh top cities, respectively, for the Portuguese version of the course. For the French course, the third top city was not French-speaking as one would expect; Mexico City came right after Paris, France, and Bukavu, Democratic Republic of the Congo.

The introductory course was the most popular in the Region of the Americas, with almost three-fourths of the total enrollments across all languages concentrated in this region. This trend might be explained in part by the language availability of the course for the first month, which was limited to English and Spanish along with the French and Chinese versions (Figure 1). Totaling 54.00% (103,207/191,130) of the total course enrollments, the Spanish version had 95.53% (98,554/103,207) of its enrollments from the Region of the Americas. With 37.03% (70,774/191,130) of the total course enrollments, the English version had 46.39% (32,841/70,774) of its enrollments from this region. The latter statistic speaks to the more even distribution of the 70,744 enrollments in the English language course across the other WHO regions: African Region 8.44% (n=5975), Eastern Mediterranean Region 12.15% (n=8597), European Region 15.82% (n=11,194), South-East Asia Region 9.24% (n=6539), Western Pacific Region 6.96% (n=4924), and the territories 0.99% (n=704). Even taking into consideration the gradual release of the language versions' impact on these statistics, it is worth noting that no other languages had similar slopes as those seen for the Spanish and English versions (Figure 3).

The European Region came second with 9.55% (18,259/191,130) of the total enrollments, mainly distributed between the English (11,194/18,259, 61.31%) and Spanish (n=3942, 21.59%) versions of the course. The Eastern Mediterranean Region comprised 6.77% (12,945/191,130) of the total enrollments, mainly distributed between the English (n=8597/12,945, 66.41%) and Arabic (n=3419, 26.41%) versions. These proportions might also be explained by the release of the Arabic version 1 month after the English version. The African and South-East Asia regions each made up about 4% of the total enrollments, with the main language versions used being English (5975/7643, 78.18%) and French (n=1255, 16.42%) for the African Region and primarily English (6539/7245, 90.25%) for the South-East Asia Region. This was similar to the Western Pacific Region, which comprised of 2.77% (5291/191,130) of the total course enrollments. The English course was accessed by 93.06% (4924/5291) of the enrollments in this region. The territories comprised 0.65% (1244/191,130) of the total course enrollments, also mainly using the English (704/1244, 56.59%) and Spanish (n=341, 27.41%) versions.

Figure 5. Users from the top 10 countries and their affiliations. GOARN: Global Outbreak Alert and Response Network; NGO: nongovernmental organization; WHO: World Health Organization.

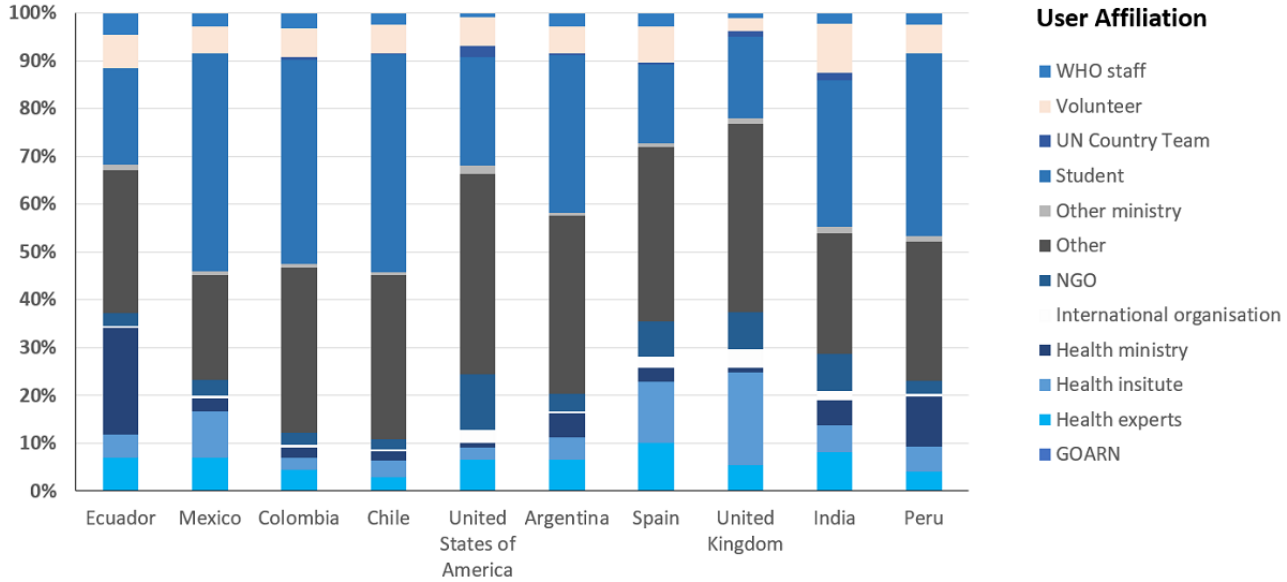


Table 1. Introductory course use overview by World Health Organization region and language used for course enrollments (N=191,130).

Region	Total, n (%)	English, n (%)	French, n (%)	Russian, n (%)	Hindi, n (%)	Spanish, n (%)	Persian, n (%)	Portuguese, n (%)	Serbian, n (%)	Turkish, n (%)	Arabic, n (%)	Chinese, n (%)
All regions	191,130 (100.00)	70,774 (37.03)	3398 (1.78)	486 (0.25)	348 (0.18)	103,207 (54.00)	716 (0.37)	5934 (3.10)	296 (0.15)	1274 (0.67)	4208 (2.20)	489 (0.26)
AFRO[a]	7643 (4.00)	5975 (3.13)	1255 (0.66)	3 (0.00)	3 (0.00)	97 (0.05)	36 (0.02)	123 (0.06)	6 (0.00)	63 (0.03)	68 (0.04)	14 (0.01)
AMRO[b]	138,503 (72.47)	32,841 (17.18)	915 (0.48)	98 (0.05)	10 (0.01)	98,554 (51.56)	193 (0.10)	4899 (2.56)	193 (0.10)	462 (0.24)	188 (0.10)	150 (0.08)
EMRO[c]	12,945 (6.77)	8597 (4.50)	284 (0.15)	15 (0.01)	10 (0.01)	137 (0.07)	264 (0.14)	56 (0.03)	21 (0.01)	112 (0.06)	3419 (1.79)	30 (0.02)
EURO[d]	18,259 (9.55)	11,194 (5.86)	821 (0.43)	354 (0.19)	5 (0.00)	3942 (2.06)	156 (0.08)	806 (0.42)	68 (0.04)	456 (0.24)	395 (0.21)	62 (0.03)
SEARO[e]	7245 (3.79)	6539 (3.42)	71 (0.04)	11 (0.01)	313 (0.16)	72 (0.04)	39 (0.02)	28 (0.01)	7 (0.00)	128 (0.07)	16 (0.01)	21 (0.01)
WPRO[f]	5291 (2.77)	4924 (2.58)	31 (0.02)	5 (0.00)	7 (0.00)	64 (0.03)	22 (0.01)	20 (0.01)	1 (0.00)	48 (0.03)	19 (0.01)	150 (0.08)
Territories	1244 (0.65)	704 (0.37)	21 (0.01)	0 (0.00)	0 (0.00)	341 (0.18)	6 (0.00)	2 (0.00)	0 (0.00)	5 (0.00)	103 (0.05)	62 (0.03)

[a]AFRO: African Region.

[b]AMRO: Region of the Americas.

[c]EMRO: Eastern Mediterranean Region.

[d]EURO: European Region.

[e]SEARO: South-East Asia Region.

[f]WPRO: Western Pacific Region.

Spanish Course Use Case

The highest Spanish course use was in Ecuador (n=36,345), Mexico (n=26,141), Colombia (n=19,733), Chile (n=11,793), and Argentina (n=11,711). The highest English course use was in the United States (n=12,250), Mexico (n=7659), Ecuador (n=5805), India (n=5296), and the United Kingdom (n=4052), with Colombia and Argentina also making it to the top 10.

Of the total users who indicated their language of preference, 61.45% (60,800/98,937) selected Spanish. This language preference helps the OpenWHO team further target the courses and prioritize languages. The other preferred languages selected were English (n=31,837, 32.18%), French (n=2369, 2.39%), Portuguese (n=1374, 2.17%), Arabic (n=1374, 1.39%), Russian (n=249, 0.25%), and Chinese (n=160, 0.16%).

The indicated language preference correlated with the language course use; however, as the option for the preferred language only included the six UN official languages and Portuguese, the OpenWHO team was not able to capture if there were any national or local languages popular in addition to the official UN languages.

The same introductory course was published on the Pan-American Health Organization's (PAHO) virtual campus (VC) in Spanish on February 11 and in Portuguese on February 28, 2020. This paper and the data exclude the 92,000 users of these identical courses on the PAHO VC, which would bring the merged worldwide user numbers even higher (Figure 4).

User Background Information

Before the COVID-19 pandemic, users were spread over the 11 distinct affiliation options provided by the platform, with a small fraction of users identifying themselves as "Other." With the epidemic accelerating into a pandemic, the largest number of the COVID-19 introductory course users selected "Other," suggesting a large number of users who were not health professionals or academics. Students were the largest identifiable group among those who indicated their affiliation. Health ministries and health experts made up 14.21% (7417/52,214), and UN country teams and WHO staff amounted to 4.00% (n=2087; Table 2).

Table 2. Users' professional affiliations for the total users who specified their professional affiliation (n=52,214).

Affiliations	Users, n (%)
Other	16,527 (31.65)
Student	14,945 (28.62)
Health ministry	4243 (8.13)
Volunteer	3330 (6.38)
Health experts	3174 (6.08)
Health institute	3141 (6.02)
Nongovernmental organization	2981 (5.71)
World Health Organization staff	1525 (2.92)
International organization	995 (1.91)
Other ministry	730 (1.40)
UN country team	562 (1.08)
Global Outbreak Alert and Response Network	61 (0.12)

When looking at the top countries and user affiliations in Figure 5, students accounted for a large proportion of the users coming from those countries. For Chile, Colombia, and Mexico, for instance, students represented 45.66% (1857/4067), 42.85% (2707/6317), and 45.70% (3764/8236), respectively, of the total enrollments. As mentioned earlier for the overall analysis, Figure 5 shows that many users were not affiliated with a specific health background. If we examine the figure, we find that the percentage of users who selected "Other" as their affiliation varied from 21.90% (1803/8236) in Mexico to 39.22% (588/1499) and 41.82% (1604/3836) in the United Kingdom and the United States, respectively. Health institutes were also represented in the affiliations for the users in the top 10 countries. The share of health institute workers was the highest for the United Kingdom at 19.48% (292/1499). Another interesting finding was the attraction of health ministry representatives, especially significant for Ecuador, as they represented 22.27% (2120/9521) of the total users for the country. Besides Ecuador, Peru (96/925, 10.37%) is the only other country reporting enrollments from health ministry professionals that reached at least 10%. Other health experts also enrolled in the course; they represented, for instance, 9.85% (174/1766) of the total enrollments reported from Spain. Volunteers were also represented in all of the top 10 countries, with India reporting the highest percentage of this specific user affiliation in 10.29% (115/1117) of its enrollments.

The course registrations also suggested that women (33,216/57,712, 57.55%) were a larger user group than men (n=24,383, 42.25%). In addition, 0.20% (113) of users identified as "Other".

When comparing the English and Spanish course age groups (Figures 6 and 7), the Spanish course had a large cohort of participants who were 70 years or older, which was much higher than the platform average or the English course. As COVID-19 was stated to be impacting older people the most, this use case comes as no surprise. On average, the English course had a younger use case than the platform average, especially in the age groups 30-39 years and 40-49 years.

Figure 6. English course use by age groups.

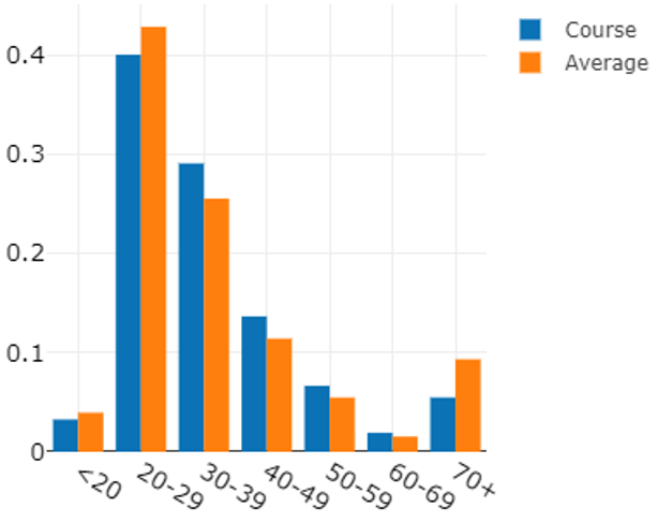
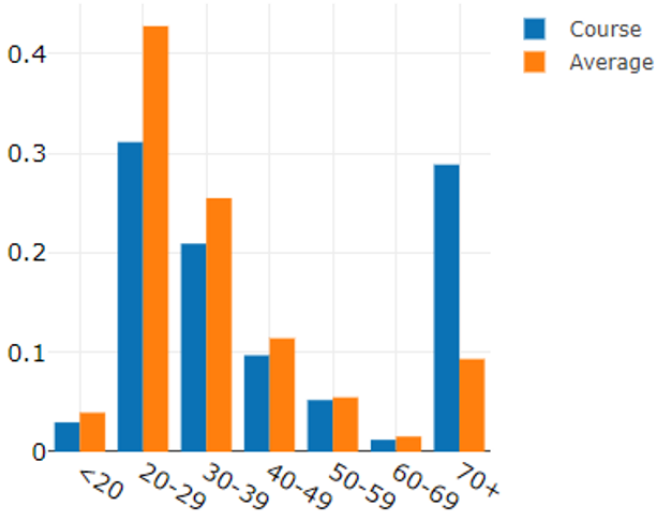


Figure 7. Spanish course use by age groups.



Spanish and English Courses and Completion Rates

A total of 36.43% (14,382/39,475) of participants that enrolled in the Spanish introductory course completed all six learning items within the course. In comparison with industry standards, which place the completion rate for massive open online courses (MOOCs) at 7.40% (3700 completions per 50,000 enrollments) [9], this rate is high. Again, in relation to the Spanish introductory course, 44.66% (17,630/39,475) of the participants completed at least 80.00% of the course material, and 49.09% (n=19,377) visited at least 60.00% of the course material. When the completion rate was calculated across all language versions of the course, this trend continued, with 21.63% (16,003/73,980) of users visiting 100.00% of the course items, 27.62% (n=20,433) visiting at least 80.00%, and 32.54% (n=24,075) of users visiting at least 60.00% of the total learning resources in the course.

Participants enrolled in the English version did not perform as well. Only 2.94% (813/27,639) of the participants who enrolled in the English course completed all course material, and only 6.61% (n=1828) completed at least 80.00% of the course material, and 16.65% (n=4603) visited at least half of all course items. Further investigation is required to determine the cause of the discrepancy between the completion rates for the English version of the course compared with the completion rate across all language versions combined. Unlike the subsequent language versions, the English introductory COVID-19 course was assembled over a period of several weeks, with new materials being made available as they were constructed and cleared by the technical experts responsible. In contrast, most of the subsequent language versions were launched as full packages with all course items available at once. This difference could begin to explain the discrepancy in completion rates, as the first set of users who enrolled in the English course would have had to return to the course at later dates to view new material as it was added.

Platform User Surge During the Early Weeks of the COVID-19 Pandemic

Since the launch of the first course related to COVID-19 on OpenWHO on January 26, 2020, the number of unique learners on the platform increased seven times, from 90,700 unique learners to 629,500 as of March 25. The introductory course brought the largest number of new learners along with the Infection Prevention and Control for COVID-19 course.

In the 2.5 years of operations prior to the coronavirus pandemic of 2020, outbreak-related learning resources were each used by thousands of users, with some courses such as eProtect occupational health and safety for Ebola and Antimicrobial Stewardship reaching up to 20,000-30,000 users over 2 years of the course's life span. The two most popular COVID-19 courses (Introduction and Infection Prevention and Control) attracted more than 200,000 learners each in less than 2 months.

Before COVID-19, there were on average some 100 course enrollments per day. During the first months of 2020, there were some 10,000-20,000 enrollments per day, with sharply increasing figures reaching up to 50,000 new learner registrations the week of the pandemic declaration (Figure 8). This testifies to the essence of OpenWHO offering health-related technical knowledge to frontline responders and the general public as an open and scalable solution for the fast distribution of lifesaving content in disease outbreaks and, in particular, during a pandemic.

OpenWHO has been working in full support of COVID-19 preparedness and response with the timely upload of learner resources, which was characterized by an accelerated process to make different language versions of the learning materials rapidly available. There was also an emphasis on quickly delivering the key available technical and operational information. Including the introductory COVID-19 course, a total of six courses were produced fully or partially in 40 different languages in the first quarter of 2020 (Figure 9).

During the early part of the coronavirus epidemic and pandemic, platform use shifted from health professionals and experts to largely non-health-related audiences. Between January 26 and

February 25, 2020, OpenWHO expanded from some 80,000 existing unique users to 160,000, doubling the number of learners. From February 26 to March 25, 2020, the unique user numbers almost quadrupled to 600,000. Including the enrollments in the same courses hosted on PAHO's VC platform, there were more than 840,000 enrollments in all of the COVID-19 courses.

After the declaration of the pandemic on March 11, 2020, the number of unique learners on the platform nearly tripled in 2 weeks, from 235,250 users to 629,500 users as of March 25. The increase from March 11 to March 25, 2020, consisted of a total of 394,250 new learners in merely 14 days. This number is more than four times higher than the 90,700 total users on the platform from 2017 to 2019 (Figure 10).

Figure 8. OpenWHO accumulated enrollments from June 30, 2017, to March 27, 2020. COVID-19: coronavirus disease.

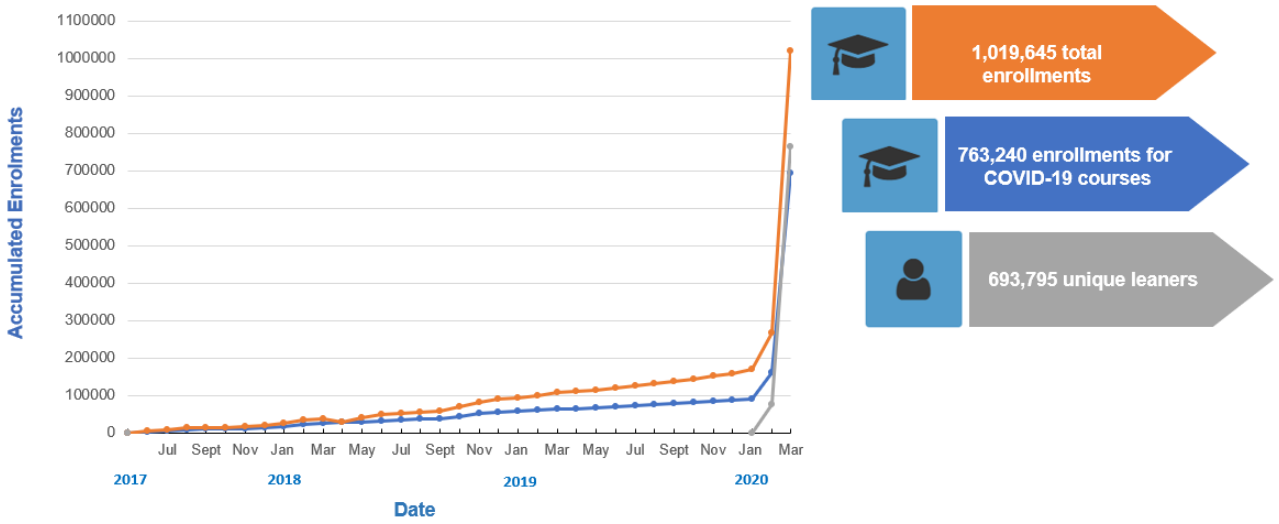


Figure 9. OpenWHO courses related to COVID-19 and language versions as of March 27, 2020. COVID-19: coronavirus disease.

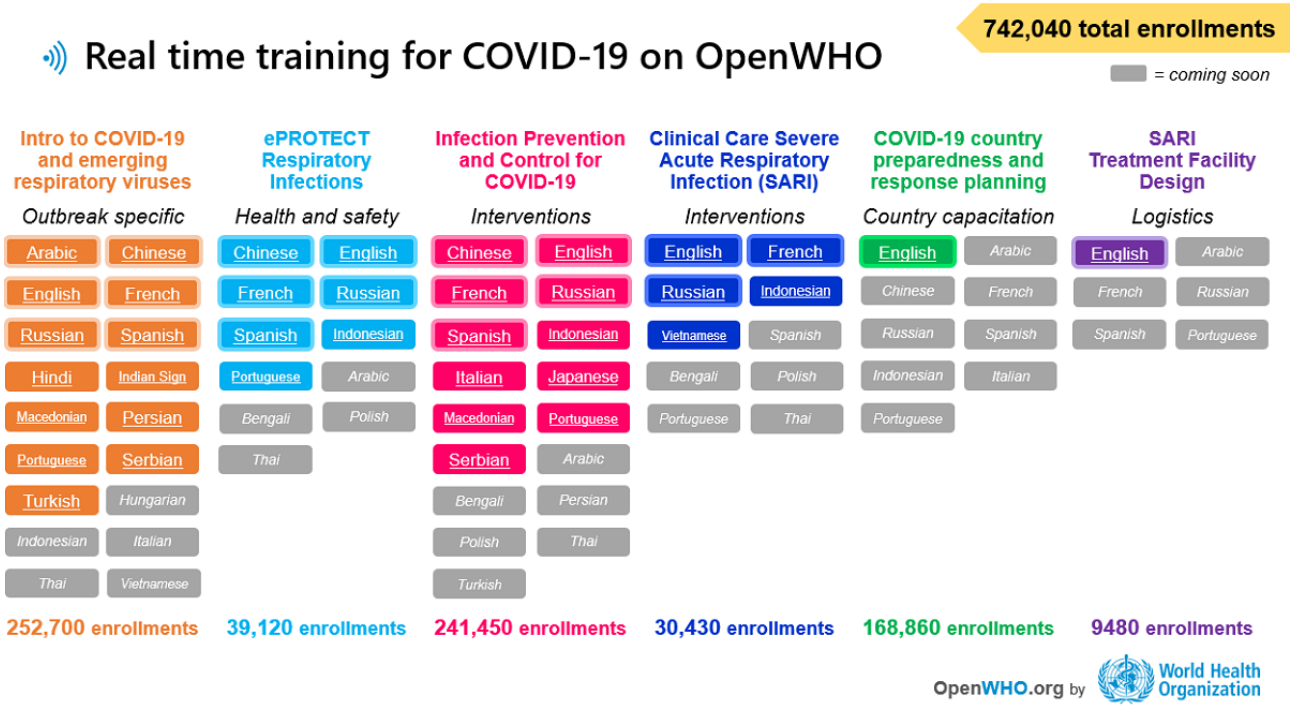
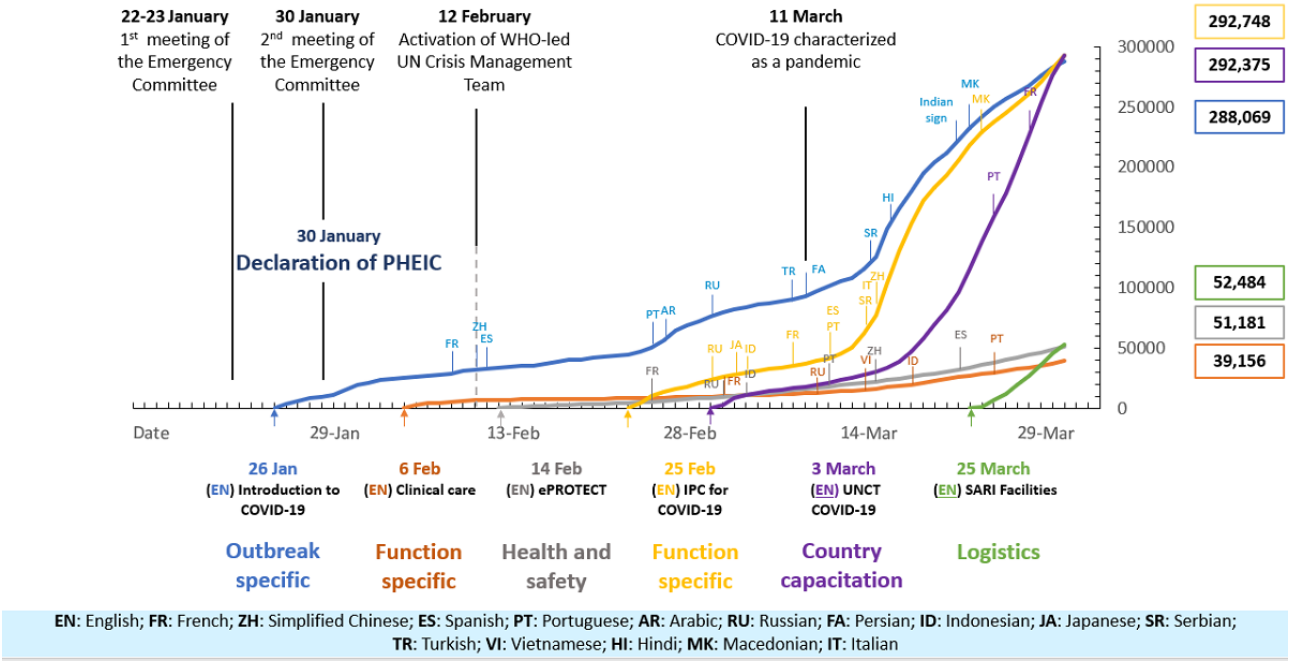


Figure 10. Timeline of all OpenWHO.org COVID-19 courses' launch and use. COVID-19: coronavirus disease; IPC: infection prevention and control; PHEIC: Public Health Emergency of International Concern; SARI: severe acute respiratory infection; UNCT: UN country teams; WHO: World Health Organization.



Discussion

Principal Findings

The OpenWHO learning resources are usually initially published in English, as this language is shared by many native and nonnative speakers across the world and is the most commonly used working language of the UN system. Recent trends seen on the platform during the pandemic, notably the popularity of the Spanish versions of the COVID-19 resources surpassing that of the original English courses, stressed the importance of a multilingual learning platform that allows learners to access information in the language they are the most comfortable using. This has been a continuous effort since the platform's inception, with courses even produced in local languages for localized emergencies. During the Ebola outbreak in the Democratic Republic of the Congo, for example, the Ebola knowledge resources for the responders course was published in Lingala and Swahili.

This multilingual approach is even more important in the middle of a pandemic, where OpenWHO is experiencing record-breaking enrollments from learners across the world who are seeking the latest WHO guidance to support the preparation and response to COVID-19. Indeed, in its first 2 months, the introductory course to COVID-19 comprised more than a quarter (n=241,749) of the accumulated total enrollments on a platform that has been running since June 2017. The analysis also indicated that substantial numbers of people were using the resource in languages other than the national languages of the countries in which they were located, suggesting that the platform provides a service that local governments may not be able to offer. With diaspora populations scattered across the world, the ability to access material in one's native language irrespective of location is important.

The geographic analysis demonstrated that the enrollment surge for this emergency-related course mainly originated from Central and South America, with Ecuador, Mexico, Columbia, Chile, and the United States rounding out the top 5 countries, reflecting a new demographic that is attracted to the platform. Interestingly, the age distribution analysis for the Spanish course, with enrollments mainly concentrated in the Region of the Americas (n=98,554/103,207, 95.49%), revealed the popularity of the course with those 70 years or older compared with the other courses hosted on the OpenWHO platform. This is consistent with the at-risk group seeking reliable knowledge on the emerging respiratory disease. Prior to the pandemic, the top countries for OpenWHO's most popular emergency courses were on the African continent, along with recurrent appearances by the United States and India, as these populous countries are big MOOC users.

This geographic shift is consistent with the nature of a pandemic. Rather than affecting specific parts of the world, as was the case for the 2014 Ebola outbreak, for example, the COVID-19 epidemic has accelerated into a pandemic reaching almost every country on the planet.

In contrast with the top emergency courses prior to the COVID-19 outbreak, the analysis also revealed that nearly 30% of users indicated that they were not affiliated with the student and professional health sectors (n=16,527/52,214, 31.65%). This shift reflects the impact a pandemic has on the audience profile, with the general public enrolling in the course to become informed on the novel coronavirus. It also suggests that OpenWHO can serve as one mechanism to help combat the "infodemic" of misinformation that has occurred during the COVID-19 outbreak [10]; research has found that many people turn to the internet for health information, including during crises, and the information they find can influence health behaviors [11-13].

OpenWHO hit the 1 million enrollments milestone on March 27, 2020. About three-quarters of the total enrollments were on the courses related to COVID-19. This reflects a massive increase in the popularity of OpenWHO and the critical role it is playing in supporting preparedness and response during this unprecedented pandemic. As the outbreak continues to evolve, new resources and language versions will be added to the platform to provide lifesaving knowledge to affected communities, and existing courses will be updated to best reflect the changing contexts.

Limitations and Future Research

This analysis was limited to the 2-month period following the launch of the introductory COVID-19 course on OpenWHO. Data such as affiliation, gender, and age were only available for 32.43% (52,214/161,007) of the users who indicated their background. The geographical data was based on the users' internet protocol addresses, which would not account for potential manipulation by virtual private networks or other factors. In addition, completion rates were measured by user visits to each of the course's learning items, as quizzes were not yet implemented due to the evolving nature of the emergency guidance.

Future research should examine the use case of additional COVID-19 courses on the OpenWHO platform that are more targeted to specific audiences, such as courses designed for clinicians and public health professionals. Data should also be analyzed against the worldwide, regional, and country-level epidemic curves for COVID-19 to identify broader use trends, including how the use case evolved throughout the pandemic, as different regions were more or less affected.

Conclusions

During health emergencies, lifesaving information must be packaged and delivered in the languages spoken by the target audiences to effectively transfer urgent knowledge [4,5]. Everyone has the right to access lifesaving knowledge, and OpenWHO is continuing to work with partners to make its resources available in as many languages as possible. The OpenWHO team has never before experienced such high levels of volunteerism for platform language production and has relied on crowdsourcing across the world to publish additional language versions. This has promoted the localization of materials into a variety of languages, helping people better protect themselves and fight the pandemic.

The OpenWHO platform offers courses on six distinct topics to support the COVID-19 response. These are products that transform the WHO guidance into learning packages that users can grasp and digest more easily. The courses have been translated and published fully or partially into 40 language versions during the first 2 months of the response, and OpenWHO has experienced an unprecedented increase in platform use. Amidst huge demand for reliable resources that offer the knowledge to understand and decipher the evolving situation, OpenWHO has served as one source of digitized information.

Acknowledgments

The OpenWHO platform operations have been made possible by the strong efforts of Oliver Stucke. The authors would also like to express their gratitude to Susan Spackman, Michael Reza Farzi, and Jana Martic from the Learning and Capacity Development Unit for providing editorial support for this paper.

The Spanish and Portuguese language collaboration has been enabled by the entire PAHO VC team, with the collaboration led by Ian Stein and Gabriel Listovsky.

No external funding and no sponsors were involved.

Conflicts of Interest

None declared.

References

1. World Health Organization. The WHO health promotion glossary URL: http://www.who.int/healthpromotion/HPG/en/ [accessed 2019-08-12]
2. Kendal E. Public health crises in popular media: how viral outbreak films affect the public's health literacy. Med Humanit 2019 Jan 19. [doi: 10.1136/medhum-2018-011446] [Medline: 30661040]
3. Hunter-Adams J, Rother H. A qualitative study of language barriers between South African health care providers and cross-border migrants. BMC Health Serv Res 2017 Jan 31;17(1):97. [doi: 10.1186/s12913-017-2042-5] [Medline: 28143514]
4. Translators Without Boarders. 2015 Apr. Does translated health-related information lead to higher comprehension? A study of rural and urban Kenyans URL: https://translatorswithoutborders.org/wp-content/uploads/2017/04/TWB_WoR_ImpactStudy_FINAL.pdf [accessed 2019-08-14]
5. Translators Without Borders. 2019 Mar. Missing the mark? People in Eastern DRC need information on Ebola in a language they understand: a rapid language needs assessment in Goma, DRC URL: https://translatorswithoutborders.org/wp-content/uploads/2019/03/DRC_Ebola_Assessment_English.pdf [accessed 2019-08-15]

6. Perera KYS, Ranasinghe P, Adikari AMMC, Balagobi B, Constantine GR, Jayasinghe S. Medium of language in discharge summaries: would the use of native language improve patients' knowledge of their illness and medications? J Health Commun 2012;17(2):141-148. [doi: 10.1080/10810730.2011.585926] [Medline: 22112212]
7. Arora A, Nguyen D, Do QV, Nguyen B, Hilton G, Do LG, et al. 'What do these words mean?': a qualitative approach to explore oral health literacy in Vietnamese immigrant mothers in Australia. Health Education Journal 2013 Feb 24;73(3):303-312. [doi: 10.1177/0017896912471051]
8. Tutu RA, Gupta S, Elavarthi S, Busingye JD, Boateng JK. Exploring the development of a household cholera-focused health literacy scale in James Town, Accra. J Infect Public Health 2019;12(1):62-69. [doi: 10.1016/j.jiph.2018.08.006] [Medline: 30219358]
9. Khalil H, Ebner M. MOOCs completion rates and possible methods to improve retention- a literature review. 2014 Jun 23 Presented at: World Conference on Educational Multimedia, Hypermedia and Telecommunications; 2014; Tampere, Finland p. 1305-1313.
10. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
11. Van de Belt TH, Engelen LJLPG, Berben SAA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct 02;15(10):e220. [doi: 10.2196/jmir.2607] [Medline: 24088272]
12. Kittler, Hobbs J, Volk LA, Kreps GL, Bates DW. The Internet as a vehicle to communicate health information during a public health emergency: a survey analysis involving the anthrax scare of 2001. J Med Internet Res 2004 Mar 03;6(1):e8. [doi: 10.2196/jmir.6.1.e8] [Medline: 15111274]
13. Kanda H, Takahashi K, Sugaya N, Mizushima S, Koyama K. Internet usage and knowledge of radiation health effects and preventive behaviours among workers in Fukushima after the Fukushima Daiichi nuclear power plant accident. Emerg Med J 2014 Oct;31(e1):e60-e65. [doi: 10.1136/emermed-2013-202955] [Medline: 24127108]

Abbreviations

COVID-19: coronavirus disease
MOOC: massive open online course
PAHO: Pan-American Health Organization
WHO: World Health Organization
VC: virtual campus

Edited by T Sanchez, G Eysenbach; submitted 03.04.20; peer-reviewed by S Lopez, D Toppenberg-Pejcic, C Basch; comments to author 10.04.20; revised version received 14.04.20; accepted 14.04.20; published 27.04.20.

Please cite as:
Utunen H, Ndiaye N, Piroux C, George R, Attias M, Gamhewage G
Global Reach of an Online COVID-19 Course in Multiple Languages on OpenWHO in the First Quarter of 2020: Analysis of Platform Use Data
J Med Internet Res 2020;22(4):e19076
URL: https://www.jmir.org/2020/4/e19076
doi:10.2196/19076
PMID:32293580

Review

Massive Open Online Course Evaluation Methods: Systematic Review

Abrar Alturkistani[1], BSc, MPH; Ching Lam[2], MEng; Kimberley Foley[1], BSc, MSc, PhD; Terese Stenfors[3], MSc, PhD; Elizabeth R Blum[3], BA, PhD; Michelle Helena Van Velthoven[2], BSc, MSc, PhD; Edward Meinert[1,2], MA, MSc, MBA, MPA, PhD

[1]Global Digital Health Unit, Imperial College London, London, United Kingdom

[2]Digitally Enabled PrevenTative Health Research Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom

[3]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Edward Meinert, MA, MSc, MBA, MPA, PhD
Digitally Enabled PrevenTative Health Research Group
Department of Paediatrics
University of Oxford
John Radcliffe Hospital
Children's Hospital
Oxford, OX3 9DU
United Kingdom
Phone: 44 7824446808
Email: e.meinert14@imperial.ac.uk

Abstract

Background: Massive open online courses (MOOCs) have the potential to make a broader educational impact because many learners undertake these courses. Despite their reach, there is a lack of knowledge about which methods are used for evaluating these courses.

Objective: The aim of this review was to identify current MOOC evaluation methods to inform future study designs.

Methods: We systematically searched the following databases for studies published from January 2008 to October 2018: (1) Scopus, (2) Education Resources Information Center, (3) IEEE (Institute of Electrical and Electronic Engineers) Xplore, (4) PubMed, (5) Web of Science, (6) British Education Index, and (7) Google Scholar search engine. Two reviewers independently screened the abstracts and titles of the studies. Published studies in the English language that evaluated MOOCs were included. The study design of the evaluations, the underlying motivation for the evaluation studies, data collection, and data analysis methods were quantitatively and qualitatively analyzed. The quality of the included studies was appraised using the Cochrane Collaboration Risk of Bias Tool for randomized controlled trials (RCTs) and the National Institutes of Health—National Heart, Lung, and Blood Institute quality assessment tool for cohort observational studies and for before-after (pre-post) studies with no control group.

Results: The initial search resulted in 3275 studies, and 33 eligible studies were included in this review. In total, 16 studies used a quantitative study design, 11 used a qualitative design, and 6 used a mixed methods study design. In all, 16 studies evaluated learner characteristics and behavior, and 20 studies evaluated learning outcomes and experiences. A total of 12 studies used 1 data source, 11 used 2 data sources, 7 used 3 data sources, 4 used 2 data sources, and 1 used 5 data sources. Overall, 3 studies used more than 3 data sources in their evaluation. In terms of the data analysis methods, quantitative methods were most prominent with descriptive and inferential statistics, which were the top 2 preferred methods. In all, 26 studies with a cross-sectional design had a low-quality assessment, whereas RCTs and quasi-experimental studies received a high-quality assessment.

Conclusions: The MOOC evaluation data collection and data analysis methods should be determined carefully on the basis of the aim of the evaluation. The MOOC evaluations are subject to bias, which could be reduced using pre-MOOC measures for comparison or by controlling for confounding variables. Future MOOC evaluations should consider using more diverse data sources and data analysis methods.

International Registered Report Identifier (IRRID): RR2-10.2196/12087

(J Med Internet Res 2020;22(4):e13851) doi:10.2196/13851

KEYWORDS

online learning; learning; computer-assisted instruction

Introduction

Massive open online courses (MOOCs) are free Web-based open courses available to anyone everywhere and have the potential to revolutionize education by increasing the accessibility and reach of education to large numbers of people [1]. However, questions remain regarding the quality of education provided through MOOCs [1]. One way to ensure the quality of MOOCs is through the evaluation of the course in a systematic way with the goal of *improvement over time* [2]. Although research about MOOCs has increased in recent years, there is limited research on the evaluation of MOOCs [3]. In addition, there is a need for effective evaluation methods for appraising the effectiveness and success of the courses.

Evaluation of courses to assess the success and effectiveness and to advise on course improvements is a long-studied approach in the field of education [4-6]. However, owing to the differences between teaching in MOOCs and traditional, face-to-face classrooms, it is not possible to adapt the same traditional evaluation methods [7,8]. For example, MOOCs generally have no restrictions on entrance, withdrawal, or the submission of assignments and assessments [7]. The methods used in Web-based education or e-learning are not always applicable to MOOCs because Web-based or e-learning courses are often provided as a part of university or higher education curricula, which are different from MOOCs per student expectations [8]. It is not suitable to directly compare MOOCs with higher education courses by using traditional evaluation standards and criteria [8].

Despite the limitations in MOOC evaluation methods, several reviews have been conducted on MOOC-related research methods, without specifically focusing on MOOC evaluations. Two recent systematic reviews were published synthesizing MOOC research methods and topics [9,10]. Zhu et al [9] and Bozkurt et al [11] recommended further research on the methodological approaches for MOOC evaluation. This research found little focus on the quality of the techniques and methodologies used [11]. In addition, a large number of studies on MOOCs examine general pedagogical aspects of the course without evaluating the course itself. Although the general evaluation of MOOC education and pedagogy is useful, it is essential that courses are also evaluated [12].

To address the gaps in MOOC evaluation methods in the literature, this systematic review aimed to identify and analyze current MOOC evaluation methods. The objective of this review was to inform future MOOC evaluation methodology.

Methods

This review explored the following research question: *What methods have been used to evaluate MOOCs?* [13]. This systematic review was conducted according to the Cochrane guidelines [14] and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (Multimedia Appendix 1) [15]. As the review only used publicly available information, an ethics review board approval was not required. The review was executed in accordance with the protocol published by Foley et al [13].

Eligibility Criteria

Eligible studies focused on the evaluation of MOOCs with reference to the course design, materials, or topics. The evaluation used the following population, intervention, comparator, outcome (PICO) framework for inclusion in the study:

- Population: learners in any geographic area who have participated in MOOCs [13].
- Intervention: MOOC evaluation methods. This is intended to be broad to include qualitative, quantitative, and mixed methods [13].
- Comparator: studies did not need to include a comparator for inclusion in this systematic review [13].
- Outcome: learner-focused outcomes such as attitudes, cognitive changes, learner satisfaction, etc, will be assessed [13].

Further to the abovementioned PICO framework, we used the following inclusion and exclusion criteria.

Inclusion Criteria

- Studies with a primary focus on MOOC evaluation and studies that have applied or reviewed MOOC evaluation methods (quantitative, qualitative, or mixed methods) [13].
- Studies published from 2008 to 2018 [13].
- All types of MOOCs, for example, extended MOOCs, connectivist MOOCs, language MOOCs, or hybrid MOOCs.

Exclusion Criteria

- Studies not in the English language [13].
- Studies that primarily focused on e-learning or blended learning instead of MOOCs [13].
- Studies that focused only on understanding MOOC learners such as their behaviors or motivation to join MOOCs, without referring to the MOOC.
- Studies that focused on machine learning or predictive models to predict learner behavior.

Search Strategy

We searched the following databases for potentially relevant literature from January 2008 to October 2018: (1) Scopus, (2) Education Resources Information Center, (3) IEEE (Institute of Electrical and Electronic Engineers) Xplore, (4) Medical Literature Analysis and Retrieval System Online/PubMed, (5) Web of Science, (6) British Education Index, and (7) Google Scholar search engine. The first search was performed in Scopus. The search words and terms for Scopus were as follows: (mooc* OR “massive open online course” OR coursera OR edx OR odl OR udacity OR futurelearn AND evaluat* OR measur* OR

compar* OR analys* OR report* OR assess* AND knowledge OR "applicable knowledge" OR retent* OR impact OR quality OR improv* OR environment OR effect "learning outcome" OR learning). The asterisks after the search terms allow all terms beginning with the same root word to be included in the search. The search terms were then adjusted for each database. The complete search strategy for each database can be found in the protocol by Foley et al [13] and in Multimedia Appendix 2. In addition, we scanned the reference lists of included studies.

Selection of Studies

Two reviewers (AA and CL) independently screened the titles and abstracts of the articles for eligibility. Selected studies were identified for full-text reading. Disagreements between the reviewers were resolved by discussions with a third reviewer (EM). Few studies (<10) were discussed with a third reviewer.

Data Extraction

The following information was extracted from each included study using a data abstraction form (Multimedia Appendix 2): (1) article title, country of the first author, and year of publication; (2) study aims; (3) evaluation: evaluation method, study design, evaluation type (evaluation of a single MOOC, multiple MOOCs, or review of a method), data collection methods, data analysis methods, and number of participants; and (4) outcome measures of the study: learner-focused outcomes and other outcomes. The studies were classified as quantitative, mixed methods, or qualitative based on the methods used. Studies were considered as mixed methods if they used a combination of qualitative or quantitative *techniques, methods, approaches, concepts, or language* in the same study [16].

Assessment of Methodological Quality

The Cochrane Collaboration Risk of Bias Tool for randomized controlled trials (RCTs) [17] and the National Institutes of Health—National Heart, Lung, and Blood Institute quality assessment tool for cohort observational studies and for before-after (pre-post) studies with no control group [18] were used to assess the methodological quality of the included studies depending on their study design.

Data Synthesis

We summarized the data graphically and descriptively. The evaluation results were reported according to the design thinking approach for evaluations that follows the subsequent order: (1) problem framing, (2) data collection, (3) analysis, and (4) interpretation [19].

Problem Framing

The evaluation-focused categories in the problem framing section were determined through discussions among the primary authors to summarize study aims and objectives. The 3 categories used in the evaluation-focused categories were defined as follows:

1. The learner-focused evaluation seeks to gain insight into the learner characteristics and behavior, including metrics such as completion and participation rates, satisfaction rates, their learning experiences, and outcomes.
2. Teaching-focused evaluation studies aim to analyze pedagogical practices so as to improve teaching.
3. MOOC-focused evaluation studies aim to better understand the efficacy of the learning platform to improve the overall impact of these courses.

Further to the evaluation-focused categories, the subcategories were generated by conducting a thematic analysis of the MOOC evaluation studies' aims and objectives. The themes resulted through an iterative process where study aims were coded and then consolidated into themes by the first author. The themes were then discussed with and reviewed by the second author until an agreement was reached.

Data Collection Analysis and Interpretation

The categories reported in the data collection sections were all representations of what the studies reported to be the data collection method. The categorization of the learner-focused parameters was done based on how the authors identified the outcomes. For example, if authors mention that the reported outcome was measuring *learners' attitudes to evaluate overall MOOC experience*, the parameter was recorded in the *learner experience* category. Similarly, if the authors mentioned that the reported outcome was evaluating what students gained from the course, the parameter was recorded as *longer term learner outcomes*.

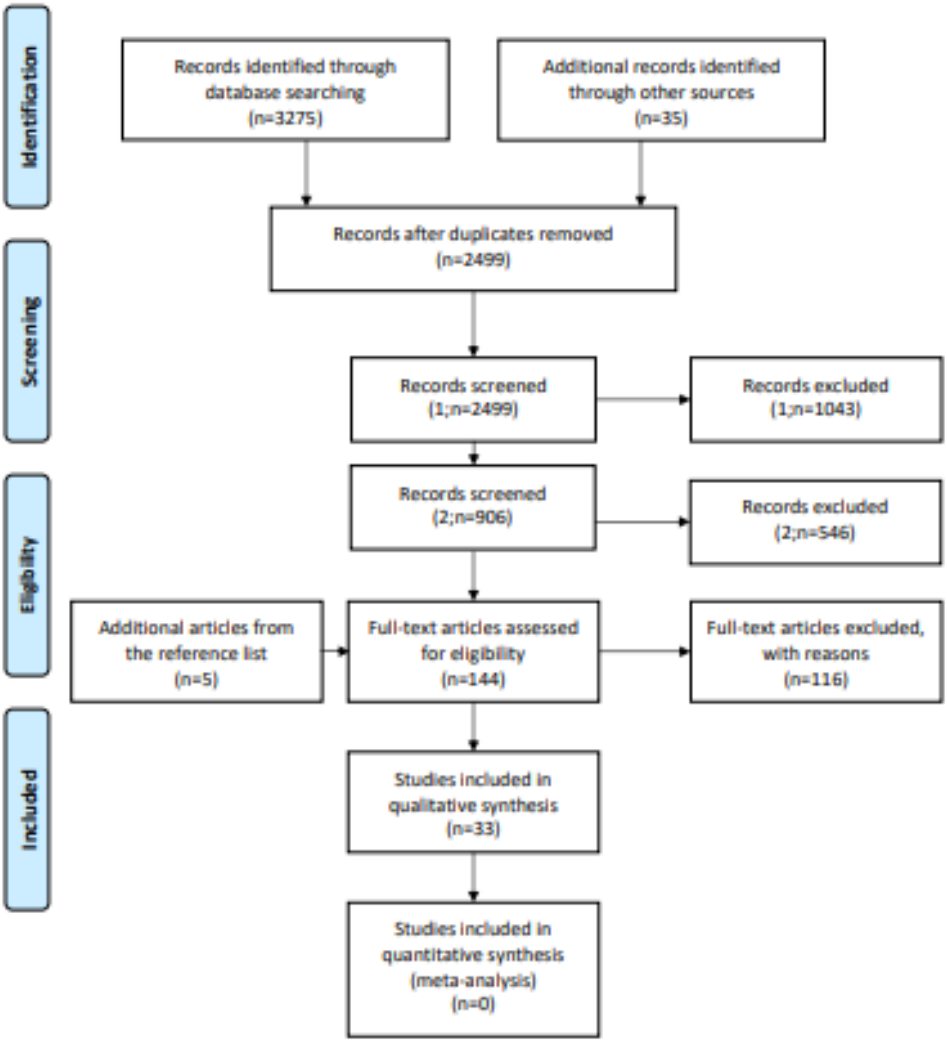
Results

In this section, we have described the search results and the methodological quality assessment results. We have then described the study findings using the following categories for MOOC evaluation: research design, aim, data collection methods, data analysis methods, and analysis and interpretation.

Search Results

There were 3275 records identified in the literature search and 2499 records remained after duplicates were removed. Records were screened twice before full-text reading. In the first screening (n=2499), all articles that did not focus on MOOCs specifically were removed (Figure 1). In the second screening (n=906), all articles that did not focus on MOOC learners or MOOC evaluation methods were removed (Figure 1). This was followed by full-text reading of 154 studies (Figure 1). An additional 5 studies were identified by searching the bibliographies of the included studies. In total, 33 publications were included in this review. There were 31 cross-sectional studies, 1 randomized trial, and 1 quasi-experimental study. The completed data abstraction forms of the included studies are in Multimedia Appendix 3.

Figure 1. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of the literature search.



Methodological Quality

The RCT included in this study [20] received a low risk-of-bias classification (Multimedia Appendix 4).

Of the 31 cross-sectional studies, 26 received poor ratings because of a high risk of bias (Multimedia Appendix 5). The remaining 5 studies received a fair rating because of a higher consideration for possible bias. In total, 2 studies that were able to measure exposure before outcomes such as studies that performed pretests and posttests [21,22], 3 studies that accounted for confounding variables [21-23], 2 studies that used validated exposure [24,25], and 2 studies that used outcome measures [23,25] received a better quality rating.

A quality assessment of the quasi-experimental study using longitudinal pretests and posttests [26] is included in Multimedia Appendix 6.

Massive Open Online Course Evaluation Research Design

In total, 16 studies used a quantitative study design, 11 studies used a qualitative study design, and 6 studies used a mixed methods study design. There was 1 RCT [20] and 1 quasi-experimental study [26]. In total, 4 studies evaluated more than 1 MOOC [27-30]. In all, 2 studies evaluated 2 runs of the same MOOC [31,32], and 1 study evaluated 3 parts of the same MOOC, run twice for consecutive years [33].

In total, 6 studies used a comparator in their methods. A study compared precourse and postcourse surveys by performing a chi-square test of changes in *confidence, attitudes, and knowledge* [34]. A study compared the average assignment and final essay scores of MOOC learners with face-to-face learners and calculated 2 independent sample *t* tests to compare the differences between learners but did not include any pre- and posttest or survey results [35]. In all, 4 studies conducted pretest and posttest analyses [20,26]. Hossain et al [20] used an RCT design and calculated the mean between-group differences of knowledge, confidence, and satisfaction comparing MOOC learners with other Web-based learners. Colvin et al [21] calculated normalized gain using item response between pretest and posttest scores and the Item Response Theory for weekly performance compared with that of on-campus learners. Rubio et al [26] compared the pretest mean and posttest mean of comprehensibility scores in a MOOC, comparing results with those of face-to-face learners [26]. Konstan et al [22] calculated knowledge test gains by performing a paired *t* test of average knowledge gains, comparing these gains with those of

face-to-face learners and (comparing 2 learner groups) the average normalized learning gains among all learners [22].

Aim of Massive Open Online Course Evaluations

The aim or objective of MOOC evaluations included in this review can be categorized into learner-focused, teaching-focused, and MOOC-focused evaluation aims (Table 1). In all, 16 studies evaluated learner characteristics and behavior and 20 studies evaluated learning outcomes and experiences. One of the least studied aspects of MOOC evaluation is pedagogical practices, which were only evaluated by 2 studies [36,37].

Table 1. The aim of the massive open online course evaluations for the included studies.

Evaluation aim focus, subcategories	Studies	Number of studies
Learner		
Learner expectations	[23,38,39]	3
Learner characteristics and behavior	[22,23,27,28,30-32,38,39,40-46]	16
Learner engagement	[27,34,43,45]	4
Participation or completion rates	[24,30,34,42,44,45]	6
Learner satisfaction	[20,47]	2
Peer interaction	[25,43,46,48]	4
Learning outcomes and experience	[21-26,31,33,35,37,39-41,43,44,46-50]	20
Knowledge retention	[22]	1
Teaching		
Pedagogical practices	[36,37]	2
MOOC[a]		
Comparison with other learning platforms	[20,22,26,38]	4
MOOC content and structure	[48]	1
Implementation of MOOC	[29,31,33,49,52]	5
Sustainability of MOOC	[49]	1

[a]MOOC: massive open online course.

Massive Open Online Course Evaluation Data Collection Methods

In all, 12 studies used 1 data source [20,24,25,27,28,31,32,38,43,47], 11 studies used 2 data sources [21,26,29,30,33,34,36,40,42,45,51], 7 studies used 3 data sources [22,35,39,41,44,46,48], 2 studies used 4 data sources [23,37], and 1 study used 5 data sources [52]. The most used data sources were surveys followed by learning management system (LMS), quizzes, and interviews (Table 2). "Other" data sources that are referred to in Table 2 include data collected from social media posts [37], registration forms [30,44], online focus groups [37], and homework performance data [21]. These data sources were used to collect data on different aspects of the evaluation.

Table 2. Studies using different data sources (N=33).

Data source	Value, n (%)
Surveys	20 (30.8)
Interviews	8 (12.3)
Learning Management System	18 (27.7)
Discussions	5 (7.7)
Quizzes	9 (13.8)
Other	5 (7.7)

In total, 8 studies collected data through interviews and had a population size ranging from 2 to 44 [23,37,39,42,43,49,51,52]. In total, 20 studies that collected data through surveys had a population size ranging from 25 to 10,392 [22-41,44-46,51,52]. In all, 18 studies that collected data through the LMS [22,24,26,29-31,33-35,39-42,44-46,48,52] had a population size made of participants or data points (eg, discussion posts) ranging from 59 to 209,871. Nine studies used quiz data [20-22,26,33,35,41,47,52]. Studies that used quiz data had a population size of 48 [20], 53 [47], 136 [41], 1080 [21], and 5255 [22]. Other data sources used did not have a clearly reported sample size for a particular source.

Table 3 shows the various data collection methods and their uses. Pre-MOOC surveys or pretests could be used for baseline data such as learner expectations [22,36,50] or learner baseline test scores [20-22,26,33], which, then, allows tests scores to be compared with post-MOOC survey and quiz data [20-22,27,33].

Table 3 explains *how* studies collected data to meet the aims of their evaluation. In general, surveys were used to collect demographic data, learner experience, and learner perceptions and reactions, whereas LMS data were used for tracking learner completion of the MOOCs.

Table 3. Data collection methods and their uses in massive open online course evaluations.

Data	Uses
Registration form	To collect demographic information [26,30]
Pre-MOOC[a] survey	To collect data on the following: demographic information [23,29,36,40,46,50,52]; learners' background [22,29,36,46] and expectations; perceptions [22,36,50]; learners' experience [40]; learners' past MOOC experience [29]; learners' self-efficacy [52], motivation [52], and goals [44,50]; assess learners' knowledge [40] and course efficacy [50]
Pretest	To collect baseline test scores for comparison with posttest scores [20-22,26,33]
Learning management system data	To collect data on the following: demographic information [24]; attendance rates [24,35,42]; completion of the different components of the MOOC [24,35,36,39,42]; quiz or assignment scores [26,35,45]; learner activity [45]
Discussion posts	Feedback about the course [46] and learner interactions [25]
Quiz, homework, or test (not specified as pre- or postquiz or test)	Grades to assess learning [21,35,41] and a weekly quiz to record learners' reaction to the tools called *digital readiness tools* of the course [47]
Post-MOOC survey	To collect demographic information [23,39-41,50,51]; to record the learning experience [22,30,35,39,41]; to record course influence [48]; to guide MOOC design [48], course feedback [30,33,39,41,46,50], perceptions [38,50], excitement [38], learners' motivation [23,32,39], learners' satisfaction [35,40,41], enjoyment of the course [40] and "Patterns and levels of participation" in the course [37], learning strategies [32]; to assess learner knowledge [40], course usefulness [39,45], course degree of perseverance [45], reasons for dropping out of the MOOC [30]; to recruit participants for research [23]; to collect course feedback [44]
Posttest	To assess learning [20,21]; to assess confidence in applying learning [20]; to assess satisfaction [20]; to calculate the difference in scores compared with pretest [22,26,33]; to assess knowledge retention 5 months post-MOOC [22]
End of MOOC quiz	To record learners' feedback in relation to the course material (whether the course helped them become *flexible learners*) [47]
Postcourse interview	Course participation and evaluation [37,42]; course effectiveness [43,49]; sustainability of the course [49]; reason for taking the course [23]; learners' motivation [23,42]; to understand learning behavior [51]; *postcourse practices* or learners' behavior [52]
Email interview	To understand learners' behavior and learning in MOOCs [37]; specify MOOC positives [39]; motivation in MOOC; challenges in MOOC [39]
Online focus group	Assessment of the course: organization, assessment, *use of technology* and *inclusive practice* [37]

[a]MOOC: massive open online course.

Massive Open Online Course Evaluation Analysis and Interpretation

In terms of the data analysis methods, quantitative methods were the only type of method used in 16 studies with descriptive and inferential statistics, the top 2 preferred methods. Qualitative analysis methods such as thematic analyses, which can include grounded theory [49], focused coding [38,39], and content analysis [25,50], were mainly used in qualitative studies.

A summary of the parameters, indicators, and data analysis used for the MOOC evaluation can be found in Table 4. Most notably, inferential statistics were used to analyze learning outcomes (Table 4) such as the comparison of means or the use of regression methods to analyze quiz or test grades. These outcomes were also used as a measure to evaluate the overall effectiveness of a MOOC by the studies. Table 4 shows how the data collection method uses mentioned in Table 3 were measured and analyzed. In general, studies focused on measuring learner engagement and learners' behavior–related indicators. Studies referred to learning in different ways such as *learning*, learning performance, learning outcome, or gain in comprehensibility depending on the learning material of the course. Other studies considered learning outcomes such as knowledge retention or *what students took away from the course*. There was a consensus that learner engagement can be measured by measuring the various learner activities in the course, whereas learner behavior was a more general term used by studies to describe the different MOOC evaluation measures. For teaching-focused evaluation, both Mackness et al [37] and Singh et al [36] used learner parameters to reflect and analyze pedagogical practices.

Table 4. Data collection method uses mentioned earlier and how they were analyzed in massive open online course evaluations.

Data collection method uses, parameters or themes reported	Data analysis methods
To measure learning outcomes	
• Learning [20,33] • Learning performance [52] • "Learning outcome" [41] • Learning [21] • Overall learner ability in the course [21] • The students' gains in comprehensibility [26] • Subject-matter knowledge [22] • Comprehensibility of learner audio recordings in a language MOOC[a] [26] • Knowledge retention [22] • Learning performance [35]	• Calculation of the mean difference between pretest and posttest scores [20,33] • Compare pretest and posttest scores using a paired *t* test [33] • Descriptive statistics [52] • "Regressing quiz and homework score on participation and MOOC experience" [41] • Calculation of normalized gain between pretest and posttest • Using the Item Response Theory analyzing pretest, posttest, and homework performance [21]A matched-pairs *t* test to measure the "gains in comprehensibility between the pretest and posttest" and an unpaired *t* test to compare the pretest and posttest means [26] • Knowledge test gains by calculating normalized learning gains when comparing pretest and posttest scores [22] • Calculation of gains in comprehensibility [26] • "A paired-samples *t* test to examine student knowledge retention as measured by the postcourse and follow-up tests" [22] • Two independent sample *t* tests to compare the quiz and assignment scores [35]
To measure learner participation or engagement	
• Contributions per week, number of tweets, "quality of posted comments and learning designs," "quality of peer feedback," "ranking of importance of course features," "comments received by those posting and sharing a scenario idea in Week 2" [35,44] • Determinants of completion [22] • Course completion rate [42] • The number of videos watched, video activity (play, stops, and full watch), the number of quizzes submitted, and discussion forum activity; reading in forums, the number of posts and comments, and dropout rate [31] • Reasons for dropping out of the course [30], total number of reads in forums, the number of forum and post comments [31] • The number of comments per participant, completed steps, and the "likes" count [34] • The frequency of viewing lectures [40] and frequency of attempting quizzes [40] • Learner course activity and course grade [45] and frequency of interaction on online forums • Satisfaction with MOOC, comfort with learning new things, and joining MOOC because of the "Love for Learning" [27]	• Descriptive statistics [30,31,34,35,42,44] and frequency analysis [45] • Logistic regression of homework and exam outcomes [22] • Regression [40] • Frequency analysis [45] and structural equation modeling [27]
To measure learner experience	
• Comparison of a Likert scale rating of "the technology quality and user-friendliness of the Web environment, the quality of instructional content, and the instructional arrangement," satisfaction with interactions with instructors, satisfaction with support received, and the satisfaction of learning needs between MOOC and onsite learners [35] • "Perceived usefulness and ease of use" of MOOC [41] • "Perceived learning experience" [41] • Learner rating of the "usefulness and relevance of the activities" [45] • Overall learner attitude [23]	• Comparison of "Likert scale items" using the Mann-Whitney U tests [35], regression analysis [41], factor analysis of factors related to *poststudy feedback* [41], frequency analysis [45], sentiment analysis of interview data [23], and descriptive statistics [30]
To measure learner expectation	
• Student expectations (theme) [50] • Whether course fulfilled expectations [30]	• Descriptive content analysis [50] • Descriptive statistics [30]
To measure learner behavior	

Data collection method uses, parameters or themes reported	Data analysis methods
• "Autonomous learning across distributed platforms, learning through diversity, learning through openness and interactivity, organizing learning through aggregation, co-creation, and creativity through remixing and repurposing, coping with uncertainty, and identity building" (themes) [37] • How learners approach "professional learning" in a MOOC, what learner behavior is exhibited by learners, and how "professionals relate their MOOC learning to their professional role" [51] • Factors predicting learner and student success [22] • Learner self-reported "assertions on learning strategies" [32]	• Qualitative descriptive [37] • Coding of interview data [51] • Ordinary least squares regression using learner demographic data and knowledge data [22] • Descriptive statistics [32]
To measure learner retention	
• Learner course activity and course grade [45] • Learner rating of course perseverance [45]	• 1-2 frequency analysis [45]
To measure long-term learner outcomes	
• Learner opinions about course effectiveness [49] • "What students took away from the MOOC" (theme) [50]	• Using grounded-theory methods of interview data [49] • Descriptive content analysis [50]
To measure social interactions	
• Learner "interaction in forums" [42] • Learner to learner interactions [25] • Learner collaboration patterns [25]	• Social network analysis [42] • Content analysis of discussions posts using the Interaction Analysis Model [25] • Social network analysis [25]
To measure learner motivation	
• "Learning motivation" [42] • Learner self-reported "assertions on motivation" [32] • "A reason for taking or completing the course" [23] • Exploring the "primary motivation for taking" the course [28]	• Descriptive qualitative [42] • Descriptive statistics [32] • Thematic analysis of interview data [23] • Emergent coding on survey data [28]

[a]MOOC: massive open online course.

Discussion

This study aimed to review current MOOC evaluation methods to understand the methods that have been used in published MOOC studies and subsequently to inform future designs of MOOC evaluation methods. Owing to the diversity of MOOC topics and learners, it is not possible to propose a single evaluation method for all MOOCs. Researchers aiming to evaluate a MOOC should choose a method based on the aims of their evaluation or the parameters they would like to measure. In general, data collection methods were similar in most evaluations, such as the use of interviews or survey data, and the analysis methods were highly heterogeneous among studies.

Massive Open Online Course Evaluation Research Design

The cross-sectional study design was used in 31 of 33 of the included studies. The cross-sectional study design was used when the aim was to investigate the factors affecting outcomes for a population at a given time point [53]. For the MOOC evaluation, this is particularly useful for observing the population of learners and for understanding the factors affecting the success and impact of a MOOC. They are relatively inexpensive to conduct and can assess many outcomes and factors at the same time. However, cross-sectional study designs are subject to *nonresponse bias*, which means that studies are only representative of those who participated, who incidentally may happen to be different from the rest of the population [53].

One of the most effective methods of evaluation used in MOOCs was the use of baseline data to compare outcomes. Studies that did pretests and posttests had a less likelihood of bias in their outcomes owing to the measurement of exposure before the measurement of outcome [18]. Even when studies used pre- and postcourse surveys or tests, they were not longitudinal in design, as such a design requires a follow-up of the same individuals and requires observing them *at multiple time points* [53]. Therefore, the use of pre- and postsurveys or tests without linking the individuals may simply represent a difference in the groups studied rather than changes in learning or learner outcomes. The advantages of this method are that it can reduce bias, and quasi-experimental studies are known as strong methods. However, the disadvantage is that although this method may work with assessing learning, such as memorizing information, it may not work to assess skill development or the application of skills.

Aim of Massive Open Online Course Evaluations

Understanding the aim behind the evaluation of MOOCs is critically important in designing MOOC evaluation methods as it influences the performance indicators and parameters to be

evaluated. More importantly, motivation for the evaluation determines the data methods that will be used. One reason for the inability to conclude a standardized evaluation method from this review is that studies differ in the aspects and purposes of why they are conducting the evaluation. For example, not all studies perform evaluations of MOOCs to evaluate overall effectiveness, which is an important aspect to consider if MOOCs are to be adopted more formally in higher education [54]. The variability in the motivation of MOOC evaluations may also explain the high variability in the outcomes measured and reported.

Data Collection Methodology

In all, 12 studies used 1 data source and 11 studies used 2 data sources (Table 3), which is not different from previous findings [10]. The results of this study also show that there is high flexibility in data collection methods for MOOC evaluations from survey data to LMS data to more distinct methods such as online focus groups [37]. The number of participants in the studies was exceedingly varied. This is due to the difference in the data collection methods used. For example, studies with data captured through the LMS, which is capable of capturing data from all of the learners who joined the course, had the highest number of learners. On the contrary, studies that used more time-consuming methods, such as surveys or interviews, generally had a lower number of participants. It is important to note that the MOOC evaluation is not necessarily improved by increasing the number of data sources but rather by conducting a meaningful analysis of the available data. Some studies preferred multiple methods of evaluation and assessment of learning. One paper argued that this allows to evaluate learning of the diverse MOOC population in a more effective way [22]. Studies should use the best data collection methods to answer their research aims and questions.

Analysis and Interpretation

In total, 16 of 33 studies used only quantitative methods for analysis (Table 4), which is in line with the general MOOC research, which has been predominated by quantitative methods [10,55]. Studies used statistical methods such as descriptive and inferential statistics for data analysis and interpretation of results. The availability of data from sources such as the LMS may have encouraged the use of descriptive statistical methods [10]. However, 17 of the 33 included studies used some form of qualitative data analysis methods either by using a qualitative study design or by using a mixed methods study design (Table 4). This may be explained by the recent (2016-2017) rise in the use of qualitative methods in MOOC research [10].

Although inferential statistics can help create better outcomes from studies, this is not always possible. For example, one study [36] mentioned a high variation between pre- and postcourse survey participant numbers and another [29] mentioned a small sample size as reasons for not using inferential statistical methods. It should be noted that using data from multiple sources and having a large sample size does not guarantee the quality of the evaluation methods.

In MOOC research, qualitative data can be useful to understand the meaning of different behaviors as quantitative data, oftentimes, cannot answer why things happened [56].

Thematic and sentiment data analysis methods seek to represent qualitative data in a systematic way. The thematic analysis seeks to organize information into themes to find patterns [57]. This is especially useful for generalizing data for a subsequent analysis. For instance, Singh et al [36], Draffan et al [34], and Shapiro et al [23] all used a thematic analysis to simplify heterogeneous responses from interviewees and participants to understand what students enjoy about the MOOCs. Focused coding and grounded theory use similar approaches to grouping qualitative data into themes based on conceptual similarity and to developing analytic narratives. Liu et al [38] used focused coding to group data from course surveys into positive and negative aspects of MOOCs for future MOOC improvement [7]. Sentiment analysis and social network analysis are both qualitative analysis strategies with a greater focus on opinion-rich data [58]. These are important strategies used in understanding the opinions of learners and converting subjective feelings of learners into data that can be analyzed and interpreted.

Outcome Measures

The outcome measures reported greatly varied among studies, which is expected, as identifying the right outcome measures is an inherent challenge in educational research, including more traditional classroom-based studies [7].

The choice of evaluation methods is highly dependent on the aim of the evaluation and the size of the MOOCs. For quantitative measures, such as completion and participation rates, metrics can be easily collected through the MOOC platform. However, these metrics alone may be insufficient to provide insights into why students fail to complete the course for future improvement. Although it may be difficult to represent the problem holistically using qualitative methods, it can be useful in providing insights from individuals who participated in the MOOCs. Mixed methods studies combine the 2 modalities to better understand metrics generated and produce greater insights for future improvement of the MOOCs.

Learning outcomes were mostly analyzed by inferential statistical methods owing to the use of pretest and posttest methods and the calculation of gains in learning. This method may be most suited for MOOCs that require knowledge retention. Learning parameters also involved a lot of comparisons, either a comparison with pre-MOOC measures or a comparison with other learners or both. Social interactions were studied in 2 of the MOOC evaluations using social network analysis methods. Although the MOOC completion rate has been often cited as a parameter for MOOC success, it can be noticed that studies started to move away from only using completion rates. For example, studies looked at completion of different steps of the MOOCs or looked at overall completion. The learning outcomes reported in this review should be used with caution as not all of them have been validated or assessed for their reliability except for a few.

Methodological Quality

In total, 26 studies with a cross-sectional design had a low-quality assessment, whereas RCTs and quasi-experimental studies received a high-quality assessment. Having a high level of bias affects the generalizability of studies, which is a common problem in most research using data from MOOCs [30,59]. The availability of high risks of bias in current MOOC evaluations requires a closer look at what were the sources of bias and what methods can be used to reduce them. The use of not validated, self-reported data sources and the lack of longitudinal data also increases the risk of bias in these studies [56]. However, although most MOOCs struggle with learner retention and MOOC completion rates [54], it is understandable that studies are not able to collect longitudinal data.

Future Directions

The scarcity of studies focusing on the evaluation of the effectiveness of particular MOOCs relative to the number of available studies on MOOCs raises some questions. For example, many studies that were excluded from this review studied MOOC learners or aspects of the MOOCs without conducting an evaluation of course success or effectiveness. As shown in this review, there is a diverse range of evaluation methods, and the quality of these evaluation studies can be as diverse. The motivation of the evaluation exercise should be the basis of the evaluation study design to design effective quantitative or qualitative data collection strategies. The development of general guidance, standardized performance indicators, and an evaluation framework using a design thinking approach can allow these MOOC evaluation exercises to yield data of better quality and precision and allow improved evaluation outcomes. To provide a comprehensive evaluation of MOOCs, studies should try to use a framework to be able to systematically review all of the aspects of the course.

In general, the adoption of a mixed methods analysis considering both quantitative and qualitative data can be more useful for evaluating the overall quality of MOOCs. Although it is useful to have quantitative data such as learner participation and dropout rates, qualitative data gathered through interviews and opinion mining provide valuable insights into the reasons behind the success or failure of a MOOC. Studies of MOOC evaluations should aim to use data collection and analysis methods that can minimize the risk of bias and provide objective results. Whenever possible, studies should use comparison methods, such as the use of pretest or posttest or a comparison with other types of learners, as a control measure. In addition, learner persistence is an important indicator for MOOC evaluation that needs to be addressed in future research.

Strengths and Limitations

To our knowledge, this is the first study to systematically review the evaluation methods of MOOCs. The findings of this review can serve future MOOC evaluators with recommendations on their evaluation methods to facilitate better study designs and maximize the impact of these Web-based platforms. However, as a lot of MOOCs are not necessarily provided by universities and systematically evaluated and published, the scope of this review can only reflect a small part of MOOC evaluation studies.

Conclusions

There is no one way of completing a MOOC evaluation, but there are considerations that should be taken into account in every evaluation. First, because MOOCs are very large, there is a tendency to use quantitative methods using aggregate-level data. However, aggregate-level data do not always tell why things are happening. Qualitative data could further help interpret the results by exploring why things are happening. Evaluations lacked longitudinal data and very few accounted for confounding variables owing to data collection challenges associated with MOOCs such as not having longitudinal data or not having enough data sources. Future studies could help identify how these challenges could be overcome or minimized.

LMS may not report useful findings on an individual level, but they should still be considered and used in MOOC evaluations. Big data in the form of learning analytics can help with decision making, predicting learner behavior, and providing a more *comprehensive* picture of the phenomena studied [60]. Studies should still consider using LMS as it can provide a valuable addition to the research, but researchers need to be careful about the depth of the findings that can be concluded from LMS-only datasets.

The use of qualitative data could help enhance the findings from the studies by explaining the phenomena. Both quantitative and qualitative methods could play a key role in MOOC evaluations.

Current MOOC evaluations are subject to many sources of bias owing to the nature of the courses being open and available to a very large and diverse number of participants. However, methods are available to reduce the sources of bias. Studies could use a comparator, such as pretest scores, or other types of learners to be able to calculate relative changes in learning. In addition, studies could control for confounding variables to reduce bias.

This review has provided an in-depth view of how MOOCs can be evaluated and explored the methodological approaches used. Exploring MOOC methodological approaches has been stated as an area for future research [10]. The review also provided recommendations for future MOOC evaluations and for future research in this area to help improve the quality and reliability of the studies. MOOC evaluations could contribute to the development and improvement of these courses.

Acknowledgments

The authors would like to thank the medical librarian, Rebecca Jones, for her guidance in the search methods and for reviewing the search strategy used in this protocol. This work was funded by the European Institute of Technology and Innovation Health (Grant No. 18654).

Authors' Contributions

AA and CL completed the screening of articles and data analysis. AA and CL completed the first draft of the manuscript. All authors reviewed and edited the manuscript for content and clarity. EM was the guarantor.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2009 checklist.
[DOCX File , 19 KB - jmir_v22i4e13851_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 14 KB - jmir_v22i4e13851_app2.docx]

Multimedia Appendix 3
Data abstraction form.
[DOCX File , 24 KB - jmir_v22i4e13851_app3.docx]

Multimedia Appendix 4
Quality assessment results of the Randomized Controlled Trial [20] using the Cochrane Collaboration Risk of Bias Tool.
[DOCX File , 14 KB - jmir_v22i4e13851_app4.docx]

Multimedia Appendix 5
Quality assessment results of cross-sectional studies using the NIH - National Heart, Lung and Blood Institute quality assessment tool.
[DOCX File , 43 KB - jmir_v22i4e13851_app5.docx]

Multimedia Appendix 6
Quality assessment results for the quasi experimental study using the Cochrane Collaboration Risk of Bias Tool for Before-After (Pre-Post) Studies With No Control Group.
[DOCX File , 15 KB - jmir_v22i4e13851_app6.docx]

References

1. Rolfe V. A systematic review of the socio-ethical aspects of massive online open courses. Eur J Open Distance E-Learn 2015;18(1):52-71 [FREE Full text] [doi: 10.1515/eurodl-2015-0004]
2. Jansen D, Rosewell J, Kear K. Quality frameworks for MOOCs. In: Open Education: from OERs to MOOCs. Berlin, Heidelberg: Springer; 2017:261-281.
3. Tahiri JS, Bennani S, Idrissi MK. Using an Analytical Formalism to Diagnostic and Evaluate Massive Open Online Courses. In: Proceedings of the 2015 10th International Conference on Intelligent Systems: Theories and Applications. 2015 Presented at: SITA'15; October 20-21, 2015; Rabat, Morocco. [doi: 10.1109/sita.2015.7358389]
4. Goos M, Salomons A. Measuring teaching quality in higher education: assessing selection bias in course evaluations. Res High Educ 2017;58(4):341-364. [doi: 10.1007/s11162-016-9429-8]
5. Gravestock P, Gregor-Greenleaf E. Higher Education Quality Council of Ontario. Toronto: Higher Education Quality Council of Ontario; 2008. Student Course Evaluations: Research, Models and Trends URL: http://www.heqco.ca/SiteCollectionDocuments/Student%20Course%20Evaluations_Research,%20Models%20and%20Trends.pdf [accessed 2020-02-07]
6. Werdell PR. Education Resources Information Center. 1967. Course and Teacher Evaluation URL: https://eric.ed.gov/?id=ED050693 [accessed 2020-02-07]
7. Breslow L, Pritchard DE, DeBoer J, Stump GS, Ho AD, Seaton DT. Studying Learning in the Worldwide Classroom Research into edX's First MOOC. Res Pract Assess 2013;8:1-25 [FREE Full text]

8. Literat I. Implications of massive open online courses for higher education: mitigating or reifying educational inequities? High Educ Res Dev 2015;34(6):1-14. [doi: 10.1080/07294360.2015.1024624]
9. Zhu M, Sari A, Lee MM. A systematic review of research methods and topics of the empirical MOOC literature (2014–2016). Intern High Educ 2018 Apr;37:31-39. [doi: 10.1016/j.iheduc.2018.01.002]
10. Zhu M, Sari A, Bonk C. A Systematic Review of MOOC Research Methods and Topics: Comparing 2014-2016 and 2016-2017. : Association for the Advancement of Computing in Education (AACE); 2018 Presented at: World Conference on Educational Media & Technology; June 25, 2018; Amsterdam, Netherlands URL: https://www.learntechlib.org/primary/p/184395/
11. Bozkurt A, Akgün-Özbek E, Zawacki-Richter O. Trends and patterns in massive open online courses: review and content analysis of research on MOOCs (2008-2015). Int Rev Res Open Distrib Learn 2017 Aug 15;18(5):118-147. [doi: 10.19173/irrodl.v18i5.3080]
12. Bali M. MOOC Pedagogy: gleaning good practice from existing MOOCs. J Online Learn Teach 2014;10(1):44-56 [FREE Full text]
13. Foley K, Alturkistani A, Carter A, Stenfors T, Blum E, Car J, et al. Massive Open Online Courses (MOOC) evaluation methods: protocol for a systematic review. JMIR Res Protoc 2019 Mar 7;8(3):e12087 [FREE Full text] [doi: 10.2196/12087] [Medline: 30843868]
14. Cochrane Training. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 URL: http://training.cochrane.org/handbook [accessed 2020-02-04]
15. Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. J Craniomaxillofac Surg 2011 Mar;39(2):91-92. [doi: 10.1016/j.jcms.2010.11.001] [Medline: 21145753]
16. Johnson RB, Onwuegbuzie AJ. Mixed methods research: a research paradigm whose time has come. Educ Res 2004;33(7):14-26. [doi: 10.3102/0013189X033007014]
17. 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]
18. National Heart, Lung, and Blood Institute (NHLBI). Study Quality Assessment Tools URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2018-11-21]
19. Patton MQ. Utilization-Focused Evaluation. Dordrecht: Springer; 2000.
20. Hossain MS, Islam MD, Glinsky JV, Lowe R, Lowe T, Harvey LA. A massive open online course (MOOC) can be used to teach physiotherapy students about spinal cord injuries: a randomised trial. J Physiother 2015 Jan;61(1):21-27 [FREE Full text] [doi: 10.1016/j.jphys.2014.09.008] [Medline: 25498151]
21. Colvin KF, Champaign J, Liu A, Zhou Q, Fredericks C, Pritchard DE. Learning in an introductory physics MOOC: All cohorts learn equally, including an on-campus class. Int Rev Res Open Distrib Learn 2014;15(4):263-283. [doi: 10.19173/irrodl.v15i4.1902]
22. Konstan JA, Walker JD, Brooks DC, Brown K, Ekstrand MD. Teaching recommender systems at large scale. ACM Trans Comput-Hum Interact 2015;22(2):1-23. [doi: 10.1145/2728171]
23. Shapiro HB, Lee CH, Roth NE, Li K, Çetinkaya-Rundel M, Canelas DA. Understanding the massive open online course (MOOC) student experience: an examination of attitudes, motivations, and barriers. Comput Educ 2017 Jul;110:35-50. [doi: 10.1016/j.compedu.2017.03.003]
24. de la Garza LY, Sancho-Vinuesa T, Zermeño MG. Atypical: Analysis of a Massive Open Online Course (MOOC) with a Relatively High Rate of Program Completers. Glob Educ Rev 2015;2(3):68-81.
25. Tawfik AA, Reeves TD, Stich AE, Gill A, Hong C, McDade J, et al. The nature and level of learner–learner interaction in a chemistry massive open online course (MOOC). J Comput High Educ 2017;29(3):411-431. [doi: 10.1007/s12528-017-9135-3]
26. Rubio F. Teaching pronunciation and comprehensibility in a language MOOC. In: Martin-Monje E, Barcena E, editors. Language MOOCs: Providing Learning, Transcending Boundaries. Berlin, Germany: De Gruyter; 2015:143-159.
27. Kaveri A, Gunasekar S, Gupta D, Pratap M. Decoding Engagement in MOOCs: An Indian Learner Perspective. In: Proceedings of the 2016 IEEE Eighth International Conference on Technology for Education. 2016 Presented at: T4E'16; December 2-4, 2016; Mumbai, India. [doi: 10.1109/t4e.2016.027]
28. Milligan C, Littlejohn A. Why study on a MOOC? The motives of students and professionals. Int Rev Res Open Distrib Learn 2017;18(2):92-102. [doi: 10.19173/irrodl.v18i2.3033]
29. Lesjak B, Florjancic V. IDEAS/RePEc. 2014. Evaluation of MOOC: Hands-On Project or Creative Use of ICT in Teaching URL: https://ideas.repec.org/h/tkp/mklp14/1147-1155.html [accessed 2018-10-22]
30. Morales M, Rizzardini RH, Gütl C. Telescope, a MOOCs Initiative in Latin America: Infrastructure, Best Practices, Completion and Dropout Analysis. In: Proceedings of the 2014 IEEE Frontiers in Education Conference. 2014 Presented at: FIE'14; October 22-25, 2014; Madrid, Spain. [doi: 10.1109/fie.2014.7044103]
31. Khalil M, Ebner M. Can learning analytics find success in didactical measurements? Results from a MOOC case study. In: Ifenthaler D, editor. Digital Workplace Learning. Cham: Springer; 2018:211-225.

32. Alario-Hoyos C, Estévez-Ayres I, Pérez-Sanagustín M, Kloos CD, Fernández-Panadero C. Understanding learners' motivation and learning strategies in MOOCs. Int Rev Res Open Distrib Learn 2017;18(3):119-137. [doi: 10.19173/irrodl.v18i3.2996]
33. Jacquet GA, Umoren RA, Hayward AS, Myers JG, Modi P, Dunlop SJ, et al. The Practitioner's Guide to Global Health: an interactive, online, open-access curriculum preparing medical learners for global health experiences. Med Educ Online 2018 Dec;23(1):1503914 [FREE Full text] [doi: 10.1080/10872981.2018.1503914] [Medline: 30081760]
34. Draffan EA, Leon M, James A, Aljaloud S, Wald M. Completion, comments and repurposing a digital accessibility MOOC. In: Miesenberger K, Kouroupetroglou G, editors. Computers Helping People with Special Needs. New York City: Springer International Publishing; 2018:138-145.
35. Chen W, Jia J, Miao J, Wu X, Wang A, Yang B. Assessing Students' Learning Experience and Achievements in a Medium-Sized Massively Open Online Course. In: Proceedings of the 2015 IEEE 15th International Conference on Advanced Learning Technologies. 2015 Presented at: ICALT'15; July 6-9, 2015; Hualien, Taiwan. [doi: 10.1109/icalt.2015.69]
36. Singh AB, Mørch AI. An analysis of participants' experiences from the first international MOOC offered at the University of Oslo. Nord J Digit Lit 2018;13(1):40-64 [FREE Full text] [doi: 10.18261/issn.1891-943x-2018-01-04]
37. Mackness J, Waite M, Roberts G, Lovegrove E. Learning in a small, task–oriented, connectivist MOOC: pedagogical issues and implications for higher education. Int Rev Res Open Distrib Learn 2013;14(4):140-159. [doi: 10.19173/irrodl.v14i4.1548]
38. Liu M, Kang J, McKelroy E. Examining learners' perspective of taking a MOOC: reasons, excitement, and perception of usefulness. Educ Media Int 2015;52(2):1-18. [doi: 10.1080/09523987.2015.1053289]
39. Liu M, Kang J, Cao M, Lim M, Ko Y, Myers R, et al. Understanding MOOCs as an emerging online learning tool: perspectives from the students. Am J Distance Educ 2014;28(3):147-159. [doi: 10.1080/08923647.2014.926145]
40. MacKay JR, Langford F, Waran N. Massive open online courses as a tool for global animal welfare education. J Vet Med Educ 2016;43(3):287-301. [doi: 10.3138/jvme.0415-054R2] [Medline: 26751911]
41. Liang D, Jia J, Wu X, Miao J, Wang A. Analysis of learners' behaviors and learning outcomes in a massive open online course. Knowl Manag E-Learn 2014;6(3):281-298 [FREE Full text] [doi: 10.34105/j.kmel.2014.06.019]
42. Li Q, Wan F. A Case Study of the Characteristics of MOOCs Completers: Taking an Online Professional Training MOOC for Example. In: Proceedings of the 2016 IEEE 16th International Conference on Advanced Learning Technologies. 2016 Presented at: ICALT'16; July 25-28, 2016; Austin, TX, USA. [doi: 10.1109/icalt.2016.2]
43. Alturkistani A, Car J, Majeed A, Brindley D, Wells G, Meinert E. Determining the Effectiveness of a Massive Open Online Course in Data Science for Health. In: International Association for Development of the Information Society (IADIS) International Conference on e-Learning. 2018 Presented at: IADIS'18; 2018; Madrid, Spain URL: https://pdfs.semanticscholar.org/9426/4917562bf2fde478cf6f8910ae349547f1b1.pdf [doi: 10.2196/preprints.10982]
44. Cross S. Open Research Online - The Open University. Milton Keynes: Open University; 2013. Evaluation of the OLDS MOOC Curriculum Design Course: Participant Perspectives, Expectations and Experiences URL: http://oro.open.ac.uk/37836/1/EvaluationReport_OLDSMOOC_v1.0.pdf [accessed 2020-02-07]
45. Warriem J, Murthy S, Iyer S. Shifting the focus from learner completion to learner perseverancevidences from a teacher professional development MOOC. In: Evidences from a Teacher Professional Development MOOC. 2016 Presented at: Proceedings of the 24th International Conference on Computers in Education; 2016; India p. 540 URL: http://www.et.iitb.ac.in/~jkmadathil/publications/ICCE2016_Perseverance.pdf
46. Hudson L, Kortuem G, Wolff A, Hudson PL. Smart Cities MOOC: Teaching Citizens How to Co-Create Smart Cities. In: Proceedings of ICT for Sustainability 2016. 2016 Presented at: ICT4S'16; August 29-September 1, 2016; Amsterdam, Netherlands. [doi: 10.2991/ict4s-16.2016.18]
47. Brunton J, Brown M, Costello E, Farrell O, Mahon C. Giving flexible learners a head start on higher education: Designing and implementing a pre-induction socialisation MOOC. In: Delgado KC, Jermann P, Pérez-Sanagustín M, Seaton D, White S, editors. Digital Education: Out to the World and Back to the Campus. Cham: Springer; 2017:10-19.
48. Lei CU, Hou X, Kwok T, Chan T, Lee J, Oh E, et al. Advancing MOOC and SPOC Development via a Learner Decision Journey Analytic Framework. In: Proceedings of the 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering. 2015 Presented at: TALE'15; December 10-12, 2015; Zhuhai, China p. 156. [doi: 10.1109/tale.2015.7386034]
49. Mee CK, Mei SL, Jano Z, Husin H. The readiness of the administrators and undergraduates in using Massive Open Online Course (MOOC) in the Mandarin subject. Soc Sci 2016;11(12):3017-3023 [FREE Full text] [doi: 10.3923/sscience.2016.3017.3023]
50. Stephens M, Jones KM. MOOCs as LIS Professional Development Platforms: Evaluating and Refining SJSU's First Not-for-Credit MOOC. J Educ Libr Inf Sci 2014;55(4):345-361 [FREE Full text]
51. Milligan C, Littlejohn A. Supporting professional learning in a massive open online course. Int Rev Res Open Distrib Learn 2014;15(5):197-213. [doi: 10.19173/irrodl.v15i5.1855]
52. Lin J, Cantoni L. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. In: Eval Program Plan.: Information and Communication Technologies in Tourism 2017. Springer, Cham; 2004 Aug Presented at: ENTER 2017; 24-26 January, 2017; Rome, Italy p. 341-347. [doi: 10.1007/978-3-319-51168-9_10]
53. Sedgwick P. Cross sectional studies: advantages and disadvantages. Br Med J 2014;348:g2276. [doi: 10.1136/bmj.g2276]

54. Khalil H, Ebner M. MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2014. Chesapeake, VA: AACE; 2014:1236-1244.
55. Veletsianos G, Shepherdson P. A Systematic Analysis and Synthesis of the Empirical MOOC Literature Published in 2013–2015. Int Rev Res Open Distrib Learn 2016;17(2):198-221. [doi: 10.19173/irrodl.v17i2.2448]
56. Hone KS, El Said GR. Exploring the factors affecting MOOC retention: a survey study. Comput Educ 2016;98:157-168. [doi: 10.1016/j.compedu.2016.03.016]
57. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks: SAGE Publications; 1998.
58. Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends Inf Retr 2008;2(1–2):1-135. [doi: 10.1561/1500000011]
59. Kidzinski L, Sharma K, Shirvani BM, Dillenbourg P. On generalizability of MOOC models. In: Proceedings of the 9th International Conference on Educational Data Mining. 2016 Presented at: EDM'16; June 29 - July 2, 2016; North Carolina, USA p. 406-411.
60. Roy S, Singh SN. Emerging Trends in Applications of Big Data in Educational Data Mining and Learning Analytics. In: Proceedings of the 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence. 2017 Presented at: CONFLUENCE'17; Jan 12-13, 2017; Noida, India. [doi: 10.1109/confluence.2017.7943148]

Abbreviations

IEEE: Institute of Electrical and Electronic Engineers
LMS: learning management system
MOOC: massive open online course
PICO: population, intervention, comparator, outcome
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 27.02.19; peer-reviewed by E Toki, O Las Vergnas, F Sendra-Portero; comments to author 20.03.19; revised version received 20.11.19; accepted 22.01.20; published 27.04.20.

Please cite as:
Alturkistani A, Lam C, Foley K, Stenfors T, Blum ER, Van Velthoven MH, Meinert E
Massive Open Online Course Evaluation Methods: Systematic Review
J Med Internet Res 2020;22(4):e13851
URL: https://www.jmir.org/2020/4/e13851
doi:10.2196/13851
PMID:32338618

Original Paper

Patients' Willingness to Share Information in Online Patient Communities: Questionnaire Study

Panpan Zhu[1], PhD; Jiang Shen[1], Prof Dr; Man Xu[2], Prof Dr

[1]College of Management and Economics, Tianjin University, Tianjin, China

[2]Business School, Nankai University, Tianjin, China

Corresponding Author:
Man Xu, Prof Dr
Business School
Nankai University
No 94 Weijin Road, Nankai District
Tianjin, 300071
China
Phone: 86 139 2036 3161
Email: td_xuman@nankai.edu.cn

Abstract

Background: Online patient communities provide new channels for users to access and share medical information. In-depth study of users' willingness to share information in online patient communities is of great significance for improving the level of information sharing among the patient community and the long-term development of communities.

Objective: The aim of this study was to build a model of factors affecting patients' willingness to share medical information from the perspective of both positive and negative utilities. Specifically, we aimed to determine the influence of online information support and privacy concerns, as well as the moderating effect of disease severity and information sensitivity of different patients on their willingness to share.

Methods: Data from 490 users with experience in online patient communities were collected through a questionnaire survey, and structural equations were applied to empirically verify the model hypotheses.

Results: Privacy concerns negatively affected the patients' willingness to share information (P<.001), whereas online information support positively affected patients' willingness to share information (P<.001), and information sensitivity negatively moderated the impact of online information support on sharing willingness (P=.01). Disease severity positively moderated the impact of privacy concerns on sharing willingness (P=.05). However, the hypotheses that information sensitivity is a negative moderator and disease severity is a positive moderator of the impact of privacy concerns on sharing willingness could not be supported.

Conclusions: To improve the level of user information sharing, the online patient community should design a safe user registration process, ensure the confidentiality of information, reduce the privacy concerns of users, and accurately identify the information needs of patients to provide personalized support services.

(J Med Internet Res 2020;22(4):e16546) doi:10.2196/16546

KEYWORDS

online patient community; information sharing; willingness to share; questionnaire; structural equation

Introduction

Background

Social media is increasingly being used by health services to encourage the growth of social support networks among people with health-related issues [1,2]. Through participation on different platforms, patients are able to provide and receive social support in various ways, without the typical barriers and constraints experienced by this population [3]. In this way, problems of information asymmetry, as well as the shortages and uneven distribution of medical resources can be alleviated [4,5]. Jackson et al [6] studied the characteristics of patients who shared medical records with others, and found that participants who shared medical records were more likely to experience better health management. To obtain information support from other patients in the community, users need to share information on personal medical treatments, medication records, or medical diagnosis. However, disclosing such

information may also have serious adverse consequences [7]. In the face of risky decisions about online information demands and privacy disclosure, patients may choose whether to share such personal information according to their judgment on the possible benefits and losses. Therefore, in-depth study of users' information-sharing willingness and the decision-making mechanisms of online patient communities is of great significance to the improvement of information sharing among patient communities and the long-term development of these communities.

Most of the existing research on patient information sharing focuses on the impacts of sharing willingness from the perspective of the value brought by online communities [8-10]. However, users usually consider both the gains and risks associated with information sharing when making decisions about to whether they should share information [11]. In this context, we here establish a model of factors affecting patients' willingness to share medical information combined with utility theory, which considers both the influence of online information support (OIS) and privacy concerns on patients' willingness to share information (WSI) from the perspectives of both the positive and negative utilities brought by information sharing to users. The information-sharing decisions of online community users depend not only on the community environment (ie, whether the community provides OIS and whether the community environment can dispel users' privacy concerns) but also on the personal attributes of each community user (ie, health status and sensitivity to information). If patients are not sensitive to the positive utility of information sharing but are sensitive to the negative utilities (ie, afraid of the risks), the probability that they will choose not to share information is high, whereas if patients are not sensitive to the risks of information sharing but are very sensitive to the positive utilities, the possibility of sharing information is high. Therefore, the main contribution of this paper is to introduce patients' information sensitivity and health status into the model as moderating variables to explore the moderating effect of such individual patient attributes on information-sharing intention.

Literature Review and Hypotheses

Information-Sharing Behaviors in Online Patient Communities

Information sharing refers to the process during which individuals share the information they own with other members in the community in various forms and routes [12]. In most cases, members of online patient communities are patients troubled by different diseases [13]; therefore, they have relatively strong empathy. Their similar situations enable them to easily obtain both the emotional and information support they need while relating to the feelings of other members of the community during the process of communication [14]. Combined with the features of online patient communities, shared information can be classified into general health information and special health information. General health information refers to common medical and health information, which is usually open to the public and can be easily obtained by the masses, such as information on doctors and hospitals, prices of medicines, and features of a specific disease. Special health information is usually concerned with privacy, which refers to personal empirical information, including medical records, treatment, and personal health conditions [15].

Utility Theory

Utility theory is often adopted when individuals make decisions [16]. Utility is defined as the degree to which a user's psychological expectation of a certain factor is satisfied or realized [17]. This judgment is completely controlled by personal preferences, interests, or the subjective consciousness: the higher the satisfaction degree, the larger the utility, and vice versa [18]. The criterion is maximizing utility by weighing costs and benefits [18], and is considered the ultimate expression of reason and egotism, reflecting the decision maker's attitude toward risks. In the utility theory formula $U_D=f(B,C)$, C refers to the potential risk of privacy disclosure caused by information sharing, which is a variable indicating negative utility to users; B refers to the positive utility obtained by users, which mainly includes the OIS that patients receive; and D refers to the decision of users on whether or not to participate in information sharing after judging the positive and negative utilities brought by this behavior.

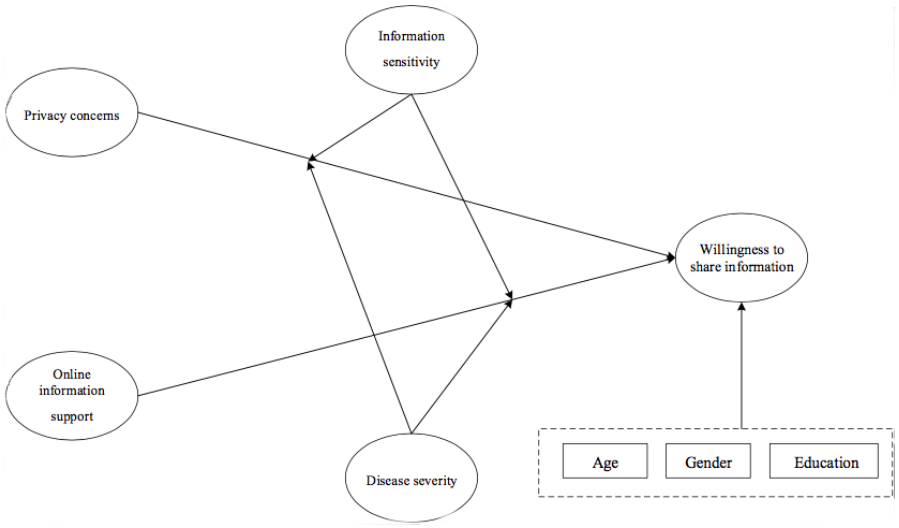
Information sharing creates value for users but also invites risks. Different individuals have different sensitivities to personal medical health information; such difference in sensitivity could result in different concerns regarding information sharing, thereby affecting their WSI. In addition, different patients have different physical health conditions. If the disease of a patient is severe, the patient may be urgently seeking OIS, which could weaken relative concerns regarding privacy. In this case, the patient is more likely to acquire the necessary information by positively participating in information exchange in communities.

Based on the above analysis, we established an influence model of OIS and privacy concerns for patients' WSI (Figure 1). The moderating functions of patients' information sensitivity and disease severity on the WSI were also considered. Specifically, we consider two problems in this model structure: (1) the influence of OIS and privacy concerns on the willingness to share medical information, and (2) the functions of information sensitivity and disease severity as moderating variables.

In this model, the information-sharing intention of patients is the dependent variable, privacy concerns and OIS are independent variables, and information sensitivity and disease severity are moderator variables. That is, privacy concerns and OIS are the main factors affecting patients' intentions to share medical information, the effects of which are regulated by disease severity and information sensitivity. Moderator variables with increasing negative effects (ie, negative moderator influence factors) affect sharing intentions, whereas moderator variables with decreasing negative effects (ie, positive moderator influence factors) also affect sharing willingness. The control variables include gender, age, and education level. Along with the different social roles and characteristics between genders, feelings and behaviors can differ according to gender regarding the same issue [19]. As a result, gender may influence information-sharing intentions. Age may also have effects on information-sharing intentions since age influences physical condition and life experience, which can result in quite different

opinions and feelings about the same issue [20]. With respect to education level, higher education levels are associated with stronger judgment and cognitive competence; thus, behavioral patterns may differ across this characteristic as well [21] to ultimately affect the information-sharing intentions of online health community users.

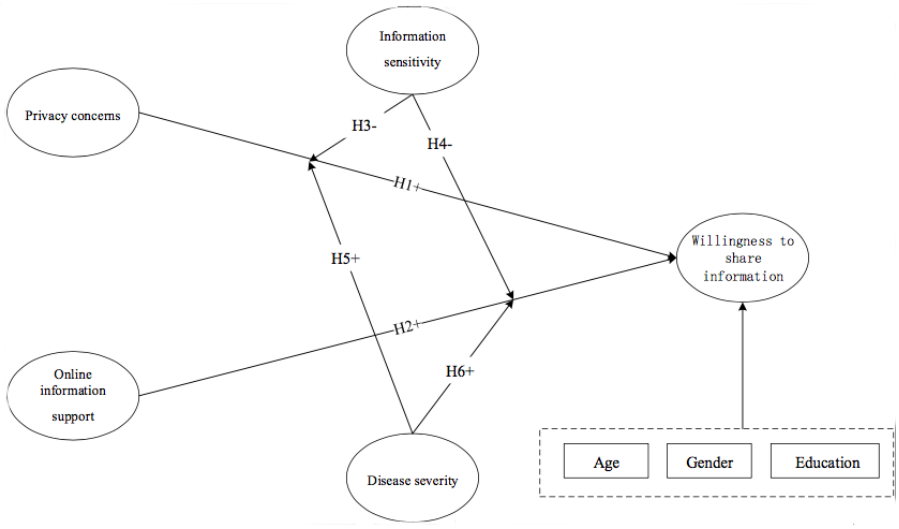
Figure 1. Model of information-sharing intentions of patients in online health communities.



Research Hypotheses

Based on the above theoretical model, six hypotheses are proposed. As shown in Figure 2, the hypotheses are composed of two parts: hypotheses of direct effects and hypotheses of moderation effects.

Figure 2. Research hypotheses on online information-sharing willingness.



Hypotheses of Direct Effects

Relationship Between Online Information Support and Patients' Willingness to Share Information

In an online patient community, patients conduct information sharing to acquire medical assistance and personalized experience information [22]. In a study of two online communities, Hargreaves et al [23] found that patients share experiences, information, and emotions, and receive empathetic support in a supportive and warm community atmosphere. Based on the theory of planned behavior and the theory of technology acceptance, Zhou [24] found that social returns and perceived usefulness can determine the intentions to share personal health information. On the basis of a technology acceptance model, Zhang [25] found that users' usefulness of acquiring perceptual information from virtual communities can affect their information-sharing behaviors. On the basis of social exchange theory, Gui and Hu [26] found that mutual benefits can significantly affect intentions to share health knowledge by conducting empirical analyses. Thus, research employing various theoretical perspectives has demonstrated that OIS has positive impacts on the information-sharing intentions of patients; in other words, support or help from other community members strengthens the intention to share information.

Thus, the following hypothesis is presented:

H1: OIS has positive impacts on patients' WSI.

Relationship Between Privacy Concerns and Patients' Willingness to Share

As electronic health records become ubiquitous in the health care industry, privacy breaches are increasing [27]. In an online patient community, patients can predict but not fully control these potential risks, thereby reducing their WSI [28]. Privacy concerns are one of the major factors affecting attitudes toward exchanging electronic health information [29,30]. The privacy of patients is prominent in health information exchange discussions, given that their potentially sensitive personal health information may be electronically shared for various health care purposes [31]. Based on a grounded theory, Holloway et al [32] found that issues of privacy and confidentiality were paramount to information sharing. In an empirical study, Zhang et al [33] found that privacy concerns have obvious negative impacts on the attitudes of online patient community users regarding information disclosure. The main objectives of different methods of health information exchange are to reduce privacy concerns [34]. Frost et al [35] studied the information-sharing preferences of cancer patients in the community and found that poor online privacy experience, age, and health status negatively affected the WSI.

Thus, the following hypothesis is presented:

> *H2: Privacy concerns have negative impacts on patients' WSI.*

Hypotheses of Moderation Effects

Information sensitivity, which is closely related to privacy concerns, refers to the extent of concern of individuals regarding information in a specific environment: the greater the extent of concern, the more sensitive the information [36]. Medical information also includes personal information that is not directly related to personal health, such as population statistics (ie, name and date of birth) and information used to identify a patient's personal identity (ie, ID card number, family address, private phone number, financial information, and insurance status). In addition, such information includes details that may help in making a correct diagnosis and adopting proper treatment measures, such as the cause for seeking medical advice, symptoms, medical history of family members, drug allergy history, and medical history, as well as disease data arising from diagnosis and treatment, such as examination results, progress notes, operation consent, anesthesia consent, operation records, and anesthesia records. Since most of this information involves personal privacy, online sharing causes concern regarding privacy disclosure: the more sensitive the information, the greater the privacy concerns.

Many people believe that if sensitive information is shared or disclosed, it will have bad consequences or even cause damage [37]. Consumers' intentions to disclose personal information are closely related to the sensitivity of such information [38]. Milne [39] found that the validity of information changes with the sensitivity of such information when collecting personal information from users. According to Okazaki et al [40], information sensitivity increases privacy concerns. Information sensitivity is frequently used as the prepositive variable of privacy concerns [41]. In a network environment, privacy concerns cause a reduction in the publishing of information by online users, whereas information sensitivity causes concerns regarding private information [42]. In this study, the information sensitivity of patients was used as a moderator variable to explain the behavioral intentions of patients regarding sharing information in online patient communities. The users' privacy concerns during information sharing will be more serious with respect to more sensitive medical information, thereby affecting sharing behavioral decisions [43].

Thus, the following hypotheses are presented:

> *H3: Information sensitivity negatively moderates the impact of privacy concerns on patients' WSI.*
>
> *H4: Information sensitivity negatively moderates the impact of OIS on patients' WSI.*

Disease severity refers to the users' perception of their own health status; that is, a subjective perception of whether one's health is good or bad will affect their information-sharing intentions. Tisnado et al [44] studied the impacts of demographic characteristics and individual health on the concordance between medical records and health information sharing. Patients who considered their health to be poor were more sensitive about sharing their personal health information compared with others; these patients worried that they may face uncontrollable risks if their medical and health information were to be disclosed [45]. Health conditions as perceived by online patient community users translate into privacy concerns regarding their medical and health information, which ultimately affect their intentions to share this information [46]. In online patient communities, the priority of patients is to improve their health, interact with other community users, and seek medical assistance. The levels of participation and activeness in the community also vary according to diversified health conditions. When they are physically healthy, they need not consult doctors or tell others about their medical treatment experiences; considering the privacy of medical information, their intentions to share relevant medical and health information may be reduced. When they are physically unhealthy, they urgently need to acquire the information support and assistance provided by online communities to improve their health [47]. However, the precondition is that to receive this assistance, they will need to illustrate their health condition to relevant helpers. Therefore, they have to actively share medical information, and thus their intentions to share relevant information increase. Consequently, the disease severity perceived by online patient community users may affect their intentions in sharing personal medical information by regulating their privacy concerns and OIS demands. In this study, disease severity was used as a moderator variable to moderate the relationship between privacy concerns and OIS demand intensity when both sharing and intending to share health and medical information so as to verify the impact on sharing behavioral patterns.

Thus, the following hypotheses are presented:

> *H5: Disease severity positively moderates the effect of privacy concerns on patients' WSI.*
>
> *H6: Disease severity positively moderates the effect of OIS on patients' WSI.*

Methods

Questionnaire Design

The questionnaire was composed of three parts: Part 1 reflects the experience of the respondents in sharing medical information with any online health community (if such experience is available), Part 2 reflects the objective cognition of the respondents regarding online medical services, and Part 3 mainly concerns the respondents' basic information and medical history.

The medical information sharing intentions, disease severity, and perceived health information sensitivity were measured according to the method of Bansal et al [48], and privacy concerns and OIS were measured according to the method of Liang et al [49]. A typical 7-point Likert scale was used to score opinions and conclusions, with dimensions of "quite disagree," "disagree," "basically disagree," "neutral," "basically agree," "agree," and "quite agree", respectively. The specific scale is shown in Table 1.

Table 1. Measurement scale.

Component and measurement number	Scale
Willingness to share (WSI)	
WSI1	I am willing to provide my medical health information to online medical websites to get proper treatment.
WSI2	I agree to let online medical websites use my medical health information.
WSI3	I do not think it is improper if my medical health information is revealed to online medical websites.
WSI4	It is highly probable that I will publish my medical health information on some online medical websites.
Privacy concerns (PC)	
PC1	I worry that my medical health information posted online may be misused.
PC2	I worry that other people can see my medical health information online.
PC3	In my opinion, personal medical health information is more important than other information.
PC4	I worry that my medical health information posted online will be used by others in an unexpected manner.
Disease severity (DS)	
DS1	My body is seldom subject to long-term pain and discomfort.
DS2	I have no chronic disease.
DS3	My overall health is good.
Online information support (OIS)	
OIS1	When I need assistance related to health issues, anyone from an online medical website can provide relevant advice.
OIS2	When I encounter any health problem, anyone from an online medical website can provide relevant information to help me overcome the problem.
OIS3	Anyone from an online medical website will help me find the cause and provide relevant suggestions for any health problem I might have.
Information sensitivity (IS)	
IS1	Medicine taken
IS2	Current health condition
IS3	Medical history
IS4	Process of seeking medical advice

Data Collection

To guarantee the validity of the questionnaire design and content, many experts and some respondents were invited to evaluate each question in the questionnaire. An internal test was conducted within a small range before issuing these questionnaires. After confirming that there were no errors, the questionnaires were issued through wjx.cn, an online platform for issuing questionnaires. Those who had experience sharing information in online health communities were invited to fill in these questionnaires. A total of 531 copies were collected, 41 of which were incomplete, unreasonable, or unqualified, leaving a total of 490 valid responses.

Data Analysis

The statistical description of the correlated variables is shown in Table 2.

Table 2. Statistical description of variables.

Variables	Sample, n (%)
Gender	
Men	168 (34.3)
Women	322 (65.7)
Age (years)	
≤20	3 (0.6)
21-30	216 (44.1)
31-40	243 (49.6)
41-50	28 (6)
51-60	0 (0)
>60	0 (0)
Education level	
Middle school and below	0 (0)
High school	10 (2)
College	43 (9)
Bachelor	357 (72.9)
Master	68 (14)
Doctorate and above	12 (3)
Duration of using the internet (years)	
≤1	0 (0)
2-4	40 (8)
5-7	123 (25.1)
8-10	171 (34.9)
>10	156 (31.8)
Duration of using online medical websites (years)	
≤1	93 (19)
2-3	278 (56.7)
4-5	101 (20.6)
6-7	15 (3)
>7	3 (0.6)
Annual frequency of getting ill	
≤1	88 (18)
2-3	291 (59.4)
4-6	93 (19)
7-10	15 (3)
>10	3 (0.6)

Results

Measurement Model Testing

Reliability Test

The inspection criteria for the reliability and validity tests were the load levels of the corresponding components, common factor loading, Cronbach alpha, composite reliability, and average variance extracted (AVE). The desired values for these indices [50] are shown in Table 3.

The measurement results for the data factor loading coefficients are shown in Table 4.

The factor loading measurements <0.7 in the questionnaire were deleted, with those remaining reaching the criteria. The results of the modified Smart PLS data analysis are shown in Table 5.

As shown in Table 5, Cronbach alpha for each component was larger than 0.77, indicating that the reliability for each question in the research design was good. The composite reliability coefficient of each component was larger than 0.78, the AVE for each component was greater than 0.57, and all components could be aggregated effectively.

Table 3. Desired values for the reliability test.

Evaluation index	Critical value
Factor loading coefficient	>0.70
Cronbach alpha	>.70
Composite reliability coefficient	>0.70
Average variance extracted	>0.50

Table 4. Factor loading coefficients for all variables.

Component and scale	Factor loading coefficient
Privacy concerns (PC)	
PC1	0.934
PC2	0.922
PC3	0.767
PC4	0.938
Online information support (OIS)	
OIS1	0.836
OIS2	0.848
OIS3	0.808
Information sensitivity (IS)	
IS1	0.671
IS2	0.951
IS3	0.850
IS4	0.756
Disease severity (DS)	
DS1	0.717
DS2	0.981
DS3	0.776

Table 5. Reliability of test results.

Variable	AVE[a]	Composite reliability	R^2	Cronbach alpha	Communality	Redundancy
Age	1	1	N/A[b]	1	1	N/A
DS[c]	0.574	0.785	N/A	0.854	0.574	N/A
EDU[d]	1	1	N/A	1	1	N/A
Gender	1	1	N/A	1	1	N/A
IS[e]	0.631	0.869	N/A	0.846	0.631	N/A
OIS[f]	0.690	0.870	N/A	0.775	0.690	N/A
PC[g]	0.731	0.913	N/A	0.889	0.731	N/A
WSI[h]	0.654	0.883	0.331	0.823	0.654	0.039

[a]AVE: average variance extracted.
[b]N/A: not applicable.
[c]DS: disease severity.
[d]EDU: education level.
[e]IS: information sensitivity.
[f]OIS: online information support
[g]PC: privacy concerns.
[h]WSI: willingness to share information.

Validity Test

In general, two measurement indices are used for the validity test for each component: the discriminant validity test requires that the correlation coefficient between two components be less than its AVE value, and the convergent validity test requires that the load value at each proposed measured variable be larger than that loaded at any other measured variable in cross-factor loading. Table 6 shows the correlation coefficients used for measuring discriminant validity; diagonal elements contain the arithmetic square root of AVE, and nondiagonal elements contain the correlation coefficient for each corresponding component. Based on these correlations, the data have favorable discriminant validity by satisfying relevant requirements. Table 7 further shows that the factor loading loaded at the corresponding component was larger than that loaded at any other latent variable, demonstrating quite good convergent validity between two components.

Table 6. Correlation coefficients.

Variable	Age	DS[a]	EDU[b]	Gender	IS[c]	OIS[d]	PC[e]	WSI[f]
Age	1	—[g]	—	—	—	—	—	—
DS	–0.120	1	—	—	—	—	—	—
EDU	–0.204	–0.085	1	—	—	—	—	—
Gender	–0.044	–0.021	–0.100	1	—	—	—	—
IS	–0.158	–0.038	0.093	0.0351	1	—	—	—
OIS	0.123	0.209	–0.116	0.032	0.132	1	—	—
PC	–0.179	0.079	0.092	–0.041	0.466	0.083	1	—
WSI	0.254	0.158	-0.009	0.043	–0.129	0.436	–0.267	1

[a]DS: disease severity.
[b]EDU: education level.
[c]IS: information sensitivity.
[d]OIS: online information support.
[e]PC: privacy concerns.
[f]WSI: willingness to share.
[g]Not applicable.

Table 7. Cross-factor loadings.

Variable	Age	DS[a]	EDU[b]	Gender	IS[c]	OIS[d]	PC[e]	WSI[f]
Age	1	–0.120	–0.204	–0.044	–0.158	0.123	–0.179	0.254
DS1	–0.164	0.410	–0.031	0.029	0.149	0.136	0.308	–0.037
DS2	–0.137	0.980	–0.089	–0.009	–0.009	0.214	0.123	0.144
DS3	–0.161	0.771	–0.018	–0.031	0.064	0.189	0.237	0.020
EDU	–0.204	–0.085	1	–0.100	0.093	–0.116	0.092	–0.009
Gender	–0.044	–0.021	–0.100	1	0.035	0.032	–0.041	0.043
IS1	–0.090	–0.071	0.080	–0.017	0.570	–0.008	0.243	0.009
IS2	–0.152	–0.059	0.049	–0.017	0.950	0.080	0.380	–0.152
IS3	–0.138	0.015	0.153	0.117	0.851	0.217	0.487	–0.063
IS4	–0.109	–0.029	0.104	0.064	0.757	0.090	0.430	–0.047
OIS1	0.080	0.101	–0.108	0.086	0.088	0.835	0.047	0.369
OIS2	0.090	0.222	–0.047	0.069	0.074	0.848	0.026	0.358
OIS3	0.137	0.199	–0.132	–0.076	0.168	0.809	0.133	0.361
PC1	–0.100	0.056	0.094	–0.054	0.447	0.102	0.934	–0.221
PC2	–0.243	0.147	0.044	–0.103	0.420	0.070	0.922	–0.271
PC3	–0.125	0.092	–0.005	0.015	0.285	0.202	0.570	0.003
PC4	–0.143	0.011	0.124	0.047	0.438	0.065	0.937	–0.247
WSI1	0.179	0.115	–0.009	0.110	0.041	0.516	–0.048	0.733
WSI2	0.184	0.185	–0.076	0.050	–0.185	0.293	–0.245	0.837
WSI3	0.260	0.161	0.025	–0.016	–0.163	0.268	–0.283	0.820
WSI4	0.195	0.053	0.025	–0.009	–0.123	0.315	–0.297	0.841

[a]DS: disease severity.

[b]EDU: education level.

[c]IS: information sensitivity.

[d]OIS: online information support.

[e]PC: privacy concerns.

[f]WSI: willingness to share

Results of Structural Model Testing

After fully verifying the reliability and validity of the questionnaire, we next analyzed the structural model to test the various proposed hypotheses based on the following values: (1) the path coefficient between each variable, (2) the significance of the *t* value, and (3) the degree of R^2 to which the dependent variable is interpreted. In line with the research model, significance verification was conducted in two steps: the first step assessed the significance of the influencing factors among the major components (namely, verification of the basic model), and the second step examined the effect of the regulated variable added to the model. The results are shown in Table 8.

Table 8. Verification of the basic model.

Hypothesis	t_{485}	*P* value	Supported
H1: online information support has positive impacts on patients' willingness to share information.	5.77	<.001	Yes
H2: privacy concerns have negative impacts on patients' willingness to share information	3.41	<.001	Yes
H3: information sensitivity negatively moderates the impact of privacy concerns on patients' willingness to share information.	0.96	.34	No
H4: information sensitivity negatively moderates the impact of online information support on patients' willingness to share information.	2.58	.01	Yes
H5: disease severity positively moderates the effect of privacy concerns on patients' willingness to share information.	1.96	.05	Yes
H6: disease severity positively moderates the effect of online information support on patients' willingness to share information.	0.68	.50	No

Discussion

Principal Findings

This study showed that privacy concerns have a negative impact on users' WSI, and the OIS provided by online communities has a positive impact on the users' WSI. The hypothesis that information sensitivity negatively regulates the impact of privacy concerns on WSI was not supported. The indirect impact of information sensitivity on WSI mainly occurs through negative regulation of the OIS offered by communities. This phenomenon may be caused by the platform's privacy protection mechanism. In most cases, community platforms include certain user privacy protection measures to protect users' personal data from being leaked. When patients participate in online communication, most of their personal information is blocked and invisible to others. Despite the involvement of relatively sensitive information, users can be free from concerns regarding disclosure of their personal identity, which greatly diminishes the effect of information sensitivity on privacy concerns. Although the privacy protection measures of the community platform dispel the inhibition effect of privacy concerns caused by patients' information sensitivity on WSI to some degree, information sensitivity still negatively regulates the positive effect of OIS on WSI. If patients are in poor health and need to obtain information from other patients in the community, their privacy concerns will be reduced to a certain extent. The mental inhibition factors of information sharing are reduced, while the weakened privacy concerns further increase the willingness to participate in information sharing. In summary, these results suggest that a community platform should make all efforts to desensitize the information shared by users while providing effective information support for users so as to reduce the negative effects of information sharing, increase the benefits of information sharing, and improve the overall information-sharing level of community users.

Theoretical and Practical Significance

The information-sharing intention model established in this study enriches and improves theoretical knowledge in the following aspects. Primarily, utility theory was applied to study users' WSI in online patient communities. Previous studies on users' willingness to share have mainly focused on aspects of motivation, mutual benefits, social exchange, and the like. Yet, by combining the features of users' information sharing in online patient communities, this study adopted utility theory to probe users' information-sharing decisions under the condition that both privacy concerns and information demands exist from the user's perspective. In addition, based on the characteristics of shared information, we further analyzed the information sensitivity and the adjustment effect of the patients' own perceived health status on their WSI. Under the combined effect of the above factors, assessment of users' WSI in this manner expands the application of utility theory to studies on information-sharing behaviors, thereby laying the foundation for further development of utility theory.

This research also has practical significance. In view of the mechanism of users' WSI, online service operators should establish a reasonable mechanism to ensure the security and confidentiality of users' medical information through reasonable and safe user registration processes and appropriate protection measures, thereby minimizing users' privacy concerns during the process. Moreover, the patients' health status information should be obtained to accurately identify individual information requirements, so that the platform can offer personalized support services while providing appropriate guidance and incentives, thereby improving the information-sharing level of community users.

Limitations

As with all empirical studies, this study also has limitations. First, online patient communities comprise a large group of users distributed in various cities across the country. Although the data of this study were obtained from the Internet, the sample size is negligible compared with the enormous group of community users. In addition, the sample data were all obtained from China (without considering a potential Western cultural background), which may affect the conclusions of the study to a certain extent. Second, the subject of information sharing in online patient communities is necessarily "human," and thus the sharing behavior strategy is inclined to be affected by multiple factors. Moreover, with constant changes in other users' sharing behaviors, future studies could focus more on the dynamic adjustment of users' sharing strategies, thereby providing a deeper understanding of the entire sharing process.

Conclusions

This study established a model to probe the factors influencing decision making on patients' WSI from the perspectives of both positive and negative utilities. The results confirm the positive and negative effects of OIS and privacy concerns on users' WSI. In addition, we showed that information sensitivity negatively moderates the impact of OIS on sharing willingness, whereas disease severity positively moderates the impact of privacy concerns on sharing willingness. Therefore, the management of the platform should focus on desensitizing the information shared by users, while providing effective information support for users to reduce the negative effects of information sharing, increase the benefits of information sharing, and improve the information-sharing level of community users. This research can help community platform operators to clarify the behavioral mechanism of users' independent disclosure of medical information, provide theoretical guidance for operators to stipulate effective measures that encourage users to voluntarily disclose medical information, and offer a reference to improve the operators' user management of online patient community services.

Acknowledgments

This study was partially funded by the National Natural Science Foundation of China (grant numbers 71971123 and 71571105) and the Fundamental Research Funds for the Central Universities.

Conflicts of Interest

None declared.

References

1. Meng F, Guo X, Peng Z, Zhang X, Vogel D. The routine use of mobile health services in the presence of health consciousness. Electron Commer R A 2019 May;35:100847 [FREE Full text] [doi: 10.1016/j.elerap.2019.100847]
2. Ellway D, Reilly R, Le Couteur A, Ward J. Exploring how people affected by methamphetamine exchange social support through online interactions on Facebook: Content Analysis. JMIR Ment Health 2019 Oct 01;6(10):e14011 [FREE Full text] [doi: 10.2196/14011] [Medline: 31573926]
3. Stetten NE, LeBeau K, Aguirre MA, Vogt AB, Quintana JR, Jennings AR, et al. Analyzing the Communication Interchange of Individuals With Disabilities Utilizing Facebook, Discussion Forums, and Chat Rooms: Qualitative Content Analysis of Online Disabilities Support Groups. JMIR Rehabil Assist Technol 2019 Sep 30;6(2):e12667 [FREE Full text] [doi: 10.2196/12667] [Medline: 31573926]
4. 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]
5. Klein DM, Fix GM, Hogan TP, Simon SR, Nazi KM, Turvey CL. Use of the Blue Button Online Tool for Sharing Health Information: Qualitative Interviews With Patients and Providers. J Med Internet Res 2015 Aug 18;17(8):e199 [FREE Full text] [doi: 10.2196/jmir.4595] [Medline: 26286139]
6. Jackson SL, Mejilla R, Darer JD, 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):e247 [FREE Full text] [doi: 10.2196/jmir.3363] [Medline: 25405911]
7. Fan K, Jiang W, Li H, Yang Y. Lightweight RFID Protocol for Medical Privacy Protection in IoT. IEEE Trans Ind Inf 2018 Apr;14(4):1656-1665. [doi: 10.1109/tii.2018.2794996]
8. Yang H, Ju X. The effects of social support and individual goal on health condition. J Manag Sci 2017;30(1):005-061.
9. Zhang X, Liu S, Deng Z, Chen X. Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. Comput Hum Behav 2017 Oct;75:797-810. [doi: 10.1016/j.chb.2017.06.028]
10. Oh S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. J Am Soc Inf Sci 2011 Nov 01;63(3):543-557. [doi: 10.1002/asi.21676]
11. Zhang X, Guo X, Wu Y, Lai K, Vogel D. Exploring the inhibitors of online health service use intention: A status quo bias perspective. Inf Manag 2017 Dec;54(8):987-997. [doi: 10.1016/j.im.2017.02.001]
12. Vaala S, Lee J, Hood K, Mulvaney S. Sharing and helping: Predictors of adolescents? willingness to share diabetes personal health information with peers. J Am Med Informatics Assoc 2018;25(2):135-141. [doi: 10.1093/jamia/ocx051] [Medline: 28575462]
13. Nambisan P, Gustafson DH, Hawkins R, Pingree S. Social support and responsiveness in online patient communities: impact on service quality perceptions. Health Expect 2015 Jan 02;19(1):87-97. [doi: 10.1111/hex.12332] [Medline: 25556856]
14. McCaig D, Elliott MT, Siew CS, Walasek L, Meyer C. Profiling Commenters on Mental Health–Related Online Forums: A Methodological Example Focusing on Eating Disorder–Related Commenters. JMIR Ment Health 2019 Apr 22;6(4):e12555 [FREE Full text] [doi: 10.2196/12555] [Medline: 31008715]

15. Lognos B, Carbonnel F, Boulze Launay I, Bringay S, Guerdoux-Ninot E, Mollevi C, et al. Complementary and Alternative Medicine in Patients With Breast Cancer: Exploratory Study of Social Network Forum Data. JMIR Cancer 2019 Nov 27;5(2):e12536 [FREE Full text] [doi: 10.2196/12536]
16. Dhami S, Wei M, al-Nowaihi A. Public goods games and psychological utility: Theory and evidence. J Econ Behav Org 2019 Nov;167:361-390. [doi: 10.1016/j.jebo.2017.11.002]
17. Le LK, Sanci L, Chatterton ML, Kauer S, Buhagiar K, Mihalopoulos C. The Cost-Effectiveness of an Internet Intervention to Facilitate Mental Health Help-Seeking by Young Adults: Randomized Controlled Trial. J Med Internet Res 2019 Jul 22;21(7):e13065 [FREE Full text] [doi: 10.2196/13065] [Medline: 31333199]
18. Andersen S, Harrison GW, Lau MI, Rutström EE. Multiattribute utility theory, intertemporal utility, and correlation aversion. Int Econ Rev 2018 May 03;59(2):537-555. [doi: 10.1111/iere.12279]
19. Andersen S, Harrison GW, Lau MI, Rutström EE. Multiattribute utility theory, intertemporal utility, and correlation aversion. Int Econ Rev 2018 May 03;59(2):537-555. [doi: 10.1111/iere.12279]
20. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
21. Reuter T, Ziegelmann JP, Wiedemann AU, Lippke S, Schüz B, Aiken LS. Planning bridges the intention–behaviour gap: Age makes a difference and strategy use explains why. Psychol Health 2010 Sep;25(7):873-887. [doi: 10.1080/08870440902939857] [Medline: 20204963]
22. Ng T, Feldman D. How broadly does education contribute to job performance? Pers Psychol 2009;62(1):89-134. [doi: 10.1111/j.1744-6570.2008.01130.x]
23. Verberne S, Batenburg A, Sanders R, van Eenbergen M, Das E, Lambooij MS. Analyzing Empowerment Processes Among Cancer Patients in an Online Community: A Text Mining Approach. JMIR Cancer 2019 Apr 17;5(1):e9887 [FREE Full text] [doi: 10.2196/cancer.9887] [Medline: 30994468]
24. Hargreaves S, Bath P, Duffin S. Sharing and empathy in digital spaces: qualitative study of online health forums for breast cancer and motor neuron disease (amyotrophic lateral sclerosis). J Med Internet Res 2018;20:e222 [FREE Full text] [doi: 10.2196/jmir.9709] [Medline: 29903695]
25. Zhou J. Factors Influencing People's Personal Information Disclosure Behaviors in Online Health Communities: A Pilot Study. Asia Pac J Public Health 2018 Feb 06;30(3):286-295. [doi: 10.1177/1010539518754390] [Medline: 29405740]
26. Zhang J. Empirical research on information acquiring and information sharing intention and behavior in virtual communities. Inform Sci 2015;8:59-64. [doi: 10.13833/j.cnki.is.2015.08.012]
27. Gui P, Hu XF. Study on factors influencing knowledge sharing willingness of health online community members based on internet word-of-mouth social exchange theory. J Modern Inf 2017;27:250-252. [doi: 10.16541/j.cnki.2095-8420.2017.27.113]
28. Walker DM, Johnson T, Ford EW, Huerta TR. Trust Me, I'm a Doctor: Examining Changes in How Privacy Concerns Affect Patient Withholding Behavior. J Med Internet Res 2017 Jan 04;19(1):e2 [FREE Full text] [doi: 10.2196/jmir.6296] [Medline: 28052843]
29. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What Predicts Patients' Adoption Intention Toward mHealth Services in China: Empirical Study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
30. Abdelhamid M, Gaia J, Sanders GL. Putting the Focus Back on the Patient: How Privacy Concerns Affect Personal Health Information Sharing Intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
31. Zhang X, Liu S, Chen X, Wang L, Gao B, Zhu Q. Health information privacy concerns, antecedents, and information disclosure intention in online health communities. Inform Manag 2018 Jun;55(4):482-493. [doi: 10.1016/j.im.2017.11.003]
32. Shen N, Sequeira L, Silver M, Carter-Langford A, Strauss J, Wiljer D. Patient Privacy Perspectives on Health Information Exchange in a Mental Health Context: Qualitative Study. JMIR Ment Health 2019 Nov 13;6(11):e13306 [FREE Full text] [doi: 10.2196/13306] [Medline: 31719029]
33. Holloway IW, Winder TJA, Lea ICH, Tan D, Boyd D, Novak D. Technology Use and Preferences for Mobile Phone-Based HIV Prevention and Treatment Among Black Young Men Who Have Sex With Men: Exploratory Research. JMIR Mhealth Uhealth 2017 Apr 13;5(4):e46 [FREE Full text] [doi: 10.2196/mhealth.6436] [Medline: 28408360]
34. Zhang X, Chen X, Hou DL. An analysis of online health information disclosure willingness influencing factors: An integrated model of TPB and privacy calculus. Information and Documentation Services. Inform Document Serv 2016;37(1):48-53. [doi: 10.3969/j.issn.1002-0314.2016.01.008]
35. Esmaeilzadeh P, Mirzaei T. The Potential of Blockchain Technology for Health Information Exchange: Experimental Study From Patients' Perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
36. Frost J, Vermeulen IE, Beekers N. Anonymity versus privacy: selective information sharing in online cancer communities. J Med Internet Res 2014 May;16(5):e126 [FREE Full text] [doi: 10.2196/jmir.2684] [Medline: 24828114]
37. Kim D, Park K, Park Y, Ahn JH. Willingness to provide personal information: Perspective of privacy calculus in IoT services. Comput Hum Behav 2019 Mar;92:273-281. [doi: 10.1016/j.chb.2018.11.022]

38. Feri F, Giannetti C, Jentzsch N. Disclosure of personal information under risk of privacy shocks. J Econ Behav Org 2016 Mar;123:138-148. [doi: 10.1016/j.jebo.2015.12.001]
39. Househ M, Grainger R, Petersen C, Bamidis P, Merolli M. Balancing Between Privacy and Patient Needs for Health Information in the Age of Participatory Health and Social Media: A Scoping Review. Yearb Med Inform 2018 Apr 22;27(01):029-036. [doi: 10.1055/s-0038-1641197] [Medline: 29681040]
40. Dinev T, Hart P. An Extended Privacy Calculus Model for E-Commerce Transactions. Inform Syst Res 2006 Mar;17(1):61-80. [doi: 10.1287/isre.1060.0080]
41. Okazaki S, Li H, Hirose M. Consumer Privacy Concerns and Preference for Degree of Regulatory Control. J Advert 2009 Dec;38(4):63-77. [doi: 10.2753/joa0091-3367380405]
42. Rahim FA, Ismail Z, Samy GN. Information Privacy Concerns in the Use of Social Media Among Healthcare Practitioners: A Systematic Literature Review. Adv Sci Lett 2014 Oct 01;20(10):2176-2179. [doi: 10.1166/asl.2014.5659]
43. Ozdemir ZD, Jeff Smith H, Benamati JH. Antecedents and outcomes of information privacy concerns in a peer context: An exploratory study. Eur J Inform Syst 2018 Feb 15;26(6):642-660. [doi: 10.1057/s41303-017-0056-z]
44. Bansal G, Zahedi F, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
45. Tisnado DM, Adams JL, Liu H, Damberg CL, Hu FA, Chen W, et al. Does the concordance between medical records and patient self-report vary with patient characteristics? Health Serv Outcomes Res Method 2006 Nov 23;6(3-4):157-175. [doi: 10.1007/s10742-006-0012-1]
46. Reisner SL, Randazzo RK, White Hughto JM, Peitzmeier S, DuBois LZ, Pardee DJ, et al. Sensitive Health Topics With Underserved Patient Populations: Methodological Considerations for Online Focus Group Discussions. Qual Health Res 2017 May 04;28(10):1658-1673. [doi: 10.1177/1049732317705355] [Medline: 29298574]
47. Bansal G, Zahedi F, Gefen D. The role of privacy assurance mechanisms in building trust and the moderating role of privacy concern. Eur J Inf Syst 2017 Dec 19;24(6):624-644. [doi: 10.1057/ejis.2014.41]
48. Meng F, Guo X, Peng Z, Lai K, Zhao X. Investigating the Adoption of Mobile Health Services by Elderly Users: Trust Transfer Model and Survey Study. JMIR Mhealth Uhealth 2019 Jan 08;7(1):e12269 [FREE Full text] [doi: 10.2196/12269] [Medline: 30622092]
49. Bansal G, Zahedi F, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decision Support Systems 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
50. Liang T, Ho Y, Li Y, Turban E. What Drives Social Commerce: The Role of Social Support and Relationship Quality. Int J Electron Commer 2014 Dec 10;16(2):69-90. [doi: 10.2753/jec1086-4415160204]

Abbreviations

AVE: average variance extracted
DS: disease severity
EDU: education level
IS: information sensitivity
OIS: online information support
PC: privacy concerns
WSI: willingness to share information

Edited by G Eysenbach; submitted 08.10.19; peer-reviewed by X Zhang, F Meng, D Huang; comments to author 12.11.19; revised version received 20.11.19; accepted 31.12.19; published 01.04.20.

Please cite as:
Zhu P, Shen J, Xu M
Patients' Willingness to Share Information in Online Patient Communities: Questionnaire Study
J Med Internet Res 2020;22(4):e16546
URL: https://www.jmir.org/2020/4/e16546
doi:10.2196/16546
PMID:32234698

Original Paper

The Relationship Between Engagement in Online Support Groups and Social Isolation Among Military Caregivers: Longitudinal Questionnaire Study

Thomas Trail[1], PhD; Esther Friedman[2], PhD; Carolyn M Rutter[2], PhD; Terri Tanielian[1], MA

[1]RAND Corporation, Arlington, VA, United States

[2]RAND Corporation, Santa Monica, CA, United States

Corresponding Author:
Thomas Trail, PhD
RAND Corporation
1200 South Hayes St
Arlington, VA, 22202
United States
Phone: 1 703 413 1100
Email: ttrail@rand.org

Abstract

Background: There is a lack of research on the effectiveness of online peer support groups for reducing social isolation and depressive symptoms among caregivers, and previous research has mixed results.

Objective: This study aimed to test whether military caregivers who joined a new online peer support community or engaged with an existing online community experienced decreased perceived social isolation and improved depressive symptoms over 6 months.

Methods: We conducted a longitudinal study of 212 military caregivers who had newly joined an online community and those who were members of other military caregiver groups. Multiple indicators of perceived social isolation and depressive symptoms were assessed at baseline and at 3 and 6 months.

Results: Compared with caregivers in the comparison group, caregivers who joined the new group experienced less perceived social isolation at 3 months (eg, number of caregivers in social network [unstandardized regression coefficients] *b*=0.49, SE 0.19, 95% CI 0.87 to 0.02), but this effect did not persist at 6 months. Those who engaged more with new or existing groups experienced less perceived social isolation over time (eg, number of caregivers in social network *b*=0.18, SE 0.06, 95% CI 0.02 to 0.27), and this relationship was mediated by increased interactions with other military caregivers (95% CI 0.0046 to 0.0961). Engagement with an online group was not associated with improvements in depressive symptoms.

Conclusions: Online communities might help reduce social isolation when members engage with the group, but more intensive treatment is needed to improve depressive symptoms.

(J Med Internet Res 2020;22(4):e16423) doi:10.2196/16423

KEYWORDS

caregivers; family caregivers; social isolation; loneliness; depression; social support; online intervention; self-help groups; veterans health

Introduction

Background

Serving as a caregiver to a relative or friend who is ill or wounded can be an isolating experience. Time spent caregiving can be significant [1] and may make it difficult to engage in social activities outside the home [2,3]. This is particularly true of those caring for ill or wounded service members or veterans: more than 70% of these caregivers reported feeling isolated, and approximately half of the caregivers reported hesitating to take their care recipient outside the home for fear of what might happen, especially when their care recipient experienced posttraumatic stress disorder (PTSD), anxiety, or depression [4]. These caregivers report experiencing a lack of connection to both the military and civilian community [5] and report wanting a forum to seek out social support and reduce social isolation [6]. An estimated 5.5 million people in the United

States serve as informal military caregivers—family and friends who provide primarily unpaid or informal care to ill or wounded service members or veterans—and those caring for veterans of the post-9/11 conflicts are potentially facing years of caring for a young veteran with a variety of physical and mental health problems [7].

Support groups have sprung up online to provide military caregivers with an opportunity to interact with their peers and receive support and guidance (eg, the Military Veteran Caregiver Network [MVCN], Hidden Heroes, and Facebook groups). However, it is unclear how well these online groups function in terms of reducing social isolation and improving mental health among military caregivers. This study explored this question by focusing on members of a new online community for military caregivers compared with caregivers who were members of existing online support groups.

Social Isolation and Psychological Health

Extensive research and theory suggest that people have a need to form meaningful social connections with others—that they have a need to belong [8]—so much so that when those connections are broken or nonexistent, people feel a sense of grief [8,9]. We define social isolation as a tangible or perceived deficit in social connections with other people. A tangible deficit in social connections is manifested in the structure or characteristics of an individual's social network of connections [10]. A perceived deficit in social connections is manifested in feelings of loneliness [10] or a lack of perceived connection to others sharing similar group membership [11-13]. A connection to others can be made in person, over the phone, or online, and the connection can be with friends, family members, or relative strangers who share a common group with the individual.

Research that examines tangible deficits in social connections considers someone to be socially isolated if they live alone, have few people in their social network, or have infrequent contact with others [10]. A recent review found that people with larger and more diverse social networks (eg, networks of friends outside of their family) have lower levels of depression [14].

People can also be socially isolated if they perceive that they have a deficit in social connections with others when they feel lonely [10,15]. Feelings of loneliness are associated with the quality, rather than the quantity, of relationships; people can feel alone even when they are surrounded by others with whom they often interact [10]. The number of social ties and feelings of loneliness are not highly correlated [16,17] and often have independent effects on mental health outcomes [18]. Numerous studies have found an association between loneliness and depression, independent of social network ties [10,18].

Finally, social connections cannot only be fulfilled through tangible or perceived connections with individuals but also through perceived connections with a group [13]. Factor analyses of the University of California, Los Angeles (UCLA) Loneliness Scale have revealed that the lack of perceived connections to a group is a distinct form of social isolation [11,19]. The lack of perceived connection to a group, also called collective connectedness, is a relatively unexplored concept in the social isolation literature [19], although some previous research has considered deficits in social identity with a group as a form of social isolation [20,21]. Greater identification with a group is typically associated with better psychological health [22], including decreased depressive symptoms [23]. Thus, greater collective connectedness with a *group* of people who share similar experiences or identities (eg, as military caregivers) could represent a decrease in social isolation and may potentially foster psychological well-being.

Online Peer Support Communities

Online forums have become popular in recent years as a way to allow peers in similar circumstances or dealing with common hardships to interact with one another, form personal bonds, and provide support to one another. One advantage of these online communities is the ability of members to elicit support from others who are dealing with similar issues but who are geographically dispersed [24,25]. Another advantage is accessibility, as users can visit online communities at their convenience. Much of a caregiver's time is occupied with caring for a friend or loved one, and caregivers may have difficulty finding someone to provide care in their absence. Accordingly, participation in in-person support groups that require caregivers to leave their home for an hour or more at a time may be impractical. Thus, providing caregivers with access to peers via online communities has the potential to positively affect a variety of outcomes, including reducing social isolation and symptoms of depression.

However, there is a lack of research on the effectiveness of online peer support groups, and previous research has shown mixed results for improving depressive symptoms [26-30]. One review of many different types of support interventions specifically for caregivers (eg, group and individual, online, and in person) found no significant changes in quality of life, caregiver burden, depressive symptoms, or other outcomes [28]. A recent systematic review of online support interventions for caregivers found that those incorporating both peer support and professional help were most effective in improving caregiver mental health [29]. In general, although online peer support groups for caregivers of people with Alzheimer disease have shown some promise [30,31], little is known about the effectiveness of online peer support interventions for caregivers of people with mental health conditions such as PTSD [32].

Furthermore, some research suggests that active engagement with the online community (eg, reading, posting, and responding to posts to the website) may be a necessary element of the success of online support communities, but other research suggests that more passive engagement is sufficient (eg, simply visiting a site). For example, a recent study of a support group for patients with anxiety and depression found that only the most actively engaged participants experienced improvements in anxiety compared with participants who did not visit the community at all [33]. However, several studies have found that just visiting the community but not posting (ie, *lurking*) is associated with similar improvements in social outcomes compared with those who post to a community website [34,35]. To our knowledge, the role of engagement in online peer support groups among caregiver population has not been investigated in prior research. This study explored the role of engagement

in the online community in decreasing social isolation and improving depressive symptoms. We examined both passive (ie, visiting the community website and spending more time on the site) and active (ie, posting to the online forum) engagement with the community.

Military Veteran Caregiver Network

Military caregivers may especially need to connect with others who face similar challenges related to injuries and trauma experienced by service members and seek specific knowledge needed to negotiate veteran health care benefits. Thus, several online communities have been established to connect military caregivers with one another. One of these, MVCN, was established in 2016 to "provide military and veteran caregivers with peer support to reduce their isolation and increase their sense of connectedness, engagement, hopefulness, wellness as well as their knowledge and skills" [36]. MVCN allows members to post and read comments or questions to a community forum moderated by program staff; join groups organized around specific topics; exchange information about relevant resources through direct messaging; and attend webchats, webinars, and monthly question and answer calls about featured topics of interest to military caregivers. At the time of this study, web-based activities were moderated by trained paid staff who were also peer military caregivers, and there were around 1200 military caregiver members. According to interviews with MVCN staff, trained moderators and well-organized content are the features that distinguish MVCN from similar online support groups. Vaughan et al [37] give a more complete analysis of the experiences of MVCN participants.

This Study

This study investigates whether joining an online support community helps reduce social isolation and improves depressive symptoms among caregivers. Joining an online community should increase interactions with their military caregiver peers, which should lead to improvements across several other measures of social isolation. Improvements in social isolation should lead to improvements in depressive symptoms.

To assess changes over time among new MVCN members, control for events outside of the study context (eg, changes in policy that would affect all military caregivers) and also control for systematic effects of participating in the study (eg, natural changes in ratings over time), we compared changes over time among new members of MVCN to a comparison group. We chose the comparison group to match the MVCN group along several dimensions. We wanted comparison group members to be similar to the members in the MVCN group, in that both groups comprised caregivers of ill or wounded service members or veterans. In addition, according to interviews with MVCN staff, most members of MVCN were joined through their association with other military caregiver organizations (eg, Wounded Warrior Project). Thus, we recruited caregivers for the comparison group who were members of other military caregiver organizations in an online presence but had not yet been recruited for membership in MVCN (eg, Hidden Heroes and Operation Family Caregiver). This provided a comparison group of military caregivers who were similarly involved in military caregiver issues and who were interested and willing to become a member of an online military caregiver group.

To test whether engagement in online support communities was associated with improved outcomes over time, we conducted a longitudinal study that included military caregivers who had newly joined the MVCN online community (the MVCN group) and military caregivers who had not joined MVCN but were members of other military caregiver groups (the comparison group). Study participants were assessed on multiple indicators of social isolation along with depressive symptoms at baseline and at 3 and 6 months following baseline. Data from the baseline survey were analyzed in a previous study describing the characteristics of military caregivers who join online support communities [38]. Data from the baseline and 6-month surveys, along with information from focus groups, were examined in a previous study describing the experiences of MVCN group members [37]. We sought to answer three research questions and offer tentative hypotheses for each question. However, the study is ultimately more descriptive than focused on hypothesis testing, and results were interpreted holistically across findings rather than based on individual tests of statistical significance.

Research Question 1: Does Joining a New Online Community Increase Connections With Other Military Caregivers and Decrease Social Isolation and Depressive Symptoms Over Time?

We hypothesized that those newly joining an online community (MVCN) would experience a greater increase in connections with other military caregivers, increase in perceived collective connectedness, decrease in perceived social isolation (ie, loneliness), and decrease in depressive symptoms over time relative to a comparison group.

Research Question 2: Does Engagement With Online Military Caregiver Communities, Including the Military Veteran Caregiver Network, Decrease Social Isolation and Depressive Symptoms Over Time?

On the basis of prior research detailing the relationship between social isolation and depressive symptoms, we hypothesized that engagement with an online community would function as a *dose* effect, with greater engagement leading to improvements in social isolation and depressive symptoms over time. As other online communities offer similar opportunities as MVCN for engagement with other military caregivers, we expected that the effect of engagement would be similar for both groups.

Research Question 3: Are the Significant Relationships Between Engagement With Online Military Caregiver Communities and Outcomes Mediated by Increased Interactions With Peers?

Although some research suggests that participating in online peer support groups can lead to improved social support and decreased depression, the role of online interactions with peers in decreasing social isolation is unclear. It is possible that online interactions are not perceived as equivalent to in-person interactions, so online interactions with caregivers may not be perceived as increasing one's social network or be related to

feelings of loneliness or collective connectedness. In addition, people can feel lonely even while interacting with others online or in person, so increased interactions with other caregivers may not mediate decreased feelings of loneliness over time. Furthermore, what does it mean to *interact* with others online? Does viewing others' posts and replies in an online forum constitute interacting with those people? Is it roughly equivalent to being with a group of peers who are holding a conversation, but not contributing to the conversation oneself? Given our theoretical model, just being in the group but not contributing should be enough to decrease feelings of social isolation, so perhaps *perceived interactions* is a more appropriate term for this phenomenon. Indeed, research on social media use has provided mixed results as to whether engaging in online social groups decreases or actually *increases* social isolation and loneliness [39]. We expected that significant relationships between passive and active engagement with online military caregiver communities and social isolation or depressive symptoms would be mediated by increased perceived interactions with other military caregivers. From a theoretical perspective, we considered perceived interaction with peers as the mechanism that accounts for the relationship between engagement and social isolation or depressive symptoms. Therefore, where there are significant relationships between engagement and interactions with peers, we tested a mediation model of the indirect effect of engagement on social isolation or depressive symptoms through the relationship with interaction with peers. A significant mediation effect would provide additional evidence that the benefits of online communities are because of the increase in participants' perceived ability to connect with their peers.

Methods

Participants

To be eligible for this study, individuals had to be adult (>18 years) military caregivers, defined here as someone who provides unpaid care and assistance for or manages the care of, a current or former member of the US military, National Guard, or Reserves who has an illness, injury, or condition for which they require outside support.

Participants in the MVCN group were recruited when they joined MVCN, and all MVCN applicants were verified to be military caregivers by virtue of their membership in other military caregiver organizations or via staff review of applicants' documentation. New members received an email welcoming them to the group, which included text explaining the study and inviting them to participate, followed by a link to a screener survey. Of those caregivers who joined MVCN during the study recruitment period (September 2016 to February 2017), 62.0% (323/521) took the screener survey. The comparison group participants were recruited based on their membership in military caregiver organizations other than MVCN and included Hidden Heroes, Operation Family Caregiver at the Rosalynn Carter Institute for Caregiving, the Caregiving Action Network, Blue Star Families, and the American Legion Auxiliary. According to participants, most of the organizations that comparison group members belonged to offered similar activities as MVCN, including a library of information and resources, webinars, interest groups, webchats, the availability of peer mentors, and the ability to serve as peer moderators for the group. Between October 2016 and April 2017, potential comparison group participants either received an email from their member organization recruiting them for the study or the study invitation was posted on a restricted-access Facebook page administered by the group. Other than text referencing the member organization, the text of recruiting materials for comparison group participants was identical.

The screener survey included questions verifying that the participant was an unpaid caregiver for an ill or wounded service member or veteran and for those recruited for the comparison group, whether they were already a member of MVCN. Participants who successfully screened into the study then read the consent form to inform them of the purpose of the study and the survey methodology. If they indicated consent by clicking the *yes* box agreeing to participate in the study, they entered their email address and were immediately sent an email that included an individualized link to the baseline survey. All procedures were conducted in compliance with the RAND Corporation's Human Subjects Protection Committee. Completion of the baseline survey was required for continued participation in the study. Owing to the study design, the comparison group participants were not restricted from joining MVCN during the study. As one goal of the study was to examine new members of online groups, and all participants completed the same baseline survey, a decision was made a priori to include those caregivers who were initially in the comparison group but joined MVCN between the baseline and 3-month surveys in the MVCN group (n=13, total). We asked these new members when they had joined MVCN. Of those who completed all 3 survey waves and were included in the current analysis (n=10), only 1 participant had joined the month before the survey; 1 participant had joined between 1 and 2 months before the survey, and 8 participants had joined between 2 and 3 months before the 3-month survey. This suggests that the majority of those who joined MVCN between baseline and the 3-month survey had similar exposure to MVCN as those who joined at baseline. An additional 9 comparison group participants joined MVCN between the 3- and 6-month surveys. As we were comparing changes from baseline to 3 and 6 months, those 9 participants were analyzed as members of the comparison group. Importantly, as MVCN is a closed group, only those who officially joined MVCN had access to the MVCN online forum and other resources provided to MVCN members. Those comparison group members who did not join MVCN could not access these resources.

Only participants who completed all 3 waves of surveys were included in the final analytic sample (212/345, 61% who completed the baseline assessment). We had powered the study to detect an average difference in depressive symptom score changes from baseline to 6 months of 1.25 points between the MVCN and comparison groups, assuming a correlation of 0.50 between measurements. This suggested a final sample of 58 participants per group. A post hoc power analysis using the smaller of the two groups (n=44) suggested an achieved power of 0.80 to detect a small effect between groups over time (ie,

where regression partial η^2=0.02). Participants were given a US $10 electronic gift card for completing the baseline survey, a US $10 gift card for completing the 3-month survey, and a US $20 gift card for completing the 6-month survey.

Measures

Participants completed the following measures at baseline, 3 months, and 6 months.

Measure of Interactions With Peers

We assessed participants' frequency of interacting with other military caregivers using a single item constructed for this study: "How often do you interact with other military caregivers, either in person or online (for example, texting, messaging, responding to social media posts)?" Response options included *I have never interacted with other military caregivers*, *once or twice a year or less*, *every few months*, *once a month*, *two or three times a month*, *once a week*, *a few times a week*, or *daily*, coded 0 to 7, respectively.

Social Isolation Measures

We assessed the number of military caregivers in the participant's social network using a measure that has been shown to be associated with psychological distress in past research [40]. Participants were asked to "Think of all the people you know, who know you, and with whom you have had regular contact in the past 6 months. This contact could be face-to-face, by phone or mail, or on the internet" and to rate the number of people they know across several different categories, including "People who are also caregivers for service members or veterans." Response categories included *none*, *1-2*, *3-4*, *5-10*, *11-20*, or *21 or more*, coded 1 to 6, respectively.

Participants' level of collective connectedness as a military caregiver was assessed using a 3-item measure of group identity centrality [41]. Identity centrality assesses the salience of group membership in one's life and the extent to which group membership is a core part of one's identity [41,42] as well as their sense of belongingness and attachment to other group members (ie, their social connectedness to other group members [20,21]). Participants were asked to rate their agreement with the following 3 items using a 5-point scale (from *strongly disagree* to *strongly agree*): "In general, being a military caregiver is an important part of my self-image," "I have a strong sense of belonging to the military caregiver community," "I have a strong attachment to other military caregivers" (Cronbach alpha=.86 at baseline).

Loneliness was measured using the 3-item short scale for measuring loneliness [17], which has been used in the National Social Life, Health, and Aging Project [43]. Participants were asked to rate how often they feel they *lack companionship*, *feel left out*, and *feel isolated* using a 4-point scale (*hardly ever*, *some of the time*, *most of the time*, or *always*). The average of the 3 items was computed to form the loneliness scale (Cronbach alpha=.89 at baseline). This measure was chosen because of its broad applicability to online relationships, whereas other loneliness measures have items that are more suitable for assessing in-person or family relationships. For example, the UCLA Loneliness Scale refers to *being alone* and feeling isolated from *those around you* [11], which could be interpreted as applying to in-person relationships only. The Social and Emotional Loneliness Scale for Adults assesses relationships with friends, family, and romantic partners [44], which are not relevant categories for online communities.

Measure of Depressive Symptoms

Depressive symptom severity was measured using the 8-item version of the Patient Health Questionnaire (PHQ-8), which is a clinically validated measure of depressive symptoms based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria for depressive disorders [45,46]. Participants rated the extent to which they had been bothered by 8 symptoms of depression in the past 2 weeks (eg, *having little interest or pleasure in doing things*) using a 4-point scale, ranging from 0 (*not at all*) to 3 (*nearly every day*). Scores were summed to create an index of depressive symptoms ranging from 0 to 24 (Cronbach alpha=.90 at baseline).

Engagement in Online Community Measures

In the 3-month survey, participants who were members of MVCN were asked a series of questions about their engagement with the MVCN community. Participants who were not members of MVCN (comparison group participants) were asked to name the online community of caregivers, which they *engage with the most (spend time at the website, post comments, attend webinars, etc)*. The same questions assessing engagement were asked of the group named by each participant.

Passive engagement with the online community was assessed by measuring the frequency of visits to the website and the time spent on the website. The frequency of visits to the community website was measured via 1 item: *Since joining [group], how often have you visited the [group] website?* Where *group* was either *MVCN network* or the online group with which comparison group participants engaged the most. Response options and numeric codes were *every day* (6), *almost every day* (5), *two or three times a week* (4),*once a week* (3), *two or three times a month* (2), *once a month or less* (1), or *I have not visited the website for [group] since joining this group* (0). The length of time spent on the site was measured by asking, "When you visit, how much time do you usually spend on the [*group*] website?" Where *group* was either *MVCN network* or the online group with which comparison group participants engaged the most. Response options and numeric codes were *less than 10 min* (1), *between 10 and 20 min* (2), *20 to 30 min* (3), *30 min to an hour* (4), or *an hour or more* (5).

Active engagement with the online community via posts to the online forum was assessed by asking, "Have you posted any comments, questions, or links to the [*group*]? Where *group* was either *MVCN network* or the online group with which comparison group participants engaged the most. Their responses were coded 1 if *yes* and 0 if *no*.

Covariate Measures

We examined several potential covariates for inclusion in the analysis, including caregiver age, gender, race, level of education, household income, hours spent providing care in a typical week, caregiver's relationship with the care recipient,

whether the care recipient served during the pre- or post-9/11 period, the number of activities of daily living and instrumental activities of daily living that the care recipient needed help performing, membership in other structured support groups, veteran's age, years spent caring for the care recipient, and whether the caregiver lived in a metropolitan area. For each of these, we examined whether the variable was significantly related to MVCN membership, any of the engagement variables, or attrition from the study at 3 and 6 months. *t* tests, chi-square tests, and regression analyses were used to test for differences, as appropriate. A conservative criterion of $P<.10$ was used in these analyses. We found that MVCN and comparison group participants differed in veteran care recipient age (mean 40.2, SD 10.6 years for MVCN and mean 45.4, SD 18.3 years for comparison care recipients). No other differences emerged between the MVCN and comparison group participants. Being a member of another online or in-person support group was associated with more visits to the MVCN or comparison group community website. In addition, males spent more time on the respective community website, as did those caregivers who helped their care recipients with more tasks and those who lived outside of metropolitan areas. Those caring for older veterans, living outside of metropolitan areas, and who were members of another support group were more likely to discontinue participation in the study over time. Thus, we controlled for these 3 variables plus caregiver gender, membership in other online support groups, and number of caregiver tasks in all analyses. Note that all analyses reported below replicate when these covariates were not included in the models. Finally, to distinguish the effects of visits to the website from time spent on the site and posts to the forum, we controlled for any visits to the website (yes/no) in the analysis of these variables.

Analytic Approach

Our analytic approach focused on assessing changes in outcomes from the baseline assessment to the 3- and 6-month assessments. We conducted two separate sets of models using multiple regression: one set of models regressed 3-month outcomes on baseline predictors controlling for baseline levels of outcomes and covariates, and the other set regressed 6-month outcomes on baseline predictors controlling for baseline levels of outcomes and covariates. For each set of models, to answer research question 1, we first analyzed changes from baseline for new members of an online community (MVCN) in comparison with those who were already members of one or more online communities other than MVCN (comparison group). To answer research question 2, we used cross-lagged models to examine changes in outcomes from baseline to 6 months as a function of engagement with the online community at 3 months (ie, we added a measure of engagement at 3 months to the 6-month models examined in research question 1). We also assessed whether this relationship differed for MVCN and comparison participants, based on an interaction between MVCN membership (vs comparison group) and 3-month engagement. Cross-lagged models allow for explanatory predictions of present outcomes based on past behaviors and are thus more likely to reflect causal effects than are regressions using cross-sectional measures. Although we conducted statistical tests for each model and included the associated *P* values for the tests, we provided 95% CIs for each result and interpreted our findings based on these CIs and the pattern of results across findings, rather than based on the statistical significance of any one test.

Finally, for research question 3, we conducted mediational analyses for models where engagement significantly predicted both interactions with other military caregivers and one or more outcomes. Engagement at 3 months was entered as the predictor variable, with changes in the frequency of interactions with other military caregivers from baseline to 6 months serving as the mediator. A difference score was calculated for changes in interactions from baseline to 6 months, and we included the baseline measure of interactions as a covariate to control for the possibility that the extent of changes in interactions was driven by baseline levels of interactions. Outcomes were measured at month 6, and changes in outcomes were assessed by including baseline levels of each outcome in the models. Mediational analyses were conducted in SAS 9.3 (SAS Institute Inc) using the INDIRECT macro, which uses asymptotic bootstrapping calculations to test for the significance of indirect effects [47]. The INDIRECT macro generated 95% bias-corrected SEs and CIs using 5000 bootstrapped samples, and mediation models were considered significant when the CI for the estimation of the indirect effect did not contain 0.

Results

Participant Characteristics

Of the 212 participants who completed all 3 waves of the study, 199 (93.9%) were female, 166 (78.3%) were non-Hispanic white, 23 (10.8%) were Hispanic, and 23 (10.8%) were in other racial/ethnic groups. The majority of caregivers were spouses or partners of the veteran for whom they provided care (189/212, 89.2%), and the remaining caregivers were mostly other family members (1 participant was a friend or neighbor of the veteran). Most caregivers were under the age of 40 years (124/212, 58.5%), with an additional 74 (34.9%) aged between 40 and 59 years. The care recipients were mostly male (188/212, 88.7%), with 125 (59.0%) serving during the post-9/11 period. Almost all veterans had been diagnosed with one or more physical health conditions (203/212, 95.8%), with 182 (85.8%) diagnosed with more than one physical health problem. In addition, 87.7% (186/211) of veterans had been diagnosed with one or more psychological health conditions, with 174 (82.1%) veterans diagnosed with PTSD. We compared those who completed all 3 waves of the survey with those who did not complete all surveys on baseline-level demographic characteristics and all variables included in this study and found no significant differences between groups. In comparison with the overall population of military caregivers studied by Ramchand et al [7], participants in this study were more likely to be female, non-Hispanic white, married to the care recipient, and caring for a veteran with one or more psychological health conditions, particularly PTSD. It is unclear whether these differences reflect the population of members of online communities or only those who participated in this study.

Means and SDs for outcome variables (at baseline) and engagement variables (at 3 months) are shown in Table 1. Mean

scores for variables at baseline were generally near the midpoint of the range of possible scores. Missing data were uniformly low (8/212 or less). Using a cutoff score of 10 or higher to indicate probable major depressive disorder on the PHQ-8 [45,46], 59.4% (126/212) of participants screened positive for probable depression at baseline, which is a higher proportion than that found in the general military caregiver population [7]. One statistically significant difference between MVCN and comparison group participants emerged on outcomes measured at baseline: MVCN participants reported more frequent interactions with other military caregivers than did comparison group members (2-tailed t_{210}=2.08; *P*=.04).

Among engagement variables measured at 3 months, 25.8% (42/163) of MVCN and 23.3% (10/43) of comparison group participants indicated that they had not visited the community website since baseline (χ^2_1=0.1; *P*=.74), and the comparison group participants visited the website more often than MVCN participants (2-tailed t_{204}=3.16; *P*=.002). Among those who visited their community's website, the average time spent on the website was around 10 and 20 min (coded as *2*). MVCN group participants were also less likely to have posted to the site than were comparison participants (χ^2_1=26.1; *P*<.001).

Table 1. Means and SDs for outcome and engagement variables by group.

Variable	Range	MVCN[a]			Comparison group		
		Baseline	3 months	6 months	Baseline	3 months	6 months
Frequency of interacting with military caregivers, mean (SD)[b]	0-7	5.08 (2.47)	5.17 (2.41)	5.17 (2.25)	4.20 (2.61)	4.02 (2.50)	4.21 (2.40)
Military caregivers in network, mean (SD)[b]	1-6	2.48 (1.50)	2.51 (1.47)	2.50 (1.49)	2.32 (1.49)	1.95 (1.08)	2.09 (1.17)
Collective connectedness with military caregivers, mean (SD)[b]	1-5	2.98 (1.20)	3.02 (1.18)	3.05 (1.20)	2.86 (1.07)	2.72 (1.00)	2.61 (1.18)
Loneliness, mean (SD)[b]	3-12	8.01 (2.36)	7.70 (2.51)	7.90 (2.33)	7.36 (2.62)	7.89 (2.39)	7.21 (2.64)
Depression severity PHQ-8[c], mean (SD)[b]	0-24	11.35 (6.66)	10.29 (6.51)	10.25 (6.29)	9.89 (5.84)	10.02 (6.10)	9.21 (6.68)
Frequency of visiting online community, mean (SD)[b]	0-6	N/A[d]	1.41 (1.32)	N/A	N/A	2.23 (2.11)	N/A
Average amount of time on site per visit, mean (SD)[b]	0-5	N/A	1.61 (1.30)	N/A	N/A	1.33 (1.15)	N/A
Posted to site, n (%)	N/A	N/A	27 (16.6)	N/A	N/A	24 (55.8)	N/A

[a]MVCN: Military Veteran Caregiver Network.

[b]Means and SDs are unadjusted for other variables in the regression models.

[c]PHQ-8: 8-item patient health questionnaire.

[d]N/A: not applicable.

Results for Research Question 1: Does Joining a New Online Community Increase Connections With Other Military Caregivers and Decrease Social Isolation and Depressive Symptoms Over Time?

Comparing social isolation outcomes for MVCN and comparison group members at 3 months, controlling for baseline levels of outcomes and covariates, yielded two statistically significant findings: compared with comparison group members, MVCN members reported a higher number of military caregivers in their social network, and MVCN members reported feeling less lonely (Table 2). These differences did not persist at 6 months. The only significant difference among social isolation measures at 6 months was that MVCN members experienced relatively greater collective connectedness with other military caregivers compared with comparison group members (Table 2).

Table 2. Regression results predicting difference from baseline to 3-month outcomes and 6-month outcomes by membership in Military Veteran Caregiver Network vs comparison group (N=212).

Outcome	MVCN[a] vs comparison group[b]					
	3-month outcomes			6-month outcomes		
	b (95% CI)	SE	ΔR^2	*b* (95% CI)	SE	ΔR^2
Frequency of interacting with other military caregivers	0.49 (−0.12 to 0.97)	0.19	0.005	0.22 (−0.33 to 0.77)	0.28	0.001
Number of military caregivers in network	0.49[c] (0.11 to 0.87)	0.19	0.020	0.33 (0.04 to 0.95)	0.21	0.009
Collective connectedness with military caregivers	0.15 (−0.12 to 0.41)	0.14	0.003	0.39[d] (0.10 to 0.69)	0.15	0.017
Loneliness	−0.67[c] (−1.33 to −0.01)	0.34	0.012	0.16 (−0.57 to 0.89)	0.37	0.001
Depression severity PHQ-8[e] score	−0.93 (−2.54 to 0.69)	0.82	0.003	0.08 (−1.78 to 1.94)	0.94	0.000

[a]MVCN: Military Veteran Caregiver Network.

[b]Models controlled for caregiver gender, membership in any other support group, number of caregiving tasks, veteran age, metropolitan residence, and baseline levels of the outcome. Larger numbers indicate an increase in MVCN participants over time relative to the comparison group participants.

[c]*P*<.05.

[d]*P*<.01.

[e]PHQ-8: 8-item Patient Health Questionnaire.

Results for Research Question 2: Does Engagement With Online Military Caregiver Communities, Including Military Veteran Caregiver Network, Increase Connections With Other Military Caregivers and Decrease Social Isolation and Depressive Symptoms Over Time?

The next set of analyses examined the relationship between engagement with online caregiver communities at 3 months with changes in outcomes from baseline to 6 months. The results are shown in Table 3. The number of visits to the online community at 3 months was significantly associated with a greater increase in the frequency of interactions with other military caregivers, the number of military caregivers in one's social network, and in collective connectedness with other military caregivers from baseline to 6 months. The average time spent on the community website at 3 months was significantly associated with a decrease in feelings of loneliness and an increase in collective connectedness from baseline to 6 months. Posting to the online community was significantly associated with an increase in the number of military caregivers in one's social network from baseline to 6 months, but no other significant relationships with posting were observed. Finally, no significant interactions between any of the engagement variables and MVCN membership were observed.

Table 3. Regression results predicting baseline to 6-month changes by 3-month engagement with online community (N=212).

Outcome	Number of visits to online community[a]			Time spent in online community[a]			Post to online community forum (yes or no)[a]		
	b (95% CI)	SE	ΔR^2	*b* (95% CI)	SE	ΔR^2	*b* (95% CI)	SE	ΔR^2
Frequency of interacting with other military caregivers	0.18[b] (0.03 to 0.33)	0.08	0.012	0.08 (−0.17 to 0.33)	0.13	0.001	0.32 (−0.25 to 0.89)	0.29	0.003
Total number of military caregivers in network	0.18[c] (0.02 to 0.27)	0.06	0.030	0.02 (−0.18 to 0.22)	0.10	0.000	0.61[c] (0.18 to 1.04)	0.22	0.026
Collective connectedness with military caregivers	0.20[d] (0.12 to 0.28)	0.04	0.057	0.15[b] (0.01 to 0.29)	0.07	0.011	0.19 (−0.11 to 0.50)	0.15	0.004
Loneliness	−0.15 (−0.35 to 0.05)	0.10	0.008	−0.42[b] (−0.75 to −0.08)	0.17	0.022	−0.04 (−0.81 to 0.73)	0.39	0.000
Depression severity PHQ-8[e] score	0.23 (−0.29 to 0.74)	0.26	0.003	−0.18 (−1.05 to 0.69)	0.44	0.001	1.24 (−0.72 to 3.21)	1.00	0.005

[a]Covariates included in the models were caregiver gender, membership in any other support group, number of caregiving tasks, veteran age, metropolitan residence, membership in Military Veteran Caregiver Network, and baseline levels of the outcome. Analysis of time on site and posts to online community also included a dummy variable for any visits to the site.

[b]*P*<.05.

[c]*P*<.01.

[d]*P*<.001.

[e]PHQ-8: 8-item Patient Health Questionnaire.

Results for Research Question 3: Are the Significant Relationships Between Engagement With Online Military Caregiver Communities and Outcomes Mediated by Increased Interactions With Peers?

The number of visits to the online community at 3 months was significantly associated with changes in the frequency of interactions with other military caregivers (the hypothesized mediator) and with the total number of military caregivers included in one's social network and collective connectedness with other military caregivers. First, examining the number of military caregivers in one's social network, analysis using the INDIRECT macro in SAS [47] replicated our earlier findings: number of visits at 3 months was significantly associated with an increase in social network connections from baseline to 6 months and with changes in interactions with other military caregivers. When included in the full model, an increase in interactions with other military caregivers was also significantly associated with changes in social network connections ([unstandardized regression coefficients] b=0.22; SE 0.05; P<.001), and the relationship between the number of visits and the changes in social network connections was reduced to marginal significance (b=0.11; SE 0.06; P=.051). Analyses confirmed that the mediation model was significant (95% CI 0.0046 to 0.0961).

Next, examining changes in collective connectedness with other military caregivers, analysis using the INDIRECT macro replicated our earlier findings: number of visits at 3 months was significantly associated with an increase in collective connectedness from baseline to 6 months and with changes in interactions with other military caregivers. When included in the full model, an increase in interactions with other military caregivers was also significantly associated with changes in collective connectedness (b=0.19; SE 0.04; P<.001), but the relationship between the number of visits and changes in collective connectedness remained significant (b=0.17; SE 0.04; P<.001). Analyses confirmed that the mediation model was significant (95% CI 0.0022 to 0.0719).

Discussion

Principal Findings and Implications

This study suggests that relative to a comparison group, military caregivers joining a new online community such as MVCN experience less social isolation and loneliness over time, that the benefits of membership in new or existing online caregiver communities are greater for those who engage more with the community, and that the relationships between engagement and social isolation are mediated by an increase in participant interactions with other military caregivers. The effects of decreased loneliness are particularly noteworthy as loneliness has been associated with an increased risk of mortality and poor mental health across a variety of indicators [10]. However, our results did not demonstrate that engagement with online communities was associated with improvements in depressive symptoms over time.

Our findings suggest that joining a new group such as MVCN may have a short-term impact on social isolation in terms of increasing the number of military caregivers in one's social network and decreasing feelings of loneliness. These gains do not appear to continue at 6 months, possibly because new MVCN members formed new connections with peers that reached saturation at 3 months. MVCN members' collective connectedness with other military caregivers showed an increase at 6 months, relative to comparison group members. As MVCN members were recruited right when they joined the group, the MVCN community was a new environment for them to meet and form bonds with other military caregivers. Comparison group members were recruited via groups to which they already belonged, so their ties to their existing groups were more established, and their collective connectedness with other caregivers was possibly more formed. Thus, the effects of joining MVCN likely reflect engagement in a new environment with presumably new people who share a similar identity and similar challenges. It is perhaps not surprising that those joining a new group experienced reduced social isolation—it demonstrates that the group is doing what it was primarily designed to do for military caregivers.

Joining MVCN was not associated with improvements in depressive symptoms. MVCN participants scored relatively high on the PHQ-8, with 61.1% (102/167) having a PHQ-8 score of 10 or above at baseline, suggesting probable major depressive disorder. A review of web-based interventions for caregivers revealed that these interventions show some promise for reducing depressive symptoms [48], and recent research evaluating the impact of an in-person support program for military caregivers demonstrated improvements in depressive symptoms among participants [49]. It is possible that joining a new online support group is not enough of an intervention on its own to reduce depressive symptoms, and these studies suggest that more intensive online or in-person interventions might be needed to successfully treat depression among military caregivers.

Engagement in online communities across the MVCN and comparison groups was associated with a decrease in social isolation at 6 months. Cross-lagged models predicting 6-month outcomes from engagement at 3 months revealed that more visits to the community at 3 months were associated with greater gains in the number of military caregivers in one's social network and increased collective connectedness with other military caregivers at 6 months. Furthermore, both these relationships were mediated by an increase in participant interactions with other military caregivers, suggesting that simply visiting an online community of peers more often can increase interactions with peers, which is then associated with reduced social isolation. In addition, spending more time on the community website was associated with increased collective connectedness and decreased feelings of loneliness at 6 months. Posting to online community forums was only associated with gains in social network connections with military caregivers at 6 months. Thus, across the three indicators of engagement with an online community of peers, we found that increased engagement was associated with decreased social isolation over time, and more active engagement in the forums through posting was not associated with additional improvements in loneliness or social network size. Although engagement was not associated

with a decrease in depression, our results support the theory that engaging with online social networks might help reduce social isolation, particularly when they enable users to make meaningful connections with others [39], which suggests that programs implementing online support groups should use strategies to encourage members to visit and spend time on the community website, at a minimum.

One aspect of the MVCN community forum that might have affected engagement with the group was the active role that online forum moderators play in that group. Users prefer online support groups that are facilitated by trained peer moderators that provide trustworthy information [50]. The MVCN community forum is facilitated by a team of trained peer moderators, and research using focus groups of MVCN members and initial survey data from this study found that the information provided by the website and forum moderators was trustworthy compared with other online forums [37]. It is possible that active engagement by trained peer moderators served to improve the impact of engagement with MVCN on members' well-being. However, this research also found that participation in MVCN was generally low and mostly passive [37], which suggests that the presence of trained peer moderators could also have decreased more active engagement with the online forum (eg, participants might not respond to posts requesting information if they think that a trained moderator will respond instead). This is not necessarily a bad trade-off because the engagement of peer facilitators might make the forum more trustworthy and decrease the spread of rumors and misinformation. Additional research is needed to determine the beneficial impact that trained peer moderators might have on participants in online support groups.

Strengths and Limitations

Although this study has several strengths, including 3 waves of data, several measures of social isolation, and the ability to conduct causal analyses using cross-lagged models, it also has several limitations. First, participants were not randomly assigned to groups, so MVCN participants could have systematically differed from comparison group participants in ways that affected our results, including their motivation or ability to engage with their online community and their proclivity for change over time. Furthermore, we did not control group membership, so comparison group participants were able to join MVCN during the study, and 10 comparison group participants included in this study had joined MVCN at the 3-month survey. We made an a priori decision to include these participants in the MVCN group because their data at 3 and 6 months would reflect similar experiences with MVCN as those who had joined MVCN at baseline. Most (8/10) of these new members had joined MVCN between 2 and 3 months before the 3-month survey. However, an additional 9 comparison group participants indicated that they had joined MVCN at the 6-month survey and were counted as comparison group participants for the purposes of this study. A post hoc analysis excluding these participants from the study did not substantially change the results, although a significant effect of the MVCN group on the number of military caregivers in one's social network emerged at the 6-month assessment, with MVCN members reporting increased numbers of military caregivers in their social network from baseline to 6 months relative to comparison group members (b=0.49; SE 0.23; 2-tailed t_{181}=2.15; P=.03; ΔR^2=0.017). This finding suggests that the decision to keep new MVCN members at 6 months in the comparison group made our results comparing the two groups more conservative, rather than biasing the results toward significant differences. However, in combination with the lack of random assignment, the results distinguishing the MVCN participants from the comparison group participants need to be interpreted with caution. In addition, we did not control engagement in online communities, so individual differences in propensity to engage with online communities could have driven the effects reported here. Finally, engagement was self-reported, so individual biases in reporting engagement could have affected the results. Using a within-person design and examining cross-lagged effects helps correct for these biases but does not completely rule them out. Future research should randomly assign participants to groups and manipulate engagement with online communities to better assess causality.

Conclusions

Although helping military caregivers to overcome feelings of social isolation is an important contribution of online support groups such as MVCN, our study indicates that more intensive efforts may be needed to improve depressive symptoms. This will likely include opportunities that promote treatment seeking and offer more structured peer or professional support through online [51] or in-person connections.

Acknowledgments

The authors wish to acknowledge the staff at the MVCN for their support and assistance in the recruitment of study participants and the caregivers who participated in the study. This study was supported by a research grant from the Bristol-Myers Squibb Foundation (BMSF; BMSF_10.01.15_) awarded to TTr and TTa.

Conflicts of Interest

None declared.

References

1. van den Berg B, Spauwen P. Measurement of informal care: an empirical study into the valid measurement of time spent on informal caregiving. Health Econ 2006 May;15(5):447-460. [doi: 10.1002/hec.1075] [Medline: 16389664]

2. Hoyert DL, Seltzer MM. Factors related to the well-being and life activities of family caregivers. J Fam Relat 1992 Jan;41(1):74-81. [doi: 10.2307/585397]
3. Seltzer M, Li L. The dynamics of caregiving: transitions during a three-year prospective study. Gerontologist 2000 Apr;40(2):165-178. [doi: 10.1093/geront/40.2.165] [Medline: 10820919]
4. National Alliance for Caregiving. 2010. Caregivers of Veterans–Serving on the Homefront: Report of Study Findings URL: http://www.caregiving.org/data/2010_Caregivers_of_Veterans_FULLREPORT_WEB_FINAL.pdf [accessed 2020-01-30]
5. Voris SE, Steinkopf J. Suffering in the shadows: interviews with wives of combat veterans suffering from post-traumatic stress disorder and/or traumatic brain injury. J Marriage Fam Rev 2018 Nov 6;55(6):493-511. [doi: 10.1080/01494929.2018.1519494]
6. Tanielian T, Ramchand R, Fisher M, Sims C, Harris R, Harrell M. Military Caregivers : Cornerstones of Support for Our Nation's Wounded, Ill, and Injured Veterans. Santa Monica, CA: RAND Corporation; 2013.
7. Ramchand R, Tanielian T, Fisher M, Vaughan C, Trail T, Epley C, et al. Hidden Heroes: America's Military Caregivers. Santa Monica, CA: RAND Corporation; 2014.
8. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 1995 May;117(3):497-529. [Medline: 7777651]
9. Watt SE, Badger AJ. Effects of social belonging on homesickness: an application of the belongingness hypothesis. Pers Soc Psychol Bull 2009 Apr;35(4):516-530. [doi: 10.1177/0146167208329695] [Medline: 19193602]
10. Cacioppo JT, Cacioppo S. Social relationships and health: the toxic effects of perceived social isolation. Soc Personal Psychol Compass 2014 Feb 1;8(2):58-72 [FREE Full text] [doi: 10.1111/spc3.12087] [Medline: 24839458]
11. McWhirter BT. Factor analysis of the revised UCLA loneliness scale. Curr Psychol 1990 Mar;9(1):56-68. [doi: 10.1007/bf02686768]
12. Hawkley LC, Browne MW, Cacioppo JT. How can I connect with thee? Let me count the ways. Psychol Sci 2005 Oct;16(10):798-804. [doi: 10.1111/j.1467-9280.2005.01617.x] [Medline: 16181443]
13. Weiss R. Loneliness: The Experience of Emotional and Social Isolation. Cambridge: MIT Press; 1973.
14. Santini ZI, Koyanagi A, Tyrovolas S, Mason C, Haro JM. The association between social relationships and depression: a systematic review. J Affect Disord 2015 May 1;175:53-65. [doi: 10.1016/j.jad.2014.12.049] [Medline: 25594512]
15. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
16. Perissinotto CM, Covinsky KE. Living alone, socially isolated or lonely-what are we measuring? J Gen Intern Med 2014 Dec;29(11):1429-1431 [FREE Full text] [doi: 10.1007/s11606-014-2977-8] [Medline: 25092012]
17. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
18. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago health, aging, and social relations study. Psychol Aging 2010 Jul;25(2):453-463 [FREE Full text] [doi: 10.1037/a0017216] [Medline: 20545429]
19. Shevlin M, Murphy S, Murphy J. The latent structure of loneliness: testing competing factor models of the UCLA loneliness scale in a large adolescent sample. Assessment 2015 May;22(2):208-215. [doi: 10.1177/1073191114542596] [Medline: 25022276]
20. Lee RM, Robbins SB. Measuring belongingness: the social connectedness and the social assurance scales. J Couns Psychol 1995 Apr;42(2):232-241. [doi: 10.1037/0022-0167.42.2.232]
21. Lee RM, Robbins SB. The relationship between social connectedness and anxiety, self-esteem, and social identity. J Couns Psychol 1998 Jul;45(3):338-345. [doi: 10.1037/0022-0167.45.3.338]
22. Haslam SA, Jetten J, Postmes T, Haslam C. Social identity, health and well‐being: an emerging agenda for applied psychology. Appl Psychol 2008 Dec 31;58(1):1-23. [doi: 10.1111/j.1464-0597.2008.00379.x]
23. Cruwys T, Haslam SA, Dingle GA, Haslam C, Jetten J. Depression and social identity: an integrative review. Pers Soc Psychol Rev 2014 Aug;18(3):215-238. [doi: 10.1177/1088868314523839] [Medline: 24727974]
24. Wright K, Muhtaseb A. Personal relationships and computer-mediated support groups. In: Webb LM, Wright KB, editors. Computer-Mediated Communication in Personal Relationships. New York, USA: Peter Lang Publishing; 2010:135-155.
25. Smith-Merry J, Goggin G, Campbell A, McKenzie K, Ridout B, Baylosis C. Social connection and online engagement: insights from interviews with users of a mental health online forum. JMIR Ment Health 2019 Mar 26;6(3):e11084 [FREE Full text] [doi: 10.2196/11084] [Medline: 30912760]
26. Griffiths KM, Calear AL, Banfield M. Systematic review on internet support groups (ISGs) and depression (1): do ISGs reduce depressive symptoms? J Med Internet Res 2009 Oct 30;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
27. Powell J, Chiu T, Eysenbach G. A systematic review of networked technologies supporting carers of people with dementia. J Telemed Telecare 2008;14(3):154-156. [doi: 10.1258/jtt.2008.003018] [Medline: 18430288]

28. Dam AE, de Vugt ME, Klinkenberg IP, Verhey FR, van Boxtel MP. A systematic review of social support interventions for caregivers of people with dementia: are they doing what they promise? Maturitas 2016 Mar;85:117-130. [doi: 10.1016/j.maturitas.2015.12.008] [Medline: 26857890]
29. Guay C, Auger C, Demers L, Mortenson WB, Miller WC, Gélinas-Bronsard D, et al. Components and outcomes of internet-based interventions for caregivers of older adults: systematic review. J Med Internet Res 2017 Sep 19;19(9):e313 [FREE Full text] [doi: 10.2196/jmir.7896] [Medline: 28928109]
30. McKechnie V, Barker C, Stott J. The effectiveness of an internet support forum for carers of people with dementia: a pre-post cohort study. J Med Internet Res 2014 Mar 28;16(2):e68 [FREE Full text] [doi: 10.2196/jmir.3166] [Medline: 24583789]
31. Barbabella F, Poli A, Andréasson F, Salzmann B, Papa R, Hanson E, et al. A web-based psychosocial intervention for family caregivers of older people: results from a mixed-methods study in three European countries. JMIR Res Protoc 2016 Oct 6;5(4):e196 [FREE Full text] [doi: 10.2196/resprot.5847] [Medline: 27713113]
32. Chien W, Norman I. The effectiveness and active ingredients of mutual support groups for family caregivers of people with psychotic disorders: a literature review. Int J Nurs Stud 2009 Dec;46(12):1604-1623. [doi: 10.1016/j.ijnurstu.2009.04.003] [Medline: 19481205]
33. Geramita EM, Belnap BH, Abebe KZ, Rothenberger SD, Rotondi AJ, Rollman BL. The association between increased levels of patient engagement with an internet support group and improved mental health outcomes at 6-month follow-up: post-hoc analyses from a randomized controlled trial. J Med Internet Res 2018 Jul 17;20(7):e10402 [FREE Full text] [doi: 10.2196/10402] [Medline: 30021711]
34. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
35. Mo PK, Coulson NS. Empowering processes in online support groups among people living with HIV/AIDS: a comparative analysis of 'lurkers' and 'posters'. Comput Hum Behav 2010 Sep;26(5):1183-1193. [doi: 10.1016/j.chb.2010.03.028]
36. Red Cross. Military and Veteran Caregiver Network: Caregivers of All Eras, All Relations, Across All Locations URL: https://www.redcross.org/get-help/military-families/services-for-veterans/military-veteran-caregiver-network.html [accessed 2020-01-30]
37. Vaughan C, Trail TE, 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]
38. Friedman E, Trail T, Vaughan C, Tanielian T. Online peer support groups for family caregivers: are they reaching the caregivers with the greatest needs? J Am Med Inform Assoc 2018 Sep 1;25(9):1130-1136. [doi: 10.1093/jamia/ocy086] [Medline: 30016449]
39. Clark JL, Algoe SB, Green MC. Social network sites and well-being: the role of social connection. Curr Dir Psychol Sci 2017 Dec 19;27(1):32-37. [doi: 10.1177/0963721417730833]
40. Flórez KR, Ghosh-Dastidar MB, Beckman R, de la Haye K, Duru OK, Abraído-Lanza AF, et al. The power of place: social network characteristics, perceived neighborhood features, and psychological distress among African Americans in the historic hill district in Pittsburgh, Pennsylvania. Am J Community Psychol 2016 Sep;58(1-2):60-68 [FREE Full text] [doi: 10.1002/ajcp.12086] [Medline: 27612324]
41. Sellers RM, Rowley SA, Chavous TM, Shelton JN, Smith MA. Multidimensional inventory of black identity: a preliminary investigation of reliability and constuct validity. J Pers Soc Psychol 1997;73(4):805-815. [doi: 10.1037//0022-3514.73.4.805]
42. Cameron JE. A three-factor model of social identity. Self Identity 2004 Jul;3(3):239-262. [doi: 10.1080/13576500444000047]
43. Cornwell EY, Waite LJ. Measuring social isolation among older adults using multiple indicators from the NSHAP study. J Gerontol B Psychol Sci Soc Sci 2009 Nov;64(Suppl 1):i38-i46 [FREE Full text] [doi: 10.1093/geronb/gbp037] [Medline: 19508982]
44. DiTommaso E, Brannen C, Best LA. Measurement and validity characteristics of the short version of the social and emotional loneliness scale for adults. Educ Psychol Meas 2016 Jul 2;64(1):99-119. [doi: 10.1177/0013164403258450]
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
46. 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]
47. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [doi: 10.3758/BRM.40.3.879]
48. 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]
49. Easom LR, Wang K, Moore RH, Wang H, Bauer L. Operation family caregiver: problem-solving training for military caregivers in a community setting. J Clin Psychol 2018 Apr;74(4):536-553. [doi: 10.1002/jclp.22536] [Medline: 29139114]

50. Plinsinga ML, Besomi M, Maclachlan L, Melo L, Robbins S, Lawford BJ, et al. Exploring the characteristics and preferences for online support groups: mixed method study. J Med Internet Res 2019 Dec 3;21(12):e15987 [FREE Full text] [doi: 10.2196/15987] [Medline: 31793893]
51. Burger F, Neerincx MA, Brinkman W. Technological state of the art of electronic mental health interventions for major depressive disorder: systematic literature review. J Med Internet Res 2020 Jan 20;22(1):e12599 [FREE Full text] [doi: 10.2196/12599] [Medline: 31958063]

Abbreviations

MVCN: Military Veteran Caregiver Network
PHQ-8: 8-item Patient Health Questionnaire
PTSD: posttraumatic stress disorder
UCLA: University of California, Los Angeles

Edited by G Eysenbach; submitted 27.09.19; peer-reviewed by M Wasilewski, AC Pinto-Bruno, S Ge, K Miller, J Taylor, K Machin; comments to author 28.10.19; revised version received 31.01.20; accepted 16.02.20; published 23.04.20.

Please cite as:
Trail T, Friedman E, Rutter CM, Tanielian T
The Relationship Between Engagement in Online Support Groups and Social Isolation Among Military Caregivers: Longitudinal Questionnaire Study
J Med Internet Res 2020;22(4):e16423
URL: http://www.jmir.org/2020/4/e16423/
doi:10.2196/16423
PMID:32324141

Review

Design, Delivery, Maintenance, and Outcomes of Peer-to-Peer Online Support Groups for People With Chronic Musculoskeletal Disorders: Systematic Review

Liam R Maclachlan[1*], PhD; Kathryn Mills[2*], PhD; Belinda J Lawford[3], PhD; Thorlene Egerton[3], PhD; Jenny Setchell[1], PhD; Leanne M Hall[1], PhD; Melanie L Plinsinga[1], MSc; Manuela Besomi[1], MSc; Pek Ling Teo[3], MHSc; Jillian P Eyles[4,5], PhD; Rebecca Mellor[1], PhD; Luciano Melo[6], BSc; Sarah Robbins[4,5], PhD; Paul W Hodges[1], DSC, MedDr, PhD; David J Hunter[4,5], MD, PhD; Bill Vicenzino[1], PhD; Kim L Bennell[3], PhD

[1]The School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia

[2]Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia

[3]Centre for Health, Exercise & Sports Medicine, Department of Physiotherapy, The University of Melbourne, Melbourne, Australia

[4]Faculty of Medicine and Health, University of Sydney, Kolling Institute of Medical Research, Institute of Bone and Joint Research, Sydney, Australia

[5]Department of Rheumatology, Royal North Shore Hospital, Sydney, Australia

[6]Sax Institute, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Kathryn Mills, PhD
Faculty of Medicine and Health Sciences
Macquarie University
Physiotherapy Ground Floor, 75 Talavera Road
Macquarie Park
Sydney, 2019
Australia
Phone: 61 2 9850 6624
Email: Kathryn.mills@mq.edu.au

Abstract

Background: Online support groups (OSGs) are one way for people with chronic diseases, their family or friends, and health professionals to communicate, gain information, and provide social support. As the number of peer-to-peer OSGs for chronic musculoskeletal conditions grows, it is important to gain insight into the different designs of groups available, who is accessing them, if and how they may be effective, and what strategies are being used to implement or increase consumer engagement.

Objective: The objectives of this systematic review of people with musculoskeletal conditions were to (1) describe the design features (functions, usage options, moderation, and expert input) of peer-to-peer OSGs, (2) describe the characteristics of the individuals using peer-to-peer OSGs, (3) synthesize the evidence on outcomes of participation, and (4) identify strategies used in the delivery and maintenance of OSGs.

Methods: A search comprising terms related to the population (people with musculoskeletal disorders) and the intervention (peer-to-peer OSGs) was conducted in 6 databases. Results were filtered from 1990 (internet inception) to February 2019. Studies identified in the search were screened according to predefined eligibility criteria using a 2-step process. Quantitative studies were appraised by 2 reviewers using the Risk Of Bias In Non-Randomized Studies of Interventions tool. Qualitative studies were appraised by 2 different reviewers using the Critical Appraisal Skills Programme checklist. Extracted data were synthesized narratively.

Results: We examined 21 studies with low to moderate risk of bias. Of these studies, 13 studies included OSGs hosted on public platforms, 11 studies examined OSGs that were conducted in English, and 6 studies used moderators or peer leaders to facilitate engagement. Studies either reported the number of OSG members (n=1985 across all studies) or the number of posts (range: 223-200,000). The majority of OSG members were females who were not full-time employees and with varied levels of education. There were no randomized controlled trials measuring the efficacy of OSGs. Qualitative and quantitative studies identified empowerment, social support, self-management behavior, and health literacy as primary constructs to measure OSG efficacy.

Neutral or marginal improvement was reported in these constructs. Sharing experiences and a greater level of engagement appeared to have an important influence on OSGs efficacy. The extent to which members posted on the website influenced engagement.

Conclusions: Across a diverse range of designs, languages, included features, and delivery platforms, peer-to-peer OSGs for chronic musculoskeletal conditions attract predominantly female participants of all ages and education levels. The level of participation of a member appears to be related to their perceived benefit, health literacy, and empowerment. Future studies are needed to identify which design and maintenance strategies have superior efficacy and whether there are concomitant improvements in health outcomes for people with chronic musculoskeletal conditions resulting from participation in OSGs.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42018090326; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018090326

(J Med Internet Res 2020;22(4):e15822) doi:10.2196/15822

KEYWORDS

social support; musculoskeletal diseases; online social networking; empowerment

Introduction

Background

Chronic musculoskeletal disorders are highly prevalent [1], the leading cause of nonfatal disease burden [2], and include conditions (such as low back pain) that are the leading cause of disability internationally [3]. Musculoskeletal disorders disrupt daily living and account for a large proportion of lost productivity in the workplace [4]. Given that there is no cure for many chronic musculoskeletal disorders, long-term self-management is a core recommendation of clinical guidelines [5,6]. Central to effective long-term management is patient education and advice relating to medication, therapeutic exercise, general physical activity, weight loss (if appropriate), and potentially beneficial physical and psychological treatments [6-8].

Another key factor in the management of musculoskeletal disorders is social support, as it may positively influence health behaviors susceptible to social influence [9]. Social support may also buffer the negative impact of low health literacy [10]. Both are essential in negotiating health care systems [11] and may impact health outcomes. For example, in individuals with hip and knee osteoarthritis, increased social support has been associated with higher levels of health-related quality of life [12]. Conversely, in those with rheumatoid arthritis, low levels of social support at the time of diagnosis have been predictive of poorer functional disability and pain outcomes 5 years later [13].

Online support groups (OSGs) are one way in which people with chronic musculoskeletal disorders can access social support and information. OSGs range from self-initiated groups on social media (eg, Facebook) to custom-developed websites run by clinicians or organizations. Their common goal is to provide opportunities for people to share experiences, advice, and support for their chronic disorders [14]. Given that a United Nations report (December 7, 2018) reported that more than 50% of the world's population now has access to the internet and that Web-based health service usage is increasing, OSGs may provide an accessible, convenient, and efficient means of augmenting social support and self-management. To date, the research pertaining to the characteristics of OSG platforms, group members, and implementation strategies is varied, and there is little focus on individuals with chronic musculoskeletal disorders. This makes it difficult to draw conclusions regarding if and how they are clinically effective or have a role in musculoskeletal health care.

Objectives

This study aimed to systematically review the literature evaluating the use of peer-to-peer OSGs for people with chronic musculoskeletal disorders. The 4 objectives of this review were to (1) describe the design features of peer-to-peer OSGs, (2) describe the characteristics of individuals involved in peer-to-peer OSGs, (3) synthesize the evidence on the effectiveness of OSGs, and (4) identify implementation strategies used in the delivery of OSGs.

Methods

Review Registration

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement was used to ensure complete reporting, and the review protocol was registered in the International Prospective Register of Systematic Reviews (CRD42018090326).

Search Strategy

The search strategy was developed in consultation with a librarian from The University of Queensland and involved 2 components: the population (people with chronic musculoskeletal disorders) and the intervention (peer-to-peer online support). The full PubMed search strategy is shown in Multimedia Appendix 1. The following electronic databases were searched: PubMed, CINAHL, EMBASE, PsycINFO, Scopus, and PubMed Central. A time-based filter was implemented, capturing all potential studies from 1990 (year inception of the internet) to February 26, 2019. The search included both keywords and subject heading terms. Supplementary searches of reference lists of included studies were undertaken.

Study Selection

Studies involving OSGs for adults (>18 years) with chronic (>3-month duration) musculoskeletal disorders (ie, disorder that primarily affects the musculoskeletal system) were considered eligible. Eligible interventions included any peer-to-peer (ie,

participants interacting) OSG (>3 participants on an online platform) with or without moderation or expert input or supervision. Observational studies, cohort studies, case-control studies, randomized controlled trials, qualitative studies, and mixed method studies were eligible for inclusion.

Studies not available in English and studies of pediatric populations and animals were excluded. Telehealth interventions, where health care consultations are delivered remotely via phone or internet, were excluded. Studies that used online peer-to-peer support as part of a combined or complex intervention were only included if the OSG component of the intervention was examined as an independent component, and data were available for extraction. In studies that investigated a range of morbidities, extracted data were limited to those from individuals with musculoskeletal disorders. Studies in which data pertaining to musculoskeletal disorders were not presented separately and could not be extracted were included if musculoskeletal disorders accounted for the majority of cases and authors could provide these data when contacted. When multiple studies were identified from the same groups of authors, they were contacted to determine whether samples used were independent or the same across studies. When no response was received, samples that were similar in terms of musculoskeletal disorder and year of recruitment were assumed to be the same and included only once in the analysis.

Using the eligibility criteria described above, a 2-step process was used for screening and selection. Titles and abstracts of all identified studies were screened by any 2 of the 4 reviewers (LM, MP, MB, and RM) using Covidence (Covidence, Melbourne, Australia). Additional reviewers (KM, JE, and TE) were asked to resolve screening disagreements. Full-text articles of all eligible studies were retrieved and screened by any 2 of the reviewers mentioned above, with conflicts resolved by discussion.

Data Extraction

The authors worked in 4 groups (1 for each research question) to extract data using custom-developed spreadsheets. For the first research question relating to the design features of OSGs, the following data were extracted: (1) presence and type of moderation or expert input; (2) functions and design features of host platforms; (3) content, frequency, and volume of member posts and information uploaded; and (4) involvement from participants. The second research question regarding member characteristics involved the extraction of demographics, roles and relationships, and health disorders. For the third research question relating to the effectiveness of OSGs, the following data were extracted: (1) the constructs by which effectiveness was measured, (2) outcome measures utilized to quantify effectiveness constructs, (3) processes and themes explaining any benefits, and (4) results of effectiveness studies or satisfaction ratings. For the fourth research question relating to implementation strategies, the following data were extracted: (1) group development and initiation strategies, (2) where the group was hosted, and (3) barriers and enablers to engagement in support groups.

Assessment of Study Quality

Qualitative studies were evaluated with the Critical Appraisal Skills Programme (CASP) checklist [15]. The CASP involves 10 questions divided into 3 sections: (1) validity of the results (questions 1 to 6), (2) reporting of results (questions 7 to 9), and (3) utility of the results (question 10). Moreover, 2 of the 3 reviewers (JS, TE, and KM) independently assessed the included qualitative studies. Conflicts were resolved through discussion until consensus was reached.

Quantitative studies were appraised with the Risk Of Bias In Non-Randomized Studies of Interventions tool (ROBINS-I) [16]. The ROBINS-I assesses 7 domains of bias divided across 3 timepoints: preintervention (confounding and selection bias), at intervention (classification of the intervention), and postintervention (deviation from the intervention, missing data, measurement error, and reporting bias). A total of 2 authors (LH and KM) performed the assessment, with any conflicts resolved until consensus was reached. The overall risk of bias was determined by the triangulation of results across all domains.

Data Synthesis

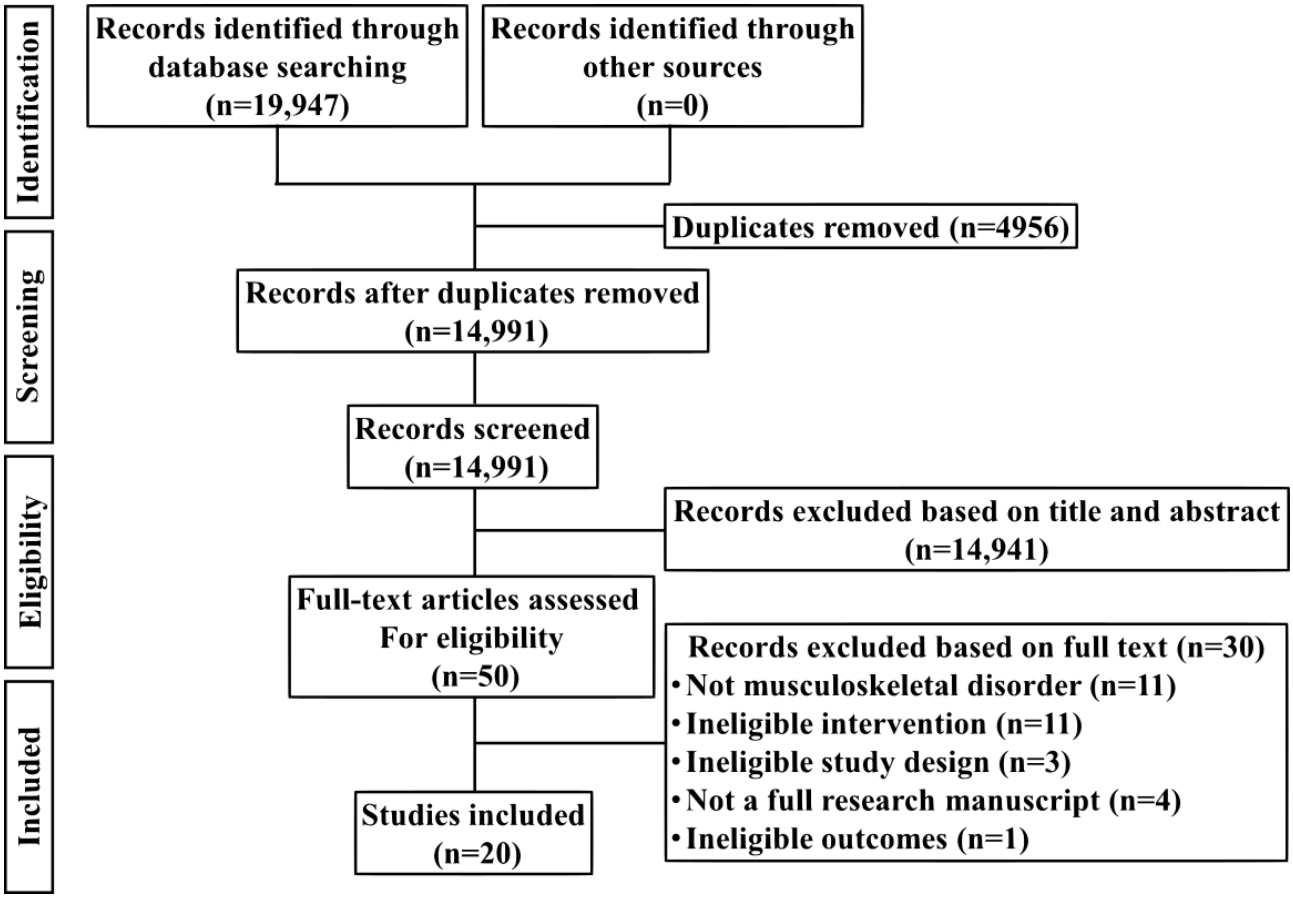
A narrative synthesis of findings was conducted because of the heterogeneity in the type of OSG, evaluation measures used, and population and designs of the included studies.

Results

Study Selection

The process of study selection is shown in Figure 1. The search yielded 19,947 articles. Following the removal of duplicates, 14,991 titles and abstracts were screened. Of these, 50 full-text articles were considered, from which 20 studies were eligible for the review.

Figure 1. Study eligibility flow.



Study Characteristics

Description of the design, sample size, and aims of the included studies is shown in Table 1. Overall, 10 studies were qualitative, 7 were quantitative, and 3 employed both qualitative and quantitative components. In terms of study design, 3 studies were prospective and the rest were cross-sectional. None of the studies were randomized controlled trials. We found 3 studies that used data from the same OSGs in the Netherlands but had different foci: forum leaders (n=32) [17], all participants (n=528) [18], or compared *posters* (people who write comments on OSG pages) with *lurkers* (people who read material without contributing posts; n=109) [19]. Moreover, 2 studies used the discourse of the same 20 members from 4 arthritis-related OSGs in the United States [20,21], and a third study by the same authors examined 1960 posts from the same 4 OSGs [22].

Study Quality

Quality Assessment of Qualitative studies

Of the 20 included studies, 13 included a qualitative component. On average, studies met 7.4 (out of 10) CASP items. Most did not articulate how interviewer perspectives may have influenced their findings. One study met 3 of the 10 criteria, implying poor methodological quality and inability to confirm the validity of findings [23]. Multimedia Appendix 2 provides full details of the quality assessment of qualitative studies. A study [24] that described the design and development of an online community without undertaking a formal qualitative evaluation was excluded from the quality assessment.

Table 1. Design, sample size, and aims of included studies.

References	Country[a]	Study design	Sample size/sample volume	Study aim
Ammerlaan et al [25]	The Netherlands	Prospective feasibility (participant survey)	12 members	To test the feasibility of the Web-based and face-to-face self-management program
Bright et al [26]	United Kingdom	Retrospective online participant survey	152 respondents	To identify the characteristics and motivations of Web-based health information seekers accessing the Web-based health community
Camerini et al [27]	Switzerland	Retrospective online participant survey	209 respondents	To evaluate the effectiveness of an internet-based patient education intervention
Hadert and Rodham [28]	United Kingdom	Retrospective, qualitative, interpretive, phenomenological analysis	374 members, 1068 posts	To investigate how and why an arthritis Web-based message board was used
Shigaki et al [29]	United States	Retrospective, qualitative	30 participants	To evaluate social interactions among individuals with rheumatoid arthritis participating in an empirically based, cognitive-behavioral, self-management, peer support program delivered in a Web-based format
Smarr et al [24]	United States	Feasibility	114 members, 448 posts	To describe the Web-based transformation of an empirically validated, clinic-based, self-management program for rheumatoid arthritis
Smedley et al [30]	United Kingdom	Retrospective qualitative content analysis	23 members, 223 posts	To explore the experiences of members in a newly launched complex regional pain syndrome discussion forum to examine how support processes become established
Smedley et al [31]	United Kingdom	Retrospective, qualitative thematic analysis	59 moderators, 790 posts	To identify and describe the activities performed by Web-based support community moderators
van Uden-Kraan et al [17]	The Netherlands	Semistructured interviews	32 participants	To explore if, and in which ways, patients feel empowered by participation in OSGs[b]
van Uden-Kraan et al [32]	The Netherlands	Retrospective online participant survey	528 respondents	To explore if lurkers in Web-based patient support groups profit to the same extent as posters do
van Uden-Kraan et al [19]	The Netherlands	Retrospective qualitative content analysis	1500 posts	To explore the extent to which potential disadvantages actually occur when participating in OSGs
van Uden-Kraan et al [18]	The Netherlands	Retrospective online participant survey	528 respondents	To explore the extent to which patients feel empowered by their participation in OSGs and what processes occurring in these groups are related to the empowering outcomes
van Uden-Kraan et al [23]	The Netherlands	Semistructured interviews	23 Web-masters	To determine the success factors of OSGs for patients and the motives and goals of people who start such groups
van Uden-Kraan et al [33]	The Netherlands	Prospective participant survey	679 respondents	To explore factors that facilitate or impede engagement in face-to-face and Web-based peer support
van der Vaart et al [34]	The Netherlands	Prospective participant survey	227 respondents	To examine current disease-related internet use and intentions to use various Web-based support services on a hospital-based interactive health communication app of patients with rheumatic diseases
Walker [35]	United States	Retrospective qualitative content analysis	292 posts	To explore how a relatively new medium of a disease-specific Facebook group is used to address needs of people affected by thoracic outlet syndrome
Willis [21]	United States	Retrospective qualitative discourse analysis (ethnomethodology)	5 members, 8231 posts	To understand how patients with arthritis use Web-based health communities to exchange disease-related information to better manage their chronic disease
Willis [20]	United States	Retrospective qualitative discourse analysis (ethnomethodology)	8231 posts	To examine self-efficacy within the computer-mediated communication of 4 Web-based health communities used by people with arthritis
Willis and Royne [22]	United States	Retrospective quantitative content analysis	1960 posts	To examine the computer-mediated communication within Web-based health communities for evidence of chronic disease self-management behaviors

References	Country[a]	Study design	Sample size/sample volume	Study aim
Xing et al [36]	United States	Retrospective content analysis and survival analysis	100,000 users, 200,000 user posts	To understand how requests for and provisions of informational support by members with different social roles influence members' continued participation in Web-based health communities

[a]Origin of online support groups when they are multinational.

[b]OSG: online support group.

Quality Assessment of Quantitative Studies

Of the 20 included studies, 10 included a quantitative component (Multimedia Appendix 3). Overall, 7 studies were rated as low risk of bias [22,27-33], and 3 studies were rated as moderate risk of bias [18,29]. Of the studies with a moderate risk of bias, 2 used multiple outcome measures to describe or quantify a single variable or concept, subsequently performing multiple analyses on a single research question [18,19].

Design Features of Online Support Groups

Characteristics of the design and features of the OSGs are described in Tables 1 and 2. Of the 20 studies, 13 (65%) used platforms that were publicly accessible, 7 (36%) [24,26,34] were private platforms designed specifically for the study, and 1 study (15%) did not report the type of platform [33]. The most common type of platform was a purpose-built website (13/20, 65%). English language platforms were used by 52% (11/20) studies [20-22,24,26,28-31,35,36], whereas 42% (8/20) studies used Dutch platforms [17-19,23,25,32-34] and 1 study used an Italian platform [27]. There were 6 studies [21,24,25,28,30,31] that reported the number of OSG members, which ranged from 12 to 374 people (Table 1). The number of posts examined for content was reported by 10 studies [20-22,24,28,30-32,35,36], ranging from 223 to 200,000 (Table 1). The average duration of membership for the platforms ranged from 4 weeks to 6 years.

Moderation of the OSG was used in 6 studies [23-25,30-32] (Table 2). Moderators were participants with musculoskeletal disorders (4 studies) [23,25,31,32], health professionals (1 study) [24], and organizers or administrators (1 study) [30]. The type of moderation consisted of supportive tasks, sharing experiences, facilitating information sharing, making announcements, administrative tasks (eg, removal of *disadvantaged* posts, monitoring members' activity, and maintaining the rules of the OSG), and leading group activities (eg, chat and discussion forums).

Table 2. Description of online support groups included for review.

References	Target population	Type of platform	Duration of OSG[a]	Language	Frequency of posts	Presence/source of moderation
Ammerlaan et al [25]	Young adults (age 16-25 years) with arthritis	Private website; planned weekly chat group (90 min)	6 weeks	Dutch	NR[b]	Yes/peer
Bright et al [26]	Adults with knee problems	Private website *KNEE-guru*	1 month	English	NR	No
Camerini et al [27]	Adults with FMS[c]	Private website; also included video and textual material on coping	Mean 167 days (SD 67.6)	Italian	NR	No
Hadert and Rodham [28]	Adults with arthritis	Public website	3 months	English	NR	No
Shigaki et al [29]	Adults with RA[d]	Private website	10 weeks	English	NR	No
Smarr et al [24]	Adults with RA	Private website with multiple shared resources (eg, education material and audio files)	Average of 10 weeks	English	NR	Yes/health professional
Smedley et al [30]	Adults with CRPS[e]	4 private forums	6 months	English	17=low frequency posters[f] (average 9.5 posts); 6=high-frequency posters	Yes/peer
Smedley et al [31]	Adults with arthritis, CRPS, Crohn disease, depression, Huntington disease, and diabetes	6 public discussion forums	NR	English	15 posts per moderator	Yes/peer
van Uden-Kraan et al [17]	Adults with arthritis, FMS, or breast cancer	9 public websites	NR	Dutch	Posters >1/day=140; 1/day=121; >1/week=96; 1/week=31; 1/month=6; and <1/month=6	No
van Uden-Kraan et al [32]	Adults with arthritis, FMS, or breast cancer	8 public websites	1 year (range 0-6 years)	Dutch	Minimum=1/day	Yes/peer
van Uden-Kraan et al [19]	Adults with arthritis, FMS, or breast cancer	8 public websites	3 months	Dutch	1 or 2 messages	No
van Uden-Kraan et al [18]	Adults with arthritis, FMS, or breast cancer	Public websites	Up to 2.5 years	Dutch	Posters >1/day=146; 1/day=139; >1/week=124; 1/week=50; 1/month=13; and <1/month=13	No
van Uden-Kraan et al [23]	Adults with arthritis, FMS, or breast cancer	10 Public websites, 13 private websites, 18 stand-alone (not embedded in organization website/forum) OSGs, and 5 patient advocacy websites	NR	Dutch	Ranged from a few messages per week to hundreds of messages daily	Yes
van der Vaart et al [34]	Individuals with rheumatic diagnosis	Private app	NR	Dutch	NR	No
Walker [35]	Adults with thoracic outlet syndrome	Public; hosted on Facebook	7 months	English	NR	NR
Willis [21]	Adults with arthritis	4 public websites	NR	English	Once every 4 days	No
Willis [20]	Adults with arthritis	4 public websites	NR	English	Only high-frequency posters participated	No

References	Target population	Type of platform	Duration of OSG[a]	Language	Frequency of posts	Presence/source of moderation
Willis and Royne [22]	Adults with arthritis	4 public websites	4 weeks	English	NR	No
Xing et al [36]	Individuals with (or associated with) FMS	Public website	Up to 6 years	English	Core group members: average of 393 posts; peripheral members: 9.58 posts	Possible

[a]OSG: online support group.

[b]NR: not reported.

[c]FMS: fibromyalgia.

[d]RA: rheumatoid arthritis.

[e]CRPS: complex regional pain syndrome.

[f]Poster: people who write comments on online support group pages.

Characteristics of Individuals Involved in Online Support Groups

Participant characteristics were reported to varying degrees of detail across studies (Table 3). Participants' age was reported in 8 studies [20,25-27,29,30,33,34], which ranged from 18 to 83 years. Of the 1370 participants in the 8 studies reporting gender [20,25-27,29,30,33,34], 1092 (80%) were female. Education history was reported in 6 studies [25-27,29,33,34]. Of the 1252 participants accounted for, 499 (39%) had a maximum of *low-tier education*, but this category was not defined by the studies' authors. Occupational status was reported in 3 studies [26,33], accounting for 1068 participants, of which 643 (60%) were unemployed. Relationship status was reported in 4 studies [26,29,33,34], with 836 of 1071 (78%) participants being married or cohabiting. All studies stated their disorder of interest. The most commonly encountered musculoskeletal disorders were unspecified types of arthritis [17-23,25,28,32,33] and fibromyalgia [17-19,23,25,27,32,36]. Moreover, 5 studies examined individuals with rheumatoid arthritis [24,25,29,33,34], and 2 studies each investigated rheumatic disease [28,34], chronic regional pain syndrome [30,31], and spondyloarthropathy [25,28].

Not all participants in each of the OSGs had a musculoskeletal problem (Table 3). Of the 15 studies that identified the roles of OSG members, 2 studies identified that health professionals were included in the group [24,32], and 3 studies included family members or acquaintances of people with the disorder [23,32,35]. Within groups of patient members, 3 studies identified that members could either be participants (n=292) or moderators or peer leaders (n=66) [25,31,32], and 2 studies separated members into active posters (core members) (n=460) or lurkers (peripheral members) (n=9429) [19,36]. *Peripheral members* were noted to post significantly less frequently (mean 9.9 posts, SD 21.6) than *core members* (mean 393.5 posts, SD 372.9) [36], and *lurkers*, as by definition, did not post at all [19].

Table 3. Characteristics of online support group users.

References	Age (years), mean (range or SD)	Gender (female/male)	Education levels	Occupation	Marital status	Motivation for joining
Ammerlaan et al [25]	22 (range: 17-25)	9/1	Vocational training: 1; advanced vocational training: 7; college/university: 2	N/A[a]	N/A	N/A
Bright et al [26]	40.1	93/59	Higher education qualifications: 114	Employed: 87; unemployed: 65	Cohabiting: 104	Emotional support (clarity regarding advice and treatments), social support (sharing experiences and information), and condition support (achieving a sense of authority)
Camerini et al [27]	49 (range: 25-74)	199/10	8 years of schooling: 36; high school/university: 163; not reported: 10	N/A	N/A	N/A
Hadert and Rodham [28]	N/A	N/A	N/A	N/A	N/A	Needing to be believed, information exchange, sharing support, and sharing emotions
Shigaki et al [36]	49.4 (range: 30.1-68.5)	28/2	Mean years of education: 15 (range: 12-20) years	N/A	Married: 19	N/A
Smarr et al [24]	N/A	N/A	N/A	N/A	N/A	N/A
Smedley et al [30]	36.6 (range: 20-54)[b]	18/5	N/A	N/A	N/A	N/A
Smedley et al [31]	N/A	N/A	N/A	N/A	N/A	N/A
van Uden-Kraan et al [17]	43 (range: 21-75)	30/2	Lower: 5; medium: 14; high: 13	Unemployed/unable to work: 25; employed: 7	Married/cohabiting: 26; not married: 6	N/A
van Uden-Kraan et al [32]	Posters[c]: 43 (SD 10.4); lurkers[d]: 47 (SD 9.9)	Posters: 392/27; lurkers: 102/7	Posters—lower: 129; medium: 170; high: 111. Lurkers—lower: 42; medium: 43; high: 24	Posters—working >20 hours: 128; working ≤20 hours: 54; unemployed: 234. Lurkers—working >20 hours: 39; working ≤20 hours: 11; unemployed:59	Posters—in a relationship: 331; single: 88. Lurkers—in a relationship: 85; single: 25	N/A
van Uden-Kraan et al [19]	38 (range: 21-65)	293/29; unknown: 25	N/A	N/A	N/A	N/A
van Uden-Kraan et al [18]	44 (range: 17-75)	494/34	Lower: 171; medium: 213; high: 135	Working >20 hours: 167; working ≤20 hours: 65; unemployed: 293	In a relationship: 415; single: 113	N/A
van Uden-Kraan et al [23]	46 (range: 24-65)	20/3	N/A	N/A	N/A	Provide information and social support
van Uden-Kraan et al [33]	54 (range: 18-75)	571/106	Lower: 404; medium: 176; high: 94	Employed: 212; unemployed: 447	Married/cohabiting: 530; single: 128	Improve mental health and past behaviors with support groups
van der Vaart et al [35]	52 (SD 11)	143/84	Lower: 61; average: 116; high: 46; unknown: 4	Employed: 119; unemployed: 106	Married/cohabiting: 183; single: 42; unknown: 2	Poor mental health and improving health literacy[e]
Walker [35]	N/A	N/A	N/A	N/A	N/A	N/A
Willis [21]	Range: 21-83	15/5	N/A	N/A	N/A	N/A
Willis [20]	Range: 21-83	15/5	N/A	N/A	N/A	N/A

References	Age (years), mean (range or SD)	Gender (female/male)	Education levels	Occupation	Marital status	Motivation for joining
Willis and Royne [22]	N/A	N/A	N/A	N/A	N/A	N/A
Xing et al [36]	N/A	N/A	N/A	N/A	N/A	N/A

[a]N/A: not applicable.

[b]Age was available for 9 participants, and the duration of symptoms was available for 14 participants.

[c]Poster: people who write comments on online support group pages.

[d]Lurker: people who read material without contributing posts to the forum.

[e]People with good health literacy were more likely to use peer support services to further improve knowledge.

Effectiveness of Online Support Groups

Overall, 10 studies reported on measures of effectiveness from OSGs [17-22,25,27,29,30]. Effectiveness was conceptualized as the development of patient empowerment [17-19], social support [25,29,30], self-management processes [20,22,27], and health literacy [21,27]. In evaluating effectiveness, none of the studies considered clinical domains (eg, pain or physical function; Multimedia Appendix 4).

Themes and processes of developing social activity, empowerment, self-management, and health literacy were explored by 4 studies using qualitative study designs [17,20,29], by 5 studies using quantitative designs [18,19,22,27,30], and by 1 study using mixed method [25]. Using semistructured interviews with 32 OSG users with arthritis, fibromyalgia, or breast cancer, van Uden-Kraan et al [17] concluded that patient empowerment was achieved by (1) being better informed, (2) feeling more confident, (3) increasing social well-being and enhanced self-esteem, and (4) acceptance and coping with chronic disease. Information and support were also found to be important themes for developing self-management and social support. Moreover, 2 studies reported that approximately one-third of all user posts contained these themes [21,30]. Sharing personal and disease experiences, particularly from disease veterans, was important in the development of a social activity, self-management plans, and improving health literacy [20,25]. This frequently included posts on drug management (29.3%) and symptom management (22.7%) [22]. Other common themes in these latter 3 effectiveness domains were seeking emotional support, positive feedback, and reinforcement from the community [20,21,25,30] (Multimedia Appendix 4).

When quantifying the effectiveness of OSG participation, participants with arthritis aged 25 years or younger reported high levels (mean 8.4, range: 6-10) of satisfaction with goal attainment, using a 10-point numerical rating scale [25]. However, a survey of 528 individuals with fibromyalgia, arthritis, and breast cancer indicated that they were neutral or in agreement (scores of 3-4 on a 5-point Likert scale) with their achievement of (1) being better informed, (2) enhancing social well-being, and (3) improving illness acceptance as a result of their participation in an OSG [18]. Further exploration of these findings revealed that people who were more engaged, evidenced by visiting the site more frequently or making more posts, experienced greater gains in health literacy, self-esteem, and self-management than those who made fewer posts or lurked [19,27] (Multimedia Appendix 4). Willis and Royne [22] reported that improvements in mobility, flexibility, pain, and energy were among the most frequently reported benefits of participation across 4 arthritis OSGs; however, they also reported significant differences between OSGs, suggesting that the perceived benefits may be specific to a group. Multimedia Appendix 4 summarizes all measures used to investigate the effectiveness of OSGs.

Implementation Strategies to Deliver Online Support Groups

Overall, 7 studies [22,36] reported on the strategies used to implement OSGs. Groups tended to be either self-initiated by an individual sufferer of the disorder or by official consumer associations [32]. The latter were embedded within pre-existing websites containing health information, being either open access or available to subscribed members only [23]. We found 2 studies that used relevant stakeholders such as program moderators and/or patients in the development and testing of their OSG [24,25] and for its delivery [25].

One study reported that a key component of OSGs was to continually promote the group and keep it alive, which took considerable time and energy [23]. Strategies to do this included moderation, augmented learning, or a small core group of individuals who posted more frequently than more peripheral users. Moderating a group took approximately 10 to 15 hours per week (unspecific group size), which was often perceived by individual moderators as onerous [23]. In addition to moderation, 2 OSGs augmented learning by scheduling weekly group chats or setting homework tasks that centered around predetermined themes [24,25].

Member engagement, or staying in the OSG, was significantly associated with starting or contributing to threads and requesting information. Xing et al [36] reported that OSG members who start or contribute to threads 1 SD more frequently than average (range: 222 posts to 373 posts depending on group roles) were 20% more likely to stay engaged with the community. Similarly, OSG members who requested information 1 SD more frequently than average (approximately 16 information requests) were 29.3% more likely to remain in the group. Responding to questions or information requests also influenced member engagement. Posted questions generally received an answer within 24 hours, though a small number (15%) of questions received no answer at all [37]. If an information request was responded to by someone other than a *core group* member (ie, a peripheral group member), the person who made the request was 11.4% more likely to leave the group [36]. Criticisms of

OSG implementation were that the discussion posts contained casual chitchat [32], and some OSGs had become *social clubs* rather than a place to exchange information and share experience [23].

Discussion

Principal Findings

This systematic review has revealed that the design features and implementation strategies used by peer-to-peer OSGs for people with chronic musculoskeletal disorders vary widely. People across a broad demographic spectrum access OSGs; some people chose to post actively, whereas others take a passive approach. Self-efficacy, health literacy, and empowerment are the constructs most commonly explored in studies investigating the effectiveness of musculoskeletal-focused OSGs. Overall, the findings stimulate discussion around optimal design and implementation of OSGs as well as how their effectiveness might best be measured. These topics are recommended for future investigation, particularly for people with chronic musculoskeletal disorders.

Comparison With Prior Work

For individuals with chronic musculoskeletal disorders, accessibility to OSGs is not influenced by whether the group is publicly or privately hosted. On the basis of the available literature, this also seems to be the case for OSGs focused on individuals with opioid addiction [38], depression [37], and asthma [39]. The majority of OSGs included in this review were hosted on public platforms. Previously, issues regarding privacy and security offered to users of public platforms have been raised [40]. A study of Facebook users comparing the amount and type of information disclosed on public and private Facebook groups indicates that private groups may be preferred, especially by people with social anxiety, because of the perception of greater control over who people are communicating with as well as greater trust and security of their information [41]. Our findings suggest that privacy and security were not barriers to participation in OSGs for people with chronic musculoskeletal disorders, and they did not influence the themes of information being shared. This may, however, have been because of the majority of group members also having the focus disorder, rather than the wider social network found on Facebook; the prevalence of the disorders within the general community; or the similarity in characteristics between group members (ie, primarily females who were not currently working). It appears that for people with chronic musculoskeletal disorders, the internet provides acceptable accessible sources of peer support for individuals seeking it, regardless of the hosting platform.

When examining the characteristics of OSG members included in this review, the majority of musculoskeletal-focused OSG members were female, not currently in full-time employment, and cohabitating or married. There is a significant association between exhibiting a preference for Web-based communication and the duration of internet usage [42]. Web-based communication is one way for people who are not working full time to maintain social activity when their peers and partners are not present. An explanation for the higher proportion of females in OSGs could be that although men use the internet more, women have been faster to adopt and are more frequent users of social networking and Web-based chat programs [43]. Furthermore, arthritis and fibromyalgia, the musculoskeletal disorders most commonly encountered in this review, are more common in females [44,45]. Another factor explaining lower male representation may be the perceived stigmatization of men sharing disease experiences on the Web [46]. Increasing representation of men in OSGs may be one way to improve self-management of disorders such as low back pain, the leading cause of years lived with disability for males since 1990 [2].

In evaluating OSG effectiveness, this review found that studies focused on constructs such as empowerment, self-efficacy, confidence, social support, and knowledge. These outcomes are consistent with those reported across multiple OSGs [47]. However, the lack of randomized controlled trials means that no causal inference can be established regarding OSG participation and change in these constructs. Self-efficacy has been identified as a foundation of chronic disease self-management [48], and multiple cohorts and observational studies of nonmusculoskeletal disorders have reported significant positive effects on self-efficacy following participation in OSGs and peer mentoring [40,47,49,50]. Findings from this review suggest that the extent to which OSG participation results in individuals with chronic musculoskeletal disorders feeling informed, confident, accepting of their disease is limited [18,19]. Furthermore, individuals who lurked, or did not actively post to OSGs, scored lower in the constructs of social well-being and self-esteem than active posters [19]. Although the direction of this relationship cannot be determined (active participation in OSGs leading to higher levels of social well-being and self-esteem, or vice versa), these results suggest that the type of participation may have a mediating effect. This has important implications for the implementation of future OSGs, as it appears that efforts must be made to engage individuals actively to contribute to posts, share stories, or ask questions [40].

One potential implementation method to promote active posting among OSG participants is the presence of a professional moderator [51]. Less than one-third of the studies included in this review reported the presence of a moderator. Of these moderators, the majority were peers who had the focus disorder. Although there does not appear to be a difference in OSG effectiveness irrespective of whether the moderator is a peer or health professional [40,51], Young et al [38] observed high attrition rates among peer moderators themselves. The time burden and onerous tasks involved in peer moderation may be one reason for this. Furthermore, when OSG member queries are not responded by peer moderators or leaders, general group attrition increased [36]. A previous review of OSGs [47] identified that attrition rates are lower with professional moderators. As such, having health professionals as moderators may be one way to address attrition rates and engagement. Health professional involvement may also help alleviate some of the time burden associated with moderating and administration for the group.

Additional implementation strategies that were investigated by studies included in this review were pretesting of OSGs before

wider release, embedding the OSG in familiar websites, and scheduling weekly events or homework. No study investigated or reported the effectiveness of these strategies. Having identified these implementation strategies, a recommended topic for future research would be comparing the success of such implementation strategies with respect to consumer engagement and efficacy.

Limitations

There are limitations that need to be considered when interpreting the findings of this review. The main limitation is that the health disorders of interest in several included studies were diverse, and in some cases, it was not possible to identify which data came from individuals with musculoskeletal disorders. There were 6 studies [17-19,23,31,32], primarily from a single research group, where data from individuals with musculoskeletal disorders could not always be distinguished from those with other chronic disorders. People with musculoskeletal disorders account for the majority of participants included in our data synthesis. Second, several included studies reported on the same group of OSGs. Although each of these studies explored different aspects of OSGs, the smaller overall sample limits generalizability. Third, many of the studies investigating the content of OSG posts only reported a summary of the most frequently occurring topics. As the general posting rate was low, this would overrepresent the attitudes and beliefs of individuals who were more actively engaged with the group. Fourth, all studies also focused on individuals who were already members of OSGs and often collected cross-sectional data, so it is impossible to determine change or development in outcomes over time. As such, it is difficult to attribute attitudes and beliefs regarding empowerment and self-efficacy to participation in OSGs or whether these were views formed before participating. Finally, as the focus of effectiveness evaluation was on attitudes and beliefs rather than health outcomes, the impact that OSGs have on clinical features and symptoms of musculoskeletal disorders could not be evaluated.

Conclusions

OSGs provide an opportunity for individuals with musculoskeletal disorders to support one another through the sharing of knowledge and experiences. Across the diverse range of designs, languages, included features, and delivery platforms, OSGs attract participation from people of all ages and education levels, although predominantly females. The level to which group members participate appears to be related to their perceived benefit in health literacy and empowerment. However, the lack of control groups in studies means that direct inferences cannot be assessed or established. Participation may be increased by strategies such as moderation or input by a health professional or expert peers, homework tasks, and scheduled weekly chats. Whether these strategies are effective requires further investigation.

Acknowledgments

Funding for this study was provided by a Program Grant (APP1091302) and Centre of Research Excellence Grant (APP1079078) from the National Health and Medical Research Council (NHMRC) of Australia. DH, KB, and PH (APP1102905) are supported by fellowships from the NHMRC.

Conflicts of Interest

DH provides consulting advice to Merck Serono, TLCBio, Pfizer, and Eli Lilly and company.

Multimedia Appendix 1
Search strategy built and conducted in PubMed.
[DOCX File , 12 KB - jmir_v22i4e15822_app1.docx]

Multimedia Appendix 2
Results of the quality assessment of qualitative study methods using the Critical Appraisal Skills Programme criteria.
[DOCX File , 15 KB - jmir_v22i4e15822_app2.docx]

Multimedia Appendix 3
Risk Of Bias In Non-Randomized Studies of Interventions for quantitative design studies.
[DOCX File , 22 KB - jmir_v22i4e15822_app3.docx]

Multimedia Appendix 4
Themes and magnitude of measure used to investigate the effectiveness of online support groups.
[DOCX File , 18 KB - jmir_v22i4e15822_app4.docx]

References

1. Cimmino MA, Ferrone C, Cutolo M. Epidemiology of chronic musculoskeletal pain. Best Pract Res Clin Rheumatol 2011 Apr;25(2):173-183. [doi: 10.1016/j.berh.2010.01.012] [Medline: 22094194]

2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
3. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
4. Duffield SJ, Ellis BM, Goodson N, Walker-Bone K, Conaghan PG, Margham T, et al. The contribution of musculoskeletal disorders in multimorbidity: Implications for practice and policy. Best Pract Res Clin Rheumatol 2017 Apr;31(2):129-144 [FREE Full text] [doi: 10.1016/j.berh.2017.09.004] [Medline: 29224692]
5. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
6. National Clinical Guideline Centre (UK). NCBI - NIH. London: National Institute for Health and Care Excellence (UK); 2014 Feb. Osteoarthritis: Care and Management in Adults URL: https://www.ncbi.nlm.nih.gov/pubmed/25340227 [accessed 2020-02-19]
7. Smolen JS, Landewé R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis 2017 Jun;76(6):960-977. [doi: 10.1136/annrheumdis-2016-210715] [Medline: 28264816]
8. National Guideline Centre (UK). NCBI - NIH. London: National Institute for Health and Care Excellence (UK); 2016 Nov. Low Back Pain and Sciatica in Over 16s: Assessment and Management URL: https://www.ncbi.nlm.nih.gov/pubmed/27929617 [accessed 2020-02-19]
9. Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003 Apr;30(2):170-195. [doi: 10.1177/1090198102251030] [Medline: 12693522]
10. Lee SD, Arozullah AM, Cho YI. Health literacy, social support, and health: a research agenda. Soc Sci Med 2004 Apr;58(7):1309-1321. [doi: 10.1016/S0277-9536(03)00329-0] [Medline: 14759678]
11. Vandenbosch J, van den Broucke S, Vancorenland S, Avalosse H, Verniest R, Callens M. Health literacy and the use of healthcare services in Belgium. J Epidemiol Community Health 2016 Oct;70(10):1032-1038. [doi: 10.1136/jech-2015-206910] [Medline: 27116951]
12. Ethgen O, Vanparijs P, Delhalle S, Rosant S, Bruyère O, Reginster JY. Social support and health-related quality of life in hip and knee osteoarthritis. Qual Life Res 2004 Mar;13(2):321-330. [doi: 10.1023/B:QURE.0000018492.40262.d1] [Medline: 15085904]
13. Evers AW, Kraaimaat FW, Geenen R, Jacobs JW, Bijlsma JW. Pain coping and social support as predictors of long-term functional disability and pain in early rheumatoid arthritis. Behav Res Ther 2003 Nov;41(11):1295-1310. [doi: 10.1016/s0005-7967(03)00036-6] [Medline: 14527529]
14. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. Comput Hum Behav 2008;24(5):1867-1883. [doi: 10.1016/j.chb.2008.02.004]
15. Critical Appraisal Skills Programme. CASP Checklists URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2019-03-01]
16. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
17. van Uden-Kraan CF, Drossaert CH, Taal E, Shaw BR, Seydel ER, van de Laar MA. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-417. [doi: 10.1177/1049732307313429] [Medline: 18235163]
18. van Uden-Kraan C, Drossaert C, Taal E, Seydel E, van de Laar M. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
19. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008 Jun 30;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
20. Willis E. Patients' self-efficacy within online health communities: facilitating chronic disease self-management behaviors through peer education. Health Commun 2016;31(3):299-307. [doi: 10.1080/10410236.2014.950019] [Medline: 26325224]
21. Willis E. The making of expert patients: the role of online health communities in arthritis self-management. J Health Psychol 2014 Dec;19(12):1613-1625. [doi: 10.1177/1359105313496446] [Medline: 23988679]
22. Willis E, Royne MB. Online health communities and chronic disease self-management. Health Commun 2017 Mar;32(3):269-278. [doi: 10.1080/10410236.2016.1138278] [Medline: 27218836]
23. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Patient-initiated online support groups: motives for initiation, extent of success and success factors. J Telemed Telecare 2010;16(1):30-34. [doi: 10.1258/jtt.2009.001009] [Medline: 20086265]

24. Smarr KL, Musser DR, Shigaki CL, Johnson R, Hanson KD, Siva C. Online self-management in rheumatoid arthritis: a patient-centered model application. Telemed J E Health 2011 Mar;17(2):104-110. [doi: 10.1089/tmj.2010.0116] [Medline: 21361817]
25. Ammerlaan J, van Os-Medendorp H, Scholtus L, de Vos A, Zwier M, Bijlsma H, et al. Feasibility of an online and a face-to-face version of a self-management program for young adults with a rheumatic disease: experiences of young adults and peer leaders. Pediatr Rheumatol Online J 2014 Mar 25;12:10 [FREE Full text] [doi: 10.1186/1546-0096-12-10] [Medline: 24666817]
26. Bright P, Hambly K, Tamakloe S. What is the profile of individuals joining the KNEEguru online health community? A cross-sectional mixed-methods study. J Med Internet Res 2016 Apr 18;18(4):e84 [FREE Full text] [doi: 10.2196/jmir.5374] [Medline: 27089531]
27. Camerini L, Camerini A, Schulz PJ. Do participation and personalization matter? A model-driven evaluation of an Internet-based patient education intervention for fibromyalgia patients. Patient Educ Couns 2013 Aug;92(2):229-234. [doi: 10.1016/j.pec.2013.04.007] [Medline: 23664234]
28. Hadert A, Rodham K. The invisible reality of arthritis: a qualitative analysis of an online message board. Musculoskeletal Care 2008 Sep;6(3):181-196. [doi: 10.1002/msc.131] [Medline: 18570188]
29. Shigaki CL, Smarr KL, Gong Y, Donovan-Hanson K, Siva C, Johnson RA, et al. Social interactions in an online self-management program for rheumatoid arthritis. Chronic Illn 2008 Dec;4(4):239-246. [doi: 10.1177/1742395308097862] [Medline: 19091932]
30. Smedley R, Coulson N, Gavin J, Rodham K, Watts L. Online social support for Complex Regional Pain Syndrome: A content analysis of support exchanges within a newly launched discussion forum. Comput Hum Behav 2015 Oct;51:53-63. [doi: 10.1016/j.chb.2015.04.040]
31. Smedley RM, Coulson NS. A thematic analysis of messages posted by moderators within health-related asynchronous online support forums. Patient Educ Couns 2017 Sep;100(9):1688-1693. [doi: 10.1016/j.pec.2017.04.008] [Medline: 28454663]
32. van Uden-Kraan C, Drossaert C, Taal E, Lebrun C, Drossaers-Bakker K, Smit W, et al. Coping with somatic illnesses in online support groups: Do the feared disadvantages actually occur? Comput Hum Behav 2008;24(2):309-324. [doi: 10.1016/j.chb.2007.01.014]
33. van Uden-Kraan CF, Drossaert CH, Taal E, Smit WM, Bernelot Moens HJ, van de Laar MA. Determinants of engagement in face-to-face and online patient support groups. J Med Internet Res 2011 Dec 7;13(4):e106 [FREE Full text] [doi: 10.2196/jmir.1718] [Medline: 22155649]
34. van der Vaart R, Drossaert CH, Taal E, van de Laar MA. Patient preferences for a hospital-based rheumatology Interactive Health Communication Application and factors associated with these preferences. Rheumatology (Oxford) 2011 Sep;50(9):1618-1626. [doi: 10.1093/rheumatology/ker161] [Medline: 21551221]
35. Walker KK. Cognitive and affective uses of a thoracic outlet syndrome Facebook support group. Health Commun 2014 Sep;29(8):773-781. [doi: 10.1080/10410236.2013.800830] [Medline: 24171492]
36. Xing W, Goggins S, Introne J. Quantifying the Effect of Informational Support on Membership Retention in Online Communities through Large-Scale Data Analytics. Comput Hum Behav 2018 Sep;86:227-234. [doi: 10.1016/j.chb.2018.04.042]
37. Griffiths KM, Calear AL, Banfield M. Systematic review on Internet Support Groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? J Med Internet Res 2009 Sep 30;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
38. Young SD, Koussa M, Lee S, Perez H, Gill N, Gelberg L, et al. Feasibility of a social media/online community support group intervention among chronic pain patients on opioid therapy. J Addict Dis 2018;37(1-2):96-101 [FREE Full text] [doi: 10.1080/10550887.2018.1557992] [Medline: 30614403]
39. Stewart M, Letourneau N, Masuda JR, Anderson S, McGhan S. Impacts of online peer support for children with asthma and allergies: It just helps you every time you can't breathe well". J Pediatr Nurs 2013;28(5):439-452. [doi: 10.1016/j.pedn.2013.01.003] [Medline: 23398896]
40. Cooper K, Schofield P, Klein S, Smith BH, Jehu LM. Exploring peer-mentoring for community dwelling older adults with chronic low back pain: a qualitative study. Physiotherapy 2017 Jun;103(2):138-145 [FREE Full text] [doi: 10.1016/j.physio.2016.05.005] [Medline: 27613081]
41. Green T, Wilhelmsen T, Wilmots E, Dodd B, Quinn S. Social anxiety, attributes of online communication and self-disclosure across private and public Facebook communication. Comput Hum Behav 2016 May;58:206-213. [doi: 10.1016/j.chb.2015.12.066]
42. Thayer SE, Ray S. Online communication preferences across age, gender, and duration of internet use. Cyberpsychol Behav 2006 Aug;9(4):432-440. [doi: 10.1089/cpb.2006.9.432] [Medline: 16901247]
43. Kimbrough AM, Guadagno RE, Muscanell NL, Dill J. Gender differences in mediated communication: Women connect more than do men. Comput Hum Behav 2013 May;29(3):896-900. [doi: 10.1016/j.chb.2012.12.005]
44. Wolfe F, Walitt B, Perrot S, Rasker JJ, Häuser W. Fibromyalgia diagnosis and biased assessment: Sex, prevalence and bias. PLoS One 2018;13(9):e0203755 [FREE Full text] [doi: 10.1371/journal.pone.0203755] [Medline: 30212526]

45. van Vollenhoven RF. Sex differences in rheumatoid arthritis: more than meets the eye. BMC Med 2009 Mar 30;7:12 [FREE Full text] [doi: 10.1186/1741-7015-7-12] [Medline: 19331649]
46. Samulowitz A, Gremyr I, Eriksson E, Hensing G. 'Brave Men' and 'Emotional Women': A theory-guided literature review on gender bias in health care and gendered norms towards patients with chronic pain. Pain Res Manag 2018;2018:6358624 [FREE Full text] [doi: 10.1155/2018/6358624] [Medline: 29682130]
47. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
48. Jackson M, Jones D, Dyson J, Macleod U. Facilitated group work for people with long-term conditions: a systematic review of benefits from studies of group-work interventions. Br J Gen Pract 2019 May;69(682):e363-e372. [doi: 10.3399/bjgp19X702233] [Medline: 30962225]
49. Olsson CA, Boyce MF, Toumbourou JW, Sawyer SM. The role of peer support in facilitating psychosocial adjustment to chronic illness in adolescence. Clin Child Psychol Psych 2016 Jul;10(1):78-87. [doi: 10.1177/1359104505048793]
50. Lewis P, Klineberg E, Towns S, Moore K, Steinbeck K. The effects of introducing peer support to young people with a chronic illness. J Child Fam Stud 2016;25(8):2541-2553. [doi: 10.1007/s10826-016-0427-4]
51. Klemm P. Effects of online support group format (moderated vs peer-led) on depressive symptoms and extent of participation in women with breast cancer. Comput Inform Nurs 2012 Jan;30(1):9-18. [doi: 10.1097/NCN.0b013e3182343efa] [Medline: 22240564]

Abbreviations

CASP: Critical Appraisal Skills Programme
NHMRC: National Health and Medical Research Council
OSG: online support group
ROBINS-I: Risk Of Bias In Non-Randomized Studies of Interventions

Edited by G Eysenbach; submitted 09.08.19; peer-reviewed by C Barton, E Willis, N Coulson, S Atanasova; comments to author 02.10.19; revised version received 27.11.19; accepted 02.02.20; published 24.04.20.

Please cite as:
Maclachlan LR, Mills K, Lawford BJ, Egerton T, Setchell J, Hall LM, Plinsinga ML, Besomi M, Teo PL, Eyles JP, Mellor R, Melo L, Robbins S, Hodges PW, Hunter DJ, Vicenzino B, Bennell KL
Design, Delivery, Maintenance, and Outcomes of Peer-to-Peer Online Support Groups for People With Chronic Musculoskeletal Disorders: Systematic Review
J Med Internet Res 2020;22(4):e15822
URL: http://www.jmir.org/2020/4/e15822/
doi:10.2196/15822
PMID:32329746

Original Paper

Optimizing the Analytical Value of Oncology-Related Data Based on an In-Memory Analysis Layer: Development and Assessment of the Munich Online Comprehensive Cancer Analysis Platform

Daniel Nasseh[1,2], PhD; Sophie Schneiderbauer[1,2], MSc; Michael Lange[1,3], Dipl Ing; Diana Schweizer[2], MSc; Volker Heinemann[1,2,4], Prof Dr; Claus Belka[1,4,5], Prof Dr; Ranko Cadenovic[6], MSc; Laurence Buysse[6], MSc; Nicole Erickson[2], MSc; Michael Mueller[7], PhD; Karsten Kortuem[7], MD; Maximilian Niyazi[4,5], Prof Dr; Sebastian Marschner[4,5], MD; Theres Fey[2], PhD

[1]Comprehensive Cancer Center Munich, Munich, Germany

[2]Comprehensive Cancer Center, Ludwig-Maximilians-Universität München, Munich, Germany

[3]Comprehensive Cancer Center, Technical University Munich, Munich, Germany

[4]German Cancer Consortium (DKTK, partner site Munich), German Cancer Research Center (DKFZ), Heidelberg, Germany

[5]Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany

[6]Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany

[7]University Eye Hospital Munich, Munich, Germany

Corresponding Author:
Daniel Nasseh, PhD
Comprehensive Cancer Center Munich
Pettenkoferstraße 8a
Munich, 80336
Germany
Phone: 49 (0)89 440077754
Email: Daniel.Nasseh@med.uni-muenchen.de

Abstract

Background: Many comprehensive cancer centers incorporate tumor documentation software supplying structured information from the associated centers' oncology patients for internal and external audit purposes. However, much of the documentation data included in these systems often remain unused and unknown by most of the clinicians at the sites.

Objective: To improve access to such data for analytical purposes, a prerollout of an analysis layer based on the business intelligence software QlikView was implemented. This software allows for the real-time analysis and inspection of oncology-related data. The system is meant to increase access to the data while simultaneously providing tools for user-friendly real-time analytics.

Methods: The system combines in-memory capabilities (based on QlikView software) with innovative techniques that compress the complexity of the data, consequently improving its readability as well as its accessibility for designated end users. Aside from the technical and conceptual components, the software's implementation necessitated a complex system of permission and governance.

Results: A continuously running system including daily updates with a user-friendly Web interface and real-time usage was established. This paper introduces its main components and major design ideas. A commented video summarizing and presenting the work can be found within the Multimedia Appendix.

Conclusions: The system has been well-received by a focus group of physicians within an initial prerollout. Aside from improving data transparency, the system's main benefits are its quality and process control capabilities, knowledge discovery, and hypothesis generation. Limitations such as run time, governance, or misinterpretation of data are considered.

(J Med Internet Res 2020;22(4):e16533) doi:10.2196/16533

KEYWORDS

oncology; database management systems; data visualization; usability

Introduction

In recent years, hospitals have been gradually transitioning from paper-based toward electronic documentation systems. The ongoing digitalization of routine data often results in the creation of large and comprehensive datasets, which can, under the right circumstances, open doors to further analysis and research [1,2]. These datasets are often considered to be "big data", not always due to their size but also due to their complexity. Therefore, analysis of such large datasets, especially in cases of secondary use, is most often the biggest challenge [3-6]. Hence, a solution for managing and extracting the knowledge hidden in the raw data in a meaningful way for all potential end users is necessary [7]. Importantly, without proper tools and methods, interest in the data can decline and a dataset's potential information content can remain unused and obscure. In general, standard statistical tools can be applied to tackle and analyze previously defined problems. However, smart business intelligence platforms such as Microsoft BI [8], QlikView [9], or, in a broader sense, SAP Hana [10] can allow for spontaneous and quick analysis of the data, and can therefore accelerate response time as well as facilitate response efforts. In 2012, a novel technology was introduced that utilizes such systems, termed "in-memory database." This technology allows for the swift handling of data without investing too much effort in data preparation in contrast to online analytical processing systems [11-13]. Aside from alleviating daily tasks, allowing users to immediately dive into data further offers transparency, and can thus support knowledge discovery, hypothesis generation, and translational research [14,15]. Therefore, we decided to create a system utilizing an in-memory database based on the software QlikView.

The development was tested within a focus group of 10 physicians at the University Hospital of the Ludwig-Maximilians-University (LMU) in Munich. The system was set up together with its partnering site, the University Hospital of the Technical University Munich, Rechts der Isar, with both sites sharing the Cancer Retrieval Evaluation and Documentation System (CREDOS) as their local tumor documentation system. These systems allow the institutes to compile and track most oncology-relevant data [16], including specific information about diagnoses (eg, grading or histology), therapies, and follow up for most oncology patients that were treated in the centers. In total, the database contains more than 1000 attributes about the patients themselves, their medical history, and tumor descriptives. The software is suitable for not only a specific tumor site but also for all solid and nonsolid tumors, resulting in a large, complex, and comprehensive oncology database. The software was established at the local sites in 2010 and now includes most tumor entities. By the end of 2018, the databases contained detailed information for more than 20,000 patients at each site.

The primary purpose of the database is to measure specific key performance indicators such as summarizing the number of cancer cases of a specific organ, which thus serve as indicators of the eligibility of the sites to become, or remain, a certified center according to the German ONKOZERT guidelines [17]. Gathering and maintaining all of the necessary information is not a trivial task and requires resources. Therefore, collecting data solely for the purpose of certifying the centers might not be worth the effort. Consequently, to increase the usability of these big datasets so that they may be harvested for further purposes, we created an analysis layer that provides visual access to this complex dataset and offers easy-to-use tools, which allow end users to immerse into, and analyze, the data.

The analysis layer was recently (October 2018) rolled out and tested within the previously mentioned focus group at the Comprehensive Cancer Center Munich

(CCCM)-LMU site. We refer to this analysis platform as the Munich Online Comprehensive Cancer Analysis platform (MOCCA). This paper describes how the system was rolled out at our site and explains the major components of this analysis platform, and should thus serve as an inspiration for other institutions interested in making their data more accessible and transparent. Focusing on how to manage and organize large sets of oncology-related data, we present a variety of innovative ideas in terms of browsing, handling, and visualizing large cohorts of medical data, while addressing challenges that arose during the development.

Methods

Administrative Arrangements

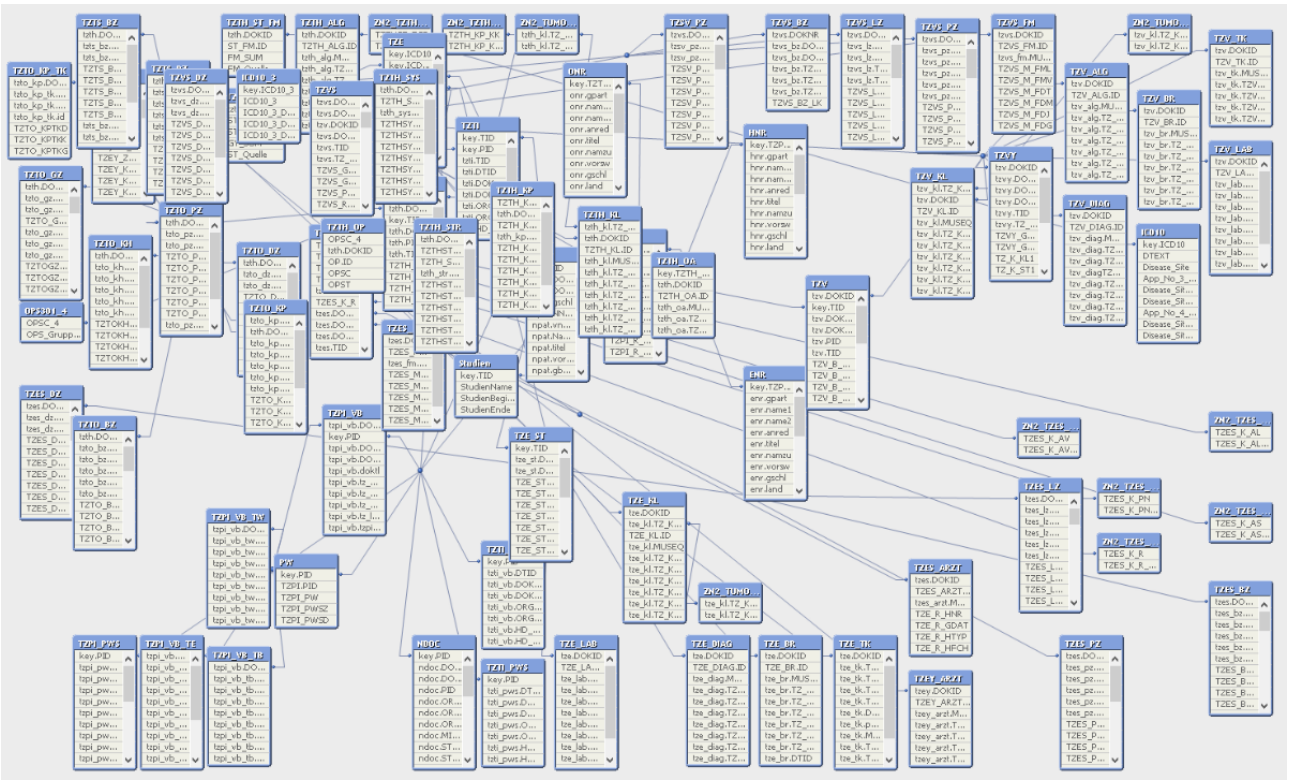
To provide general access to the data, the first step was to set up a server within the clinical intranet, which continuously runs an instance of the QlikView Enterprise Edition. The administrative user interface of the Enterprise Edition allows for data loading routines. Thus, a daily routine was established and implemented, which imports the whole CREDOS dataset into the MOCCA system. QlikView was chosen because it comes with a toolbox, enabling the construction of the contents that can be saved within a proprietary data container (*.qvw file) [9]. The toolbox also offers an intuitive user interface and its own scripting (partly Structured Query Language-based) and data-handling language. The program files themselves can not only be viewed within the developer toolbox but can also automatically be transformed into a fully functioning website and Web view [18]. Thus, the developer does not have to create an HTML framework or any other Web content, but can rather concentrate on the contents' objects themselves (eg, tables, graphs). End users can then comfortably access the resulting Web view via the URL and a Web browser without installing any additional software.

To control the Web access, we set up a connection to the hospital's active directory using a lightweight directory access protocol (LDAP) [19]. Consequently, users are required to register with their hospital intranet login to view the contents but do not have to create new accounts or passwords. To conform with laws of privacy protection as well as issues of governance, users can only view contents that they are allowed to see according to the hospital's policy. This assures that not every user who has access to the intranet can arbitrarily access the MOCCA platform. Instead, a user has to possess a single user license, which is a technical requirement before entering the platform. Detailed information about the permission system is described in the Permission System section below.

Embedding the Tumor Documentation Data

The next major step involved in setting up the platform was the conversion of the CREDOS data model into a new QlikView data model. This involved interconnecting all tables by common, unique identifiers, similar to a standard relational database. We tried to stick as closely to the original data model of CREDOS as possible, and only incorporated changes when needed. When completed, this step enabled the system to fully unveil the in-memory database capabilities after loading the data. At this point, an end user can select, via a mouse click, a specific data field within the Web view (eg, male or female within the gender field). After selecting the relevant data field, all other data tables react in real time and restrict the contents to the selected cohort. In this way, it is possible, in a matter of seconds, to create different data cohorts constricted by arbitrary filters. Most normal business intelligence or query systems offer query interfaces where different filters can be applied. However, this interface is not necessary in QlikView since all data tables and fields displayed within the Web view can directly be used as a cohort filter. QlikView offers an additional functionality through which it shows the data model. A current view of the data model in our latest release is shown in Figure 1, which highlights the complexity of what is involved in this seemingly "simple" step.

Figure 1. Screenshot of the current data model illustrating its complexity. The tables are framed in blue, displaying all of their included fields. The tables are linked via specifically chosen key attributes (eg, patient, tumor, or document ID).



We first organized both the data and the analysis layer in the same *.qvw data file. Subsequently, we decided to strictly separate the data modeling and analysis contents into two segregated QV files. The primary cleaning and preprocessing of the CREDOS data model into QlikView is done in the base module (base.qvw), while more advanced preprocessing as well as further analyses are done in the second module (analysis.qvw). This allowed us to simultaneously improve the data model along with the analytical tools and modules.

Creating the Analytics Frontend

As mentioned above, the CREDOS data model is complex, comprising more than 1000 individual data fields, which themselves can have hundreds of different data values (eg, the data field diagnosis can have hundreds of different International Statistical Classification of Diseases and Related Health Problems [ICD]-10 codes [20]). As our main aim was to make this platform accessible for clinicians, innovative ideas on how to manage this large degree of complexity in an accessible and useful manner were necessary. We therefore structured the analysis tool into six data-related categories with their respective subcategories, which are listed in Textbox 1.

Textbox 1. Data-related categories and structure of the analysis tool.

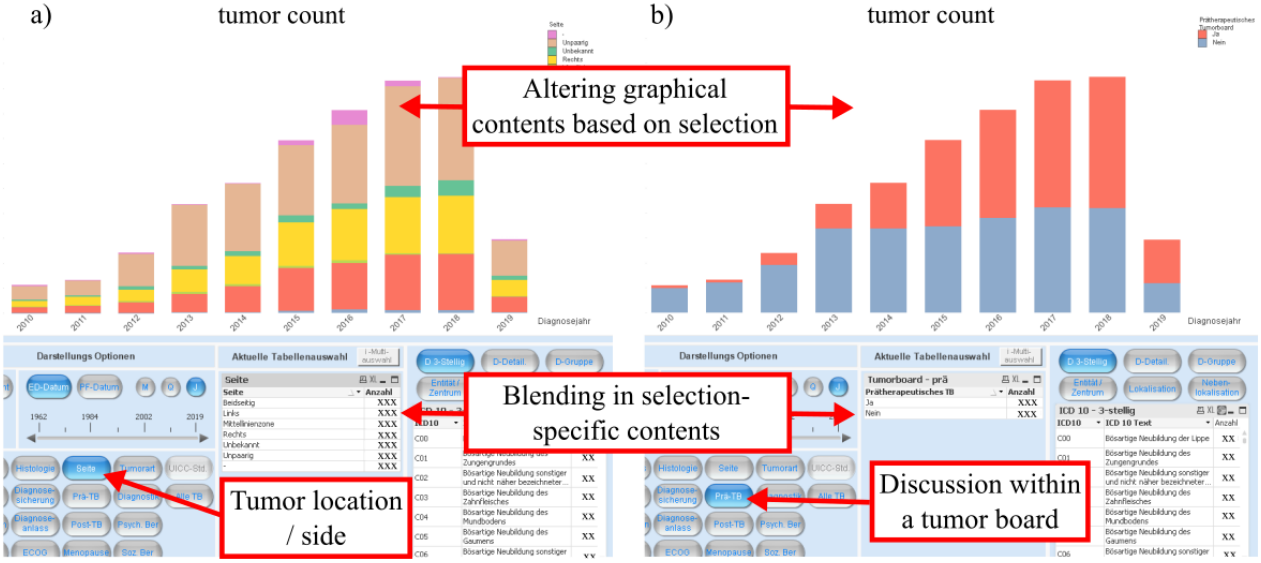
- First assessment
  a. Diagnosis
  b. TNM (tumor-node-metastasis; important tumor staging system)
  c. Classifications
- Patient-based/related data
  a. Cohort view
  b. Single patient view
- Therapies
  a. General information
  b. Operations
  c. Systemic therapies
  d. Radiation
- Progression (follow up)
- Trial metadata
- Survival

For most categories (including subcategories), we implemented detailed and comprehensive views, including embedded tools that enable easily browsing and visualizing category-specific contents. The views were designed as different tabs, analogous to a standard website navigation bar (see Multimedia Appendix 1). Although extensive contents were created for most of these categories and subsequently included in the first version of the platform, the single patient view, as well as the trial metadata view, remain in a conceptual stage and have therefore been deactivated for the first release.

For each of the remaining categories shown in Textbox 1, our goal was to fill all related data attributes into a QlikView tab, which takes up one screen (optimized for a 1680 × 1050-pixel screen resolution). Owing to the overwhelming amount of data attributes, loading the tabs with data tables would have exhausted the pixel space within the tabs multiple times. Thus, to overcome the complexity of storing a lot of information within a small space while simultaneously ensuring that it appears in an easily comprehensible manner, we enabled the end user to hide or display specific data objects such as tables or graphs by utilizing variables, which are themselves controllable (eg, by buttons) within the front end.

As shown in Figure 2, 18 buttons are displayed, which refer to 18 different data attributes related to the first diagnosis of a patient (eg, “vital status,” “histology,” or “side of tumor”). By clicking these buttons, a table showing the selected contents of these data fields as well as a corresponding graph are displayed. Although it would be possible to show the tables of all 18 attributes at the same time, we believe it makes more sense to provide control to the end users as to what they want to focus on. This results in a dramatic reduction in complexity of onscreen contents since most of the data are hidden at most times, and because users have an easier time reading the data they feel is relevant. This technique has been applied not only for this example but also for numerous contents within the entire platform. Figure 2 also illustrates how we utilized extensive numbers of easily readable charts to increase the end users’ experience by data visualization. As most data are interconnected within the data model, the integrated in-memory database allows the user to click in any chart directly, and instantly alters all other graphical objects and tables within the platform accordingly.

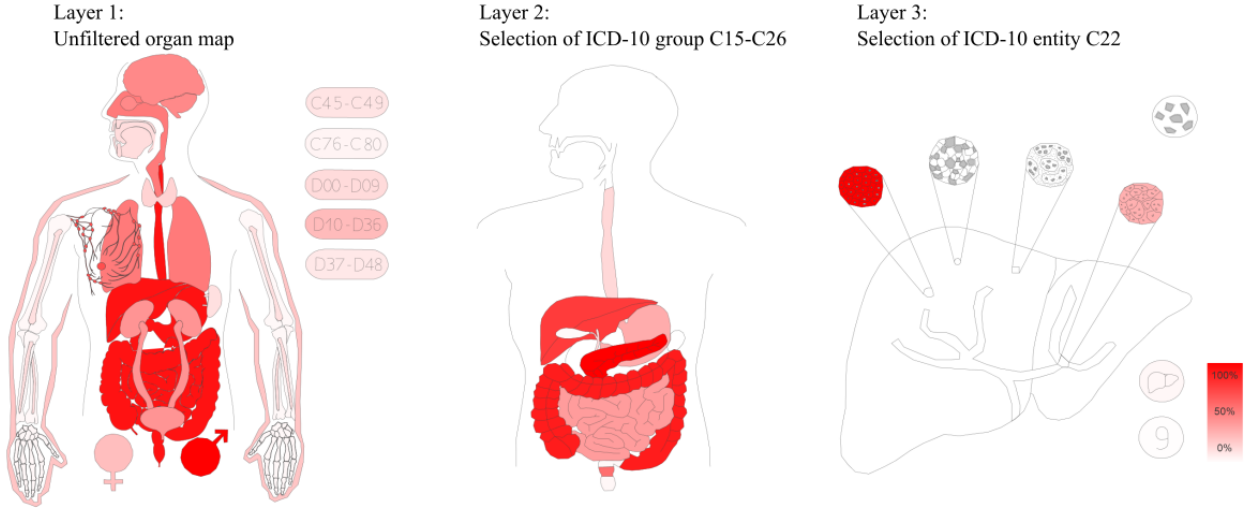
Figure 2. The image (screen language in German) displays two different selections within the first assessment (diagnosis) view. By simply clicking on different buttons, in this case "tumor location/side" in (a) and "discussion of a chosen case within a tumor board" in (b), it is possible to specifically select and examine the number of cancer cases with the chosen feature (in the chart stratified by date of diagnosis).



With regard to the graphical objects, in addition to reactive standard descriptive techniques such as bar charts or pie charts, we used a QlikView add-on (svgReader) [21], which allowed us to incorporate vector graphics and to color code them according to the selected data cohorts as well as underlying formulas such as relative abundance. Therefore, we were able to create innovative and interactive modules that increased the usability, readability, and interest of the data contents. For example, we created scalable vector graphics (SVGs) for the geographic locations of patients, a three-layer organ map including all oncology-relevant ICD-10 diagnosis codes, a map displaying the spread of metastases, and a two-layer map showcasing the areas of radiation. Within the three-layer organ map, we specifically combined the SVG map technique with the content-hiding technique described above.

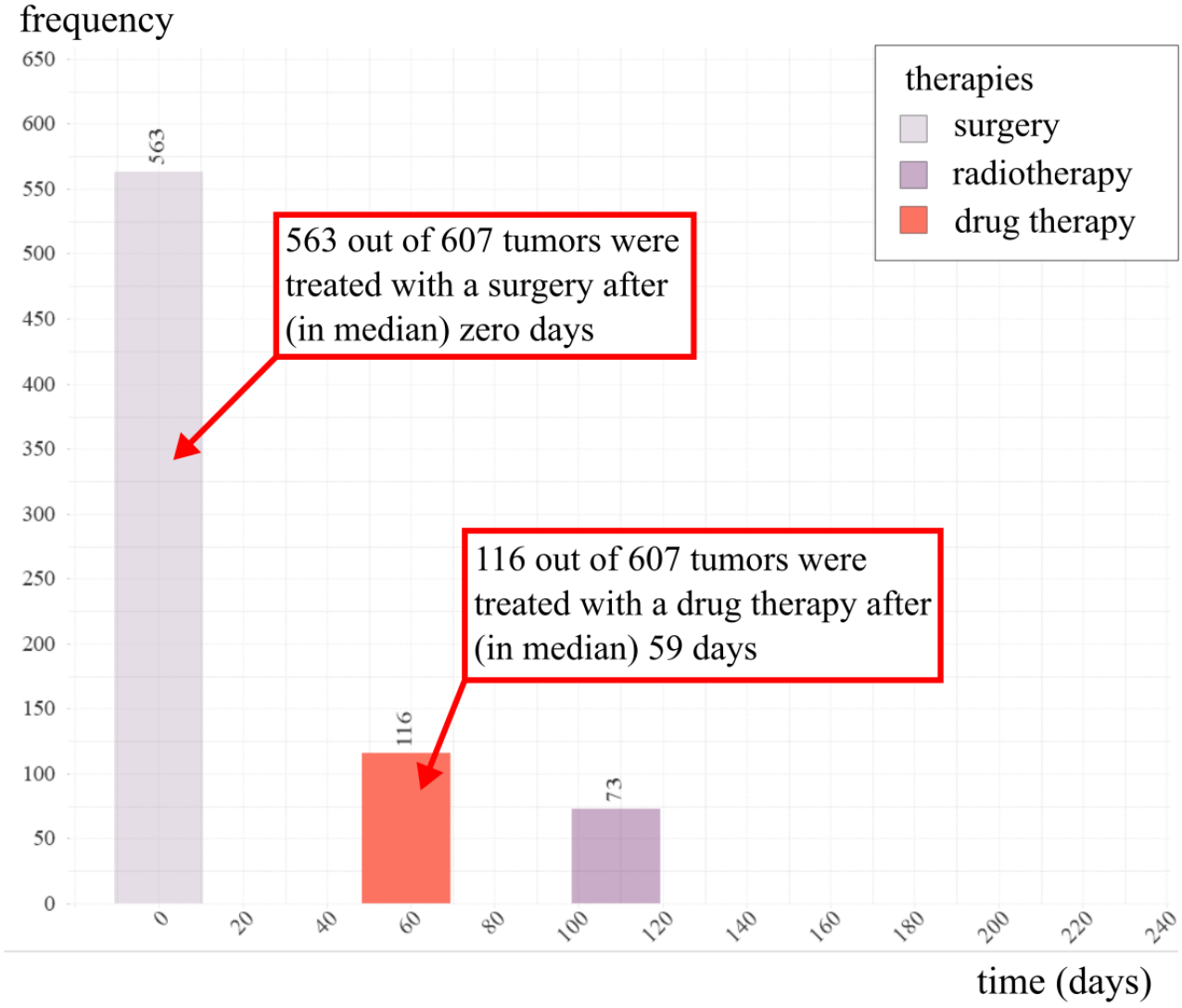
Figure 3 shows the first level of this three-level organ map. When a user clicks on a component (ie, an organ group) they switch to the specific ICD-10 group, which also sets a selection filter to this group for the whole platform. Accordingly, the image changes from the first layer to the chosen ICD-10 group. From this second layer of detail, the user can select a single organ (third layer), which then displays the components of the organ. The segments of the graphics are colored according to the relative abundance of tumor cases; for example, Figure 3 layer 1 shows that gastrointestinal tumors were predominantly treated (or rather documented) at our hospital. In general, for all of these maps, a more saturated color indicates a higher degree of documentation (linear in proportion to the data). In other words, if an organ consisted of two segments and only one segment had 100 documented cases while the other segment only had 50 documented cases within the system, the first segment would be fully color-saturated (100% on the red scale), while the other segment would only be color-saturated by 50%.

Figure 3. SVG-based organ map. Based on the diagnosis code (ICD-10), the map displays the relative amount of documented cases to each other as well as selected (via mouse clicks) ICD-10 groups (layer 2), or even specific organs (layer 3). In this example, the gastrointestinal subgroup had been selected by clicking within layer 1 (C15-C26), followed by clicking on the liver within layer 2, thereby restricting the cohort within module to C22* (liver) and displaying the relative abundance of affected segments within the liver (layer 3). A fully saturated color indicates the most commonly documented segments within the SVG, whereas lower levels of saturation linearly correspond to the amount of documentation for the associated segments. SVG: scalable vector graphic; ICD-10: International Statistical Classification of Diseases and Related Health Problems-10.



Next, we created two more complex, integrated modules. The first module displays therapy timelines and the second focuses on survival. The survival module and its implementation were previously presented at the International Conference on Informatics, Management, and Technology in Healthcare 2019 [22]. In terms of visualizing the sequential flow of therapies, it is possible to display, on a time axis, the median start time after diagnosis of specific therapies (eg, first radiation or surgery) of a chosen patient cohort. This provides an overview of typical patterns, and is meant to create scientific interest, while providing general information about the chosen cohort and how the patients were treated. Figure 4 shows an example of such a pattern for a selected cohort, where most patients typically started out with a surgery, followed by a drug therapy in some cases (59 days), and concluded by radiotherapy (109 days). Similar to all contents within the system, this module reacts to any frontend selections (clicks) and changes the pattern according to the chosen cohort within seconds.

Figure 4. Median start at which a tumor of a chosen cohort has been treated with a specific therapy. The x-axis shows the time (in days), while the y-axis displays how many tumors have been treated in this cohort. In this example, for a cohort of 607 tumors, 563 have been treated with surgery (median after zero days), 116 with drug therapy (median after 59 days), and 73 with radiotherapy (median after 109 days).



Permission System

As mentioned above, when granting access to the platform, we had to conform to European as well as to local Bavarian laws of privacy protection [23,24]. First, data access in the initial release of this platform had to be restricted to administrators as well as local clinicians. Giving access to external scientists or partners is currently not planned, as full anonymization of data, which would be required in such a case, does not seem feasible [23]. Technically, we utilized QlikView's capability to lock specific data field selections within a module on a per-user basis [25].

The Bavarian Law of Hospitals (Bayerisches Krankenhausgesetz (BayKRG) Art 27– Datenschutz (4)) served as the legal basis of our permission system [24]. This law states that physicians within a hospital are allowed to work with patient data (even in terms of research and without given consent) as long as they were involved in the process of care of that patient. According to the law, the physicians themselves are allowed to grant data access to other clinicians within the hospital as long as the data remain within the hospital. Hence, if a clinician asks for access to the platform, after signing the terms of agreement, we restrict their view to only patients formally in their care (or care unit). This was achieved by restricting the view based on filters within the data fields "organ (ICD-10)" and "treatment center." As an example, a physician in the field of pulmonology will only be able to view data and cases pertaining to the patients who visited the lung center. This same physician will not be able to view liver tumors, for example, even though both entities exist within our dataset.

This system functions for clinicians working for organ-specific treatment centers (eg, a women's hospital) or clinicians who are involved in the care of patients suffering from tumors of a specific organ. However, this system does not take into consideration clinicians working in interdisciplinary areas such as radiology. Therefore, we also included the possibility to restrict according to specific types of therapy (radiation, surgery, or drug therapy). Consequently, a radiologist with given permission would only be enabled to view data of patients who had indeed received radiotherapy and had been treated at the radiologist's center. Hence, our system provides the physicians no additional information than they would normally be allowed

to access. However, instead of having to sift through the information in all of the individual doctors' reports, they are now able to directly access the aggregated data extracted from these reports. Hence, QlikView basically facilitates analysis, and helps with visualizing the cohort for which they are already responsible.

Such permission arrangements explain how the system has currently been rolled out and how it has been accepted by the privacy protection commissioner of the hospital. We here turn to describing how the permission system should be extended in the future. According to the Bavarian Law of Hospitals, clinicians are also allowed to share data (eg, for research purposes) within the hospital. This is more of a governance problem and has not yet been implemented within our current system. However, for the sake of scientific progress, clinicians interested in organ-specific data should be allowed to request permission to access relevant data even for cases in which they were not part of the original patient care. Representatives of the organ center would be members of a committee that could initially process such requests. The request will then be referred to a board consisting of the initial committee members along with organ center–independent members of the overreaching comprehensive cancer center. If all parties of the committee accept the request, extended data access will be granted. However, extended data access would only be given in a pseudonymized form, since full data access seems only reasonable for clinicians directly involved with a patient's care.

The permission system, including the not-yet implemented extended data access, is organized via a permission table, which directly controls the contents that may be shown to a given user and within which timeframe. Table 1 provides an extract of our current permission table. In this example, the user "dnasseh" was given access to all patients of the lung tumor center, which all have a diagnosis code of C34. In contrast, the user "sopsch" would be allowed to see all radiated cases. As sopsch would be interested in liver cases as well, after a proposal, a committee would have granted them with extended permissions for liver-related data (C22) in a pseudonymized form (extended permissions are not yet implemented and are in a conceptual status).

Table 1. Example of a permission table for user access.

User	Diagnosis (ICD-10)[a]	Center	Radiation	Operation	System-therapy	Pseudo
dnasseh	C34	LTC[b]	*[c]	*	*	*
sopsch	*	*	STR[d]	*	*	*
sopsch	C22	*	*	*	*	PSD[e]

[a]ICD-10: International Statistical Classification of Diseases and Related Health Problems-10.

[b]LTC: lung tumor center.

[c]* represents a wildcard, meaning the user has full rights to view the contents within this column.

[d]STR: radiation.

[e]PSD: pseudonymized.

Results

We established a platform that is accessible through the clinical intranet via a Web browser and does not require the installation of additional software at the end users' sites. The data within the platform are updated daily, and provide preprocessed, compact visual access to the vast majority of the CREDOS contents. The platform can only be accessed after a single user license has been acquired. Based on this, the data from cohorts that the users can view are limited by a permission system, which was developed in parallel to the technical implementation. A nonlegal contract describes the rules for licensing and accesses to the platform. Before each login, a disclaimer has to be ticked (see Multimedia Appendix 1).

We structured the contents into six main categories, five of which (first assessment, patient baseline data, therapies, progression, survival) have been included in the first rollout version. To facilitate understanding of this complex system, Multimedia Appendix 1 shows screenshots (with labels translating the contents) of the layout for all of these categories and its included modules. Due to concerns of privacy protection, we blurred sensitive information within these screenshots.

As it is hard to describe the dynamic analytical possibilities of the platform, we provide a 15-minute-long commented video in Multimedia Appendix 2, displaying the real-time assembly of arbitrary data cohorts, dynamic interconnectivity between any data tables or data objects, as well as possibilities of the visual modules and its interaction with the other objects of the platform. This video was compiled to give an accurate impression about what the platform is capable to do.

Discussion

Feedback on the System

One of the major benefits of the MOCCA system is transparency. Until creation of the framework, clinicians themselves did not have the option to directly access aggregated CREDOS data. Instead, they had to assess individual patient records or send a request for help to the local information technology team. This process restricts interest in the data. Hence, for most physicians, the CREDOS dataset is comparable to a black box that is primarily used by documentation clerks and the information technology department. Thus, most clinicians were not aware of the rich contents of the database. The system was rolled out and evaluated at one of the partnering

sites (CCCM-LMU). After providing the doctors access to the platform within our prerollout phase, we received extensive feedback about the contents of our database.

This feedback reflected the high quality of the data, along with areas of further improvement for some aspects. Along these lines, due to the richness of charts and graphs, it is very easy to spot incorrectly documented information. Since a doctor can browse through the data, they will quickly realize if any data are missing or not documented in a correct manner. As an example, some wrongly documented dates could easily be spotted as they showed up within the therapy time chart as a negative time value on the x-axis. Spotting inconsistencies is of high relevance, since high data quality is one of the requirements for clinical research and is a precondition for clinical trials or network activities such as those of the national Network Genomic Medicine for lung cancer, which locally relies on correct, complete, and valid tumor documentation data [26]. For example, we received feedback that the method of documenting our radiotherapies was not fully in accordance to standards set for radiologists. Hence, we were able to address these issues within our database. Since the initial implementation of the MOCCA system, we have received valuable and extensive feedback about our data contents, not only from information technology or tumor document specialists but also directly from the physicians.

Improving the data quality itself is only one of the purposes of MOCCA; it can also support the control of processes of routine care and identify potential risks. Regarding this aspect of quality control, clinicians did not have the right tools to directly and quickly assess whether quality of care and associated processes were acceptable. As the centers are certified once a year by OnkoZert [17], they have to prepare and display specific key performance indicators such as the number of patient cases introduced to the tumor boards. Since substantial attention is focused on this event, quality problems that occur between these events might be identified too late and pose problems when attempting to overcome these shortcomings. Naturally, controlling should not be, and is not, limited to the annual certification events. The different organ centers regularly request information about the CREDOS contents (eg, amount of cases with a specific therapy). This information can be supplied by the information technology staff or documentation clerks. However, with more classical statistics tools, scripts would have to be written and altered for every analysis request. Hence, a capable business intelligence platform such as the MOCCA system provides a more direct approach, while avoiding unnecessary waiting times and communicative misunderstandings. The process is further supported by the in-memory technology, which allows real-time analyses of arbitrary data cohorts [11] and automatic daily data updates to ensure that the information available to interested physicians is always current.

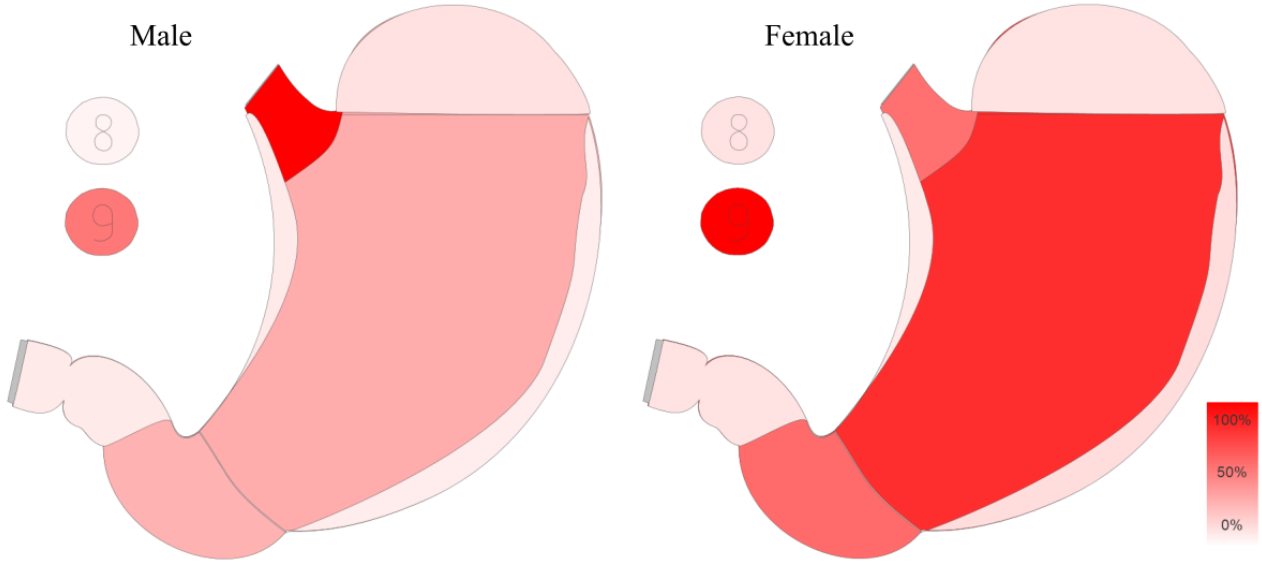
Strengths and Limitations

When critically assessed, a potential shortcoming of the system is misinterpretation caused by visual inspection of data, without considering the influences of confounding factors, sample size, and sample bias [27]. For example, when viewing the documented total occurrences of cancer cases of the CCC-LMU (Figure 2), it can be noted that the number of tumor cases per year has been rising. However, this does not mean that the occurrence of tumor cases is rising on a population level, but rather that the completeness of documentation or the frequency of patients seeking treatment for these types of tumors has increased throughout the years. As long as it is understood that MOCCA is not an epidemiological overview, but rather a local view on the present documented cancer data, we believe that a system like MOCCA can support analysis of local data and create exciting research opportunities.

Although our system presents and analyzes data in graphic detail, we recommend that the results and conclusions mined from our system should always be examined with the support of statisticians, medical computer scientists, and in comparison to larger datasets (eg, epidemiological registries) [28]. In fact, in line with Bauer et al [29], we included a disclaimer within the system that warns of misinterpretation and indicates the necessity to consult information technology or statistical professionals [29] (see Multimedia Appendix 1). Additionally, the system offers tooltips, text fields, and popups explaining most of the contents, and even provides direct references when using formulas, mostly within the survival module [22]. Since precautionary methods might be ignored, we set up another security measure in which we only hand out licenses after a direct tutorial session, discussing not only the benefits but also the limitations and dangers of the system.

In addition to the means and methods with which this system can support research, the system can also quickly and easily provide numbers for formulating scientific proposals. Furthermore, it can be used to discover as-yet-unknown information, also referred to as knowledge or data discovery, that can contribute to creating research ideas by quickly browsing through the data (hypothesis generation) [14,30]. For example, the organ map displays the relative occurrence of a tumor type in a specific location, which was treated at the center. As we are able to alter the data cohort arbitrarily, we can switch from male patients to female patients in real time, which changes the coloration of the organ map (Figure 5).

Figure 5. Comparison of two different cohorts within the organ map module. Different patterns of occurrence for female and male stomach cancer cases are evident. The color saturation is linear to the occurrence, with 100% saturation being the most affected segment.



The benefits of such data visualization and business intelligence have been previously discussed in multiple contexts [31-33]. However, this is one more example for the effectiveness of such systems, which is based on the balance shift between perception and cognition [34]. In this way, we can quickly realize the most significant changes within both graphics. We provide an example in Figure 5, in which it appears easy to identify that the cardia (opening of the stomach) is more affected among males than among females. Such a finding can result in the formulation of new research ideas and support hypothesis generation [30]. In this context, the second big limitation within our system is governance. Physicians employed at a cooperating center (at the same site) should not be able to examine data mined from patients of another center. This is particularly tricky when granting rights to interdisciplinary physicians like radiologists, whose patients would have also been treated at other organ centers. Hence, publications using data derived from multiple sites can result in conflicts. To address and deflate this concern, it is important to point out that MOCCA only facilitates an aggregated data view, but will not give more information than physicians can already find when looking into the individual digital patient files of their aggregated system for whom they were granted access according to hospital policy. Consequently, all information within CREDOS could also be found within a well-made shadow database. The difference is that the CCCM at both sites comprises staff that are professional documentation specialists, whereas associated centers do not necessarily have these resources. Often, these shadow databases are created with the support of students who lack the experience, education, and training for creating a clean dataset. However, this should not be generalized, and having individual databases is important for research, especially since these databases might contain additional information not found within CREDOS. Nevertheless, a business intelligence tool such as MOCCA can only improve these already existing systems and possibly make them, or large parts of them, redundant, resulting in a reduction of necessary capacities as well as an increase of professionalism in terms of documentation, data quality, and transparency. In this regard, we received positive feedback in our test phase when some of the physicians realized that one of their databases, which was created parallel to our system and is also used for annual audits, might be superfluous since all data needed can be supplied by the larger CREDOS system. This could have a direct impact and release bound capacities, reducing the need for each individual center to create their own systems.

In terms of measurable benefit, as the system was only recently released, the physicians have not yet utilized it for extensive research projects. Nevertheless, it serves as a quick help for everyday routine requests and is already an essential part of our annual audits. In general, we can summarize that the feedback of the prerollout was primarily positive, and it is safe to say that it generated interest in, and led to voluntary confrontation of doctors with the data, which in turn opened the door to translational interaction, improved data quality, and possibly research. An objective measurement about the benefit of the tool (eg, based on surveys, citations, or login frequencies) is planned for the future.

Conclusions and Prospects

As for future perspectives, due to the success and mostly positive feedback of our testing rollout, we aim to launch MOCCA at our partnering site (CCCM-Technical University Munich). In terms of contents, we would like to add further standard statistics and integrate some additional modules such as a single patient view. Moreover, these innovative data management techniques, and the handling of permissions and data governance, which is known as a major hurdle for many health-related projects, should serve as an inspiration for similar projects at other sites both nationally and internationally [35].

Acknowledgments

The project was supported by funding from Deutsche Krebshilfe.

Authors' Contributions

The analysis layer was conceptualized, designed, and partly implemented by DN. The initial data model was created by ML. Implementation of the analysis layer as well as the data model was completed by SS. DS provided support with the survival module. The manuscript was majorly written by DN and SS, and NE refined the manuscript content both linguistically and from a larger perspective. Server access was supported by KK and MM. SM and MN contributed to the project as important beta testers (physicians). RC and LB (students) created vector graphics. With leading positions within the CCCM-LMU, VH, TF, and CB supported the idea by providing the main authors with the time to commit to this work. As the CCCM-LMU coordinator, TF supported DN in creating the governance system, which was implemented by SS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview over all components (with translations) of the given MOCCA system.
[PDF File (Adobe PDF File), 4075 KB - jmir_v22i4e16533_app1.pdf]

Multimedia Appendix 2
Video summarizing the capabilities of the platform.
[MP4 File (MP4 Video), 153704 KB - jmir_v22i4e16533_app2.mp4]

References

1. Topol E. The Creative Destruction Of Medicine: How The Digital Revolution Will Create Better Health Care. New York: Basic Books; 2020.
2. Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med 2009;48(1):38-44. [Medline: 19151882]
3. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3. [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
4. Burghard C. Big data and analytics key to accountable care success. IDC Health Insights 2012:1-9.
5. Feldman B, Martin E, Skotnes T. Dr Bonnie 360° - Business Development for Digital Health. 2012. Big data in healthcare hype and hope URL: https://pdfs.semanticscholar.org/241e/3e0657ff97d5e2beafaf5eb87085eef40cf8.pdf?_ga=2.21148283.1937729036.1582708542-1457234903.1582708542 [accessed 2020-02-29]
6. Fernandes L, O'Connor M, Weaver V. Big data, bigger outcomes: Healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis. J AHIMA 2012 Oct;83(10):38-43; quiz 44. [Medline: 23061351]
7. Girard J, Girard J. Defining knowledge management: Toward an applied compendium. Online J Appl Knowl Manag 2015;3(1):1-20 [FREE Full text]
8. Lachev T. Applied Microsoft Power Bi (3rd Edition): Bring Your Data To Life!. Atlanta: Prologika; 2018.
9. QlikQ. 2014 Jan. What makes QlikView Unique - A Qlik White Paper URL: http://www.bitechnology.com/wp-content/uploads/2017/01/WP-What-Makes-QlikView-Unique_EN.pdf, [accessed 2018-11-16]
10. Färber F, Cha SK, Primsch J, Bornhövd C, Sigg S, Lehner W. SAP HANA database. SIGMOD Rec 2012 Jan 11;40(4):45. [doi: 10.1145/2094114.2094126]
11. Berkowitz B, Simhadri S, Christofferson P, Mein G. Google Patents. 2002. In-memory database system URL: https://patents.google.com/patent/US6457021B1/en [accessed 2020-02-26]
12. Sevilla M. CapGemini. 2011 Sep 27. OLAP databases are being killed by In-Memory solutions URL: https://www.capgemini.com/2011/09/olap-databases-are-being-killed-by-in-memory-solutions/ [accessed 2020-02-26]
13. Chamoni P, Gluchowski P. Entwicklungslinien und Architekturkonzepte des On-Line Analytical Processing [Development guidelines and architecture concepts of On-Line Analytical Processing]. In: Analytische Informationssysteme: Business Intelligence-technologien Und -Anwendungen (German Edition). Berlin, Heidelberg: Springer; 2020:143-176.
14. Fredriksson K. http://drmartin.dbxl.de. 2011 Jul. Data Discovery - BI im Zeitalter von Apps und Social Media Data Discovery - BI in the age of apps and social media URL: https://web.archive.org/web/20160528145832/http://drmartin.dbxl.de/Ausgabe-72#NL2245 [accessed 2018-11-19]
15. van der Lans RF. Data virtualization for business intelligence systems: revolutionizing data integration for data warehouses. San Francisco: Morgan Kaufmann; 2012.

16. Voigt W, Steinbock R, Scheffer B. CREDOS 3.1 ein Baukasten zur Tumordokumentation für Epidemiologische-, Klinische,-Tumorspezifische-und Zentrums-register integriert in das Kis Sap/r3 Is-h: Po344 [CREDOS 3.1 a modular system for Tumor documentation for Epidemiological, Clinical, Tumor-Specifical and and Center-Register integrated into the HIS SAP/R3 IS-H]. Onkologie 2010;33:52.
17. OnkoZert. URL: http://www.onkozert.de/ [accessed 2020-02-26]
18. Redmond S. Qlikview Server And Publisher. Birmingham: Packt Publishing; 2014.
19. Zeilenga K. IETF Request for Comments. 2006. Lightweight directory access protocol (ldap): Technical specification road map URL: http://www.ietf.org/rfc/rfc4510.txt [accessed 2020-02-26]
20. DIMDI. 2019. ICD-10-GM Version 2019, Systematisches Verzeichnis, Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision, Stand: 21.September 2018 URL: https://www.dimdi.de/dynamic/de/klassifikationen/downloads/?dir=icd-10-gm/version2019 [accessed 2019-03-26]
21. Munz B. Github. svgReader-QV11 URL: https://github.com/brianwmunz/svgReader-QV11 [accessed 2020-02-26]
22. Schneiderbauer S, Schweizer D, Fey T, Nasseh D. Dynamic Survival Analysis Using In-Memory Technology. Stud Health Technol Inform 2019 Jul 04;262:79-82. [doi: 10.3233/SHTI190021] [Medline: 31349270]
23. Rumbold JMM, Pierscionek B. The Effect of the General Data Protection Regulation on Medical Research. J Med Internet Res 2017 Feb 24;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
24. Bayern.Recht. Bayrisches Krankenhausgesetz [Bavarian law of hospitals] URL: http://www.gesetze-bayern.de/Content/Document/BayKrG-27 [accessed 2020-02-26]
25. QlikQ. Qlikview Security Concept URL: https://help.qlik.com/en-US/qlikview/November2017/Subsystems/Client/Content/Security.htm [accessed 2019-05-24]
26. Büttner R, Wolf J, Kron A, Nationales Netzwerk Genomische Medizin. [The national Network Genomic Medicine (nNGM): Model for innovative diagnostics and therapy of lung cancer within a public healthcare system]. Pathologe 2019 May;40(3):276-280. [doi: 10.1007/s00292-019-0605-4] [Medline: 31101971]
27. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986 Sep;15(3):413-419. [doi: 10.1093/ije/15.3.413] [Medline: 3771081]
28. Strigini L. limiting the dangers of intuitive decision making. IEEE Softw 1996;13(1):101-103. [doi: 10.1109/52.476293]
29. Bauer CR, Knecht C, Fretter C, Baum B, Jendrossek S, Rühlemann M, et al. Interdisciplinary approach towards a systems medicine toolbox using the example of inflammatory diseases. Brief Bioinform 2017 May 01;18(3):479-487 [FREE Full text] [doi: 10.1093/bib/bbw024] [Medline: 27016392]
30. Ligeti B, Pongor S. Biomedical data mining, data-driven hypothesis generation in cancer research. URL: http://pongor.itk.ppke.hu/sites/default/files/pdf/biomed_data_mining.pdf [accessed 2020-02-26]
31. Kortüm K, Müller M, Hirneiß C, Babenko A, Nasseh D, Kern C, et al. [Smart eye data: Development of a foundation for medical research using Smart Data applications]. Ophthalmologe 2016 Jun;113(6):469-477. [doi: 10.1007/s00347-016-0272-2] [Medline: 27222127]
32. Stadler JG, Donlon K, Siewert JD, Franken T, Lewis NE. Improving the Efficiency and Ease of Healthcare Analysis Through Use of Data Visualization Dashboards. Big Data 2016 Jun;4(2):129-135. [doi: 10.1089/big.2015.0059] [Medline: 27441717]
33. Bonney W. Applicability of Business Intelligence in Electronic Health Record. Procedia - Social and Behavioral Sciences 2013 Feb;73:257-262. [doi: 10.1016/j.sbspro.2013.02.050]
34. Few S. Data visualization for human perception. In: Encyclopedia Of Human-Computer Interaction, 2nd edition. Denmark: Interaction Design Foundation; 2014.
35. Souhami R. Governance of research that uses identifiable personal data. BMJ 2006 Aug 12;333(7563):315-316 [FREE Full text] [doi: 10.1136/bmj.333.7563.315] [Medline: 16902204]

Abbreviations

CCCM: Comprehensive Cancer Center Munich
CREDOS: Cancer Retrieval Evaluation and Documentation System
ICIMTH: International Conference on Informatics, Management and Technology in Healthcare
LDAP: lightweight directory access protocol
LMU: Ludwig-Maximilians University, Munich
LTC: Lung Tumor Center
MOCCA: Munich Online Comprehensive Cancer Analytics
PSD: Pseudonymized
STR: Radiation
SVG: scalable vector graphics

Edited by G Eysenbach; submitted 11.10.19; peer-reviewed by P Morakis, J Brooke; comments to author 28.12.19; revised version received 10.01.20; accepted 24.01.20; published 17.04.20.

Please cite as:
Nasseh D, Schneiderbauer S, Lange M, Schweizer D, Heinemann V, Belka C, Cadenovic R, Buysse L, Erickson N, Mueller M, Kortuem K, Niyazi M, Marschner S, Fey T
Optimizing the Analytical Value of Oncology-Related Data Based on an In-Memory Analysis Layer: Development and Assessment of the Munich Online Comprehensive Cancer Analysis Platform
J Med Internet Res 2020;22(4):e16533
URL: https://www.jmir.org/2020/4/e16533
doi:10.2196/16533
PMID:32077858

Original Paper

Developing an Intranet-Based Lymphedema Dashboard for Breast Cancer Multidisciplinary Teams: Design Research Study

Anna Janssen[1,2,3], BA, MPub, PhD; Candice Donnelly[1,2,3], BSc, MPH; Judy Kay[4], PhD; Peter Thiem[5], BSc, BA; Aldo Saavedra[5,6], BSc (Hons), PhD; Nirmala Pathmanathan[7,8], BSc (Med), MBBS; Elisabeth Elder[7,8], MBBS, PhD; Phuong Dinh[8,9], MBBS; Masrura Kabir[7,8], MASc; Kirsten Jackson[1,2,3], BA, MPH; Paul Harnett[2,8,9], MBBS, PhD; Tim Shaw[1,3], BSc (Hons), PhD

[1]Research in Implementation Science and eHealth Group, Faculty of Health Sciences, The University of Sydney, Sydney, Australia
[2]Sydney West Translational Cancer Research Centre, Sydney, Australia
[3]Charles Perkins Centre, The University of Sydney, Sydney, Australia
[4]Faculty of Engineering, The University of Sydney, Sydney, Australia
[5]Sydney Informatics Hub, The University of Sydney, Sydney, Australia
[6]Faculty of Health Sciences, The University of Sydney, Sydney, Australia
[7]Westmead Breast Cancer Institute, Sydney, Australia
[8]Sydney Medical School, The University of Sydney, Sydney, Australia
[9]Crown Princess Mary Cancer Centre, Westmead Hospital, Sydney, Australia

Corresponding Author:
Anna Janssen, BA, MPub, PhD
Research in Implementation Science and eHealth Group
Faculty of Health Sciences
The University of Sydney
Charles Perkins Centre (D17)
Sydney, 2006
Australia
Phone: 61 9036 9406
Email: anna.janssen@sydney.edu.au

Abstract

Background: A large quantity of data is collected during the delivery of cancer care. However, once collected, these data are difficult for health professionals to access to support clinical decision making and performance review. There is a need for innovative tools that make clinical data more accessible to support health professionals in these activities. One approach for providing health professionals with access to clinical data is to create the infrastructure and interface for a clinical dashboard to make data accessible in a timely and relevant manner.

Objective: This study aimed to develop and evaluate 2 prototype dashboards for displaying data on the identification and management of lymphedema.

Methods: The study used a co-design framework to develop 2 prototype dashboards for use by health professionals delivering breast cancer care. The key feature of these dashboards was an approach for visualizing lymphedema patient cohort and individual patient data. This project began with 2 focus group sessions conducted with members of a breast cancer multidisciplinary team (n=33) and a breast cancer consumer (n=1) to establish clinically relevant and appropriate data for presentation and the visualization requirements for a dashboard. A series of fortnightly meetings over 6 months with an Advisory Committee (n=10) occurred to inform and refine the development of a static mock-up dashboard. This mock-up was then presented to representatives of the multidisciplinary team (n=3) to get preliminary feedback about the design and use of such dashboards. Feedback from these presentations was reviewed and used to inform the development of the interactive prototypes. A structured evaluation was conducted on the prototypes, using Think Aloud Protocol and semistructured interviews with representatives of the multidisciplinary team (n=5).

Results: Lymphedema was selected as a clinically relevant area for the prototype dashboards. A qualitative evaluation is reported for 5 health professionals. These participants were selected from 3 specialties: surgery (n=1), radiation oncology (n=2), and

occupational therapy (n=2). Participants were able to complete the majority of tasks on the dashboard. Semistructured interview themes were categorized into engagement or enthusiasm for the dashboard, user experience, and data quality and completeness.

Conclusions: Findings from this study constitute the first report of a co-design process for creating a lymphedema dashboard for breast cancer health professionals. Health professionals are interested in the use of data visualization tools to make routinely collected clinical data more accessible. To be used effectively, dashboards need to be reliable and sourced from accurate and comprehensive data sets. While the co-design process used to develop the visualization tool proved effective for designing an individual patient dashboard, the complexity and accessibility of the data required for a cohort dashboard remained a challenge.

(J Med Internet Res 2020;22(4):e13188) doi:10.2196/13188

KEYWORDS

eHealth; clinical informatics; human-centered design; data visualization

Introduction

Background

Over the last decade, the quantity of clinical data collected within the health sector has increased exponentially. In parallel, the adoption of digital health, such as electronic health records (EHRs) to collect and aggregate clinical data, has increased. The widespread use of EHRs has the potential to make clinical data more readily accessible to individual health professionals. It also presents opportunities for effective downstream use of clinical data, including quality improvement activities [1], self-directed performance review, personalized professional development [2], and timely clinical research [3].

Despite the proliferation of EHRs, quality and completeness of the data remains a challenge [4,5]. The literature suggests that poorly designed and implemented EHRs contribute to the collection of low-quality clinical data [6]. Another barrier to quality data collection is the lack of interoperability between digital health systems resulting in duplication of data entry and data access issues [7,8], which can result in both cost and workload inefficiencies [9]. Strategies to counteract these barriers, include the redesign of EHR systems to encourage health professionals to enter data at the point of care delivery [10]. Beyond this, there is a need to motivate health professionals to record data consistently and accurately.

One recognized approach to motivate health professionals to collect accurate, high-quality data is to make data visible and useful for clinical practice. However, there is currently a paucity of research on providing health professionals with data in a meaningful way for care delivery. A significant portion of the data entered by health professionals is utilized for mandated reporting and gives limited immediate value to health professionals [4]. The use of clinical dashboards that visually represent such data could provide greater value to health professionals for informing clinical decision making and could also enable performance review [11,12].

Currently, there is no literature on the use of clinical dashboards for data feedback to health professionals specializing in the delivery of breast cancer care. This is surprising given the global burden of breast cancer. In Australia, it is estimated that over 17,210 people will be diagnosed with breast cancer in 2020 [13]. During 2009-2013, individuals diagnosed with breast cancer had a 90% chance of 5-year survival, highlighting the importance of improving the quality of life for patients following breast cancer treatment [13]. In Australia, as in many other countries, breast cancer treatment plans are developed by multidisciplinary teams (MDTs); a team of health professionals, such as breast surgeons, medical oncologists, radiation oncologists, pathologists, radiographers, nurses, and occupational therapists (OTs), together delivers treatment across the care continuum [8].

The research described in this paper focuses on creation of a prototype dashboard for a breast cancer MDT. The specific clinical focus selected for the prototype dashboard was the treatment side effect, lymphedema. This is defined as excess fluid accumulation in a limb causing significant reduction in the quality of life [14]. Of patients treated for breast cancer, approximately 20% will undergo an axillary dissection. Up to 3% of patients undergoing sentinel lymph node biopsy and 10-15% of patients who receive axillary radiotherapy treatment develop lymphedema [15]. The data relevant to this cohort of patients at risk of lymphedema typically come from multiple heterogenous data sources. Therefore, a clinical dashboard that integrates and represents these multiple sources of breast cancer data could assist in the early identification of patients at risk of developing lymphedema, which could significantly improve the quality of life for a large number of people in Australia and globally.

There is little published literature on clinical dashboard use to visualize aggregated data sets to health professionals. A literature review reported how medical dashboards offering health professionals immediate access to critical patient information can improve adherence to quality of care guidelines and may help improve patient outcomes [12]. Another review of the literature indicated that the use of visualization tools in intensive care unit could decrease time spent on gathering data and improve compliance with safety guidelines [16]**.** However, further high-quality detailed research studies are needed to provide evidence of their efficacy and establish guidelines for their design.

Aims

Dashboards have been effectively used in other industries such as the learning sciences, for feeding back data to both learners and educators to enable more personalized education and training [12-18]. The core of effective design of such dashboards is to follow best practice in user-centered design, including research into user needs, and iterative design and evaluation. The aim of this study was to develop a prototype clinical

dashboard for breast cancer MDTs through a co-design methodology and test the prototypes with members of the MDT.

Methods

Study Design

The study was informed by a co-design framework [19]**.** This actively engaged end users throughout the project cycle. This process was based on the identification of clinical champions, who shared ownership and support for the methodologies and solutions developed. The study site was the breast cancer department of a major metropolitan hospital in New South Wales, Australia, with a case load of approximately 450 new breast cancer patients per year.

Data integrated and utilized for this study were sourced from routinely collected clinical data sets, including a bespoke breast cancer Structured Query Language EHR for the patient's administrative treatment and follow-up data and a bioimpedance spectroscopy machine extract for lymphedema data.

Qualitative methods (described in the *Evaluation* section of this paper) were used to evaluate dashboard static mock-ups and the interactive prototypes.

Permission to conduct this study was received from the Western Sydney Local Health District Human Research Ethics Committee.

Co-Design Process

This exploratory phase of the project aimed to identify methods for improving accessibility of EHR data through visualization platforms such as dashboards. To determine what clinical data were both clinically meaningful and feasible to visualize in a dashboard, 2 focus groups were held. All members of a breast cancer MDT at the study site were invited to the focus groups by the Chair of the MDT. A breast cancer consumer was invited through the National Breast Cancer Foundation. Both focus groups were attended by the consumer representative (n=1) and health professionals (n=33) across a range of disciplines including surgical, medical, and radiation oncology and nursing.

During the first focus group, clinical areas and requirements for data visualization were identified. The aim of the second focus group was to identify an appropriate example clinical area for the focus of a prototype dashboard. MDT members nominated lymphedema as a candidate clinical focus because of its clinical relevance for MDT members and the existing collection of data sources relevant to lymphedema diagnosis and management.

Changes to patients' lymphedema index (L-Dex) can be monitored using a bioimpedance spectroscopy machine that measures the level of extracellular fluid taken [20]. This L-Dex reading is available to the health professional at the time of the assessment. Health professionals that treat lymphedema, such as OTs, have access to the history of patients' L-Dex measurements within a paper record. However, OTs have limited access to breast cancer treatment information and rely on patient recall. Conversely, members of the MDT, such as oncologists and nurses, have limited access to L-Dex measurements and treatment information outside of the individual clinician's specialty.

The translation of the requirements identified in the focus group into a visualization dashboard required close consultation with subject matter experts from the study site. The project team convened an Advisory Committee to oversee the development of the prototype dashboards (Table 1). The Advisory Committee met fortnightly for the duration of the project.

These meetings facilitated building a mutual trust and a shared language between the project team and clinical members. Committee members provided advice on the identification, mapping, and access of relevant data sources, as well as issues surrounding data quality and data completeness and appropriate visualization styles. Committee members also established a set of high-level goals that drove the development of the dashboard interfaces. These were used to create a set of tasks for the evaluation of the prototype dashboard (Table 2). It was noted that some of the tasks pertained to the cohort of breast cancer patients, while others were specific to individual patient data. For this reason, the Advisory Committee decided to create 2 dashboards, 1 to interrogate data of a cohort of patients and 1 to investigate individual patient data.

A 2-step process was used to create and refine the design of the dashboards: (1) static mock-up dashboards were created to gain early feedback on the preliminary dashboard designs and (2) interactive prototype dashboard were created for usability testing. The prototype front end was developed using JavaScript with jQuery, using Store.js for data storage, and Highcharts for information visualization. These tools were considered to be most appropriate for prototype development as they were all well-supported open source libraries that minimized a technical risk for the project.

Table 1. Advisory Committee members.

#	Role
1	Breast surgeon
2	Pathologist
3	Medical oncologist
4	Data manager
5	Radiation oncologist
6	5 × project team members (data scientists, implementation scientists, and health service researchers)

Table 2. The tasks participants were asked to complete and the completion rate for each task.

ID	Task description	SEQ[a,b] (median)	Noncomplete tasks
Individual patient dashboard			
1a	You want to understand the procedures this patient has undergone over the course of their treatment. How do you find that out from this dashboard?	2	0
1b	Can you please tell me the number of nodes resected in this patient?	1	0
1c	Can you please tell me the name of the surgeon that performed the first surgical procedure for this patient?	2	0
2a	How do you find that out from this dashboard?	1	0
2b	Can you please tell me the BMI[c] for this patient?	1	0
2c	Can you please tell me the date the first L-Dex[d] reading was taken for this patient?	1	0
3a	How do you assess the progress of a patient that has already developed lymphedema from this dashboard?	2	0
3b	Can you please tell me whether this patient's L-Dex readings were on the left or right arm?	1	0
Cohort data dashboard[e]			
1	You want to identify the proportion of patients with lymphedema that had more than 10 nodes resected? How do you find that out from this dashboard?	1	4
2	You want to identify the proportion of patients within the organization that currently have or that have had lymphedema. How do you find that out from this dashboard?	2	0
3	You want to identify the proportion of patients within the organization that have no data at all. How do you find that out from this dashboard?	1	0
4	You want to identify the proportion of patients within the organization that are having ongoing treatment for lymphedema. How do you find that out from this dashboard?	1	0
5a	You want to identify the proportion of patients within the organization that have recovered from lymphedema. How do you find that out from this dashboard?	1	0
5b	Can you please tell me how many users these data are based on?	1	0
6	You want to identify the proportion of patients within the organization that are having ongoing treatment for lymphedema and have a BMI in the overweight range. How do you find that out from this dashboard?	3	—[f]

[a]SEQ: Single Ease Question.

[b]Participants (n=5) were asked to rank each task after the completion of a 7-point Single Ease Question scale, where 1=very easy and 7=very hard. No participants rated the tasks as hard or very hard, though as indicated in the table some tasks could not be completed for Dashboard 2.

[c]BMI: body mass index.

[d]L-Dex: lymphedema index.

[e]Participants found it easier to complete tasks on Dashboard 2 as they progressed through the session and became more familiar with the structure of the dashboard.

[f]This task was only completed by 3 participants.

The static mock-up dashboards (refer to Multimedia Appendix 1 to see Static Mock-up: Individual Patient Dashboard and Cohort Dashboard) were created over a period of 3 months, with feedback sessions as each version was developed for clinicians within the breast MDT (n=3): a breast surgeon, a radiation oncologist, and an OT. In each review session, a facilitator familiar with the dashboards worked face to face with 1 clinician during a 60- to 90-min face-to-face session. During the presentations, each clinician was given an opportunity to provide feedback on the presentation of data and to identify aspects needing improvement. This feedback was reviewed by the development team and informed the next design iteration. This iterative development took 6 months to develop (Figures 1-3). During this time, iterations of the prototype were reviewed by the Advisory Committee.

Figure 1. The final prototype of the individual patient dashboard visualizes data for individual patients who have been screened for lymphedema. The prototype dashboard presents a deidentified patient record populated with clinical data. A pseudonym is used for the patient name and medical record number (MRN). This patient has 5 lymphedema index (L-Dex) readings that were taken between January 2012 and April 2014. All the readings are in the normal range for this patient and have been taken on the right side of the body (indicated with green, as opposed to blue for left). The patient had 1 surgery in April. In this figure, the user has clicked on the expand icon (+) next to Surgery 1 to expand the box and see additional details about this procedure. Radiotherapy 1 shows an unexpanded procedure. BMI: body mass index.

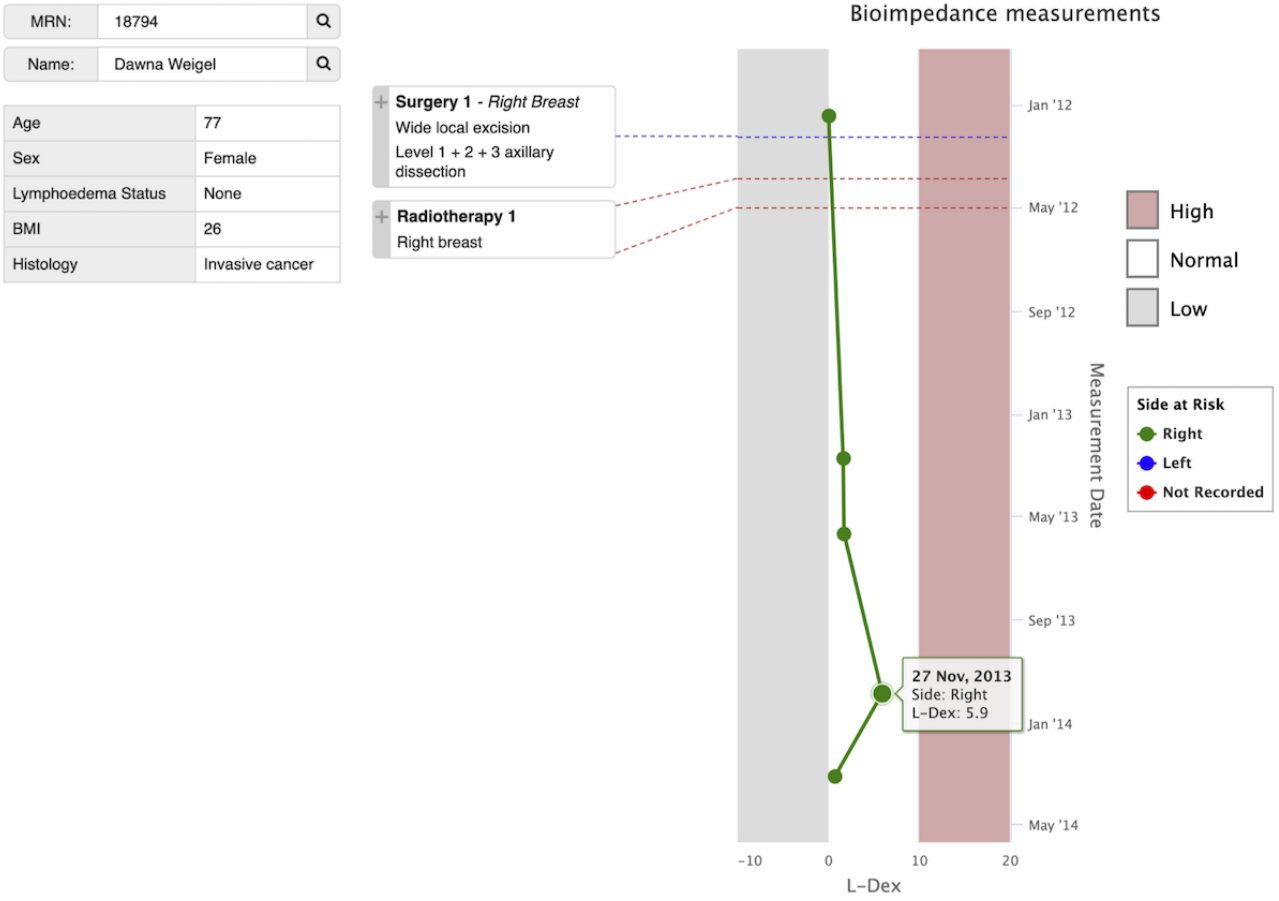


Figure 2. The final prototype of the cohort data dashboard visualizes data for the group of patients who have been screened for lymphedema and presents it to the user in a single dashboard. The dashboard is interactive and by default displays a comprehensive overview of all the cohort data available to the user. In this figure, the dashboard is showing data for all patients who have been diagnosed with lymphedema (indicated in purple) and all patients that have had a lymphedema index (L-Dex) measure <10 (indicated in green). BMI: body mass index.

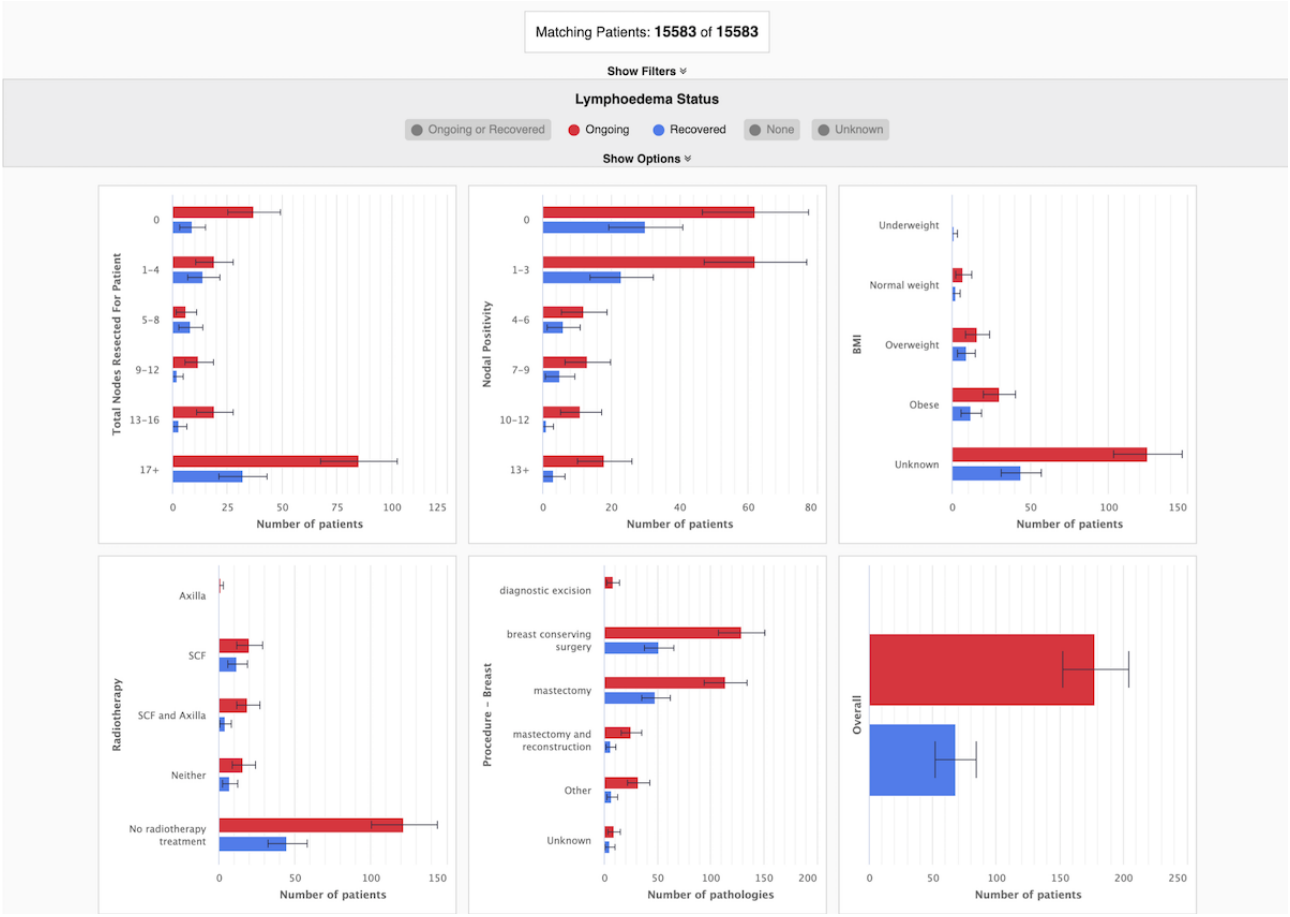
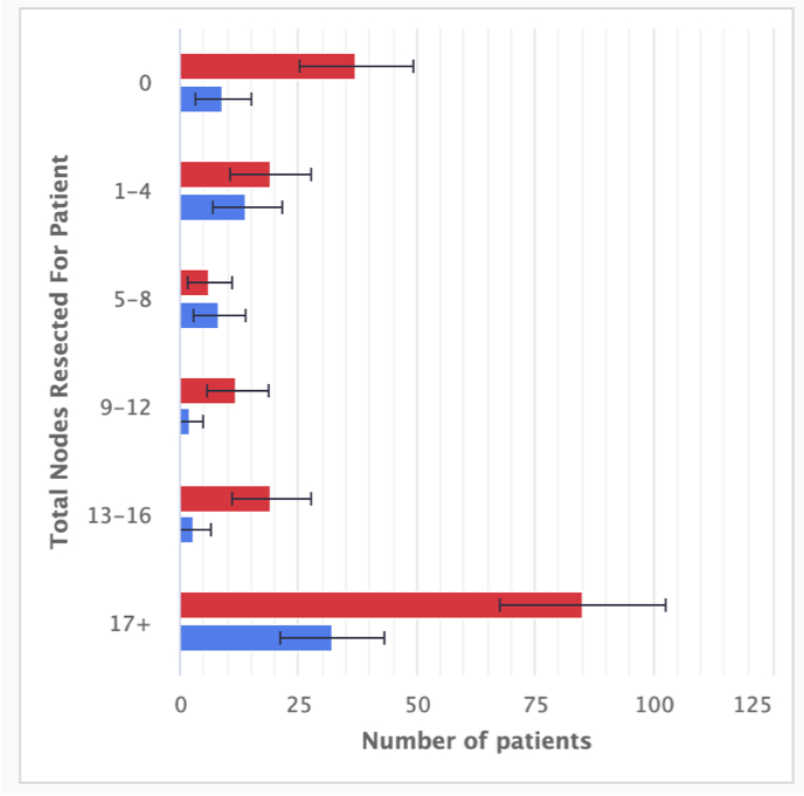


Figure 3. A close-up of one chart on the Cohort Data Dashboard: Nodes Resected.



Evaluation

The prototype dashboards were evaluated by a purposeful sampling of health professionals (n=5) with expertise supporting breast cancer patients with lymphedema. The evaluation consisted of a Think Aloud Protocol (TAP), where participants worked though the set of concrete tasks (Table 2). After each task, they answered the Single Ease Question (SEQ) [19] with 7-point ranking. The evaluation sessions were conducted by a researcher experienced in the methods used. Participants only had access to the dashboard during the evaluation sessions, as access was provided by the researcher conducting the session. Participants were not paid to participate in the evaluation sessions. The study design made use of TAP for its rich qualitative information about all aspects of use, usability, and experience and the SEQ because it is efficient, which is important for time-poor health professionals. Participants were then asked to complete a semistructured interview to explore their experiences using the dashboards. Each evaluation took between 60 and 90 min to complete.

Data from the SEQ component of the evaluation were aggregated and analyzed by a member of the research team to identify how easy the dashboards were for participants to use. Data from the recording of the semistructured interview component of the evaluation were transcribed and anonymized for evaluation by the research team. A content analysis was undertaken to categorize the transcript data. Categorization of the transcript data was undertaken by a consensus process among 3 researchers. Each transcript was read through by the 3 researchers and line-by-line coding was undertaken to ensure full inclusion of all possible data. Codes were grouped by categories and subcategories comparatively among the 3 researchers until consensus was reached. Exemplar quotations were identified and aligned with relevant categories.

Results

Participant Demographics

A total of 5 health professionals participated in the evaluation of the 2 prototypes. Participants were selected from 3 specialties: surgery (1/5, 20%), radiation oncology (2/5, 40%), and occupational therapy (2/5, 40%). Of this cohort, 2 participants had been involved in the evaluation of the static mock-ups. None of the participants had input into the creation of the tasks for the TAP (Table 2).

Current Clinical Practice

Interviewees consistently noted that the current process for retrieving data on lymphedema patients required access to multiple data sources. This could include multiple electronic databases and digitized or nondigitized clinical notes to find pertinent data about the patient. Interviewees also noted that if data were unavailable from the databases, it was common practice to liaise with another health professional involved in delivering care to the patient to find out additional information. One interviewee highlighted that there are multiple locations which patient data needed to be entered into, cross-referenced, and accessed from:

> *I have my work emails, I have [the radiation oncology eMR], I have letters I need to review on [the eMR], I have data that I enter in to my own iPad because, I enter data in to our [breast cancer] database...So I feel like I open up a lot of websites just to do my daily job.* [BD 1.4]

All interviewees stated that they did not currently have regular access to cohort data on lymphedema patients. The only instances when cohort data were available was when an individual actively sought it out, such as during a research project.

Individual Patient Dashboard

All interviewees were interested and enthusiastic about the individual patient dashboard:

> *This is brilliant. This is exactly what we wanted when we designed all of this.* [BD 1.2]

Interviewees did not have easy access to data on how an individual patient progressed through lymphedema screening and breast cancer treatment. The dashboard function in presenting the patient's treatment journey, in addition to presenting L-Dex measurements over time, was particularly well received:

> *...it was quite clear and I could see exactly what surgeries they [the patient] had. I didn't have to click on multiple buttons to get there. And you can find out when they had radiotherapy. So a lot here that was clear.* [BD 1.3]

Interviewees commented positively about color use. For example, the following comment was about red indicating when an individual patient's L-Dex measurement was moving outside the normal range:

> *I like this red zone 'cause for us who don't know the L-Dex exact measurements, it's good to know. This tells me anything more than 10 presumably is high risk.* [BD 1.5]

One interviewee commented on the way that patient data were scaled to ensure the user could see the whole time period where data were available. This resulted in different patients' data displayed across different scales:

> *The scale is now changed here compared to the other ones. That's a little bit confusing.* [BD 1.1]

Multiple interviewees commented that they would have liked additional data incorporated, particularly chemotherapy data (which were not available in the prototype because of medical oncology transitioning to a new information management system at the time of the project). For future deployments, it will be important to support such augmentation. One participant noted:

> *Yeah, I would have thought that she would have had chemotherapy as well, which doesn't show at all.* [BD 1.1]

Overall, interviewees were enthusiastic about the application of the dashboards in clinical practice. In addition, each interviewee identified a range of applications they could utilize the dashboard for, reflecting their diversity of clinical specialties

and priorities. Interviewees felt the individual patient dashboard would have particular value as a tool to augment the clinical decision support process. This was because the dashboard provides a means of getting a quick overview of the patient's pathway through treatment:

> *It would be for two things, one is to get a quick visualization of what management and what assessment have been done for the patient and, secondly to see the progression.* [BD 1.2]

In addition, 1 interviewee suggested that the individual patient dashboard may have value to facilitate patient education around their treatment:

> *I guess for a patient to visualize, sometimes just numbers don't make a lot of sense to them, but to actually see something on a graph can be helpful.* [BD 1.3]

Cohort Data Dashboard

Interviewees were generally excited to have the opportunity to see cohort data on their patients. However, all commented on how much more complex the cohort dashboard was compared with the individual patient dashboard. For example:

> *I mean, if you're looking at this [the cohort dashboard], it's a little bit more complex information than the single patient, because the single patient just hits you without any... You don't need to work anything out, it just tells you what it is straight away. This one, I needed to get my head around what we were actually looking at.* [BD 1.2]

Interviewees pointed to the inconsistency in the x-axis in the body mass index graph which is different from the others. Similarly, the y-axis is inconsistent for the 2 graphs about nodes (resected and positivity). Although they saw this as a minor issue, they explained that they wanted to be able to compare information across the charts. One interviewee also noted that having to do calculations of what the charts were saying was a barrier to use:

> *So I had to highlight a couple of extra things, I suppose the main thing would be I'd have to do a calculation of that, minus one, to the, minus 1.5 up to the 5.9.* [BD 1.1]

Finally, interviewees commented on instances where terminology used to describe data did not reflect how clinicians routinely conveyed information relating to lymphedema identification and treatment. This issue occurred for both the individual patient dashboard and the cohort data dashboard. Interviewees understood what the terminology used in the prototype dashboards meant, so it was not a barrier to use, merely an area for future improvement. An example of this was:

> *It's interesting, just using the word recovered, it's a tricky one, because they have no clinically overt symptoms but they are considered to be subclinical, like as in their system is perhaps, like they've got ongoing risk and it may actually turn up again, so yeah, recovered just makes it sound like it's…* [BD 1.1]

As was the case with the individual patient dashboard, interviewees identified issues around the completeness of the clinical data. The quality of the data was viewed as a problem for the long-term usability of the dashboard, even if the interface was user-friendly and engaging. The interviewees frequently drew on their clinical expertise to question the accuracy of the data or identify data points which they felt were missing or did not make sense. There were numerous data points on the cohort dashboard that interviewees did not expect based on their clinical expertise. This was perceived to reduce the level of trust in visualizations, which would limit the likelihood of continued use in clinical practice. The issues could be resolved during the process of refining the prototype for implementation into the clinical setting.

Interviewees suggested that the cohort dashboard would have value as a tool for helping patients understand treatment and lymphedema. Unlike the individual patient dashboard, the cohort data dashboard was viewed as valuable for helping patients understanding the outcomes of the cancer center.

> *If I see a patient and the patient asks me, "What is your outcome?" Or they want to know, "What's my survival?" And I say, "Our centre here is excellent and stage three gives you that." That [the cohort dashboard] might be useful.* [BD 1.5]

Furthermore, some interviewees indicated that the cohort data dashboard may be useful for research and feedback and to support interaction among different health professionals.

Discussion

Principal Findings

The findings of this study demonstrated that it was feasible to use routinely collected data and visualization tools to facilitate clinical decision making and monitor care delivery. This finding builds on the existing literature which has shown that there is considerable interest from health professionals in improving access to routinely collected clinical data [21]. Further, findings from the study do not just demonstrate feasibility of visualizing data but also highlight a number of considerations for designing visualization tools to meet the needs of health professionals in clinical practice.

The individual patient dashboard was successful. Key features of the interface are as follows:

- The side-by-side access to the standard medical record information
- The visualization of the patient trajectory over their full cancer journey

This interface was designed for frequent use by diverse clinicians and the MDT. Our evaluation indicates that it was easy to understand and use. All participants completed all 9 tasks without assistance. The SEQ scores indicate that participants considered all tasks as easy, with median scores for all 1 (very easy) or 2 (easy). User comments point to small refinements but confirm that the overall design is effective.

The cohort data dashboard is far more complex but understandable, with feedback from participants demonstrating

enthusiasm about having access to cohort data. The design of this dashboard was driven by the aspirations to understand many dimensions of the data. The design team was aware that visualizing the cohort data was extremely complex but concluded that it would be valuable to gain insights from evaluations at this stage. While 4 of the 5 participants had difficulties with the first task, they were successful in completing the next 6 tasks. The seventh task had high noncompletion rates as it was affected by time pressures on participants. The SEQ scores also indicate perceived high ease of use, with median scores of 1 (very easy) for all but the first 2 tasks, reflecting the start-up learning. Overall, the dashboard is promising for in-depth use by individuals and teams who review these big-picture outcomes infrequently, perhaps months apart. For such intermittent use, we envisage that it may be helpful to add scaffolding to support exploration of key aspects as well as a history mechanism to enable clinicians and administrators to track progress in management and changes.

The study described in this paper revealed that health professionals, seeing the aggregated data for the first time, could identify that there are problems in the underlying data in terms of both its completeness and accuracy. Collecting high-quality clinical data is an acknowledged challenge in the literature [8]. On the one hand, this highlights the need to carefully consider the potential biases in the information displayed. On the other hand, our dashboards have the potential to be a starting point for tackling this problem because both dashboards made data omissions and some errors more visible. They also have the potential to help if they consistently provide value from accurate data so that busy clinicians see value in creating higher-quality records. In instances where data were missing from the prototypes, clinicians made inferences regarding what they expected to have happened to the patient. As highlighted in the evaluation, health professionals drew on their clinical expertise to critically analyze the data presented in visualization tools. A key consideration for future dashboard development is the investment in identifying data sources required to populate the final dashboard and ensuring all relevant sources are incorporated.

The process for developing the dashboards highlighted that we should explore the design of our dashboards for the case of lymphedema in the context of breast cancer. This choice was driven by both the priorities of the health professionals involved in the project and the availability of key data sources, including from medical records and a separate store of lymphedema data. A complex and iterative design process was used to identify the dashboard priorities and refine them to be fit for purpose in the clinical setting. It began with focus group sessions involving 33 members of MDTs and a consumer, then months of fortnightly consultations with a team of 10 to refine the choice of problems, data, and to inform and then refine the design of the dashboards. We believe that we could streamline this for future dashboards, drawing on the lessons from this work. The work described in this article aggregated and made available practice data on lymphedema the first time in the MDT setting. The enthusiasm of the evaluation participants for this information highlights the potential power of such work.

Limitations

This study is limited due to the small sample size (n=5) used to evaluate the final dashboards. The 5 participants were part of the original focus group meetings due to the commitment of 90 min of clinical time to participate in the evaluation, which may have led to bias in the findings. In addition, the study is limited by the incomplete data set available for developing the prototype dashboards. The data were incomplete as patients received care across different institutions, and data were only available relating to treatment delivered in the organization where the study was conducted. Finally, although participants were undertaking a TAP under artificial conditions, the health professionals were under time pressure to get the tasks done.

Future researchers exploring the use of dashboards for use in health care may wish to explore questions around how users focusing on certain aspects of data such as performance measures may affect their use of all features in a tool. In addition, future research is warranted on the types of data presented in the dashboard and the balance between presenting data that are clinically relevant and data that are easily measurable.

Conclusions

Health professionals have a considerable level of interest in tools for increasing the accessibility of their routinely collected clinical data using visualization tools. However, there is currently little research into the design of such tools or strategies for implementing them into clinical workflow. Next steps for implementing dashboards into routine clinical practice include the identification of metrics that are highly relevant to clinicians and teams, rather than metrics just easily measurable. In addition, to implement dashboards into practice, it is necessary to not just understand the type of data that has value for presentation in dashboards but investigate when and how they are most useful to health professionals.

A central consideration when designing data visualization tools for health professionals is ensuring they present data in a manner which can be understood and actioned quickly and easily by end users. Furthermore, it is important that iteration is used to review and refine the quality of clinical data being presented to ensure it aligns with the priorities of the health professionals using it. Finally, the creation of visualization tools that meet the needs of health care teams is an interdisciplinary process which requires collaboration between domain experts, data scientists, developers, and user interface designers.

Acknowledgments

The authors wish to acknowledge the Sydney West Translational Cancer Research Centre, the Westmead Breast Cancer Institute, and the Sydney Informatics Hub (SIC) for their support to this study. They would also like to acknowledge Josh Stretton for his involvement in the project. Finally, they would like to thank the breast cancer MDT members for their participation in the study.

Conflicts of Interest

The authors declare no financial conflicts of interest. TS, CKC, KJ, AJ, AS, PT, and JK were part of the project team that developed and evaluated the dashboard. The project team members who developed the dashboard were not involved in the evaluation sessions.

Multimedia Appendix 1
Early mock-ups of the static individual and cohort dashboards that were used to get initial feedback from users.
[PNG File , 281 KB - jmir_v22i4e13188_app1.png]

References

1. Miller RH, Sim I. Physicians' use of electronic medical records: barriers and solutions. Health Aff (Millwood) 2004;23(2):116-126. [doi: 10.1377/hlthaff.23.2.116] [Medline: 15046136]
2. Kayyali B, Knott D, Kuiken SV. Digital Strategy Consultants. 2013. The Big-Data Revolution in US Health Care: Accelerating Value and Innovation URL: https://digitalstrategy.nl/wp-content/uploads/E2-2013.04-The-big-data-revolution-in-US-health-care-Accelerating-value-and-innovation.pdf [accessed 2018-10-22]
3. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol 2017 Jan;106(1):1-9 [FREE Full text] [doi: 10.1007/s00392-016-1025-6] [Medline: 27557678]
4. de Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015 Feb;53:162-173 [FREE Full text] [doi: 10.1016/j.jbi.2014.10.006] [Medline: 25463966]
5. Kohli R, Tan SS. Electronic health records: how can IS researchers contribute to transforming healthcare? MIS Quart 2016;40(3):553-573 [FREE Full text] [doi: 10.25300/MISQ/2016/40.3.02]
6. Xylinas E, Rouprêt M, Kluth L, Scherr DS, Shariat SF, Upper Tract Urothelial Carcinoma Collaboration. Collaborative research networks as a platform for virtual multidisciplinary, international approach to managing difficult clinical cases: an example from the Upper Tract Urothelial Carcinoma Collaboration. Eur Urol 2012 Nov;62(5):943-945. [doi: 10.1016/j.eururo.2012.08.022] [Medline: 22921717]
7. Shah GH, Leider JP, Luo H, Kaur R. Interoperability of information systems managed and used by the local health departments. J Public Health Manag Pract 2016;22(Suppl 6, Public Health Informatics):S34-S43 [FREE Full text] [doi: 10.1097/PHH.0000000000000436] [Medline: 27684616]
8. Jalil R, Ahmed M, Green JS, Sevdalis B. Factors that can make an impact on decision-making and decision implementation in cancer multidisciplinary teams: an interview study of the provider perspective. Int J Surg 2013;11(5):389-394 [FREE Full text] [doi: 10.1016/j.ijsu.2013.02.026] [Medline: 23500030]
9. Celi LA, Marshall JD, Lai Y, Stone DJ. Disrupting electronic health records systems: the next generation. JMIR Med Inform 2015 Oct 23;3(4):e34 [FREE Full text] [doi: 10.2196/medinform.4192] [Medline: 26500106]
10. Kool M, van der Sijp JR, Kroep JR, Liefers G, Jannink I, Guicherit OR, et al. Importance of patient reported outcome measures versus clinical outcomes for breast cancer patients evaluation on quality of care. Breast 2016 Jun;27:62-68. [doi: 10.1016/j.breast.2016.02.015] [Medline: 27026219]
11. Zeleke B. ProQuest Search. 2012. Clinical Dashboard for Medication Reconciliation URL: https://search.proquest.com/openview/f5a34e7d20cd351daa9ead31925e71dd/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2019-10-03]
12. Dowding D, Randell R, Gardner P, Fitzpatrick G, Dykes P, Favela J, et al. Dashboards for improving patient care: review of the literature. Int J Med Inform 2015 Feb;84(2):87-100. [doi: 10.1016/j.ijmedinf.2014.10.001] [Medline: 25453274]
13. Australian Institute of Health and Welfare. Cancer Incidence Projections, Australia 2011 to 2020. Australia: AIHW; 2012.
14. Langer I, Guller U, Berclaz G, Koechli OR, Schaer G, Fehr MK, et al. Morbidity of sentinel lymph node biopsy (SLN) alone versus SLN and completion axillary lymph node dissection after breast cancer surgery: a prospective Swiss multicenter study on 659 patients. Ann Surg 2007 Mar;245(3):452-461. [doi: 10.1097/01.sla.0000245472.47748.ec] [Medline: 17435553]
15. Taghian NR, Miller CL, Jammallo LS, O'Toole J, Skolny MN. Lymphedema following breast cancer treatment and impact on quality of life: a review. Crit Rev Oncol Hematol 2014 Dec;92(3):227-234. [doi: 10.1016/j.critrevonc.2014.06.004] [Medline: 25085806]
16. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR Hum Factors 2018 May 31;5(2):e22 [FREE Full text] [doi: 10.2196/humanfactors.9328] [Medline: 29853440]

17. Sluijter J, Otten M. Business Intelligence (BI) for Personalized Student Dashboards. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. 2017 Presented at: LAK'17; March 13-17, 2017; Vancouver, British Columbia, Canada p. 562-563 URL: https://dl.acm.org/citation.cfm?id=3029458 [doi: 10.1145/3027385.3029458]
18. Bodily R, Verbert K. Trends and Issues in Student-Facing Learning Analytics Reporting Systems Research. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. 2017 Presented at: LAK'17; March 13-17, 2017; Vancouver, British Columbia, Canada p. 309-318 URL: https://dl.acm.org/citation.cfm?id=3027403 [doi: 10.1145/3027385.3027403]
19. Sauro J, Dumas JS. Comparison of Three One-Question, Post-Task Usability Questionnaires. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; Boston, MA, USA p. 1599-1608 URL: https://dl.acm.org/citation.cfm?id=1518946 [doi: 10.1145/1518701.1518946]
20. Barrio AV, Eaton A, Frazier TG. A prospective validation study of bioimpedance with volume displacement in early-stage breast cancer patients at risk for lymphedema. Ann Surg Oncol 2015 Dec;22(Suppl 3):S370-S375 [FREE Full text] [doi: 10.1245/s10434-015-4683-0] [Medline: 26085222]
21. Robinson TE, Janssen A, Harnett P, Museth KE, Provan PJ, Hills DJ, et al. Embedding continuous quality improvement processes in multidisciplinary teams in cancer care: exploring the boundaries between quality and implementation science. Aust Health Rev 2017 Jul;41(3):291-296. [doi: 10.1071/AH16052] [Medline: 27372543]

Abbreviations

EHR: electronic health record
L-Dex: lymphedema index
MDT: multidisciplinary team
OT: occupational therapist
SEQ: Single Ease Question
TAP: Think Aloud Protocol

Edited by CL Parra-Calderón; submitted 19.12.18; peer-reviewed by K Stasiak, B Loo Gee, T Martin, D Salvi, D Attai; comments to author 12.02.19; revised version received 06.05.19; accepted 30.08.19; published 21.04.20.

Please cite as:
Janssen A, Donnelly C, Kay J, Thiem P, Saavedra A, Pathmanathan N, Elder E, Dinh P, Kabir M, Jackson K, Harnett P, Shaw T
Developing an Intranet-Based Lymphedema Dashboard for Breast Cancer Multidisciplinary Teams: Design Research Study
J Med Internet Res 2020;22(4):e13188
URL: https://www.jmir.org/2020/4/e13188
doi:10.2196/13188
PMID:32314968

Original Paper

Cloud-Based System for Effective Surveillance and Control of COVID-19: Useful Experiences From Hubei, China

Mengchun Gong[1*], MD; Li Liu[2*], MD; Xin Sun[3], MD; Yue Yang[2], PhD; Shuang Wang[4], PhD; Hong Zhu[1,2], PhD

[1]Institute of Health Management, Southern Medical University, Guangzhou, China

[2]Nanfang Hospital, Southern Medical University, Guangzhou, China

[3]Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, Chengdu, China

[4]Institutes for Systems Genetics, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Hong Zhu, PhD
Nanfang Hospital
Southern Medical University
North 1838, Guangzhou Ave, Baiyun District
Guangzhou
China
Phone: 86 130 7687 721
Email: zhnfyy@yeah.net

Abstract

Background: Coronavirus disease (COVID-19) has been an unprecedented challenge to the global health care system. Tools that can improve the focus of surveillance efforts and clinical decision support are of paramount importance.

Objective: The aim of this study was to illustrate how new medical informatics technologies may enable effective control of the pandemic through the development and successful 72-hour deployment of the Honghu Hybrid System (HHS) for COVID-19 in the city of Honghu in Hubei, China.

Methods: The HHS was designed for the collection, integration, standardization, and analysis of COVID-19-related data from multiple sources, which includes a case reporting system, diagnostic labs, electronic medical records, and social media on mobile devices.

Results: HHS supports four main features: syndromic surveillance on mobile devices, policy-making decision support, clinical decision support and prioritization of resources, and follow-up of discharged patients. The syndromic surveillance component in HHS covered over 95% of the population of over 900,000 people and provided near real time evidence for the control of epidemic emergencies. The clinical decision support component in HHS was also provided to improve patient care and prioritize the limited medical resources. However, the statistical methods still require further evaluations to confirm clinical effectiveness and appropriateness of disposition assigned in this study, which warrants further investigation.

Conclusions: The facilitating factors and challenges are discussed to provide useful insights to other cities to build suitable solutions based on cloud technologies. The HHS for COVID-19 was shown to be feasible and effective in this real-world field study, and has the potential to be migrated.

(J Med Internet Res 2020;22(4):e18948) doi:10.2196/18948

KEYWORDS

COVID-19; cloud system; syndromic surveillance; clinical decision support; stakeholders involvement; pandemic; medical informatics

Introduction

The outbreak of the coronavirus disease (COVID-19) in China and many other countries has put huge pressure on the health care system [1]. One method of controlling the communicable diseases is the use of a surveillance system to track the exposed and infected individuals, as well as clinical outcomes [2-6]. However, traditional surveillance systems have limitations in terms of timeliness, spatial resolution, and scalability [7]. Meanwhile, reporting from these systems tends to be national

or regional with insufficient information about diseases at the community or city level, which caused low efficiency for the social distancing and quarantine measures [2,8]. This is particularly true for COVID-19 surveillance for the Hubei, China, where many cases and isolated populations have challenged systems of manual reporting and tracking [9-12].

In response to this significant challenge, we developed the Honghu Hybrid System (HHS) as a pilot for COVID-19 surveillance and control, which was successfully deployed within 72 hours in Honghu in the Hubei province, a city 145 kilometers (90 miles) away from Wuhan (the capital city of the Hubei province) with a population of over 900,000 people. The HHS integrated data from both traditional sources such as case report systems and diagnostic labs, as well as nontraditional sources including structured electronic medical records and social media on mobile devices. The real time acquisition and analysis of highly resolved digital data provide detailed information on symptoms, psychological status, contact history, social behavior, and the physical environment [4,6,13].

Methods

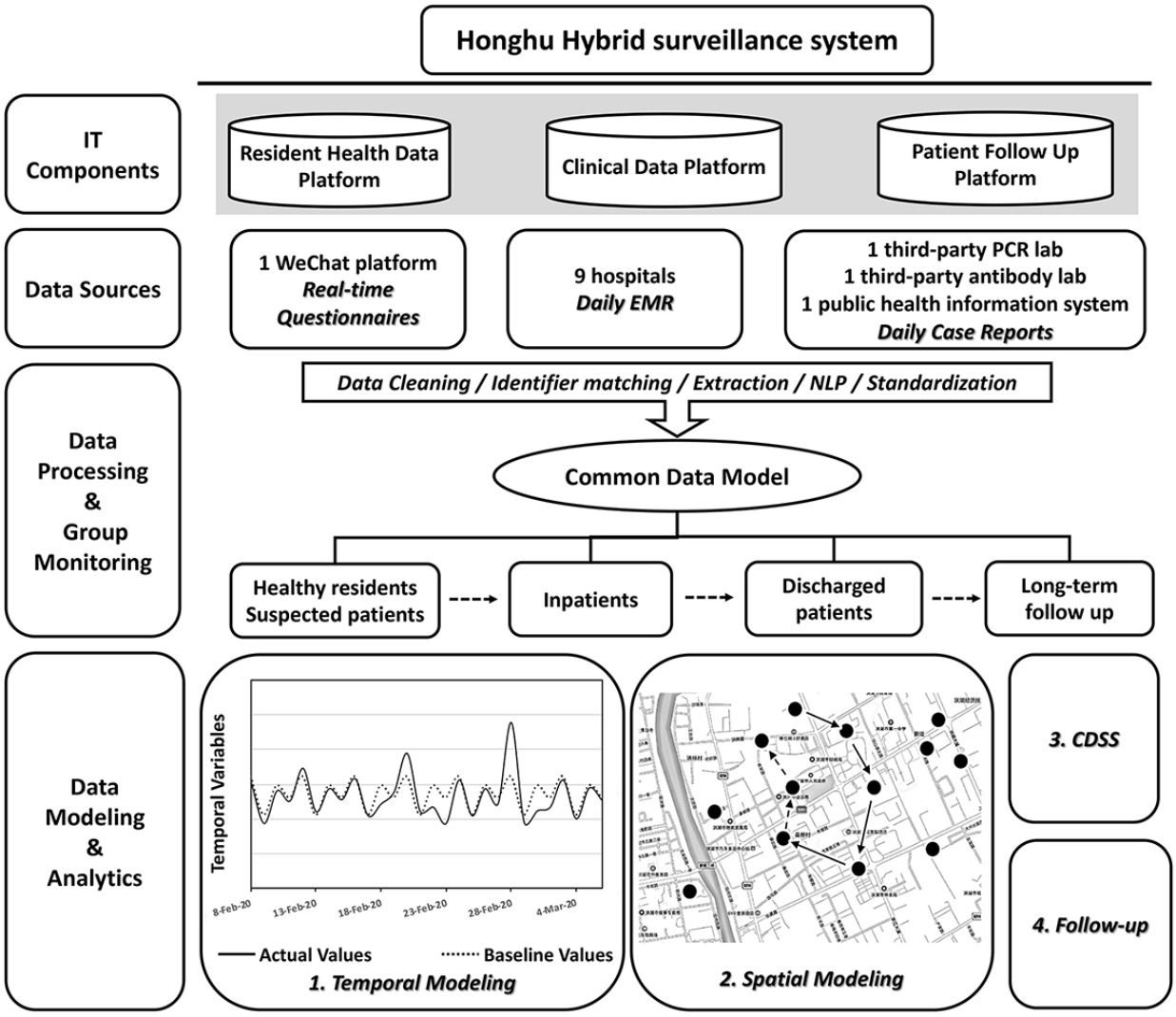
Environment and Hardware

Cloud-based hardware provided an efficient solution to solve the problems unique to the COVID-19 epidemic and effectively mitigated issues such as shortage of local technical support, unavailability of experts and physical hardware due to blocked transportation, rapidly changing needs related to functionality, and connections with multiple sources across different platforms. See Multimedia Appendix 1 for details of the virtual machine settings.

Data Collection

This system collected daily structured electronic medical record data from nine hospitals; real time information about symptoms and personal contact history from the WeChat platform (one of the largest mobile social network apps in China with more than 1 billion monthly active users); and daily reported case diagnosis information from one third-party polymerase chain reaction lab, one third-party antibody lab, and one public health information system (Figure 1). For the data collection, we leveraged existing health information systems inside the nine hospitals and developed a novel mini program for the WeChat platform software development kit for symptom reporting and spatial data collection.

Figure 1. Schematic representation of data streams, processing, and analytics in the Honghu Hybrid System for the coronavirus disease surveillance and control. CDSS: clinical decision support system; EMR: electronic medical records; IT: information technology; NLP: natural language processing; PCR: polymerase chain reaction.



Data Processing

The data feeds were normalized temporally and spatially and then loaded into a common data model that had been built for the storage, management, and analysis of the integrated COVID-19 data. Vocabulary control was implemented based on the Systematized Nomenclature of Medicine-Clinical Terms synonyms in Chinese for symptoms and the disease itself [14]. Logical Observation Identifiers Names and Codes were adopted to code-related tests. The International Classification of Diseases (ICD) (ie, ICD-10 Clinical Modification) codes were used for the diseases based on the coding standards released by the National Health Commission of China [15,16]. The elements for the data model and recommended synonyms for COVID-19 were collected from 54 Chinese experts, including clinical doctors, researchers, public health professionals, and informaticians.

Ethics Review, Privacy Protection, and Data Security

The Nanfang Hospital Ethics Committee approved this study. All the users on the syndromic surveillance system consented sharing the necessary information for using this system. Data security and privacy protection were particularly emphasized with administrative as well as technical support [17] throughout the deployment of the system. Standard security settings and software (eg, firewalls, data encryption) were implemented. A virtual private network was used for remote technical support and data analytics. To protect the privacy of patients, we set different levels of access to the data. Aggregated reports were the main form for data sharing. The expert team in the Honghu Municipal COVID-19 Control Headquarter had the highest level of data access. They used the data to triage the patients, coordinate social workers in communities, and allocate the health care resources. The remote analytics team had no access to the identity of the patients. The different levels of access to the data were designed and set up at the beginning of the project by the experts in data security and privacy protection, who have been certified by the Health Insurance Portability and Accountability Act of 1996 security certification. An independent auditor was responsible for monitoring any violation of the rules, as well as reporting and correcting them. Written informed consent was obtained from the inhospital patients before enrollment when data were collected

retrospectively. It is important to note that all authorized personnel received training on data security and privacy protection, and additionally, signed legally binding affidavits. In addition, the team worked closely with the cloud computing service provider to separate sensitive data from the rest of the data.

Results

Syndromic Surveillance on Mobile Devices

From January 26, when the first case was diagnosed in Honghu, to March 16, 2020, when all the patients with confirmed COVID-19 were discharged, there were in total 383 COVID-19 cases, of which 13 died, unfortunately. Under the huge pressure, HHS was initiated on February 14 and successfully deployed within 72 hours. The accumulated number of self-reports reached 17.5 million by March 16 and the maximum daily active reports reached 900,000 person-times.

Syndromic surveillance was implemented on a mobile phone–based social media platform targeting different groups of individuals. This included the general population, inhospital and discharged patients, people with higher risk of infection (ie, those with travel history to Wuhan, contact history with confirmed cases, or under medical observation in isolation sites), and health care professionals (ie, doctors, nurses, public health experts, and social workers). See Textbox 1 for details of the questionnaire for the general population who was quarantined at home. The items included in the questionnaire had been proposed by infectious disease experts from Nanfang Hospital. Their suggestion was then discussed and modified in a meeting joined by local government staff, local clinical doctors, public health professionals, and informaticians. The items concerning job-seeking support were added to the questionnaire on March 2 when the government was evaluating the release of the restriction on work and production in Honghu. Social workers followed up with over 10,000 positive reports (eg, “I had a temperature of over 37.3 °C” or “I had a severe cough today”) via phone call or home visits. More than 30 individuals were assisted in going to the fever clinic for further screening and then quarantined. This was an active surveillance mechanism initiated by the city residents through new information channels and was effective in COVID-19 screening. The high coverage (over 95% of the residents) and daily active reports (over 600,000 person-times) demonstrated the feasibility of intense monitoring during the COVID-19 epidemic. The stable trends of positive reports (0.10%-0.12% after the daily reports exceeded 500,000 persons) provided the strong evidence that the countermeasures in place were effective in preventing the local outbreaks.

Textbox 1. Real time questionnaire on WeChat.

Demographic characteristics

- Name
- ID number
- Gender
- Address (semi-structured)
- Phone number (free text)

Spatial information

- Current location (structured)

Epidemiological exposure

- Have you visited or stayed in communities with confirmed cases in the last 14 days? (Yes/No)
- Have you been in contact with confirmed cases in the last 14 days？ (Yes/No)
- Have you been in contact with residents from Wuhan City or from the community with case reports or who have had respiratory symptoms in the last 14 days? (Yes/No)
- Have you participated in small-scale gatherings with more than 2 cases reported in the last 14 days? (Yes/No)

Physical conditions

- Today's physical condition: good/cough/runny nose/chest tightness/diarrhea/muscle soreness? (multiple choices)
- Today's body temperature?

Psychological conditions

- Did you feel nervous, fearful or anxious today? (Yes/No)

Community support

- Are there public health officials to visit you for investigation? (Yes/No)
- Were there any social workers who called you or came to your home to help solve the reported problems? (Yes/No)

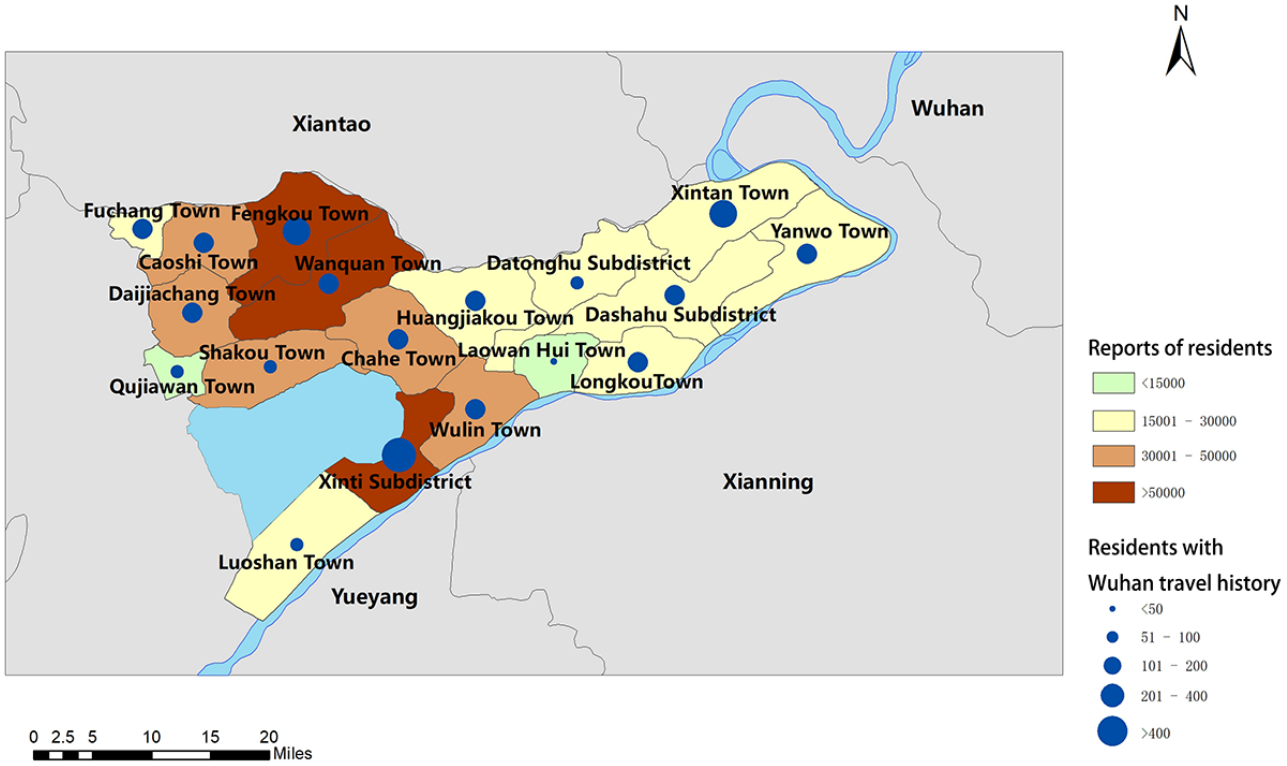
Job-seeking support

- Are you willing to work locally? (free text)
- Are you willing to participate in free skill training? (free text)

Policy-Making Decision Support

Monitoring the fluctuation and trends analysis of the syndromic surveillance data supported policy-related decision making. Due to the large population size, stability and fluctuation of the trends provided strong evidence for local authorities to evaluate the effectiveness of disease management and make timely adjustments accordingly. Spatial analyses also played a critical role. Clustering of exposed residents inside a community, indicated by the concentration of the patients on the map (Figure 2), illustrated high risk for local outbreaks and would then trigger home visits by social workers automatically.

Figure 2. Honghu health reports heat map.



Clinical Decision Support and Prioritization of Resources

A clinical decision support system based on an inhospital mortality prediction system was built for patients with COVID-19 to improve the clinical care, decrease death risk, and prioritize limited medical resources. Based on the Multilobular Infiltration, Hypo-Lymphocytosis, Bacterial Coinfection, Smoking History, Hyper-Tension and Age (MuLBSTA) [18] scoring system, which is a partially validated prediction system for the inhospital mortality of patients with COVID-19, 36 patients out of the 383 cases were classified as high-risk (MuLBSTA score ≥12). They were either relocated to the single hospital in the area that had an intensive care unit or screened with important biochemical markers more frequently [18,19]. To the best of the authors' knowledge, this is the first practical use of a mortality risk prediction system specifically for patients with COVID-19.

Follow-Up for Discharged Patients

We used the social media platform to register the discharged patients and required the patients to report their symptoms daily in the 2 months after discharge. After the follow-up system was initiated, 100% coverage was achieved within 3 days. The reported recurrence of symptoms such as high fever was linked with home visits by social workers inside communities and readmission to hospitals.

Discussion

The system had strengthened the checkpoints across the entire chain of COVID-19 control, including early discovery through symptom surveillance covering almost the whole population, early report by active information channel directly connected with follow-ups, early isolation strengthened by spatial tracking, and early treatment enabled by the clinical decision support system.

The Effects of HHS for COVID-19

One powerful function of the hybrid system was the enhancement of the full spectrum for the management of COVID-19. The first step was the deployment of the syndromic surveillance platform, covering a large population through their mobile devices, to direct and empower the implementation of public health countermeasures. Geospatial data were also used to analyze the potential close contact history and monitor the local small-scale outbreaks. The second step was to use the inhospital mortality prediction system to direct the clinical interventions of patients and to allocate the limited resources. The third step was the seamless coverage of patients at different stages, including high-risk of infection, isolation under medical observation, admitted, critically ill, discharged, or follow-up cases. All of these management actions helped decision making for controlling the disease effectively in Honghu, which was under pressure due to the large numbers of people who traveled back from Wuhan (over 50,000), and the severe lack of medical staff and public health experts at the onset of the outbreak.

Facilitating Factors for the Adoption of HHS

The most important factor contributing to the successful building and adoption of HHS was the deep involvement of multiple stakeholders. Experts, including those from public health, clinical medicine, medical informatics, data analytics, decision science, and other backgrounds, provided methodology guidance, clinical expertise, and research support. This helped to convince the residents who had high faith in the professionals from the national-level medical institutes such as Nanfang

Hospital of Southern Medical University. Strong coordination by the local government, including the mobilization of social workers, public media exposure of the HHS gateway, and connecting services including transportation and social services, increased the coverage to over 95%, which has been traditionally difficult for a city with a population close to 1 million. Active participation and engagement by the residents were also crucial. The willingness to share their health status data during the COVID-19 epidemics was the basis for the accomplishment of the platform. Last but not least, the technology companies allocated enough resources to the local sites and provided mature technologies and protocols for implementation. During the 72 hours of building the system, a team of over 40 experts were involved, and they took shifts to make sure that progress was made every hour. It required delicate coordination and project management, dedication from each team member, and strong support on the organizational level to make sure that the contribution was appropriately recognized.

Relevant Issues for the Migration

Internet infrastructure and technology availability are the key issues. The high penetration rate of WeChat in China provided the foundation for successful implementation of HHS [20]. Similar social media platforms such as Line and Facebook may also be suitable solutions in different countries or areas, based on an evaluation of the penetration and the active use. Cost-effectiveness analysis is also important for decision making when building new systems. Using existing open platforms also helped control the costs, which was about 3 million CNY (US ~$431,000) for HHS, including cloud-based virtual resource, labor, and other costs. The effectiveness can be more significant in areas where the infection rate is high and social distancing and traffic control are making traditional syndromic surveillance impossible [21]. Resources for implementation of the new system, especially when connecting with hospital information systems, and further data analytics would also be challenging and require coordination and collaboration between local and external resources.

Limitations of the Study

This study has several limitations. The lack of internal or external validation of the system was caused by the emergent situation during the epidemics, the ethics dilemma, and limited resources. The sample size was not large enough to draw statistical inferences because the scope of the system was within the city of Honghu. The proposed HHS strategy can be adapted by other researchers, innovators, or authorities to develop localized or national solutions to respond to the urgent situation and help to prioritize health care resource allocation, but the clinical effectiveness of the statistical methods used in this system still requires additional studies in the future. The HHS has several deficiencies. The national ID of each participant was inputted by users manually and mistakes may exist. This can be corrected by collecting the photocopies of the ID card and automatically extracting the ID number. Moreover, because of the limited time, the system was not friendly enough for users, as mentioned in the feedback review from some government staff, and did not support access from mobile devices. Further development can help improve these existing limitations.

Conclusion

Based on the field study in Honghu city, the HHS has been observed to be effective and feasible for COVID-19 surveillance and control. It helped strengthen the checkpoints on the full chain of COVID-19 control, including “early test, early report, early isolation, and early treatment” during the outbreak and the long-term follow-up after the epidemics. As we are still in the early stage of applying informatics systems for tackling the emerging COVID-19 pandemic, it is worth mentioning that the statistical methods used in this study require further analysis to confirm its clinical effectiveness and appropriateness. The integrated informatics technologies, cost-effective solutions, and fast deployment provided the base for its replication in areas where COVID-19 is outbreaking and for similar disease pandemics in the future.

Acknowledgments

The authors appreciate Dr Douglas Fridsma, Dr Huang Kun, and Dr Shi Wenzhao for careful comments on earlier drafts of this work. The authors would like to thank Digital China Health Technologies co ltd, the Key Lab of Information Network Security of Ministry of Public Security (The Third Research Institute of Ministry of Public Security) under funding C19609, and Hangzhou Nuowei Information Technology for their support. We thank all patients and citizens of Honghu involved in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Honghu project information technology infrastructure configuration.
[DOC File , 41 KB - jmir_v22i4e18948_app1.doc]

Multimedia Appendix 2
Informed consent form used in this study in the Chinese language, as well as the corresponding English translation based on Google Translate for non-Chinese readers as a reference.
[DOCX File , 33 KB - jmir_v22i4e18948_app2.docx]

References

1. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A novel coronavirus emerging in China - key questions for impact assessment. N Engl J Med 2020 Feb 20;382(8):692-694. [doi: 10.1056/NEJMp2000929] [Medline: 31978293]
2. 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 Dec 01;214(suppl_4):S380-S385 [FREE Full text] [doi: 10.1093/infdis/jiw376] [Medline: 28830112]
3. Houlihan CF, Whitworth JA. Outbreak science: recent progress in the detection and response to outbreaks of infectious diseases. Clin Med (Lond) 2019 Mar;19(2):140-144 [FREE Full text] [doi: 10.7861/clinmedicine.19-2-140] [Medline: 30872298]
4. Abidi SSR, Abidi SR. Intelligent health data analytics: a convergence of artificial intelligence and big data. Healthc Manage Forum 2019 Jul;32(4):178-182. [doi: 10.1177/0840470419846134] [Medline: 31117831]
5. Dong X, Boulton ML, Carlson B, Montgomery JP, Wells EV. Syndromic surveillance for influenza in Tianjin, China: 2013-14. J Public Health (Oxf) 2017 Jun 01;39(2):274-281. [doi: 10.1093/pubmed/fdw022] [Medline: 26968483]
6. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. J Infect Dis 2016 Dec 01;214(suppl_4):S375-S379 [FREE Full text] [doi: 10.1093/infdis/jiw400] [Medline: 28830113]
7. Salathé M. Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health. J Infect Dis 2016 Dec 01;214(suppl_4):S399-S403 [FREE Full text] [doi: 10.1093/infdis/jiw281] [Medline: 28830106]
8. Nelli L, Ferguson HM, Matthiopoulos J. Achieving explanatory depth and spatial breadth in infectious disease modelling: integrating active and passive case surveillance. Stat Methods Med Res 2019 Jun 18:962280219856380. [doi: 10.1177/0962280219856380] [Medline: 31213191]
9. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506. [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
10. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020 Feb 22;395(10224):565-574. [doi: 10.1016/S0140-6736(20)30251-8] [Medline: 32007145]
11. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
12. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
13. Payne PRO, Lussier Y, Foraker RE, Embi PJ. Rethinking the role and impact of health information technology: informatics as an interventional discipline. BMC Med Inform Decis Mak 2016 Mar 29;16:40. [doi: 10.1186/s12911-016-0278-3] [Medline: 27025583]
14. SNOMED. SNOMED CT coronavirus content 2020 URL: https://confluence.ihtsdotools.org/display/snomed/SNOMED%2BCT%2BCoronavirus%2BContent [accessed 2020-03-23]
15. LOINC. 2020 Feb 05. Coronavirus surveillance: Regenstrief team creating unique codes for disease tracking URL: https://loinc.org/news/coronavirus-surveillance-regenstrief-team-creating-unique-codes-for-disease-tracking/ [accessed 2020-03-23]
16. Centers for Disease Control and Prevention. 2020 Feb 20. ICD-10-CM official coding guidelines - supplement coding encounters related to COVID-19 coronavirus outbreak URL: https://www.cdc.gov/nchs/data/icd/ICD-10-CM-Official-Coding-Gudance-Interim-Advice-coronavirus-feb-20-2020.pdf [accessed 2020-03-23]
17. Gong M, Wang S, Wang L, Liu C, Wang J, Guo Q, et al. Evaluation of privacy risks of patients' data in China: case study. JMIR Med Inform 2020 Feb 05;8(2):e13046. [doi: 10.2196/13046] [Medline: 32022691]
18. Guo L, Wei D, Zhang X, Wu Y, Li Q, Zhou M, et al. Clinical features predicting mortality risk in patients with viral pneumonia: the MuLBSTA score. Front Microbiol 2019;10:2752. [doi: 10.3389/fmicb.2019.02752] [Medline: 31849894]
19. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513. [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
20. CINIC. China Internet Network Information Center. 2019. Internet statistics URL: https://cnnic.com.cn/IDR/ [accessed 2020-03-23]
21. Feldstein LR, Ellis EM, Rowhani-Rahbar A, Hennessey MJ, Staples JE, Halloran ME, et al. Estimating the cost of illness and burden of disease associated with the 2014-2015 chikungunya outbreak in the U.S. Virgin Islands. PLoS Negl Trop Dis 2019 Jul;13(7):e0007563. [doi: 10.1371/journal.pntd.0007563] [Medline: 31323020]

Abbreviations

COVID-19: coronavirus disease
HHS: Honghu Hybrid System
ICD: International Classification of Diseases

MuLBSTA: Multilobular Infiltration, Hypo-Lymphocytosis, Bacterial Coinfection, Smoking History, Hyper-Tension and Age

Edited by G Eysenbach; submitted 29.03.20; peer-reviewed by MA Bahrami, D Gunasekeran, Ó Brito Fernandes, I Mircheva; comments to author 05.04.20; revised version received 09.04.20; accepted 09.04.20; published 22.04.20.

Please cite as:
Gong M, Liu L, Sun X, Yang Y, Wang S, Zhu H
Cloud-Based System for Effective Surveillance and Control of COVID-19: Useful Experiences From Hubei, China
J Med Internet Res 2020;22(4):e18948
URL: http://www.jmir.org/2020/4/e18948/
doi:10.2196/18948
PMID:32287040

Original Paper

Artificial Intelligence-Based Differential Diagnosis: Development and Validation of a Probabilistic Model to Address Lack of Large-Scale Clinical Datasets

Shahrukh Chishti[1], BTech; Karan Raj Jaggi[1], MBBS, MS, DNB, MCh; Anuj Saini[1], MBBS, MMST; Gaurav Agarwal[1], BTech; Ashish Ranjan[1], MBBS, MD

1mg Technologies Pvt Ltd, Gurgaon, India

Corresponding Author:
Karan Raj Jaggi, MBBS, MS, DNB, MCh
1mg Technologies Pvt Ltd
Motorola Building, MG Road
Gurgaon, 122001
India
Phone: 91 9654601958
Email: karan.jaggi@1mg.com

Abstract

Background: Machine-learning or deep-learning algorithms for clinical diagnosis are inherently dependent on the availability of large-scale clinical datasets. Lack of such datasets and inherent problems such as overfitting often necessitate the development of innovative solutions. Probabilistic modeling closely mimics the rationale behind clinical diagnosis and represents a unique solution.

Objective: The aim of this study was to develop and validate a probabilistic model for differential diagnosis in different medical domains.

Methods: Numerical values of symptom-disease associations were utilized to mathematically represent medical domain knowledge. These values served as the core engine for the probabilistic model. For the given set of symptoms, the model was utilized to produce a ranked list of differential diagnoses, which was compared to the differential diagnosis constructed by a physician in a consult. Practicing medical specialists were integral in the development and validation of this model. Clinical vignettes (patient case studies) were utilized to compare the accuracy of doctors and the model against the assumed gold standard. The accuracy analysis was carried out over the following metrics: top 3 accuracy, precision, and recall.

Results: The model demonstrated a statistically significant improvement (P=.002) in diagnostic accuracy (85%) as compared to the doctors' performance (67%). This advantage was retained across all three categories of clinical vignettes: 100% vs 82% (P<.001) for highly specific disease presentation, 83% vs 65% for moderately specific disease presentation (P=.005), and 72% vs 49% (P<.001) for nonspecific disease presentation. The model performed slightly better than the doctors' average in precision (62% vs 60%, P=.43) but there was no improvement with respect to recall (53% vs 56%, P=.27). However, neither difference was statistically significant.

Conclusions: The present study demonstrates a drastic improvement over previously reported results that can be attributed to the development of a stable probabilistic framework utilizing symptom-disease associations to mathematically represent medical domain knowledge. The current iteration relies on static, manually curated values for calculating the degree of association. Shifting to real-world data–derived values represents the next step in model development.

(J Med Internet Res 2020;22(4):e17550) doi:10.2196/17550

KEYWORDS

artificial intelligence; medical diagnosis; probabilistic modeling; Bayesian model; machine learning

Introduction

The World Health Organization (WHO) advocates a minimum doctor:population ratio of 1:1000; although this prescribed ratio has been attained in most of the Western world, 44% of the WHO member states report less than 1 physician per 1000 patients [1]. The situation is particularly grim in South Asia and Africa, with a ratio as low as 0.01 physicians per 1000 individuals in Malawi [2]. The WHO estimates a global shortage of 12.9 million health care workers by 2035 [3]. Such acute shortages in the health care system necessitate the development of low-cost, deployable, and scalable tools that can be integrated into multiple health care delivery models.

Several machine-learning and deep-learning algorithms have been applied to facilitate clinical diagnosis, but such tools often require large clinical datasets for training. Lack of availability of such datasets and inherent problems such as overfitting often necessitate the development of innovative solutions. We here introduce a probabilistic model for medical diagnosis that has been developed from the ground up. This method utilizes a stable probabilistic framework that was subsequently adapted to local disease patterns in India. The model was developed and validated against differential diagnoses made by six doctors with respect to various symptom presentation scenarios.

Methods

Model Development

For model development, we focused on infectious diseases as a major contributor to patient morbidity and mortality in developing countries, with most patients presenting with fever as their primary symptom [4]. The 15 most common causes of fever in India (Multimedia Appendix 1) were identified through national epidemiological data and were independently verified by internal medicine and infectious disease specialists. These 15 diseases represent a bulk of the patient load and were used for developing the framework of the probabilistic model and subsequent accuracy testing. This approach allows for the construction of diagnostic tools that provide high levels of diagnostic accuracy while retaining the inherent scalability provided by mathematical constructs.

In a medical consultation, the objective of the doctor-patient interaction is to gather evidence to formulate a provisional diagnosis based on the presenting symptoms. The addition of every new symptom results in the probability of disease being modified; that is, the prior probabilities are updated with the addition of new evidence. Thus, the science behind reaching a clinical diagnosis mimics a probabilistic framework.

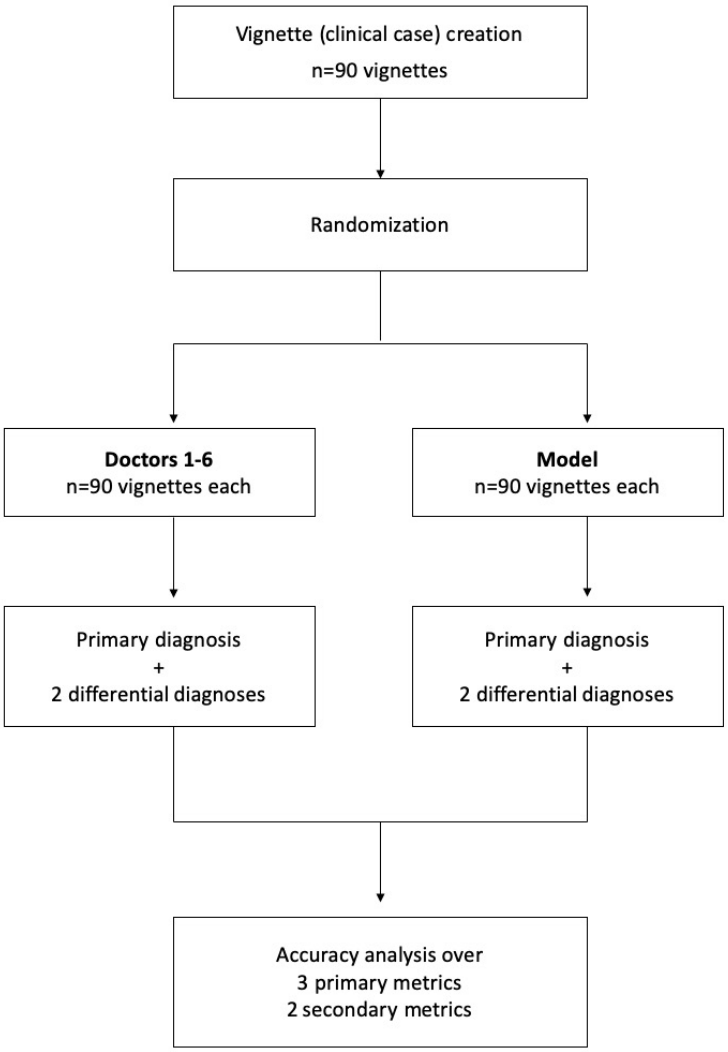
We developed a probabilistic model of diagnostic assessment, which was simplified to a multiclass classification problem. This required diseases to be distributed over a probabilistic distribution based on the presented evidence. The mathematical (Bayesian) interpretation of this distribution is a set of numbers corresponding to each possible disease. These numbers are representative of the probability of occurrence of a disease given the set of symptoms [5]. In this way, a list of differential diagnosis is generated.

To define the model, the following assumptions were considered: (1) medical history can be scientifically objectified, leading to definite universal symptom characteristics; (2) all possible diseases are included, which ensures a definite class of disease/condition; and (3) a single disease/condition is responsible for the set of presented symptoms. The mathematical framework for the model is described in detail in Multimedia Appendix 2.

Validation

An overview of the validation process is presented in Figure 1. Validation of the developed model was performed through a set of clinical vignettes (patient case summaries). To create the clinical vignettes, each of the 15 chosen diseases was stratified into three real-world clinical scenarios: highly specific disease presentation, moderately specific disease presentation, and nonspecific disease presentation. Two independent internal medicine specialists with a minimum clinical experience of 15 years each were entrusted with the creation of said vignettes, resulting in a total of 90 (15×3×2) clinical vignettes. The vignettes developed were subsequently verified by a third senior specialist to reduce any potential effects of selection bias.

Figure 1. Study design.



Each clinical vignette was modeled on the standard medical history format containing patient demographic details (age, sex, and date of presentation), presenting complaints, present illness, and medical history, including family history and personal history. A primary diagnosis and two differential diagnoses with percentage surety values were provided in each vignette and were used for accuracy analysis. A sample clinical vignette is shown in Multimedia Appendix 3.

The clinical vignettes were randomized, assigned a code, and subsequently provided to 6 independent doctors of varying clinical experience (0-10 years). All participating doctors were provided with a brief detailing the clinical context prior to initiation of the study. For each vignette, the participating doctors were asked to provide one primary diagnosis and two differential diagnoses along with percentage surety values.

To calculate the performance of the probabilistic model, the clinical vignettes were fed into the model and were used to generate a ranked list of diagnoses with percentage surety values. Percentage surety values were utilized as an arbitrary representation of the level of confidence for each of the differential diagnoses, with the sum being 100% for each vignette. The top diagnosis from this list was assumed as the primary diagnosis for accuracy analysis.

Accuracy Analysis

The performances of the model and the 6 doctors were compared against the assumed gold-standard diagnosis of the clinical vignettes under each of the three previously defined clinical scenarios: highly specific, moderately specific, and nonspecific disease presentation. Each of the chosen metrics provided a different perspective about the diagnostic accuracy.

The simplest method for accuracy analysis involves determination of the percentage of cases where the primary diagnosis (model or doctor) matches with the primary diagnosis of the clinical vignettes. Although this approach is useful in accuracy analysis for patients presenting with classical symptoms, it is not as suitable for patients with nonspecific presentations of the disease. Additionally, as all probabilistic models utilize only patient history for evaluation and do not take into account physical examination and relevant investigations, arriving at a single diagnosis is often medically unsound. Alternatively, top-3 accuracy, as an extension of the primary diagnostic accuracy, and standardized performance metrics such as precision and recall are useful in such scenarios, and were thus chosen as the main metrics for the current study.

The top-3 accuracy is an extension of the primary diagnostic accuracy and is defined as the percentage of cases where the

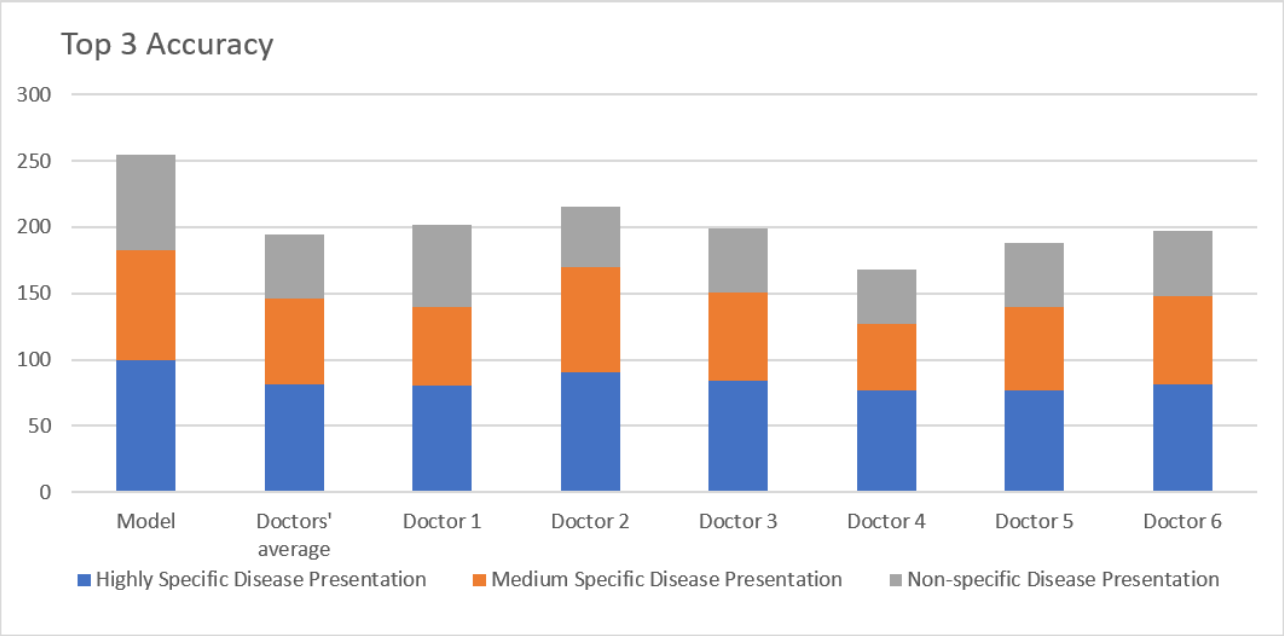
list of three differential diagnoses developed contains the primary diagnosis of the clinical vignettes [6]. Thus, the top-3 accuracy denotes not only the diagnostic accuracy but also the safety of the model in situations faced by the patient. Precision, measured as the positive predictive value, is defined as the percentage of correct predictions (true positives/true positives+false positives). Precision is a measure of exactness and can be understood as the percentage of times a disease was correctly predicted among the total number of times its prediction was made. Precision was determined for each of the 15 individual diseases and an aggregate precision value was calculated by averaging over the frequency of each of the diseases. Recall, measured by sensitivity, is defined as the percentage of correct identification (true positives/true positives+false negatives). It is the percentage of times a disease was correctly predicted among the total number of times the disease was present. Recall was determined for each individual disease and an aggregate recall value was calculated. An upper-tailed *t* test was used for calculation of statistical significance.

Results

For each of the defined metrics, an overall analysis of the performance of the model and doctors in comparison to the gold standard was performed. Additionally, comparative analysis was performed following stratification of the clinical vignettes based on the degree of specificity. Moving from highly specific to nonspecific vignettes denotes a progressive decrease in conclusive medical evidence available for diagnosis. In accordance, the performance of both the model and the participating doctors declined when moving down the spectrum of disease specificity.

The top-3 accuracy of the model (Figure 2) was 85% in comparison to the doctor average of 67% (*P*=.002). The statistical significance of this advantage was retained across all three categories of clinical vignettes, 100% vs 82% (*P*<.001) for highly specific disease presentation, 83% vs 65% for moderately specific disease presentation (*P*=.005), and 72% vs 49% (*P*<.001) for nonspecific disease presentation.

Figure 2. Top 3 accuracy. Comparison of model with doctors' average and individual doctor's performances over three scenarios: highly specific, moderately specific and nonspecific disease presentations.



The precision for the model (Figure 3) was 62% in comparison to the doctor average of 60%. However, this difference was not statistically significant (*P*=.43). Stratification of the results revealed that the model performed better than doctors in highly specific vignettes (78% vs 75%, *P*=.04) and moderately specific vignettes (61% vs 50%, *P*=.09) but not for clinical vignettes with low specificity (20% vs 39%, *P*<.001). The overall recall for the model (Figure 4) was 53% in comparison to the doctor average of 56% (*P*=.27). Results obtained following segmentation of the clinical vignettes followed similar patterns, with the model faring better than doctors in highly specific (81% vs 75%, *P*=.008) and moderately specific (60% vs 52%, *P*=.13) vignettes, but not for nonspecific vignettes (17% vs 36%, *P*<.001).

Figure 3. Precision. Comparison of model with doctors' average and individual doctor's performances over three scenarios: highly specific, moderately specific and nonspecific disease presentations.

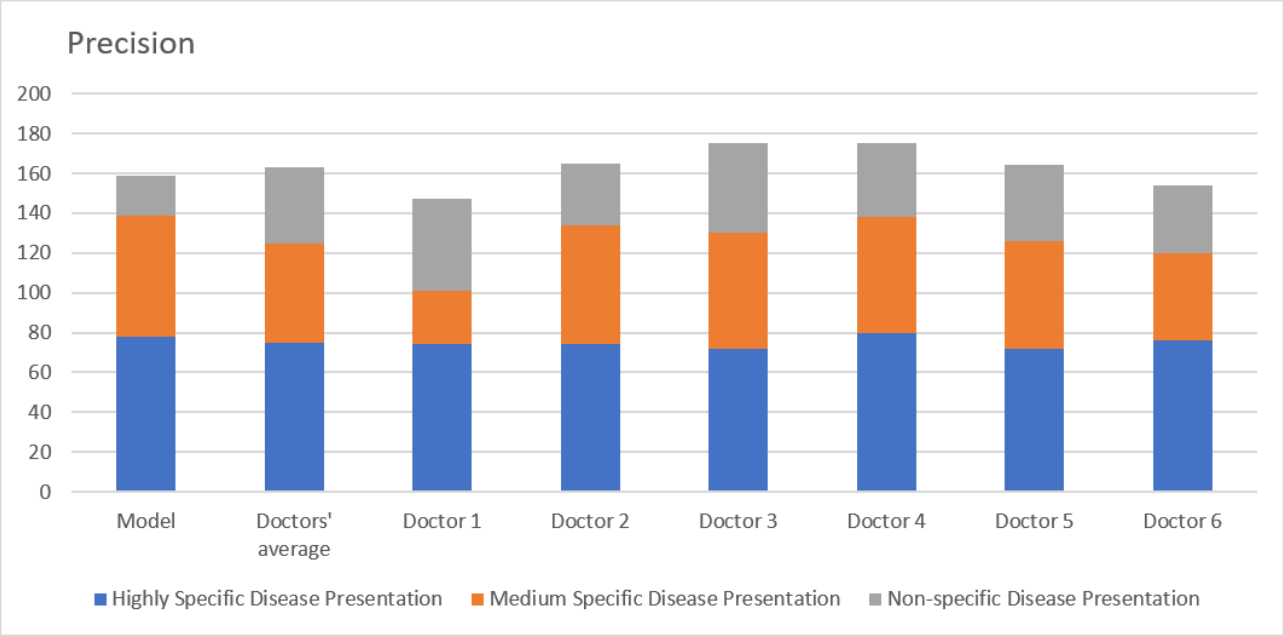
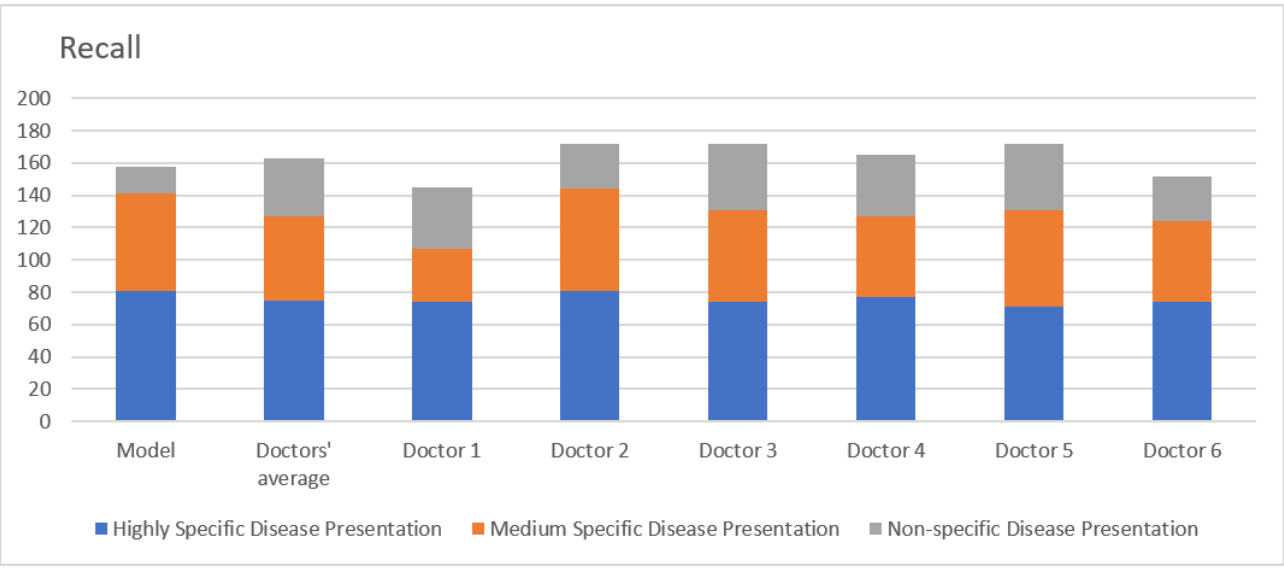


Figure 4. Recall. Comparison of model with doctors' average and individual doctor's performances over three scenarios: highly specific, moderately specific and nonspecific disease presentations.



As discussed above, the top-3 accuracy, precision, and recall take only the primary diagnosis into account, and thus additional metrics such as the Jaccard similarity index and cosine similarity are of greater clinical relevance. Additional details about the chosen metrics are provided in Multimedia Appendix 4. Performance analysis on the basis of these metrics revealed that the model consistently outperformed doctors for both the complete set of clinical vignettes as well as in each of the three subcategories of vignettes. The overall performance of the model on the basis of the Jaccard similarity index was higher than that of the average doctor (56% vs 47%, *P*=.02). The performance differential between the model and the doctors was relatively narrow in the highly specific vignettes (62% vs 56%, *P*=.02) but widened considerably in the moderate (62% vs 48%, *P*=.008) and low (43% vs 31%, *P*=.01) specificity vignettes. An overview of the same is demonstrated in Figure 5.

The cosine similarity of the model (Figure 6) was 72% in comparison to the doctors' average of 64% (*P*=.002). Performance following stratification of the vignettes revealed similar patterns to those found for the Jaccard similarity index, with a relatively narrow differential in the highly specific vignettes (79% vs 73%, *P*=.004) that widened in the moderate (72% vs 63%, *P*=.04) and low (65% vs 53%, *P*=.002) specificity of vignettes.

Figure 5. Jaccard Similarity. Comparison of model with doctors' average and individual doctor's performances over three scenarios: highly specific, moderately specific and nonspecific disease presentations.

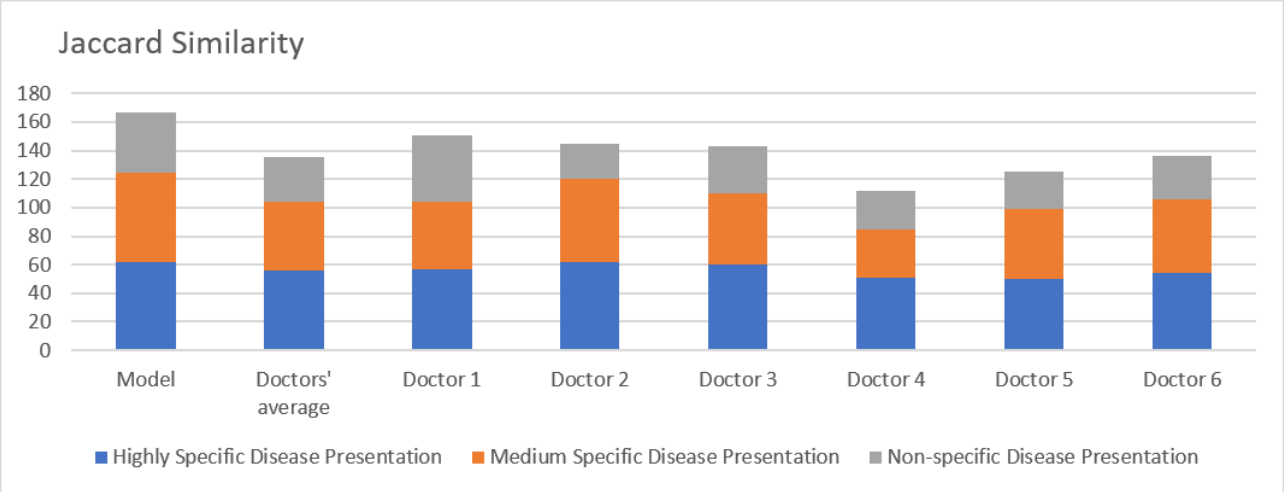
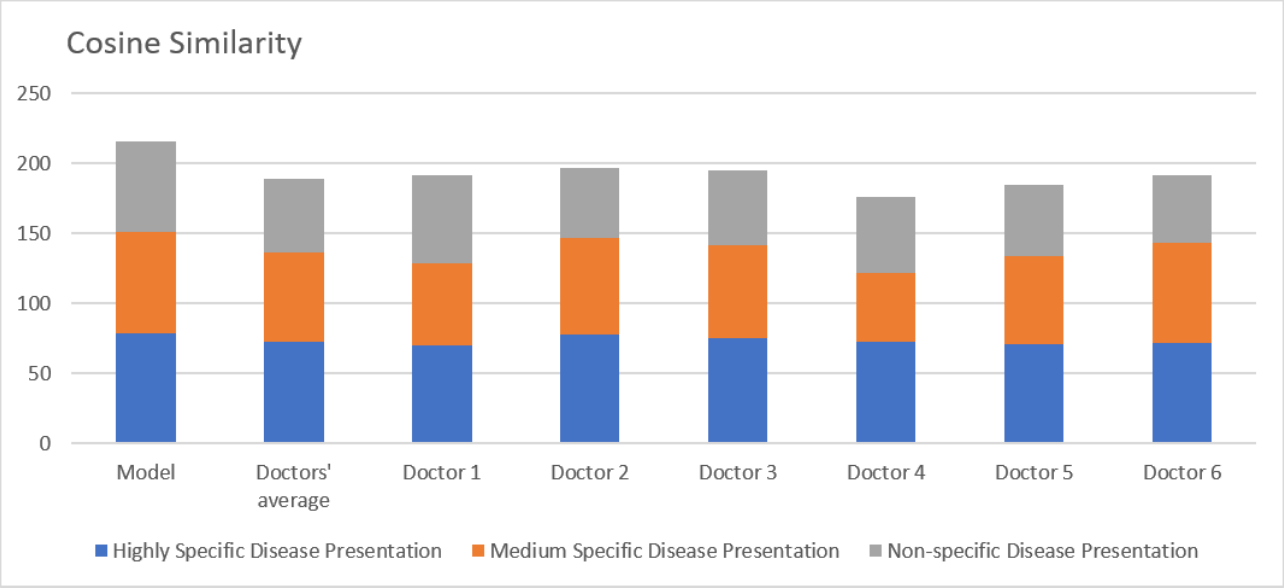


Figure 6. Cosine Similarity. Comparison of model with doctors' average and individual doctor's performances over three scenarios: highly specific, moderately specific and nonspecific disease presentations.



Discussion

Principal Findings

Historically, the art of medical diagnosis has relied upon the three pillars of patient history, physical examination, and investigative reports. Hampton et al [7] reported that a detailed and thorough patient history was sufficient to reach a clinical diagnosis in 83% of patients presenting to the medical outpatient department. This statistic is particularly relevant in settings where a patient's physical examination and investigative tests cannot be performed. Advances in statistical analysis tools have allowed for the creation of probabilistic models serving as both diagnostic and prognostic tools for a myriad of clinical scenarios [8]. These models have been predominantly deployed in patients faced with the need for self-diagnosis, leading to the development of the common moniker "symptom checker" [9].

For the purpose of accuracy analysis, most authors advocate performance testing of both the model and doctors against an assumed gold standard. Evans et al [10] analyzed vignette methodology and concluded that vignettes are a powerful tool to study physicians' clinical judgement, and can be strongly reflective of clinicians' real-world behavior. This approach ensures that the accuracy analysis of the model is not unfairly hampered by errors in diagnosis by individual doctors. Additionally, as both the model and the doctors are fed the same information, over a statistically significant number, the quality of the clinical vignettes in question ceases to affect the accuracy analysis as the performance of both the model and doctors increases or decreases according to the level of information contained in the vignette. Razzouk et al [11] compared the performance of a decision support system for the diagnosis of schizophrenia disorders against an expert using 38 clinical vignettes prepared from outpatient charts. In a retrospective study, Ronicke et al [12] prepared clinical vignettes by extracting information from medical records, which were fed into the model (Ada DX) after anonymization. The second method of accuracy analysis involves a direct comparison of the model performance against diagnosis by the doctors, thereby

assuming the doctors' response as the gold standard. However, this approach does not take into account errors in judgement of the doctors themselves, which is currently estimated to vary between 10% and 15% [13]. This method also heavily relies on the assumption that the doctors participating in the study have sufficient clinical expertise, thus negating the effects of error over a statistically significant sample size. A recent study conducted by Berry et al [14] compared the diagnostic accuracy of the three most commonly used symptom checkers (WebMD, iTriage, FreeMD) in comparison to that of the doctors for patients presenting with abdominal pain, and found the accuracy of the top diagnosis to be 14.3%, with the top-3 accuracy being 36.7%. However, it is notable that only 49 patients (statistically insignificant) were included in the study and the level of experience of the practitioners was not taken into consideration [14].

Interestingly, once the sample size increases, the effects of diagnostic errors by individual doctors decreases, thereby presenting a fairer representation of the performance of the model. Bisson et al [15] compared the diagnostic accuracy of a web-based symptom checker for 328 patients (163 men and 165 women) presenting with knee pain, which revealed an accuracy of 58% for the model.

In the current study, the performance of both the model and doctors was compared to the assumed gold standard. In our opinion, this approach is statistically sound and should be considered for the primary performance evaluation of developed models, with a one-to-one analysis reserved for specific indications once the prerequisites of a statistically significant sample size and demonstrated clinical experience of the participating doctors have been met. The chosen metrics reflect not only the diagnostic accuracy but also the capability of the model to adapt to varying clinical scenarios and disease presentations [16].

The performance of the present model was superior to that of the panel of doctors across the entire gamut of chosen primary and secondary metrics. Several studies have failed to demonstrate such levels of clinical performance, with all trained models trailing in diagnostic accuracy achieved by clinicians. Semigran et al [9] used 45 clinical vignettes and reported 51.2% accuracy for the top-3 accuracy of an online symptom checker in comparison to 84.3% for doctors. A similar study conducted by Shen et al [6] for ophthalmologic diagnosis found the top-3 accuracy of the model to be a mere 38%. Davies et al [17] found that web-based symptom checkers listed degenerative cervical myelopathy as a differential diagnosis in only 45% of symptom composites tailored from 31 recognized symptoms in the literature. In addition, Ronicke et al [12] showed that their model (Ada DX) suggested a correct diagnosis in the top 5 suggestions in 53.8% of cases and as the top diagnosis in 37.6% of cases.

Although Razzouk et al [11] reported the model's accuracy to be 66%-82% for diagnosing schizophrenia, it is important to mention that their model was constructed to diagnose only one disease and was evaluated for diagnosis of that particular disease.

Strengths and Limitations

The current study represents the first probabilistic model that consistently outperformed trained medical professionals. This jump in diagnostic accuracy can by and large be attributed to building a model from the ground up. Various authors have utilized preconstructed or preconfigured models either directly or after small modifications. Shen et al [6] and Davies et al [17] used readily available web-based symptom checker tools such as WebMD, Healthtools, AARP, Healthline, or Netdoctor to study the accuracy of such models against the diagnostic performance of practicing physicians. However, in the present study, the probabilistic model was constructed from scratch. The mathematical framework was closely modeled to mimic the science behind arriving at a clinical diagnosis. Development of symptom-disease associations offers a novel approach to mathematically represent medical domain knowledge. Additionally, utilizing local epidemiological trends and disease profiles to develop these associations resulted in a high degree of accuracy. These symptom-disease relevance associations were manually curated by a team of doctors. This exercise was carried out over several iterations with extensive feedback from various medical experts. The model in the study represents the 6th iteration.

Delving deeper into the model performance after stratification of the clinical vignettes revealed some features that merit special mention. In cases where the clinical vignettes contain a large amount of specific clinical data (ie, highly specific and moderately specific disease presentation), the probabilistic model performed at par or better than the doctors across the gamut of chosen metrics (see Multimedia Appendix 5). However, this advantage was lost in the case of nonspecific disease presentation, with the model significantly trailing the panel of doctors in precision and recall values. These values serve as a reminder that even high-performing models fail in rare instances such as for patients with nonspecific disease presentations. Various studies have demonstrated similar results. In a review article, Mishra et al [18] lists unusual/atypical/silent disease presentation (nonspecific disease presentation) under the category "no fault errors." They conclude that such errors are due to limitations of present medical knowledge and can only be reduced by furtherance of medical research and technological advancements. These findings are particularly relevant from a clinical context as they reinforce the fact that all such diagnostic tools should not be developed as a replacement for doctors but only to serve as clinical decision support systems. In a systematic review, Garg et al [19] reported that among 97 randomized and nonrandomized controlled trials studying the role of computerized clinical decision support systems in clinical practice, 64% of studies demonstrated improved practitioner performance. This approach will ensure standardization of care while retaining the inherent safety provided by a thorough clinical evaluation by a trained medical professional.

The present study suffers from a few limitations. First, the major limitation lies in the small sample size. Accuracy analysis was performed by comparing the performance of the model and the doctors over 90 clinical vignettes. Second, the present study was based on a small dataset (ie, patients presenting with fever

in a developing country). This problem statement has been used to develop and subsequently test a probabilistic model. Further large-scale studies are required to prove the clinical relevance of this model across different clinical scenarios. The authors have attached the relevant study data in the appendices and encourage researchers to independently replicate the results on a statistically significant sample size. Lastly, a claim can be made that clinical vignettes might not represent a true visualization of the disease presentation, and accuracy analysis on actual patient data might prove to be beneficial [20].

Future Scope

The current study represents a proof of concept of a probabilistic clinical decision support system. The present iteration of the model relies on static, manually curated values for calculating the degree of association. Shifting to real-world data–derived values represents the next step in model development. This migration would not only enhance diagnostic accuracy but also provide the ability to adapt to sudden changes in the disease environment in real time, an invaluable asset in disease epidemic prediction. Incorporation of investigative reports in the current framework—although challenging—is the key to a major jump in diagnostic accuracy. Accuracy testing for the current model has been performed through clinical vignettes that represent artificial textbook cases of disease presentations. Testing on real patient data would offer additional insights into the performance of the model.

Conclusions

The present research demonstrates a drastic improvement over previously reported results that can be attributed to the development of the current model while keeping the local patient presentation and disease profile in mind instead of utilizing an off-the-shelf approach. This approach provides a greater degree of diagnostic accuracy than previous models. In addition, extensive involvement of practicing clinicians during the development phase is essential for the creation of a solution with demonstrable accuracy and clinical relevance. Importantly, these results are based on a relatively narrow dataset with the aim of developing a proof of concept. Therefore, additional large-scale clinical trials need to be conducted before these models can be deployed universally. In the interest of patient safety, the authors suggest positioning of all such tools as clinical decision support systems rather than as a substitute for trained medical doctors.

Acknowledgments

Advice provided by Dr Varun Gupta was a great help in the study. Assistance provided by Dr Sarthak Chakravarty was greatly appreciated. We would like to thank our external advisors Dr Neha Gupta, Dr Anoop Purkayastha, Dr Waseem Mirza, Dr Sheetal Kaul, and Dr Shweta Karn for their expert opinions. We would like to extend our special thanks to the six doctors for participating in the study. We wish to acknowledge the help provided by our colleagues Harshad Purandare, Bhavik Grover, and Suprabho Dhenki. This research was supported by 1mg Technologies Private Limited, and all of the participating doctors were compensated for their time and effort.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Fifteen most common causes of fever in India.
[DOC File , 34 KB - jmir_v22i4e17550_app1.doc]

Multimedia Appendix 2
Mathematical framework.
[DOC File , 35 KB - jmir_v22i4e17550_app2.doc]

Multimedia Appendix 3
Sample clinical vignette.
[DOC File , 35 KB - jmir_v22i4e17550_app3.doc]

Multimedia Appendix 4
Secondary metrics.
[DOC File , 31 KB - jmir_v22i4e17550_app4.doc]

Multimedia Appendix 5
Detailed results.
[DOC File , 57 KB - jmir_v22i4e17550_app5.doc]

References

1. Global Health Observatory (GHO) Data.: World Health Organization; 2019. Density of physicians (total number per 1000 population), latest available year URL: https://www.who.int/gho/health_workforce/physicians_density_text/en/ [accessed 2019-12-16]
2. WHO Global Health Workforce Statistics. World Health Organization URL: http://www.who.int/hrh/statistics/hwfstats/ [accessed 2019-12-16]
3. Third Global Forum on Human Resources for Health Report. 2013. A Universal Truth: No Health Without a Workforce URL: https://www.who.int/workforcealliance/knowledge/resources/GHWA-a_universal_truth_report.pdf?ua=1 [accessed 2019-12-16]
4. Herbrich R. Learning Kernel Classifiers. Cambridge, MA: MIT Press; 2002.
5. Joyce J. Stanford Encyclopedia of Philosophy. Bayes' Theorem URL: https://plato.stanford.edu/entries/bayes-theorem/ [accessed 2019-12-16]
6. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a Popular Online Symptom Checker for Ophthalmic Diagnoses. JAMA Ophthalmol 2019 Jun 01;137(6):690-692. [doi: 10.1001/jamaophthalmol.2019.0571] [Medline: 30973602]
7. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975 May 31;2(5969):486-489 [FREE Full text] [doi: 10.1136/bmj.2.5969.486] [Medline: 1148666]
8. Cahan A, Gilon D, Manor O, Paltiel O. Probabilistic reasoning and clinical decision-making: do doctors overestimate diagnostic probabilities? QJM 2003 Oct 18;96(10):763-739. [doi: 10.1093/qjmed/hcg122] [Medline: 14500863]
9. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
10. Evans SC, Roberts MC, Keeley JW, Blossom JB, Amaro CM, Garcia AM, et al. Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. Int J Clin Health Psychol 2015 May;15(2):160-170 [FREE Full text] [doi: 10.1016/j.ijchp.2014.12.001] [Medline: 30487833]
11. Razzouk D, Mari J, Shirakawa I, Wainer J, Sigulem D. Decision support system for the diagnosis of schizophrenia disorders. Braz J Med Biol Res 2006 Jan;39(1):119-128 [FREE Full text] [doi: 10.1590/s0100-879x2006000100014] [Medline: 16400472]
12. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis 2019 Dec 21;14(1):69 [FREE Full text] [doi: 10.1186/s13023-019-1040-6] [Medline: 30898118]
13. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013 Oct;22 Suppl 2:ii21-ii27 [FREE Full text] [doi: 10.1136/bmjqs-2012-001615] [Medline: 23771902]
14. Berry A, Cash B, Mulekar M, Wang B, Melvin A, Berry B. Symptom Checkers vs. Doctors, the Ultimate Test: A Prospective Study of Patients Presenting with Abdominal Pain. Gastroenterology 2017 Apr 02;152(5):S852-S853 [FREE Full text] [doi: 10.1016/S0016-5085(17)32937-2]
15. Bisson LJ, Komm JT, Bernas GA, Fineberg MS, Marzo JM, Rauh MA, et al. How Accurate Are Patients at Diagnosing the Cause of Their Knee Pain With the Help of a Web-based Symptom Checker? Orthop J Sports Med 2016 Feb 19;4(2):2325967116630286 [FREE Full text] [doi: 10.1177/2325967116630286] [Medline: 26962542]
16. Brown JB. Classifiers and their Metrics Quantified. Mol Inform 2018 Jan 23;37(1-2):1700127 [FREE Full text] [doi: 10.1002/minf.201700127] [Medline: 29360259]
17. Davies BM, Munro CF, Kotter MR. A Novel Insight Into the Challenges of Diagnosing Degenerative Cervical Myelopathy Using Web-Based Symptom Checkers. J Med Internet Res 2019 Jan 11;21(1):e10868 [FREE Full text] [doi: 10.2196/10868] [Medline: 30300137]
18. Mishra D, Gupta P, Singh T. Teaching for Reducing Diagnostic Errors. Indian Pediatr 2017 Jan 15;54(1):37-45 [FREE Full text] [doi: 10.1007/s13312-017-0994-5] [Medline: 28141563]
19. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
20. Randhawa H, Jiwa A, Oremus M. Identifying the components of clinical vignettes describing Alzheimer's disease or other dementias: a scoping review. BMC Med Inform Decis Mak 2015 Jul 15;15:54 [FREE Full text] [doi: 10.1186/s12911-015-0179-x] [Medline: 26174660]

Abbreviations

WHO: World Health Organization

Edited by G Eysenbach; submitted 19.12.19; peer-reviewed by K Makker, R Dua; comments to author 22.01.20; revised version received 30.01.20; accepted 01.02.20; published 28.04.20.

Please cite as:
Chishti S, Jaggi KR, Saini A, Agarwal G, Ranjan A
Artificial Intelligence-Based Differential Diagnosis: Development and Validation of a Probabilistic Model to Address Lack of Large-Scale Clinical Datasets
J Med Internet Res 2020;22(4):e17550
URL: https://www.jmir.org/2020/4/e17550
doi:10.2196/17550
PMID:32343256

Original Paper

Development, Implementation, and Evaluation of a Personalized Machine Learning Algorithm for Clinical Decision Support: Case Study With Shingles Vaccination

Ji Chen[1], MS; Sara Chokshi[1], DrPH; Roshini Hegde[1], MPH; Javier Gonzalez[2], BA; Eduardo Iturrate[3], MD; Yin Aphinyanaphongs[1], MD, PhD; Devin Mann[1,2], MD, MS

[1]Department of Population Health, New York University School of Medicine, New York, NY, United States

[2]Medical Center Information Technology, New York University Langone Health, New York, NY, United States

[3]Clinical Informatics, New York University School of Medicine, New York, NY, United States

Corresponding Author:
Ji Chen, MS
Department of Population Health
New York University School of Medicine
145 E 32nd St
New York, NY, 10016
United States
Phone: 1 6465012505
Email: ji.chen@nyulangone.org

Abstract

Background: Although clinical decision support (CDS) alerts are effective reminders of best practices, their effectiveness is blunted by clinicians who fail to respond to an overabundance of inappropriate alerts. An electronic health record (EHR)–integrated machine learning (ML) algorithm is a potentially powerful tool to increase the signal-to-noise ratio of CDS alerts and positively impact the clinician's interaction with these alerts in general.

Objective: This study aimed to describe the development and implementation of an ML-based signal-to-noise optimization system (SmartCDS) to increase the *signal* of alerts by decreasing the volume of low-value herpes zoster (shingles) vaccination alerts.

Methods: We built and deployed SmartCDS, which builds personalized user activity profiles to suppress shingles vaccination alerts unlikely to yield a clinician's interaction. We extracted all records of shingles alerts from January 2017 to March 2019 from our EHR system, including 327,737 encounters, 780 providers, and 144,438 patients.

Results: During the 6 weeks of pilot deployment, the SmartCDS system suppressed an average of 43.67% (15,425/35,315) potential shingles alerts (appointments) and maintained stable counts of weekly shingles vaccination orders (326.3 with system active vs 331.3 in the control group; P=.38) and weekly user-alert interactions (1118.3 with system active vs 1166.3 in the control group; P=.20).

Conclusions: All key statistics remained stable while the system was turned on. Although the results are promising, the characteristics of the system can be subject to future data shifts, which require automated logging and monitoring. We demonstrated that an automated, ML-based method and data architecture to suppress alerts are feasible without detriment to overall order rates. This work is the first alert suppression ML-based model deployed in practice and serves as foundational work in encounter-level customization of alert display to maximize effectiveness.

(J Med Internet Res 2020;22(4):e16848) doi:10.2196/16848

KEYWORDS

clinical decision support; machine learning; alert fatigue; implementation science

Introduction

Background and Significance

The potential effectiveness of clinical decision support (CDS) alerts as a scalable tool for promoting evidence-based care for vaccine administration has led to their frequent use by most health care systems seeking to maximize vaccination rates [1-4]. CDS alerts to prompt evidence-based practices have been extensively studied and shown to work best when delivered at an appropriate time and place in the clinical workflow, that is, when the clinician is prepared to receive the information [5-8]. Successful CDS alerts have led to a reduction in prescribing brand-name antibiotics [9], improved lipid management in renal transplant patients [10], improved compliance with guidelines for treating HIV [11-13], reduced ordering of tests when costs were displayed [14], and age-specific alerts that reduce inappropriate prescribing in the elderly [15-20].

Although CDS tools are effective reminders of best practices, their effectiveness is blunted by the context in which they are deployed; *alert fatigue* (clinician desensitization driven by overwhelming number and quality of safety alerts) [21] is the result of an ever-growing number of alerts in the electronic health record (EHR), leading to clinicians commonly ignoring or failing to respond appropriately to alerts. Alert fatigue resulting from an excess of poor-quality alerts (eg, alerts firing at inappropriate times or for inappropriate patients) contributes to clinicians' perceptions that the bulk of alerts are likely clinically insignificant regardless of their clinical message. As a result, clinicians now override most medication alerts [22-24] and are becoming increasingly desensitized to alarms [25-27]. Although there is limited consensus on how to measure alert fatigue and its unintended consequences, data show that alert fatigue is significantly impacting the clinician experience and patient care [28-31]. At our large academic health system, the number of active interruptive alerts for providers grew from 13 in 2012 to 107 in 2018, an increase of more than 800%. In December 2018, our providers ordered the shingles vaccine in response to just 6.43% (2219/34,531) of the alerts, indicating that our clinicians view a majority of these alerts as inappropriate. Consequently, an improved EHR experience for clinicians has become an institutional priority for many health systems, including our own.

Individual-level factors, including clinicians' bias toward ignoring alerts and poor signal detection resulting from the overwhelming number of alerts, and poor alert reliability add to the degraded effectiveness of CDS and user experience [32,33]. To optimize a CDS system means to optimize the *signal-to-noise* ratio of alerts by increasing the *signal*, decreasing the *noise* created by an abundance of inappropriate, poorly timed alerts or both. To this end, prior work in medication alerts and monitoring alarms have implemented advanced interventions that use rules to surface or suppress alerts, intending to improve CDS alert signal. A study using basic rules to deactivate irrelevant alerts and manually alter other alert frequencies based on severity decreased the override rates from 33.6 to 4.6 per 100 orders [34]. Similar severity ranking showed success in increasing alert acceptance rate by 50%, despite a 60% increase in alert events [35]. Other studies attempting to reduce noise, however, achieved limited or mixed results [36-45].

Research indicates that delivering alerts at the appropriate time and place in the clinical workflow is key to effective CDS [5-8]. Prior work to optimize CDS tools focuses on manual approaches [4,46]; these have proven to be time consuming, difficult to maintain, and static, limiting scalability. Optimization and incorporation of more sophisticated rules to surface or suppress alerts achieves limited reduction [36-41]. Machine learning (ML) is a powerful tool for identifying patterns in complex data by using past data to predict future performance. The use of ML in health care has proliferated over the past 10 years in a variety of use cases. ML applied to EHR data specifically shows signs of promise as a tool for improving safety and quality of care; its application to problems such as predicting readmission and sepsis shows the ability of ML ability to better target alerts to the appropriate user and use case [47-49]. An EHR-integrated ML algorithm is a potentially powerful tool to improve the quality of care by increasing the signal-to-noise ratio of alerts to positively impact clinicians' interactions with these alerts. To date, the informatics literature lacks both prospective evaluation of signal-to-noise optimization interventions as well as detailed accounts of operational steps necessary to implement ML models in clinical care. Using the shingles vaccination alert as our initial use case, we leveraged historical EHR interaction data (clicks), patient and provider sociodemographic data to (1) build and train an ML model that can predict the likelihood of provider interaction with the shingles vaccination alert and (2) establish the data architecture necessary to deploy the model in a live environment.

Objective

The objective of this case study was to describe the development, implementation, and prospective evaluation of a novel, ML-based, CDS signal-to-noise optimization (SmartCDS) system that suppresses low-value vaccination alerts applied to a shingles vaccination CDS alert.

Methods

Setting

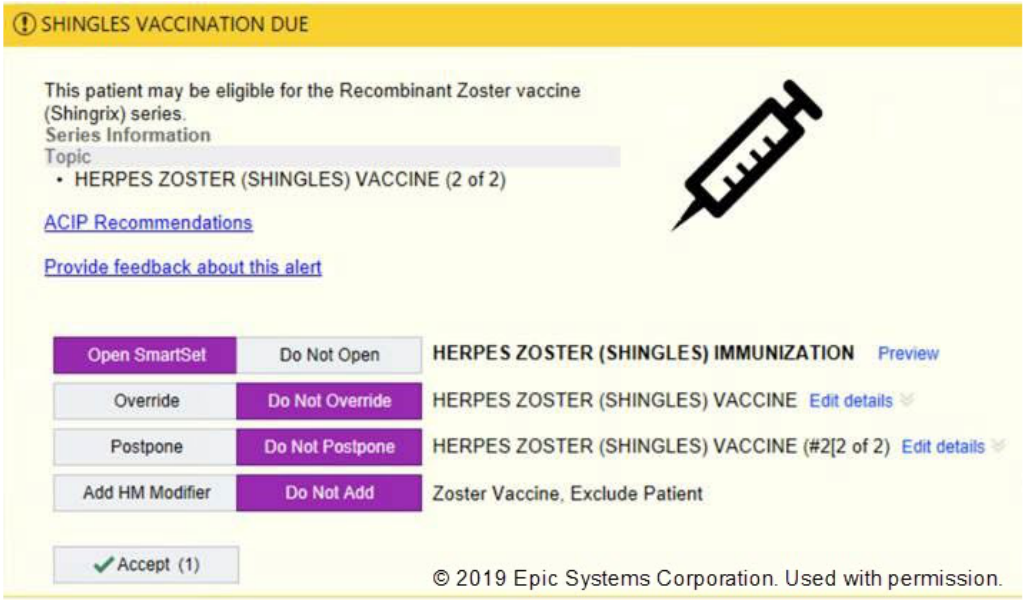
This work was conducted within a large urban academic hospital system with approximately 1300 beds over several satellite locations. In the fiscal year 2016, 3584 doctors and 4899 nurses treated approximately 38,000 inpatient admissions, 5.8 million outpatient visits, and 150,000 emergency department visits.

Data

This study uses all data from January 1, 2017, to March 11, 2019, to maintain consistency with the shingles alert content and its clinical setting. The dataset includes a total of 695,311 shingles alerts presented to 780 providers over 327,737 encounters, covering 144,438 unique patients. The overall alert interaction rate (any action toward acknowledging the shingles alert in an encounter) during this period was 16%, and the overall order rate of the shingles vaccine in response to the alert was 5%. The alert response options are illustrated in Figure 1—providers may choose from four different actions: *open SmartSet* to sign vaccine orders for targeted patients, health

maintenance *override*, *postpone* or customize health maintenance modifier based on refusal, and deferral or other decisions made by patients; the alert appears on a side tab located at the right side of the EHR interface and may also be ignored or closed when no action is taken.

Figure 1. Screenshot of the shingles vaccination electronic health record alert.



Alert Suppression Model Construction

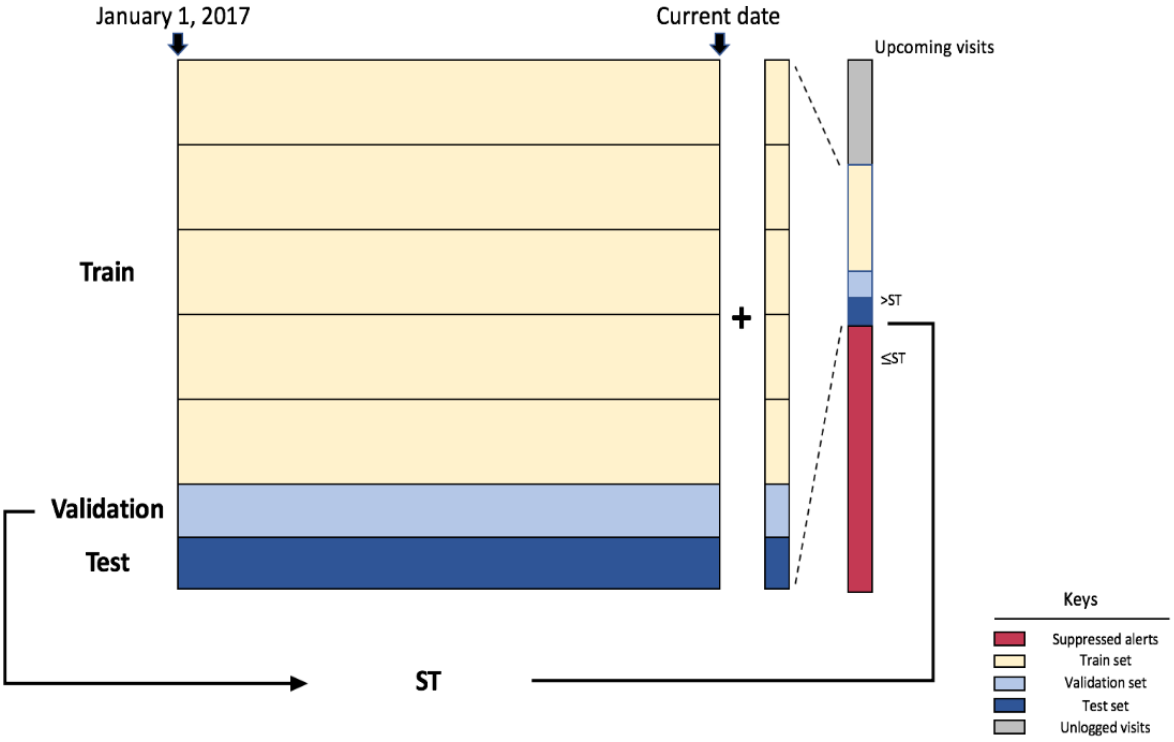
Feature Construction

A retrospective data query was performed to extract data related to the shingles vaccine alert. Key data elements to be extracted from the EHR were determined using a combination of descriptive analysis and clinical expertise. After data cleaning, historical changes were analyzed in alert usage to determine the optimal period from which data can be extracted for model training (two years of data from January 1, 2017, to December 31, 2018). Using data from our alert system, the average response rates for the alerts as well as the providers' interaction history with the alerts were examined for the purpose of determining an appropriate protocol for assigning one unique provider to each alert encounter. Initial analyses demonstrated a large variation among clinicians with regard to the frequency of interaction with the alerts (0%-92%), prompting our team to construct variables for an individual clinician's activity history, which was expanded to several short-term and long-term activity history variables capturing response rates, alert volume, and demographic variables for both clinicians and patients. The features that affect clinician's response include (1) clinician-level demographics, clinical roles, and specialties; (2) response rate to previous shingles alert (both short-term and long-term); and (3) the number of recent encounters. The patient-level data included were patient demographics and history of targeted shingles alert responses and shingles vaccine orders by clinicians. In addition, a binary flag indicating walk-in visits and scheduled office visits was included as the architecture did not capture walk-in visits in our pilot implementation.

Machine Learning Model

The model was designed as a binary classification task. The target labels were built based on whether an alert instance was interacted with or whether a follow-up order for shingles vaccination was placed in each primary care visit. The data were split randomly based on individual clinicians into 80%, 10%, and 10% sets for model training, validation, and testing, respectively, as illustrated in Figure 2. XGBoost was employed as our ML algorithm, with learning rate=0.3, maximum tree depth=0.6, minimum child weight=1, no subsampling, negative log loss, and early stopping (with a maximum of 50 rounds). The validation set was used to monitor the model training through early stopping to derive the operational score threshold and evaluate the model performance; the test set was used to evaluate the effectiveness of the score threshold and the generalizability of the trained model retrospectively. To evaluate the performance of the model, we obtained a sample of nearly 65,000 primary care visits. We reported a highly effective model, adopting individual profiling of providers to reduce the number of clinically insignificant alerts, with average area under receiver operating characteristic of 0.919 and average area under precision-recall curve of 0.562 using 5-fold cross-validation. Our simulation found that of the 50.00% (6490/12,980) lowest ranked vaccination alerts, 99.77% (6475/6490) have been ignored by providers if not suppressed. Given that the corresponding estimated order reduction via nested cross-validation was deemed conservative at 1%, a 50% suppression threshold was selected in collaboration with clinical stakeholders [50]. As a result, the model that relies on personal history for features is updated daily to incorporate the latest data and update the 50% score threshold during prospective implementation. Upcoming appointments are used for ongoing training, making the training window ongoing. The patients' appointments for initial primary care visits are excluded from suppression.

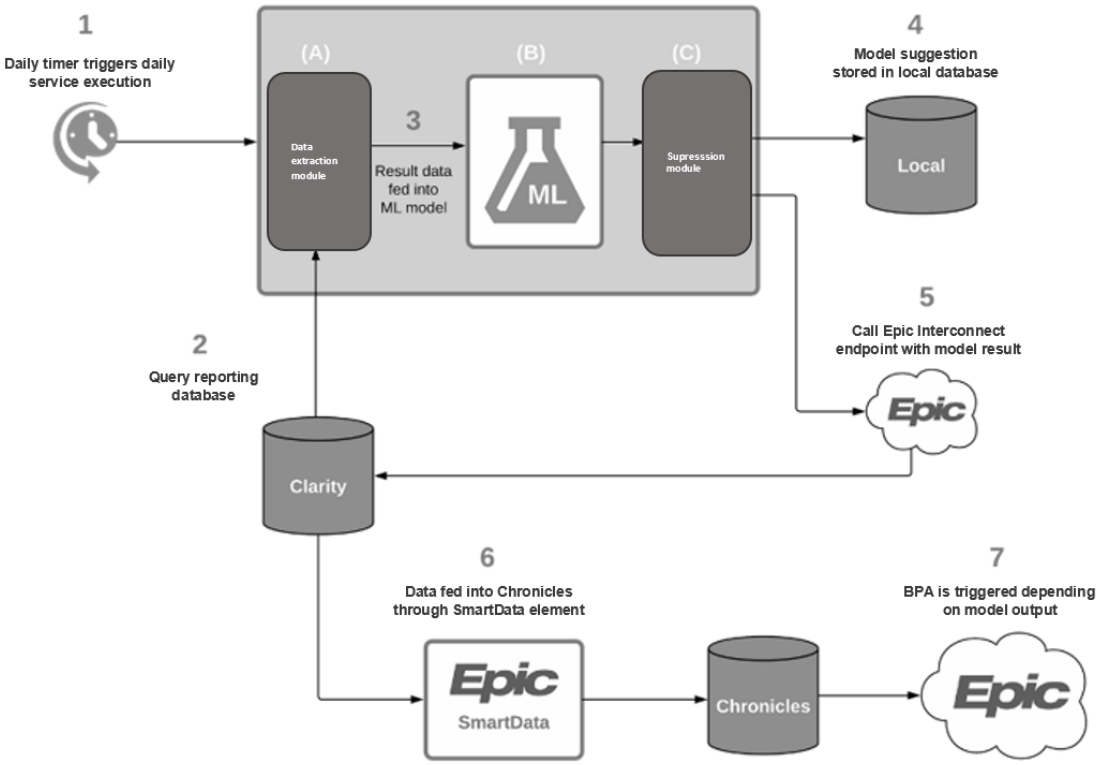
Figure 2. Experimental design. All data from January 1, 2017, to March 11, 2019, are used to train the model. Data are divided based on clinicians into 80%, 10%, 10% splits as train, validation, and test set, respectively. Each day, the data from the previous day are added to the dataset and the model is retrained and evaluated on the updated validation set to derive the 50% suppression score threshold. Predictions are made on upcoming visits (appointments) following the shingles best practice advisory (BPA) eligibility in the next day, and BPA instances are suppressed if the predicted score is lower than the threshold. Upcoming visits, which are logged into the shingles BPA log in the electronic health record system, are used for training in the future. In this design, the training window is always growing, with January 2017 as the start date. ST: score threshold.



Pilot Design and Evaluation

A pilot study was designed over 6 weeks (Figure 3) in biweekly cycles (alternating turning the model on for one week and turning the model off for another week) to verify that the data distribution in the training/validation set was applicable to that in production. In the pilot, key statistical measures were examined with the model both turned on and turned off to compare prospective model performance with estimates generated in retrospective evaluation. The provider response and follow-up orders associated with suppressed shingles alerts cannot be measured; therefore, prospective model performance was evaluated using the percentage of daily suppressed alerts, daily alert response rate, weekly shingles vaccination order count, and alerts per order rate. Weekly aggregated measures were employed because of weekly patterns detected in the clinical setting (eg, Wednesdays and weekend days featured lower alert volume).

Figure 3. SmartCDS system architecture. This illustrates the SmartCDS machine learning (ML) implementation flow. (1) A timer (Cron job) is configured to run every day and invoke the SmartCDS service. (2) The data extraction module will query the reporting database (Epic Clarity) and feed (3) encounter, provider, user demographic, and best practice advisory data to the ML model. (4) The model output is then both stored in a local database for further analysis and pushed (5) to Epic through an Epic Interconnect Web Services endpoint. From here, information about what alert per encounter should be suppressed is written (6) into the Epic event database (Chronicles) through a SmartData element. An alert rule will inspect (7) these data to allow or suppress the alert being fired. BPA: best practice advisory; ML: machine learning.



Results

Architecture for Signal-to-Noise Optimization System Deployment

After the construction of ML model for alert suppression (see Methods), we built a new data architecture to operationalize the model (Figure 3). The overall signal-to-noise optimization (SmartCDS) system was broken into three components: (1) the *data extraction module*, which identifies planned visits for the next day, with the intent to identify upcoming vaccine alerts to suppress, and queries the EHR to extract the variables required to run the ML module; (2) the suppression ML model itself (the *ML module* built as described above); and (3) the *suppression module*, which leverages a series of application programming interface calls to the EHR to communicate the alerts that should to be suppressed. The data extraction module queries the EHR and extracts features that should then be passed to the ML module.

The steps, related tasks, and timeline for the development and operationalization of the system are detailed in Table 1.

Table 1. Signal-to-noise optimization (SmartCDS) system development.

Step	Task	Timeline
Alert suppression model construction	• Retrospective data query: Manually retrieved historical data (best practice advisory alert log, shingles vaccine order log, and patient/provider demographics) from our EHR[a] databases • Data cleaning and initial analysis: Aggregated data and conducted analyses to determine the average response rate for alerts • Construction of predictive variables: Used long-term and short-term (1 month) personal interactive history of providers and clinician/patient demographics • Model training and performance evaluation: Used constructed variables to build model and evaluated performance using a presplit (based on clinicians) training set (80%), validation set (10%), and test set (10%) from historical data • Predictive variables and refinement of retrained models: Iterated predictive variable construction, model training, and performance evaluation on the validation set to determine the optimal predictive variables to train the model • Optimization of model parameters: Optimized model performance by fine-tuning built-in model parameters	3 months
Aggregation of production data	• Virtual table creation: Retrieved live data from our EHR databases (alert log, shingles vaccine order log, and patient/provider demographics) • Storage of interim, preprocessed data on local database: Created repository to track results of each run (eg, errors)	2 weeks
Web service endpoint configuration	• Web service isolation: Determined which Web service to call within our EHR's interoperability Web Application Programing Interface • Endpoint rule creation (part 1): Built rule that reacts to data sent to endpoint • Suppression rule creation (part 2): Built rule that can determine whether to suppress the alert or not	3 weeks
Machine learning script optimization	• Data query: Incorporated live data into model • Feature engineering and storage to local database: Updated daily additive dataset as model is retrained with most recent log • Model training: Trained model daily to incorporate live data • Threshold setting: Used training and validation datasets to simulate the predicted relationship of alerts suppressed and orders missed • Storage of model prediction results in local database: Recorded predictions, model scores, dates, and the corresponding score threshold for each upcoming vaccine alert	1 month
Docker image and container formation	• Configuration of Web service setup: Installed required modules, packages, and drivers as well as tested Web service endpoint • Cron job setup: Defined frequency and timing of specific system functions • Logging: Monitored and recorded system function, including errors	1 week
Reporting and dashboard development	• Production of relevant data elements and storage in local database: Daily report of summary statistics as tables and plots recorded • Report delivery and cadence: Daily logging on status of the pipeline; email sent upon fatal errors	2 weeks

[a]EHR: electronic health record.

Aggregation of Production Data and Configuration of a Web Service Endpoint

With the alert suppression model built and data aggregated to support production, we worked with our institutional EHR team to create predefined and operationally approved queries to build easy-to-access views of our variables of interest in the EHR database. This enabled and automated the data extraction needed to operationalize the SmartCDS system. We then established a local database to serve as a repository for monitoring and tracking data runs, reports, and system errors per best practices. To complete our work on creating the technical capacity necessary to implement the SmartCDS system, we worked with the enterprise information technology team to determine the appropriate Web service to *call*; we then created the rules necessary to appropriately respond to the data sent to that endpoint and, if appropriate, suppress the target alert (in this case, the shingles vaccination alert).

Optimization of Machine Learning Script and Docker Image and Formation of Container

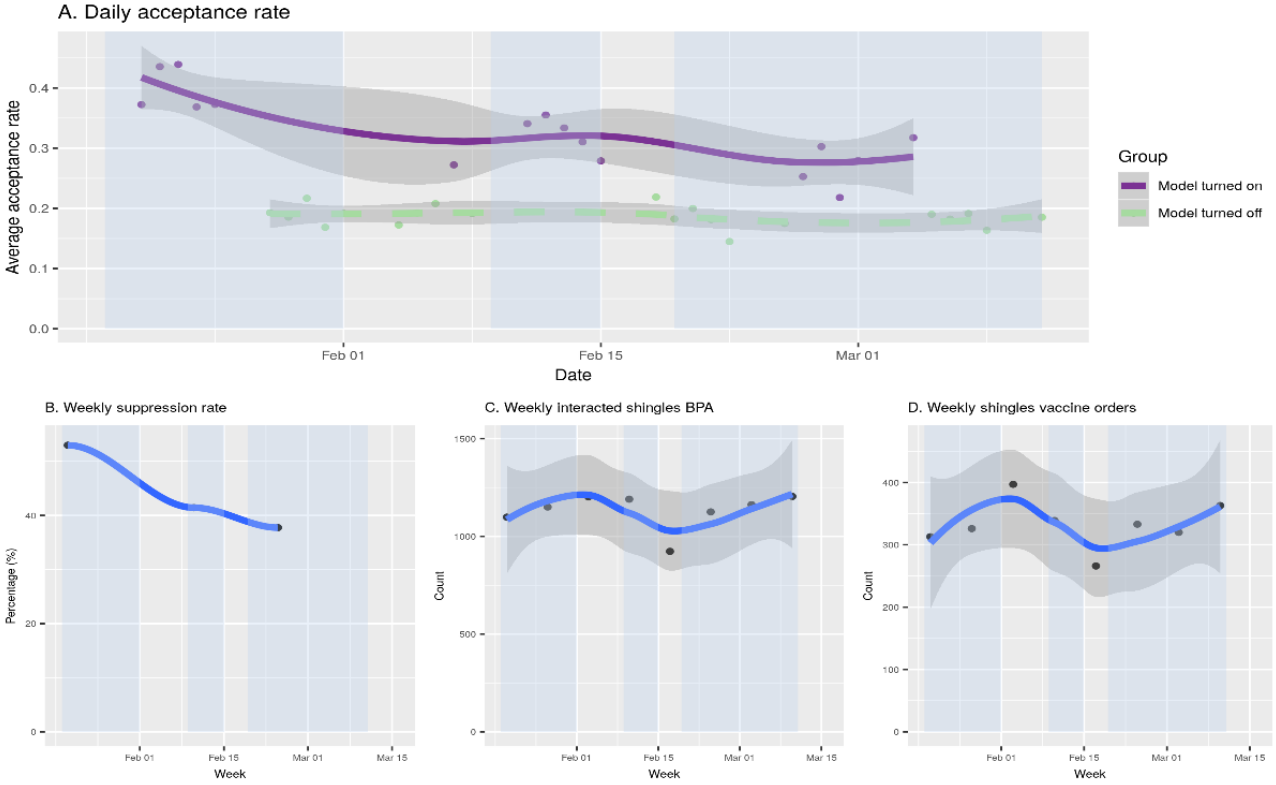
Once built, we validated the SmartCDS system with the shingles alert. Predictions are made on upcoming appointments by applying the model to our predefined views, modifying the ML script, and generating model predictions (suppress yes or no), which are saved in a local database and applied to suppress an alert with a predicted score less than the threshold (additional details in Methods). The system was designed to be modular and orthogonal with regard to call frequency, instrumentation, and configuration, allowing for easy adaptation to new environments. Under these parameters, we formed Docker containers (standard units of software that package code and all its dependencies, so each application runs quickly and reliably

from one computing environment to another) to appropriately configure the Web service, define the frequency and timing of functionality, and log the system's normal and error events.

Reporting Dashboard Development

A dashboard was developed to ensure that the system was running properly with a stable performance of the model from a safety and operational perspective and to monitor process outcomes of interest. The dashboard features process outcomes of interest (eg, suppression percentage, daily order counts, and daily alert volume) and factors in timing and frequency of report delivery based on feedback from clinical and operational stakeholders. Daily logging and weekly monitoring reports (Figure 4 and Multimedia Appendix 1) were constructed to enable detection of abnormal model behavior related to data shifts or model failures. If, on any date, the alert-related volume diverges from previous patterns, the alert log stream along with the related EHR data can be examined to locate the source of the anomaly.

Figure 4. Summary reports of the shingles model from January 19, 2019, to March 11, 2019. (A) Smoothed curves of daily aggregated acceptance (response) rates to the shingles best practice advisory (BPA; alert), 95% CI shown in shaded areas, respectively (model turned on is shown in the purple solid curve and model turned off is shown in the light green dotted curve). Weekends are not included because of the large variation resulting from low BPA volumes on weekends. Vertical shaded areas annotate the 6-week trial time period. (B) Weekly averaged suppression percentage of the shingles alert. (C) Weekly count of interacted shingles BPA. (D) Smoothed curve of weekly shingles vaccine order counts.



Pilot Results

Daily Alert–Related Volumes

We leveraged the 6 weeks of data (January 19-March 11, 2019) to compare the volume difference in shingles alert count, interacted alert count, and order count between weeks with the model turned on and turned off. We observed 42.2% (3541.0 with the model turned on vs 6123.7 with the model turned off) reduction in the alert count, no significant reduction in the interacted alert count (one-sided two sample *t* test; *P*=.20) or in the order count (one-sided two sample *t* test; *P*=.38) during the 6-week biweekly cycle with the model turned on and model turned off (Table 2).

Table 2. Shingles alert–related volume in 2019. Each statistic is the weekly average of the corresponding group during the 6-week biweekly cycle except that alerts per order rates were calculated as bulk averages within each 3-week group, respectively.

Group	Time range (weeks)	Alert count, mean (SD)	Interacted alert count[a], mean (SD)	Order count[b], mean (SD)	Accumulated alerts per order rate	Accumulated interacted alerts per order rate
Model turned off	3	6123.7 (232.9)	1162.3 (22.8)	331.3 (5.1)	18.5	3.4
Model turned on	3	3541.0 (669.4)	1118.3 (71.6)	326.3 (23.6)	10.9	3.5

[a]P=.20.

[b]P=.38.

Alerts per Order Rate and Signal-to-Noise Ratio

Our 6-week pilot deployment of the system in the live environment indicates an alert suppression rate of 43.7% out of 35,315 appointments (Figure 4), with stable shingles vaccine order volume (no statistically significant difference between active and inactive suppression) slightly lower than the predefined 50% threshold. Initial inspection showed that, on an average, walk-in visits had a higher alert ignored rate (91%) compared with scheduled office visits (87%) in 2017 and 2018. As the model only operated on scheduled appointments in this study and the activity history has the highest weight toward a suppression decision, a slightly lower suppression rate than 50% was expected.

The ratio of alerts fired to orders placed with the model turned on was almost half of that of the ratio with the model turned off, whereas the ratio of the interacted alerts per order placed remained the same. By mapping the average orders placed as the average power of *signal* and the average count of ignored alert with no follow-up orders as the average power of *noise*, the signal-to-noise ratio changed from 5.7% to 10.1%, a 78% increase. Furthermore, by mapping the interacted alerts (including follow-up orders) as *signal* and the ignored alerts with no follow-up actions as *noise*, the signal-to-noise ratio changed from 23.4% to 46.1%, a 97% increase.

Discussion

Principal Findings

This paper describes the steps and considerations involved in the development and implementation of an ML model for suppressing low-value alerts in the EHR for the shingles vaccination. As predicted in our simulation, validation of this signal-to-noise optimization (SmartCDS) system demonstrated substantial reduction in the shingles vaccine alerts at a limited vaccine ordering expense. The rate of daily alert interaction among individual clinicians during the 6-week pilot was higher with the model turned on vs the model turned off. This result was expected because of the 42.2% lower volume of shingles alerts observed with stable daily alert interactions. Interestingly, the overall interaction rate gradually decreased over the 6-week cycle (Figure 4). This finding is consistent with the findings that responsiveness to alerts tends to decrease over time [29]. During the 6-week pilot, the profile of the providers who accepted the alerts did not change, indicating that the profile of patients who are offered the vaccination did not change either. This will be confirmed in our follow-up studies. To date, our literature review indicates that our SmartCDS system is the first to develop an ML-based system to suppress clinically insignificant alerts or alerts unlikely to be accepted and to prospectively evaluate the system in a large-scale health care system. Relevant literature to date has been limited to retrospective studies focused on identifying false-positive or clinically insignificant physiologic monitor alarms (false alarms). In 2015, Physionet opened a challenge to reduce false arrhythmia alarms using a subset of the Medical Information Mart for Intensive Care II waveform database [51]. The best models showed that by allowing 30 seconds of delay, false alarms can be better distinguished from true alarms; the best models were able to achieve 80% reduction in false alarms, missing 1% of true alarms. Studies focusing on pulse oximetry to reduce peripheral capillary oxygen saturation (SpO_2) false alarms, intracranial pressure alarms, and general vital sign monitoring alarms found mixed results ranging from 25% to 47% in alarm reduction, with 0% to 5% false-negative rates [42-45]. A more recent study showed that, by increasing delayed time within 3 min for alarms with physiologic monitoring waveforms, as well as including electrocardiography, SpO_2, and arterial blood pressure, an ML model can achieve slightly better performance but fails to stably generalize to unseen data [52].

The development of a robust reporting structure allows for the logging and monitoring of the system and its impact on clinical outcomes, which are necessary to ensure the stability and safety of the system. Future work will involve gathering feedback from front-line stakeholders to support the adaptation of the signal-to-noise optimization system to other alerts, enabling the system to ingest real-time data as well as further development of a reporting dashboard with effective, user-centered data displays, and a systematic process for establishing organizationally acceptable thresholds for alert suppression.

Limitations

During the pilot implementation and evaluation, the model only operated on scheduled office visits because of infrastructure gaps restricting the ability to incorporate walk-in visits. We are working to address this gap to be able to assess the effectiveness and impact of this model on a global level. On the other hand, it is possible that clinicians will start to adjust to the volume change in the shingles alert delivery, leading to less responsiveness and less ordering. As potential external or systematic biases, such as seasonal effects, could lead to

inaccurate observations and conclusions, we will implement a more comprehensive statistical evaluation after updating the infrastructure to systematically address these potential biases.

Conclusions

Our model presented high discriminatory power in the initial prospective evaluation of shingles alert interactions. Our approach was effective in suppressing unnecessary alerts, with limited reduction in overall order volume. This work also provides potential evidence of increase in interactions and orders (eg, an increase in signal-to-noise ratio) by decreasing noise (eg, suppression). In addition, the process built to operationalize this new ML tool may prove to be a useful model for enabling the deployment of this type of tool across many use cases. Future efforts include applying this approach globally to other EHR alerts and comprehensive randomized controlled trials.

Acknowledgments

The authors thank Dr Simon Jones for providing insights and helping validate the statistical evaluation approaches.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Daily count of the shingles best practice advisories (BPAs; alerts), interacted shingles BPAs, and shingles BPA follow-up vaccine orders from January 19, 2019, to March 11, 2019. Each bar in all three plots represents one day - purple (dark) if the model was turned on and light green with a black border if the model was turned off, the shaded areas annotate excluded dates from downstream statistical analysis. BPAL best practice advisory.

[PNG File , 74 KB - jmir_v22i4e16848_app1.PNG]

References

1. McDonald CJ, Hui SL, Tierney WM. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. MD Comput 1992;9(5):304-312. [Medline: 1522792]
2. Dexter PR, Perkins S, Overhage JM, Maharry K, Kohler RB, McDonald CJ. A computerized reminder system to increase the use of preventive care for hospitalized patients. N Engl J Med 2001 Sep 27;345(13):965-970. [doi: 10.1056/NEJMsa010181] [Medline: 11575289]
3. Kim DK, Riley LE, Hunter P, Advisory Committee on Immunization Practices. Recommended immunization schedule for adults aged 19 years or older, United States, 2018. Ann Intern Med 2018 Feb 6;168(3):210-220. [doi: 10.7326/M17-3439] [Medline: 29404596]
4. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007;14(2):141-145 [FREE Full text] [doi: 10.1197/jamia.M2334] [Medline: 17213487]
5. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. J Am Med Assoc 1998 Oct 21;280(15):1339-1346. [doi: 10.1001/jama.280.15.1339] [Medline: 9794315]
6. Overhage JM, Tierney WM, Zhou X, McDonald CJ. A randomized trial of 'corollary orders' to prevent errors of omission. J Am Med Inform Assoc 1997;4(5):364-375 [FREE Full text] [doi: 10.1136/jamia.1997.0040364] [Medline: 9292842]
7. Dexheimer JW, Talbot TR, Sanders DL, Rosenbloom ST, Aronsky D. Prompting clinicians about preventive care measures: a systematic review of randomized controlled trials. J Am Med Inform Assoc 2008;15(3):311-320 [FREE Full text] [doi: 10.1197/jamia.M2555] [Medline: 18308989]
8. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]
9. Bernstein SL, Whitaker D, Winograd J, Brennan JA. An electronic chart prompt to decrease proprietary antibiotic prescription to self-pay patients. Acad Emerg Med 2005 Mar;12(3):225-231 [FREE Full text] [doi: 10.1197/j.aem.2004.09.021] [Medline: 15741585]
10. Garthwaite EA, Will EJ, Bartlett C, Richardson D, Newstead CG. Patient-specific prompts in the cholesterol management of renal transplant outpatients: results and analysis of underperformance. Transplantation 2004 Oct 15;78(7):1042-1047. [doi: 10.1097/01.tp.0000137340.22880.c8] [Medline: 15480172]
11. Safran C, Rind DM, Davis RM, Currier J, Ives D, Sands DZ, et al. An electronic medical record that helps care for patients with HIV infection. Proc Annu Symp Comput Appl Med Care 1993:224-228 [FREE Full text] [Medline: 8130466]
12. Safran C, Rind D, Davis R, Ives D, Sands D, Currier J, et al. Guidelines for management of HIV infection with computer-based patient's record. Lancet 1995 Aug 5;346(8971):341-346. [doi: 10.1016/s0140-6736(95)92226-1] [Medline: 7623532]
13. Safran C, Rind D, Sands D, Davis R, Wald J, Slack W. Development of a knowledge-based electronic patient record. MD Comput 1996;13(1):46-54, 63. [Medline: 8569464]

14. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. N Engl J Med 1990 May 24;322(21):1499-1504. [doi: 10.1056/NEJM199005243222105] [Medline: 2186274]
15. Simon S, Smith D, Feldstein A, Perrin N, Yang X, Zhou Y, et al. Computerized prescribing alerts and group academic detailing to reduce the use of potentially inappropriate medications in older people. J Am Geriatr Soc 2006 Jun;54(6):963-968. [doi: 10.1111/j.1532-5415.2006.00734.x] [Medline: 16776793]
16. Shah N, Seger A, Seger D, Fiskio JM, Kuperman GJ, Blumenfeld B, et al. Improving acceptance of computerized prescribing alerts in ambulatory care. J Am Med Inform Assoc 2006;13(1):5-11 [FREE Full text] [doi: 10.1197/jamia.M1868] [Medline: 16221941]
17. Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, Hanley J, et al. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. Can Med Assoc J 2003 Sep 16;169(6):549-556 [FREE Full text] [Medline: 12975221]
18. Gaikwad R, Sketris I, Shepherd M, Duffy J. Evaluation of accuracy of drug interaction alerts triggered by two electronic medical record systems in primary healthcare. Health Informatics J 2007 Sep;13(3):163-177. [doi: 10.1177/1460458207079836] [Medline: 17711879]
19. Smith DH, Perrin N, Feldstein A, Yang X, Kuang D, Simon SR, et al. The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. Arch Intern Med 2006 May 22;166(10):1098-1104. [doi: 10.1001/archinte.166.10.1098] [Medline: 16717172]
20. Seidling HM, Schmitt SP, Bruckner T, Kaltschmidt J, Pruszydlo MG, Senger C, et al. Patient-specific electronic decision support reduces prescription of excessive doses. Qual Saf Health Care 2010 Oct;19(5):e15. [doi: 10.1136/qshc.2009.033175] [Medline: 20427312]
21. Hussain MI, Reynolds TL, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. J Am Med Inform Assoc 2019 Oct 1;26(10):1141-1149. [doi: 10.1093/jamia/ocz095] [Medline: 31206159]
22. Isaac T, Weissman JS, Davis RB, Massagli M, Cyrulik A, Sands DZ, et al. Overrides of medication alerts in ambulatory care. Arch Intern Med 2009 Feb 9;169(3):305-311. [doi: 10.1001/archinternmed.2008.551] [Medline: 19204222]
23. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
24. Carspecken CW, Sharek PJ, Longhurst C, Pageler NM. A clinical case of electronic health record drug alert fatigue: consequences for patient outcome. Pediatrics 2013 Jun;131(6):e1970-e1973. [doi: 10.1542/peds.2012-3252] [Medline: 23713099]
25. Hu X, Sapo M, Nenov V, Barry T, Kim S, Do DH, et al. Predictive combinations of monitor alarms preceding in-hospital code blue events. J Biomed Inform 2012 Oct;45(5):913-921 [FREE Full text] [doi: 10.1016/j.jbi.2012.03.001] [Medline: 22465785]
26. Voepel-Lewis T, Parker ML, Burke CN, Hemberg J, Perlin L, Kai S, et al. Pulse oximetry desaturation alarms on a general postoperative adult unit: a prospective observational study of nurse response time. Int J Nurs Stud 2013 Oct;50(10):1351-1358. [doi: 10.1016/j.ijnurstu.2013.02.006] [Medline: 23474081]
27. Bonafide CP, Lin R, Zander M, Graham CS, Paine CW, Rock W, et al. Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. J Hosp Med 2015 Jun;10(6):345-351 [FREE Full text] [doi: 10.1002/jhm.2331] [Medline: 25873486]
28. Korniewicz DM, Clark T, David Y. A national online survey on the effectiveness of clinical alarms. Am J Crit Care 2008 Jan;17(1):36-41. [Medline: 18158387]
29. Embi PJ, Leonard AC. Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study. J Am Med Inform Assoc 2012 Jun;19(e1):e145-e148 [FREE Full text] [doi: 10.1136/amiajnl-2011-000743] [Medline: 22534081]
30. Genco EK, Forster JE, Flaten H, Goss F, Heard KJ, Hoppe J, et al. Clinically inconsequential alerts: the characteristics of opioid drug alerts and their utility in preventing adverse drug events in the emergency department. Ann Emerg Med 2016 Feb;67(2):240-8.e3 [FREE Full text] [doi: 10.1016/j.annemergmed.2015.09.020] [Medline: 26553282]
31. Joint Commission. Medical device alarm safety in hospitals. Sentinel Event Alert 2013 Apr 8(50):1-3. [Medline: 23767076]
32. McCoy A, Thomas E, Krousel-Wood M, Sittig D. Clinical decision support alert appropriateness: a review and proposal for improvement. Ochsner J 2014;14(2):195-202 [FREE Full text] [Medline: 24940129]
33. Ong M, Coiera E. Evaluating the effectiveness of clinical alerts: a signal detection approach. AMIA Annu Symp Proc 2011;2011:1036-1044 [FREE Full text] [Medline: 22195164]
34. Simpao AF, Ahumada LM, Desai BR, Bonafide CP, Gálvez JA, Rehman MA, et al. Optimization of drug-drug interaction alert rules in a pediatric hospital's electronic health record system using a visual analytics dashboard. J Am Med Inform Assoc 2015 Mar;22(2):361-369. [doi: 10.1136/amiajnl-2013-002538] [Medline: 25318641]
35. Cornu P, Steurbaut S, Gentens K, van de Velde R, Dupont AG. Pilot evaluation of an optimized context-specific drug-drug interaction alerting system: A controlled pre-post study. Int J Med Inform 2015 Sep;84(9):617-629. [doi: 10.1016/j.ijmedinf.2015.05.005] [Medline: 26051616]

36. Paterno MD, Maviglia SM, Gorman PN, Seger DL, Yoshida E, Seger AC, et al. Tiering drug-drug interaction alerts by severity increases compliance rates. J Am Med Inform Assoc 2009;16(1):40-46 [FREE Full text] [doi: 10.1197/jamia.M2808] [Medline: 18952941]
37. Beccaro MA, Villanueva R, Knudson KM, Harvey EM, Langle JM, Paul W. Decision support alerts for medication ordering in a computerized provider order entry (CPOE) system: a systematic approach to decrease alerts. Appl Clin Inform 2010;1(3):346-362 [FREE Full text] [doi: 10.4338/ACI-2009-11-RA-0014] [Medline: 23616845]
38. Guzek M, Zorina OI, Semmler A, Gonzenbach RR, Huber M, Kullak-Ublick GA, et al. Evaluation of drug interactions and dosing in 484 neurological inpatients using clinical decision support software and an extended operational interaction classification system (Zurich Interaction System). Pharmacoepidemiol Drug Saf 2011 Sep;20(9):930-938. [doi: 10.1002/pds.2197] [Medline: 21774031]
39. Czock D, Konias M, Seidling HM, Kaltschmidt J, Schwenger V, Zeier M, et al. Tailoring of alerts substantially reduces the alert burden in computerized clinical decision support for drugs that should be avoided in patients with renal disease. J Am Med Inform Assoc 2015 Jul;22(4):881-887. [doi: 10.1093/jamia/ocv027] [Medline: 25911673]
40. Eppenga WL, Derijks HJ, Conemans JM, Hermens WA, Wensing M, de Smet PA. Comparison of a basic and an advanced pharmacotherapy-related clinical decision support system in a hospital care setting in the Netherlands. J Am Med Inform Assoc 2012;19(1):66-71 [FREE Full text] [doi: 10.1136/amiajnl-2011-000360] [Medline: 21890873]
41. Harinstein LM, Kane-Gill SL, Smithburger PL, Culley CM, Reddy VK, Seybert AL. Use of an abnormal laboratory value-drug combination alert to detect drug-induced thrombocytopenia in critically Ill patients. J Crit Care 2012 Jun;27(3):242-249. [doi: 10.1016/j.jcrc.2012.02.014] [Medline: 22520497]
42. Connor C, Gohil B, Harrison M. Triggering of systolic arterial pressure alarms using statistics-based versus threshold alarms. Anaesthesia 2009 Feb;64(2):131-135 [FREE Full text] [doi: 10.1111/j.1365-2044.2008.05738.x] [Medline: 19143688]
43. Borowski M, Siebig S, Wrede C, Imhoff M. Reducing false alarms of intensive care online-monitoring systems: an evaluation of two signal extraction algorithms. Comput Math Methods Med 2011;2011:143480 [FREE Full text] [doi: 10.1155/2011/143480] [Medline: 21461385]
44. Scalzo F, Hu X. Semi-supervised detection of intracranial pressure alarms using waveform dynamics. Physiol Meas 2013 Apr;34(4):465-478. [doi: 10.1088/0967-3334/34/4/465] [Medline: 23524637]
45. Schmid F, Goepfert MS, Franz F, Laule D, Reiter B, Goetz AE, et al. Reduction of clinically irrelevant alarms in patient monitoring by adaptive time delays. J Clin Monit Comput 2017 Feb;31(1):213-219. [doi: 10.1007/s10877-015-9808-2] [Medline: 26621389]
46. van de Velde S, Heselmans A, Delvaux N, Brandt L, Marco-Ruiz L, Spitaels D, et al. A systematic review of trials evaluating success factors of interventions with computerised clinical decision support. Implement Sci 2018 Aug 20;13(1):114 [FREE Full text] [doi: 10.1186/s13012-018-0790-1] [Medline: 30126421]
47. Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital readmissions. J Biomed Inform 2015 Aug;56:229-238 [FREE Full text] [doi: 10.1016/j.jbi.2015.05.016] [Medline: 26044081]
48. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2018 Jan 26;8(1):e017833 [FREE Full text] [doi: 10.1136/bmjopen-2017-017833] [Medline: 29374661]
49. Osheroff JA. Improving Medication Use and Outcomes with Clinical Decision Support: A Step by Step Guide. Chicago, IL: HIMSS Publishing; 2009.
50. Chen J, Aphinyanaphongs Y, Mann D, Iturrate E, Chokshi S, Hegde R. Personalized Clinical Decision Support (CDS) using Machine Learning: A Case Study with Shingles. In: Proceedings of the AMIA 2019 Clinical Informatics Conference. 2019 Presented at: AMIA'19; May 1-2, 2019; Atlanta, GA URL: https://www.amia.org/cic2019/presentations
51. Clifford G, Silva I, Moody B, Li Q, Kella D, Shahin A, et al. The PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU. Comput Cardiol (2010) 2015 Sep;2015:273-276 [FREE Full text] [doi: 10.1109/CIC.2015.7408639] [Medline: 27331073]
52. Chen L, Dubrawski A, Wang D, Fiterau M, Guillame-Bert M, Bose E, et al. Using supervised machine learning to classify real alerts and artifact in online multisignal vital sign monitoring data. Crit Care Med 2016 Jul;44(7):e456-e463 [FREE Full text] [doi: 10.1097/CCM.0000000000001660] [Medline: 26992068]

Abbreviations

CDS: clinical decision support
EHR: electronic health record
ML: machine learning
SpO_2: peripheral capillary oxygen saturation

Edited by G Eysenbach; submitted 30.10.19; peer-reviewed by A Al, P Marier-Deschenes; comments to author 25.12.19; revised version received 19.02.20; accepted 21.02.20; published 29.04.20.

Please cite as:
Chen J, Chokshi S, Hegde R, Gonzalez J, Iturrate E, Aphinyanaphongs Y, Mann D
Development, Implementation, and Evaluation of a Personalized Machine Learning Algorithm for Clinical Decision Support: Case Study With Shingles Vaccination
J Med Internet Res 2020;22(4):e16848
URL: http://www.jmir.org/2020/4/e16848/
doi:10.2196/16848
PMID:32347813

Original Paper

Feasibility of Social Media–Based Recruitment and Perceived Acceptability of Digital Health Interventions for Caregivers of Justice-Involved Youth: Mixed Methods Study

Johanna Bailey Folk[1], PhD; Anna Harrison[1,2], PhD; Christopher Rodriguez[1], MS; Amanda Wallace[1,3], MD; Marina Tolou-Shams[1], PhD

[1]Department of Psychiatry, University of California, San Francisco, CA, United States

[2]San Francisco Veterans Affairs Medical Center, San Francisco, CA, United States

[3]New York-Presbyterian/Columbia University Irving Medical Center, New York, NY, United States

Corresponding Author:
Johanna Bailey Folk, PhD
Department of Psychiatry
University of California
1001 Potrero Avenue
San Francisco, CA, 94110
United States
Phone: 1 6282062309
Email: Johanna.folk@ucsf.edu

Abstract

Background: Caregiver involvement is critical for supporting positive behavioral health and legal outcomes for justice-involved youth; however, recruiting this population into clinical research studies and engaging them in treatment remain challenging. Technology-based approaches are a promising, yet understudied avenue for recruiting and intervening with caregivers of justice-involved youth.

Objective: This mixed methods study aimed to assess the feasibility of recruiting caregivers of justice-involved youth using social media into clinical research and to understand caregivers' perceptions of the acceptability of digital health interventions.

Methods: Caregivers of justice-involved youth were recruited through paid Facebook advertisements to participate in a Web-based survey. Advertisement design was determined using Facebook A/B split testing, and the advertisement with the lowest cost per link click was used for the primary advertisement campaign. Survey participants were offered the option to participate in a follow-up qualitative phone interview focused on the perceived feasibility and acceptability of digital health interventions.

Results: Facebook advertisements were successful in quickly recruiting a diverse set of caregivers (80/153, 52.3% female; mean age 43 years, SD 7; 76/168, 45.2% black, 34/168, 20.2% white, and 28/168, 16.7% Latinx; and 97/156, 62.2% biological parents); cost per click was US $0.53, and conversion rate was 11.5%. Survey participants used multiple social media platforms; 60.1% (101/168) of the participants indicated they would participate in a digital health intervention for caregivers of justice-involved youth. Survey respondents' most preferred intervention was supportive and motivational parenting messages via SMS text message. Of the survey respondents, 18 completed a phone interview (12/18, 67% female; mean age 45 years, SD 10; 10/18, 56% black, 7/18, 39% white, and 1/18, 6% Latinx; and 16/18, 89% biological parents). Interview participant responses suggested digital health interventions are acceptable, but they expressed both likes (eg, alleviates barriers to treatment access) and concerns (eg, privacy); their most preferred intervention was video-based family therapy.

Conclusions: Recruiting and intervening with caregivers of justice-involved youth through social media and other digital health approaches may be a feasible and acceptable approach to overcoming barriers to accessing traditional in-person behavioral health care.

(J Med Internet Res 2020;22(4):e16370) doi:10.2196/16370

KEYWORDS

caregivers; telemedicine; mobile health; juvenile delinquency; social media

Introduction

Background

Each year in the United States, more than 920,000 youth under the age of 18 years are arrested [1], and juvenile courts process nearly 1 million delinquency cases [2]. Involvement in the juvenile justice system is associated with various adverse outcomes, including substance use [3], psychiatric disorders [4], and sexual risk behavior [5,6]. Family factors (eg, parent-child conflict and coercive parenting) are consistently associated with youth delinquency [7-10]. As family-based interventions effectively improve outcomes among justice-involved adolescents [11-13], involving family members, particularly primary caregivers, in therapeutic interventions may be an effective way to support positive youth outcomes [14].

Despite the importance of caregiver involvement, recruiting broad samples of caregivers into clinical research remains challenging. Caregivers are typically *not* court ordered to treatment along with their youth. Resources required to attend treatment (eg, time and reliable transportation) may be scarce. Families often feel overwhelmed by the number of required services related to their youth's court involvement (eg, court appointments, drug screens, and mandated treatment) [15]. Therefore, caregiver treatment attendance is often poor, and study attrition for clinical trials of family-based interventions is high [16]. The most effective behavioral health treatments include caregivers [17,18], but engaging face-to-face with researchers and clinicians may not be feasible.

New strategies are needed to engage caregivers of justice-involved youth into clinical research and treatment. Technology offers multiple promising avenues. In particular, digital mobile health (mHealth) technology is an efficacious, low-cost way of reaching underserved, vulnerable populations to engage them into and/or deliver quality care [19]. mHealth circumvents many barriers to treatment participation reported by caregivers of justice-involved youth [20], allowing for instantaneous, portable access and direct communication with providers [21]. mHealth effectively increases patient communication, monitoring, and education to reduce the burden of diseases associated with poverty and improves access to health services, clinical diagnosis, and treatment adherence [22]. Social media, specifically Facebook, is one mHealth approach that has been successful in recruiting and intervening with hard-to-reach populations [23-25], including sexual minorities [26] and young adult veterans [27].

Caregivers of justice-involved youth are interested in mHealth treatment [28,29], although little is known about what specific approaches would be most acceptable and beneficial for them. One pilot study with 5 caregivers of justice-involved youth found preliminary evidence for the feasibility and acceptability of a text messaging intervention that sent appointment reminders and motivational messages [29]. To our knowledge, no prior study has used social media to engage or intervene with caregivers of justice-involved youth. Thus, little is known about (1) the feasibility of engaging caregivers of justice-involved youth in clinical research using social media and (2) whether caregivers would consider participating in a range of digital health interventions. Engaging caregivers in clinical research and treatment through mHealth technology may expand access to family-based treatments and thereby promote health equity.

This Study

This mixed methods study aimed to assess the feasibility of using social media to recruit caregivers of justice-involved youth into research studies and to understand caregivers' perceptions of the acceptability of digital health interventions. Caregivers were recruited through Facebook to complete a Web-based survey and an optional follow-up phone interview. We hypothesized Facebook would be a feasible recruitment tool, and caregivers would be open to participating in a range of digital health interventions. The overall goal was to lay the groundwork for future studies on delivering caregiver-focused and family-focused interventions through digital health platforms such as social media.

Methods

Participants and Procedure

Advertising Strategy

Facebook advertisements were optimized for link clicks and presented to adults (1) ages 28 years and older, (2) using Facebook in English, and (3) residing in 10 US metropolitan areas. The maximum daily budget was US $80 per day. See Multimedia Appendix 1 for additional details regarding pilot testing of advertisements and advertising strategy.

Quantitative Survey

Participants were a convenience sample of 168 caregivers of justice-involved youth. Eligible participants were (1) the caregiver of a youth (ages 10–17 years) who had been arrested, detained, or court involved during the past 12 months; (2) had access to a computer or mobile device with the internet; and (3) proficient in English.

Participants were recruited through Facebook Ad Manager. Advertisements were direct promotions for the survey website; individuals who clicked the advertisement were directed to an external website (Research Electronic Data Capture) containing a Web-based screening questionnaire. Eligible participants were directed to a study information page containing the consent form. A Facebook page where participants could review this material was maintained and reachable from the advertisements. Those who provided informed consent were directed to the Web-based survey. The open survey was presented across 10 screens with nonrandomized items and a varying number of items per page (influenced by skip logic); participants were free to skip any items and could not review or change answers from prior pages. To receive the US $15 electronic Amazon gift card incentive, participants entered an email address which, after checking to prevent multiple entries from the same individual, was disconnected from survey responses. We set a predetermined sample size goal of 150 participants based on available resources. All study procedures were approved by the University of California, San Francisco, Institutional Review Board (#18-25987).

Qualitative Interview

At the end of the survey, participants were asked if they would be willing to participate in a follow-up phone interview. Those who agreed (54/168, 32.1%) provided contact information, which was immediately disconnected from survey responses; 18 qualitative interviews were completed, at which point saturation (ie, when no new themes arose) had been reached [30]. Most participants expressing interest in the phone interview identified as female, so quota sampling (ie, aims to create a sample that represents certain characteristics of a population [31]) was used to ensure adequate male representation. Interviews were approximately 45-min long and participants received a US $50 electronic gift card.

Measures

Quantitative Survey

Sample Characteristics

Caregivers identified their own and their justice-involved youth's gender, race, ethnicity, and marital status.

Youth Justice Involvement, Behavioral Health Needs, and Treatment

Caregivers reported their youth's past year justice involvement (eg, probation and detention), behavioral health needs (eg, mental health diagnosis), and participation in behavioral health treatment.

Technology and Social Media Usage

Caregivers reported their access to technology (eg, smartphone and computer) and social media (eg, Facebook and Twitter) use.

Acceptability of Digital Health Interventions

Caregivers reported their willingness to participate in digital health interventions for caregivers of justice-involved youth (1=*I would definitely not participate* to 5=*I would definitely participate*). Participants were asked whether they would participate in any of the following interventions and to rate which three were most interesting: (1) receiving supportive/motivational parenting messages via SMS text message; (2) supportive/motivational parenting posts on social media platforms (eg, Instagram and Snapchat); (3) online support community; (4) private groups on Facebook connecting caregivers with one another; (5) private groups on Facebook connecting caregivers and where they can chat with/contact a mental health professional; (6) individual therapy sessions with a therapist through video chat (eg, FaceTime and Skype); (7) family therapy sessions, including the caregiver, their child, and other family members, through video chat; (8) support group meetings with other caregivers through video chat; and (9) other (please specify).

Qualitative Interview

In-depth semistructured phone interviews were conducted in a private research laboratory by the first (JF) and third (CR) authors, who were trained and experienced in qualitative interviewing. Descriptive phenomenological interviews [32] asked participants about the (1) impact of their youth's justice involvement on different aspects of their lives (eg, work and relationships), (2) family engagement in behavioral health services as recommended or mandated by the court, (3) technology and social media use, and (4) opinions of three hypothetical digital health interventions (primary outcome). The semistructured interview format was selected to allow respondents to express complex thoughts without restriction and to allow interviewers to probe when clarification and/or depth was necessary [33]. All interviews were audio recorded using a digital tape recorder and transcribed nonverbatim, excluding nonverbal cues for a more comprehensible transcript. To ensure the accuracy of the transcript, one researcher (CR) reviewed each transcript while listening to the audio recording.

Sample Characteristics

Caregivers reported their gender, age, race, ethnicity, marital status, and relationship to the justice-involved youth. They also reported their youth's age and current justice involvement.

Technology and Social Media Usage

Caregivers reported their access to technology, most frequently used social media platform, reasons for using social media (eg, connect with family), and online support group participation.

Digital Health Intervention Vignettes

Participants were presented with three vignettes of hypothetical digital health interventions during the qualitative interview. They included (1) private groups on Facebook connecting caregivers of youth involved in the justice system with one another with the option to chat with/contact a mental health professional; (2) family therapy sessions, including the caregiver, their child, and other family members, through video chat; and (3) receiving appointment reminders and supportive/motivational parenting messages via SMS text message. Participants were asked to share their likes and dislikes about each intervention and any potential barriers to or benefits from participating.

Coding and Analysis

Inductive Thematic Analysis was selected as the analytic approach to allow themes to emerge from raw data and diminish researcher bias [34]. The initial coding scheme was informed by the semistructured interview guide and interview transcripts, ensuring the authenticity of the participant's perspective. Before data analysis, three researchers (JF, CR, and AW) employed a co-coding pilot with three randomly selected interview transcripts. Codes from the initial coding scheme were independently applied to interview transcripts and then compared to address inconsistencies (eg, different codes applied to the same section and interpreting the same code differently) and track similarities. The coding scheme was finalized after four iterations of the co-coding pilot. Revising ambiguous codes to improve intercoder reliability [35,36] preserved the integrity of the data and enhanced the rigor of the study.

On completion of the co-coding pilot, the three coders independently coded the remaining 15 interview transcripts; each transcript was coded by two of the researchers, and any disagreements were discussed to achieve a final agreement code. All interview transcripts were coded using Atlas.ti 8.0

(ATLAS.ti Scientific Software Development GmbH, Germany) [37]. Memos were written after various coding iterations to continuously develop assertions (ie, claims about the data supported by direct evidence) and propositions (ie, inferences that provide potential relationships or possible explanations of the explored theory) and again after all interviews were coded to organize common themes among and relationships between data [38]. Memoing further enhanced rigor by allowing researchers to reflect on biases that could influence the participant's perspective and make adjustments accordingly.

Results

Facebook Advertisements

Pilot testing ran from November 4, 2018, to November 8, 2018 (4 days). Advertisements were shown to 37,630 Facebook users, resulting in 461 unique link clicks. A total of US $318.08 was spent on advertisement pilot testing. Cost per click varied by advertisement (see Multimedia Appendix 1), with the most successful advertisement (woman comforting a teenager with his head down) costing US $0.47 per click and the least successful (juvenile detention center) costing US $0.91 per click.

The single most successful advertisement was used for the primary campaign. The campaign was funded from November 10, 2018, to December 3, 2018 (23 days), at which point we surpassed our target sample of 150 caregivers. Advertisements were presented to 500,208 Facebook users, and of those, 3394 clicked the advertisement link. The cost per click was US $0.53. Of the users who clicked the advertisement, 389 completed the Web-based screener (11.5% conversion rate), 235 were eligible, 185 consented to participate, and 168 completed the survey. The total primary campaign advertising cost was US $1802.72, which translated to an advertising cost of US $10.73 per participant.

Quantitative Survey

Sample Characteristics

Caregivers were 52.3% (80/153) female, on average 43 years old (SD 7), and racially/ethnically diverse (76/168, 45.2% black; 34/168, 20.2% white; 28/168, 16.7% Latinx; and 26/168, 15.5% other). Caregivers were biological parents (97/156, 62.2%), step-parents (32/156, 20.5%), or nonfamilial foster (2/156, 1.3%) parents to predominantly male (109/159, 68.6%) justice-involved youth who were on average, 15 (SD 1) years old. (*N*s differ throughout results based on missing data.)

Youth Justice Involvement, Behavioral Health Needs, and Treatment

During the past year, youth were arrested twice on average (SD 2, range 1-10); 79.2% (122/154) youths appeared in court related to their arrest, and 67.1% (104/155) of youth were found delinquent by the court. During the past year, 48.1% (74/154) of youth spent time in a juvenile detention center or court-ordered residential placement, 45.1% (69/153) of youth were on probation, 23.2% (35/151) of youth were on electronic monitoring, and 16.4% (24/146) of youth had their case transferred from juvenile to adult court.

Of 151 caregivers, 54 (35.8%) reported they had ever been told their justice-involved youth had a mental health diagnosis. The most commonly reported diagnoses were attention deficit hyperactivity (23/168, 13.7%), depressive (19/168, 11.3%), anxiety (17/168, 10.1%), and bipolar (13/168; 7.7%) disorders. Posttraumatic stress (9/168, 5.4%) and substance use (4/168, 2.4%) disorders were reported at lower rates. Behavioral health treatment utilization was high, with 67.5% (104/154) of caregivers reporting their youth had ever participated in any type of behavioral health treatment (eg, psychiatric medication, residential treatment, and crisis center). More than one-third (50/133, 37.6%) of caregivers reported their youth had ever received private professional help from a psychiatrist, psychologist, social worker, or psychiatric nurse; of the 54 caregivers who had ever been told their youth had a mental health diagnosis, 27 (50%) of these youth had ever received private professional help from a psychiatrist, psychologist, social worker, or psychiatric nurse. School-based counseling was the most commonly used service (53/127, 41.7%), followed by psychiatric medication (37/146, 25.3%), community mental health centers (25/134, 18.7%), in-home counseling (25/135, 18.5%), therapeutic foster care (23/140, 16.4%), and residential treatment centers (21/140, 15.0%). Outpatient drug or alcohol clinic services were sought by 10.9% (15/138) of youth, and 14.0% (20/143) of youth had a history of inpatient alcohol/drug treatment or detoxification unit.

Technology and Social Media Usage

Most caregivers owned a smartphone/tablet (160/167, 95.8%) and had regular computer access (145/167, 86.8%). Caregivers endorsed using multiple social media platforms, most commonly Facebook (131/168, 78.0%), Instagram (107/168, 63.7%), YouTube (104/168, 61.9%), and Twitter (83/168, 49.4%). Snapchat, Pinterest, LinkedIn, and WhatsApp were each used by less than 35%; 2.4% (4/168) of caregivers reported they did not currently use social media. Of 156 caregivers, 44 (28.2%) were participating in an online support group and 58% (22/38) of those reported the group was specifically for caregivers of justice-involved youth.

Acceptability of Digital Health Interventions

Most caregivers were open to participating in digital health interventions, with 60.5% (101/167) caregivers indicating they would *probably* or *definitely* participate in an online intervention specifically. When asked about specific types of digital health interventions, caregivers' most preferred option (66/161, 41.0% rated first choice) was receiving supportive/motivational parenting messages via SMS text message. The second highest was viewing supportive/motivational parenting posts on social media platforms (36/161, 22.4% rated first choice), followed by private groups on Facebook connecting caregivers of justice-involved youth with one another (16/161, 9.9% rated first choice). Caregivers were also willing to participate in an online support community (78/168, 46.4%), Facebook groups connecting caregivers with mental health professionals (76/168, 45.2%), and video-based sessions for individual therapy (50/168, 29.8%), family therapy (44/168, 26.2%), or support group meetings with other caregivers (40/168, 23.8%).

Qualitative Interview

Sample Characteristics

The subset of 18 caregivers was 67% (12) female, on average 45 years old (SD 10), and racially/ethnically diverse (black: 10/18, 56%; white: 7/18, 39%, Latinx: 1/18, 6%; and other: 2/18, 11%). Caregivers were all familial and largely biological (16/18, 89%) parents to justice-involved youth who were on average 16 years old (SD 3). At the time of the interview, 56% (10/18) of justice-involved youth had ongoing court appointments, 50% (9/18) were on probation, 33% (6/18) had pending charges, and 6% (1/18) were detained.

Technology and Social Media Usage

All caregivers owned a smartphone/tablet, and most had regular computer access (15/18, 83%). Caregivers endorsed using multiple social media platforms, most commonly Facebook (18/18, 100%), Instagram (13/18, 72%), and Twitter (11/18, 61%). Snapchat and WhatsApp were used by less than 35% of the sample. Most caregivers (17/18, 94%) reported using social media to connect with family and friends, with few (4/18, 22%) connecting with people they met online. Half of the caregivers endorsed online support group membership, with only one indicating this group was specifically for parents of children on probation.

Digital Health Intervention Vignettes

Caregivers were highly receptive to participating in the hypothetical digital health interventions. Perspectives on each proposed intervention are presented in order of preference.

Video Family Therapy

Almost all caregivers (17/18, 94%) reported they would participate in video-based family therapy, primarily because of convenience. Some liked that sessions could work around a caregiver's schedule balancing work and family needs (n=4), whereas others liked saving time through eliminating the need for transportation (n=9):

> *I like the fact that it could be in home. You know, if I'm having a really tough day, like we can do a session on Skype or, you know, I really like that it could come both ways without having to go to the office and... just fit in with the lifestyle of a busy person who has a family and a whole lot going on because 24 hours in a day might seem large but it's so small when you have to cram an hour here, an hour there, hour here, hour there, four hours here, six hours there. It just seems like it would help a lot.* [Native American female, 32 years]

Some liked the idea of video-based family therapy because it would deliver professional support or treatment from which their youth could benefit (n=5). Others liked that video-based family therapy would expand the reach of mental health services to under-resourced populations (n=2):

> *[D]ispersing information on a wide scale, it would be great. You could affect a great community, you know, at one time. Everybody having somewhat the same concerns. You can reach a greater audience as opposed to scheduling appointments.* [Black male, 56 years]

Only one caregiver stated they would not participate in video-based family therapy, specifically because he was not computer savvy (black male, age 55 years). However, most participants (n=10) noted aspects of this modality they disliked. In all, 5 of 6 males expressed concerns, compared with 5 of 12 females. Males expressed a range of dislikes, yet females primarily expressed concerns about the therapist's credibility and their youth's willingness to participate. Women of color expressed concerns about being able to trust the therapist leading the session is reliable or credentialed (n=3):

> *I guess like knowing whether or not the person is legit or not. I don't know, like when you go to a doctor's office they have all their credentials on the wall, you know... what if it was just some person you didn't know but was pretending to be?* [Multiracial female, 36 years]

Other concerns included how privacy would be maintained and whether sessions would be confidential (n=3), feeling like they would be stigmatized if they included extended family members in sessions (n=1) and a loss of intimacy between the therapist and the client because of the use of a video platform (n=1):

> *...to me, video is fine for informal conversation, casual conversation. Some people use it for business communication, but again, when you're dealing with somebody's health or somebody's life, I think that there's a lot that could be lost, because in dealing with lives you want the best opportunity.* [Black male, 56 years]

Caregivers mentioned several barriers to participating in video therapy sessions. Technological barriers included the potential of losing access to the internet connection (n=2) and concerns about ease of use (n=1). Caregivers expressed concerns about scheduling, related to their own availability (n=3), provider availability (n=1), and after-hours support (n=1). Other barriers included financial costs (n=1) and distractions at home (n=1). Some caregivers mentioned their youth's willingness to participate would be the biggest barrier (n=4):

> *I think, honestly, it would work for me as the parent of the juvenile, my son, but I feel like my son would not do it. I feel like video chat and him sitting down? He won't do it... I know my son won't do it and even if I did get my son to sit down and try and do it, he wouldn't talk.* [Black female, 38 years]

Private Facebook Group

Most (16/18, 89%) caregivers said they would participate in a private Facebook group to connect with other caregivers of justice-involved youth and a mental health professional. Almost all (n=15) caregivers agreed providing a space for social support made the group highly appealing:

> *Maybe, you know, there's helpful ideas or even just like I said, having somebody to talk to, knowing somebody is going through—going through, or have gone through what you've gone through before. It's*

> *good. I mean, I had positive experience with it—with the miscarriage group because I didn't know anybody that had had one. And I didn't—I didn't know how to deal with it—what to do, you know? And when you connect with people that are going through the same thing or have gone through it, like sometimes you just need to hear that, you know, to feel better. Like okay, I'm not the only one.* [White female, 35 years]

Many (n=12) caregivers shared they would like to participate in the group because it offers the opportunity to speak to a mental health professional—someone who can serve as a credible resource for knowledge acquisition about mental health care. Half (n=9) of the caregivers expressed a preference for communicating with the mental health professional through video rather than phone or chat/Facebook messenger:

> *You can learn a lot from a person just by looking at them and seeing them, you know, see if you believe in them. Again, it's all through the eyes. But, yeah, in the beginning, absolutely and on video. Once you get to know them and you're pretty comfortable with them and their words resonate with you, then sure, then you can scale it back a little bit and do, you know, instant messaging, you know, things like that next, what have you, emails.* [Latinx male, 50 years]

More than half (n=10) of the caregivers liked that their friends and family would never know they are a part of the group because of Facebook privacy settings. A primary concern, however, was whether they could trust the other caregivers to maintain privacy (n=8) as well as personally needing time to build trust (n=4):

> *Because I don't know who any of these people are. So it's like if I post something that I wanted to be within that group of members, I don't want to see something like screenshotted or any of my information exposed to everybody on the universe of Facebook or any type of social media group.* [White female, 37 years]

Several caregivers expressed concerns about other caregivers being unsupportive of the group as a unit (n=1), dominating conversations (n=1), and judging or attacking other group members (n=4). Caregivers (n=4) suggested a moderator would be useful to manage the group and remedy some of these concerns. Caregivers noted other potential barriers, including availability in their own schedules (n=5); issues with technology, specifically fearing Facebook hacks (n=1); losing internet connection (*n*=1); feeling as though online groups lack intimacy (n=1); and preferring in-the-moment responses (n=1).

Text Messaging Intervention

Most interview participants (14/18, 78%) stated they would enroll in a text messaging intervention where they receive appointment reminders and supportive/motivational parenting messages. Caregivers identified the primary benefit as the convenience of receiving reminders about appointments (n=14):

> *I think they should have had this all along. I think that would—this would—would help thousands of people in the system going through different things. I think it's great. I think, uh, that would eliminate a lot of missed - missed show ups at court. A lot of times people, you know, have issues or forget, oh, I thought it was this date, that date. I think it's great. I think it would manage a lot of people's families and help folks a lot better.* [Black female, 47 years]

Caregivers liked that the system was technology based, rather than on paper (n=3), and several caregivers shared they receive similar reminders through a current health care provider and have found these beneficial (n=3):

> *Oh, the reminders have been lifesaving. I mean, it has totally saved me in getting to an appointment. It has totally reminded me of what the specific—specificity of the appointment was for. And so it's allowed me to make that appointment and get there on time and be prepared. So I think it's a great benefit.* [Black male, 56 years]

Several caregivers, predominantly female, also expressed concerns. Some participants felt the reminders were unnecessary (n=3), instead preferring just the supportive/motivational messages (n=1). Some also had concerns about receiving too many text messages (n=3) or their own availability to see and respond to the messages (n=3).

Caregivers shared concerns about whether the system would function properly, including whether (1) information (eg, appointment times) would be accurate (n=1), (2) the system would update for rescheduled appointments (n=1), (3) the alert would go through on their phone (n=1), and (4) they would be properly removed after service use ended (n=1). Additional privacy concerns included who would be managing the system and the type of information they would have access to (n=1), having identifying information in text messages (n=1), or having messages pop-up that others could see (n=1):

> *I would wonder who's managing that type of stuff and what other things can access through that information...Like are they just going to know that there is an appointment that day, or are they going to know what it's about, or have any details of the case? Just making sure that that type of information is secure.* [Multiracial female, 36 years]

Several caregivers noted technology could be a barrier, including their phone not working consistently (n=2), having limitations to one's data/texting plan (n=1), and disliking the use of text messaging (n=1).

Discussion

Main Findings

This mixed methods study examined the feasibility of recruiting caregivers of justice-involved youth into research through social media and their perceptions regarding the acceptability of digital health interventions. The results suggest recruitment of this population through Facebook is highly feasible, and caregivers of justice-involved youth are receptive to a wide range of digital health interventions.

Recruitment was highly successful, yielding a diverse sample. Within 3 weeks, more than 3000 individuals clicked the advertisement link; cost per click was US $0.53, conversion rate was 11.5%, eligibility was 43%, and cost per participant was US $10.73. Our success was comparable with other studies, where the median cost per click for advertisements was US $0.51, conversion rate was 4% (range 0.06-29.50), eligibility was 61% (range 17-100), and cost per participant was US $14.41 [39]. This is promising, given justice involvement continues to be stigmatized, making in-person recruitment challenging. Furthermore, the cost per participant is much lower than that of in-person recruitment, which requires significant staff time; this lends promise to collecting data from diverse samples even when under budgetary constraints, as with this study.

Caregivers who participated in the survey and interview were open to participating in digital health interventions and expressed a wide range of preferences. More than half of survey respondents reported they would definitely or probably participate in a digital health intervention for caregivers of justice-involved youth. The most preferred option was a text messaging intervention that provides regular supportive/motivational parenting messages directly to their mobile phone. Caregivers' least preferred interventions were video-based individual or family therapy. In contrast, qualitative interview participants preferred video-based family therapy over the text messaging intervention. It is possible the additional details provided during the qualitative vignettes about each proposed intervention resulted in more openness to participating in the video-based therapy. Alternatively, those who elected to participate in the telephone interview might be more open to interventions involving reciprocal communication with a professional.

Qualitative interview participants endorsed many positive attitudes toward possible digital health interventions. Caregivers predominantly liked the hypothetical interventions because they could benefit from social support from peers with similar experiences, professional support from a licensed clinician, or both. They also liked that the digital delivery of interventions could resolve barriers to accessing care, primarily related to transportation and difficulty scheduling. Despite these noteworthy benefits, caregivers also shared concerns. Primary concerns surrounded privacy and information sharing, especially regarding their youth, and willingness of their youth to participate in the interventions. Concerns were also raised about technology's functionality and reliability, the credibility of the involved mental health professional, and behavior of other participants in group-based interventions.

Strengths and Limitations

This study has several notable strengths and limitations that can guide future research. Strengths include the mixed methods approach, nationwide sampling, and diversity in the caregivers recruited in terms of age, gender, and race/ethnicity. Obtaining perspectives from diverse caregivers through multiple methods allowed for a more comprehensive (ie, nationwide survey) and nuanced (ie, in-depth interviews) investigation of the use of mHealth technology to recruit and intervene with caregivers of justice-involved youth. The study provides support for using social media to engage caregivers of justice-involved youth in both research and treatment, as well as several possible acceptable avenues for intervention. The use of these strategies and proposed interventions has the potential to expand these families' access to treatment and promote health equity.

As with most internet-based data collection, a key study limitation involves verifying respondent identity. We relied on self-report of status as a caregiver of a justice-involved youth, so it is possible some respondents mischaracterized themselves to gain study entry. Given the incentive was minimal and was only provided for those who completed the full survey, it seems unlikely this would have motivated individuals to falsely identify themselves as caregivers of justice-involved youth. We were also not able to track internet protocol (IP) addresses, so it is possible some respondents attempted the survey more than once; we combatted this by removing cases where the email was identical for multiple surveys (very small percentage of cases). Future research should consider requesting the verification of identity and youth's justice involvement through official records, although this is highly sensitive information and given mistrust of researchers is common, this could limit the willingness of caregivers to participate. When possible, tracking IP addresses in future studies could provide an additional method of preventing multiple entries from the same individual.

Our study also relied on recruitment from a single social media platform, so we did not reach caregivers who use social media platforms at the exclusion of Facebook. We used Facebook because in the United States, 75% of parents use Facebook [40], and approximately one-third of Facebook users are between the ages of 35 and 54 years, the age of most caregivers of justice-involved youth [41]. Furthermore, sociodemographic characteristics of participants recruited through Facebook tend to mirror those recruited through more traditional methods or national statistics [39], and with the rise of *smart* mobile devices, racial/ethnic and economic disparities in social media use have decreased substantially [42]. Facebook also offers an easy to use advertisement platform through which to conduct research. Future studies should consider expanding recruitment to other social media sites (eg, Twitter) to reach a wider range of caregivers.

Future Directions

The results of this study suggest recruiting and intervening with caregivers of justice-involved youth through social media is a feasible and acceptable approach. Caregivers expressed willingness to participate in a wide range of digital health interventions. Their preferences and concerns varied, however, suggesting the need for a range of interventions to increase access to care for this population. Work is underway to develop and evaluate a text messaging system that provides appointment reminders and motivational messages [29] as well as to adapt an existing in-person family-based intervention [13] to be delivered via telehealth. In future development of social media–based interventions, researchers and practitioners should consider ways to address privacy concerns (eg, security and type of platform), use of a moderator to manage group-based intervention discussions, and ways to demonstrate practitioner credibility (eg, proof of licensure). Interventions involving a

clinician or moderator may also require nontraditional work hours, as many caregivers communicated the benefit of scheduling flexibility as they juggle competing demands.

Acknowledgments

The University of California, San Francisco, Psychiatry Department provided funding through a pilot award (principal investigator: AH). Partial salary support was received through grants by the National Institute of Mental Health (T32MH018262) and the National Institute on Drug Abuse (K24DA046569; R01DA035231). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health. The authors would like to thank the caregivers who participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Advertisement testing.
[DOCX File , 220 KB - jmir_v22i4e16370_app1.docx]

References

1. Office of Juvenile Justice and Delinquency Prevention. Statistical Briefing Book URL: https://www.ojjdp.gov/ojstatbb/ [accessed 2019-07-14]
2. Office of Juvenile Justice and Delinquency Prevention. Easy Access to Juvenile Court Statistics URL: https://www.ojjdp.gov/ojstatbb/ezajcs/ [accessed 2019-07-15]
3. Belenko S, Knight D, Wasserman GA, Dennis ML, Wiley T, Taxman FS, et al. The juvenile justice behavioral health services cascade: a new framework for measuring unmet substance use treatment services needs among adolescent offenders. J Subst Abuse Treat 2017 Mar;74:80-91 [FREE Full text] [doi: 10.1016/j.jsat.2016.12.012] [Medline: 28132705]
4. Teplin LA, Abram KM, McClelland GM, Dulcan MK, Mericle AA. Psychiatric disorders in youth in juvenile detention. Arch Gen Psychiatry 2002 Dec;59(12):1133-1143 [FREE Full text] [doi: 10.1001/archpsyc.59.12.1133] [Medline: 12470130]
5. Bryan AD, Schmiege SJ, Broaddus MR. HIV risk reduction among detained adolescents: a randomized, controlled trial. Pediatrics 2009 Dec;124(6):e1180-e1188. [doi: 10.1542/peds.2009-0679] [Medline: 19901006]
6. Dembo R, Belenko S, Childs K, Wareham J, Schmeidler J. Individual and community risk factors and sexually transmitted diseases among arrested youths: a two level analysis. J Behav Med 2009 Aug;32(4):303-316 [FREE Full text] [doi: 10.1007/s10865-009-9205-8] [Medline: 19224357]
7. Burt SA, McGue M, Krueger RF, Iacono WG. Environmental contributions to adolescent delinquency: a fresh look at the shared environment. J Abnorm Child Psychol 2007 Oct;35(5):787-800. [doi: 10.1007/s10802-007-9135-2] [Medline: 17505878]
8. Folk JB, Brown LK, Marshall BD, Ramos LM, Gopalakrishnan L, Koinis-Mitchell D, et al. The prospective impact of family functioning and parenting practices on court-involved youth's substance use and delinquent behavior. J Youth Adolesc 2020 Jan;49(1):238-251. [doi: 10.1007/s10964-019-01099-8] [Medline: 31399895]
9. Lee TR, Goddard HW. Developing family relationship skills to prevent substance abuse among high-risk youth. Fam Relat 1989;38(3):301-305. [doi: 10.2307/585056]
10. Waschbusch DA. A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. Psychol Bull 2002 Jan;128(1):118-150. [doi: 10.1037/0033-2909.128.1.118] [Medline: 11843545]
11. Farrington DP, Welsh BC. Family-based prevention of offending: A meta-analysis. Aust N Z J Criminol 2003;36(2):127-151 [FREE Full text] [doi: 10.1375/acri.36.2.127]
12. Kumpfer KL, Alvarado R, Whiteside HO. Family-based interventions for substance use and misuse prevention. Subst Use Misuse 2003;38(11-13):1759-1787. [doi: 10.1081/ja-120024240] [Medline: 14582577]
13. Tolou-Shams M, Dauria E, Conrad SM, Kemp K, Johnson S, Brown LK. Outcomes of a family-based HIV prevention intervention for substance using juvenile offenders. J Subst Abuse Treat 2017 Jun;77:115-125 [FREE Full text] [doi: 10.1016/j.jsat.2017.03.013] [Medline: 28476263]
14. Herrman H, Humphreys C, Halperin S, Monson K, Harvey C, Mihalopoulos C, et al. A controlled trial of implementing a complex mental health intervention for carers of vulnerable young people living in out-of-home care: the ripple project. BMC Psychiatry 2016 Dec 7;16(1):436 [FREE Full text] [doi: 10.1186/s12888-016-1145-6] [Medline: 27927174]
15. Tolou-Shams M, Harrison A, Conrad SM, Johnson S, Brown LK. Challenges to conducting adolescent HIV prevention services research with court-involved youth. Child Youth Serv Rev 2017;83:201-208. [doi: 10.1016/j.childyouth.2017.10.042]
16. Cunningham PB, Henggeler SW. Engaging multiproblem families in treatment: lessons learned throughout the development of multisystemic therapy. Fam Process 1999;38(3):265-281. [doi: 10.1111/j.1545-5300.1999.00265.x] [Medline: 10526764]

17. Elkington KS, Belmonte K, Latack JA, Mellins CA, Wasserman GA, Donenberg GR, et al. An exploration of family and juvenile justice systems to reduce youth HIV/STI risk. J Res Adolesc 2015 Dec 1;25(4):700-716 [FREE Full text] [doi: 10.1111/jora.12163] [Medline: 26539022]
18. van der Pol TM, Hoeve M, Noom MJ, Stams GJ, Doreleijers TA, van Domburgh L, et al. Research Review: The effectiveness of multidimensional family therapy in treating adolescents with multiple behavior problems - a meta-analysis. J Child Psychol Psychiatry 2017 May;58(5):532-545. [doi: 10.1111/jcpp.12685] [Medline: 28121012]
19. Batastini AB. Improving rehabilitative efforts for juvenile offenders through the use of telemental healthcare. J Child Adolesc Psychopharmacol 2016 Apr;26(3):273-277. [doi: 10.1089/cap.2015.0011] [Medline: 26288198]
20. Yonek JC, Dauria EF, Kemp K, Koinis-Mitchell D, Marshall BD, Tolou-Shams M. Factors associated with use of mental health and substance use treatment services by justice-involved youths. Psychiatr Serv 2019 Jul 1;70(7):586-595. [doi: 10.1176/appi.ps.201800322] [Medline: 31138054]
21. Ryu S. Book review: mHealth: new horizons for health through mobile technologies: based on the findings of the second global survey on eHealth (global observatory for eHealth series, volume 3). Healthc Inform Res 2012;18(3):231-233. [doi: 10.4258/hir.2012.18.3.231]
22. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17(Suppl 1):82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
23. Oesterle S, Epstein M, Haggerty KP, Moreno MA. Using Facebook to recruit parents to participate in a family program to prevent teen drug use. Prev Sci 2018 May;19(4):559-569 [FREE Full text] [doi: 10.1007/s11121-017-0844-7] [Medline: 29116552]
24. Ramo DE, Rodriguez TM, 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]
25. Pedersen ER, Kurz J. Using Facebook for health-related research study recruitment and program delivery. Curr Opin Psychol 2016 May;9:38-43 [FREE Full text] [doi: 10.1016/j.copsyc.2015.09.011] [Medline: 26726313]
26. Kaysen D, Davis KC, Kilmer J. Use of social networking sites to sample lesbian and bisexual women. Addict Newsl 2011;18:14-15 [FREE Full text] [doi: 10.1037/e523322013-003]
27. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using Facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015 Jun 1;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]
28. Bath E, Tolou-Shams M, Farabee D. Mobile health (MHealth): building the case for adapting emerging technologies for justice-involved youth. J Am Acad Child Adolesc Psychiatry 2018 Dec;57(12):903-905 [FREE Full text] [doi: 10.1016/j.jaac.2018.08.012] [Medline: 30522732]
29. Tolou-Shams M, Yonek J, Galbraith K, Bath E. Text messaging to enhance behavioral health treatment engagement among justice-involved youth: qualitative and user testing study. JMIR Mhealth Uhealth 2019 Apr 5;7(4):e10904 [FREE Full text] [doi: 10.2196/10904] [Medline: 30950808]
30. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
31. Acharya AS, Prakash A, Saxena P, Nigam A. Sampling: why and how of it? Indian J Med Spec 2013;4(2):330-333. [doi: 10.7713/ijms.2013.0032]
32. Lopez KA, Willis DG. Descriptive versus interpretive phenomenology: their contributions to nursing knowledge. Qual Health Res 2004 May;14(5):726-735. [doi: 10.1177/1049732304263638] [Medline: 15107174]
33. Barriball L, While A. Collecting data using a semi-structured interview: a discussion paper. J Adv Nurs 1994 Feb;19(2):328-335. [doi: 10.1111/j.1365-2648.1994.tb01088.x] [Medline: 8188965]
34. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006;27(2):237-246. [doi: 10.1177/1098214005283748]
35. Campbell JL, Quincy C, Osserman J, Pedersen OK. Coding in-depth semistructured interviews: problems of unitization and intercoder reliability and agreement. Sociol Methods Res 2013;42(3):294-320. [doi: 10.1177/0049124113500475]
36. Hruschka DJ, Schwartz D, St John DC, Picone-Decaro E, Jenkins RA, Carey JW. Reliability in coding open-ended data: lessons learned from HIV behavioral research. Field Methods 2004;16(3):307-331. [doi: 10.1177/1525822x04266540]
37. Friese S. Qualitative Data Analysis with ATLAS.ti. United Kingdom: Sage Publications Ltd; 2019.
38. Miles MB, Huberman AM, Saldaña J. Qualitative Data Analysis: A Methods Sourcebook. Third Edition. Thousand Oaks, CA: Sage Publications, Inc; 2013.
39. Whitaker C, Stevelink S, Fear N. The Use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
40. Duggan M, Lenhard A, Lampe C, Ellison NB. Pew Research Center. 2015 Jul 16. Parents and Social Media: Mothers Are Especially Likely to Give and Receive Support on Social Media URL: https://www.pewinternet.org/wp-content/uploads/sites/9/2015/07/Parents-and-Social-Media-FIN-DRAFT-071515.pdf [accessed 2019-07-15]

41. Harrison AJ, Tolou-Shams M. Drug use and mental health of caregivers of youth in the juvenile drug court: implications for holistic family-based intervention. 2017 Presented at: 79th Annual Scientific Meeting of the College on Problems of Drug Dependence; June 17-22, 2017; Montreal, Quebec URL: https://cpdd.org/wp-content/uploads/2019/09/058425-Program-Book-2017-Que%CC%81bec-FinalEllen6-5-17.pdf
42. Anderson M. Pew Research Center. 2015 Oct 29. Technology Device Ownership: 2015 URL: https://www.pewinternet.org/wp-content/uploads/sites/9/2015/10/PI_2015-10-29_device-ownership_FINAL.pdf [accessed 2019-07-15]

Abbreviations

IP: internet protocol
mHealth: mobile health

Edited by M Focsa; submitted 23.09.19; peer-reviewed by M Krausz, D Hansen, J Parsons; comments to author 05.11.19; revised version received 30.12.19; accepted 03.02.20; published 30.04.20.

Please cite as:
Folk JB, Harrison A, Rodriguez C, Wallace A, Tolou-Shams M
Feasibility of Social Media–Based Recruitment and Perceived Acceptability of Digital Health Interventions for Caregivers of Justice-Involved Youth: Mixed Methods Study
J Med Internet Res 2020;22(4):e16370
URL: http://www.jmir.org/2020/4/e16370/
doi:10.2196/16370
PMID:32352388

Original Paper

Quality of Information Regarding Repair Restorations on Dentist Websites: Systematic Search and Analysis

Philipp Kanzow[1], MSc, Dr med dent, Dr rer medic; Amelie Friederike Büttcher[2], MEd; Annette Wiegand[1], Prof Dr med dent; Falk Schwendicke[3], MDPH, Prof Dr med dent

[1]Department of Preventive Dentistry, Periodontology and Cariology, University Medical Center Göttingen, Göttingen, Germany

[2]Division of Medical Education Research and Curriculum Development, University Medical Center Göttingen, Göttingen, Germany

[3]Department of Operative and Preventive Dentistry, Charité – Universitätsmedizin Berlin, Berlin, Germany

Corresponding Author:
Philipp Kanzow, MSc, Dr med dent, Dr rer medic
Department of Preventive Dentistry, Periodontology and Cariology
University Medical Center Göttingen
Robert-Koch-Str 40
Göttingen, 37075
Germany
Phone: 49 551 3922877
Email: philipp.kanzow@med.uni-goettingen.de

Abstract

Background: Repairing instead of replacing partially defective dental restorations represents a minimally invasive treatment concept, and repairs are associated with advantages over complete restoration replacement. To participate in the shared decision-making process when facing partially defective restorations, patients need to be aware of the indications, limitations, and advantages or disadvantages of repairs. Patients are increasingly using the internet to gain health information like this online.

Objective: We aimed to assess the quality of German-speaking dentist websites on repairs of partially defective restorations.

Methods: Three electronic search engines were used to identify German-speaking websites of dental practices mentioning repairs. Regarding information on repairs, websites were assessed for (1) technical and functional aspects, (2) comprehensiveness of information, and (3) generic quality and risk of bias. Domains 1 and 3 were scored using validated tools (LIDA and DISCERN). Comprehensiveness was assessed using a criterion checklist related to evidence, advantages and disadvantages, restorations and defects suitable for repairs, and information regarding technical implementation. Generalized linear modeling was used to assess the impact of practice-specific parameters (practice location, practice setting, dental society membership, and year of examination or license to practice dentistry) on the quality of information. An overall quality score was calculated by averaging the quality scores of all three domains and used as primary outcome parameter. Quality scores of all three domains were also assessed individually and used as secondary outcomes.

Results: Fifty websites were included. The median score of quality of information was 23.2% (interquartile range [IQR] 21.7%-26.2%). Technical and functional aspects (55.2% [IQR 51.7%-58.6%]) showed significantly higher quality than comprehensiveness of information (8.3% [IQR 8.3%-16.7%]) and generic quality and risk of bias (3.6% [IQR 0.0%-7.1%]; $P<.001$/Wilcoxon). Quality scores were not related to practice-specific parameters ($P>.05$/generalized linear modeling).

Conclusions: The quality of German-speaking dentist websites on repairs was limited. Despite sufficient technical and functional quality, the provided information was neither comprehensive nor trustworthy. There is great need to improve the quality of information to fully and reliably inform patients, thereby allowing shared decision making.

(J Med Internet Res 2020;22(4):e17250) doi:10.2196/17250

KEYWORDS

evidence-based dentistry; internet; restoration repair; shared decision making

Introduction

Repairs of partially defective restorations represent a minimally invasive treatment concept and are associated with a number of advantages over complete restoration replacement. In recent years, numerous studies have focused on the repair behavior of dentists, dental students, and dental educators [1,2]. Both retrospective [3-6] and prospective [7] clinical studies as well as a wide range of repair protocols based on numerous in vitro studies [8] are available, and even economic evaluations [9] have been published. The acceptance of dentists, and presumably also patients, toward repairs can be regarded as high, with patient acceptance having been reported to range from 89% to 93% (these numbers are based on interviewing dentists, however) [10-13].

To allow patients to participate in the shared decision-making process when facing partially defective restorations, both patients and dentists need to be aware of the indications, limitations, and advantages or disadvantages of repairs. For patients, such information will often come from their dentist (eg, during a consultation). Increasingly, however, patients may actively assess information like this online (eg, on their dentist's website) [14,15]. Besides information, patients might also look online for a dentist able to deliver the requested care. Ideally, the information provided on dentists' websites regarding treatments (like repairs) should be unbiased and comprehensive, allowing patients to come to an informed decision instead of being misinformed or biased. Until now, whether dentists' websites allow patients to gain such comprehensive and trustworthy information on restoration repair has not been assessed.

Therefore, this study aimed to assess the quality of German-speaking dentists' websites presenting information on repair restorations across three domains: (1) technical and functional aspects, (2) comprehensiveness of repair-specific information, and (3) generic quality and risk of bias. The null hypothesis was that practice-specific parameters do not impact website information quality.

Methods

The reporting of this study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and the Enhancing Transparency in Reporting the Synthesis of Qualitative Research (ENTREQ) statements [16,17].

Search Strategy

Three electronic search engines (google.de, bing.de/yahoo.de, ask.com) were used. Searches were performed on April 28 and 29, 2019, using different search strategies, as google.de offers only limited options to combine multiple search terms with Boolean operators. Search terms represent different combinations of the German words for repair restoration(s), composite(s), and dentist(s) (Multimedia Appendix 1). A computer running macOS 10.14.4 (Apple Inc) connected to the internet in Germany was used. Cookies and browser history of Firefox Quantum 66.0.3 (Mozilla Foundation) were cleared and the default setting of each search engine was used.

In total, 2864 webpages were displayed as "most relevant" sites (google.de: 1299, bing.de/yahoo.de: 1295, ask.com: 270), and the search was not expanded beyond this number of displayed webpages assuming saturation. Also, advertisements (ie, websites from page owners paying a fee to have their website prominently displayed) were not additionally assessed.

For the purpose of this study, only websites from dentists were included. Patients are likely to be looking at these sources only while searching for a person able to deliver appropriate dental care. Therefore, websites from or associated with dental laboratories or supply and materials companies, forums and blogs operated by nondentists, dental regulatory bodies, dental schools and clinics, research agencies, or otherwise public bodies were excluded. Notably, however, patients may well find these informative, too.

The remaining 820 webpages were screened in full text. Webpages containing irrelevant information were excluded, leaving 74 webpages that were potentially eligible. Finally, after removal of duplicates, 50 websites fulfilled the inclusion criteria: (1) page freely accessible, (2) German language, (3) posted by a dental practice or practice cooperation, (4) mentions repairs. Websites containing multiple eligible webpages (ie, published under the same domain or published from the same practice) were jointly assessed as one website. The full search workflow is shown in Multimedia Appendix 2.

Data Extraction

The following parameters were collected from the websites, if available: (1) practice name, (2) URL, (3) country, (4) practice location (rural, town [<100,000 inhabitants], or city [≥100,000 inhabitants]), (5) practice setting (single practitioner, multiple dentists, or practice cooperation), (6) dentist's gender (female, male, mixed [in case of multiple dentists or practice cooperation]), (7) dental society memberships, (8) year of examination or approbation, and (9) information regarding repairs (Multimedia Appendix 3). Information regarding dental society memberships and year of examination or approbation were cross-referenced from dental societies' member information pages (ie, German Society of Dentistry and Oral Medicine [DGZMK], Swiss Dental Association [SSO]) or curriculum vitae published elsewhere (ie, in dentists' dissertations and public profiles at the social networking sites XING or LinkedIn), if information was not already listed on dentist websites. In case of multiple dentists or practice cooperations, the average years of examination or approbation was used.

Outcomes

Website quality regarding information on repairs was systematically assessed across three different domains: (1) technical and functional aspects (Table 1), (2) comprehensiveness of information (Table 2), and (3) generic quality and risk of bias (Table 3). Assessment was independently performed by two authors (PK, AFB). Discrepancies were resolved through discussion.

The established and validated LIDA instrument (version 1.2) [18] was used to assess items in domain 1 and DISCERN instrument [19] was used in domain 3. In dentistry, such tools

have been successfully applied to evaluate the quality of information on websites regarding dental caries [20,21], periodontitis [22], root canal treatment versus implant placement [23], and orthodontics [24-29]. For this study, both LIDA and DISCERN have been slightly modified to uniformly score all domains on an ordinal scale as 0 (never or no), 1 (sometimes or partially), and 2 (mostly, always, or yes).

To assess items in domain 2, a structured checklist with 6 subdomains was developed by the authors focusing on the evidence (2.1); advantages (2.2) and disadvantages (2.3) of repair restorations; restorations (2.4) and defects (2.5) suitable for repairs; and technical implementation of repairs (2.6). The same 3-point ordinal scale was used. As the number of items within each domain differed, an overall quality score was calculated by averaging the quality scores (relative percentages) of all three domains, assuming them to be of equivalent importance. This score was used as primary outcome parameter. Quality scores (relative percentages) of all three domains were also assessed individually and used as secondary outcomes.

Table 1. Subdomains regarding technical and functional aspects (domain 1) were assessed using the modified LIDA instrument (version 1.2) [18].

Subdomain and item	Median (IQR[a]; min-max[b])
1.1 Accessibility	
Does it work on a range of browsers?[c]	2 (2-2; 1-2)
Is the information available full text without registration, log-in or subscription?	2 (2-2; 2-2)
1.2 Usability	
Is there a clear statement of who this website is for?	2 (2-2; 0-2)
Is the level of detail appropriate to their[d] level of knowledge?	0 (0-0; 0-2)
Is the layout of the main block of information clear and readable?	2 (1-2; 1-2)
Is the navigation clear and well structured?	2 (2-2; 1-2)
Can you always tell your current location in the site?	2 (1-2; 0-2)
Is the colour scheme appropriate and engaging?	1.5 (1-2; 0-2)
Is the same page layout used throughout the site?	2 (2-2; 2-2)
Do navigational links have a consistent function?[e]	1 (1-2; 0-2)
Is the site structure (categories or organisation of pages) applied consistently?	2 (1-2; 0-2)
Does the site provide an effective search function?[f]	0 (0-0; 0-1)
Does the site provide effective browsing facilities?	2 (1-2; 0-2)
Does the design minimize the cognitive overhead?	1 (1-2; 0-2)
Does the site support the normal browser navigational tools?	2 (2-2; 2-2)
Can you use the site without third party plugins?	2 (2-2; 2-2)
Can the user make an effective judgment of whether the site applies to them?	2 (2-2; 1-2)
Is the website interactive?	1 (0-1; 0-2)
Can the user personalise their experience of using the site?	0 (0-0; 0-2)
Does the website integrate nontextual media?	0 (0-0; 0-2)
1.3 Reliability	
Does the site respond to recent events?	0 (0-1; 0-2)
Can users submit comments on specific content?	0 (0-0; 0-2)
Is site content updated at an appropriate interval?	0 (0-1; 0-2)
Is it clear who runs the site?	2 (2-2; 2-2)
Is it clear who pays for the site?	0 (0-0; 0-2)
Is there a declaration of the objectives of the people who run the site?	2 (1-2; 0-2)
Does the site report a clear content production method?	0 (0-0; 0-2)
Is this a robust method?	0 (0-0; 0-1)
Can the information be checked from original sources?	0 (0-0; 0-2)

[a]IQR: interquartile range.

[b]min-max: minimum and maximum score.

[c]Apple Safari 12.1, Firefox Quantum 66.0.3 for Mac, Google Chrome 74.0.3729 for Mac, and Microsoft Internet Explorer 11 were tested.

[d]The patients.

[e]As part of this item, websites were screened for broken links using a free online tool [30].

[f]Search terms “repariert,” “reparieren,” “Reparatur,” “Reparaturen,” “reparaturfähig,” “Reparaturfähigkeit,” “Füllungsreparatur,” “Füllungsreparaturen,” “Füllungserweiterung,” “Füllungserweiterungen,” “Reparaturfüllung,” “Reparaturfüllungen” were tested.

Table 2. Subdomains regarding treatment-related aspects (domain 2) were assessed.

Subdomain and item	Median (IQR[a]; min-max[b])
2.1 Evidence of repair restorations	0 (0-0; 0-2)
Are success rates or annual failure rates of repairs and replacements listed?	
Are guidelines or scientific recommendations discussed?	
Is literature cited?	
2.2 Advantages of repair restorations	0 (0-0; 0-2)
Is preservation of tooth substance (less invasive, less traumatic) mentioned?	
Is risk reduction of treatment-related complications (eg, potentially harmful effects to the pulp or iatrogenic damage of neighboring teeth) discussed?	
Are reduced costs mentioned?	
Is reduced treatment time mentioned?	
2.3 Disadvantages of repair restorations	0 (0-0; 0-0)
Are disadvantages of repair restorations discussed?	
2.4 Restorations suitable for repair	1 (1-1; 0-2)
Are amalgam restorations mentioned?	
Are composite restorations mentioned?	
Are ceramic restorations mentioned?	
Are full-metal restorations mentioned?	
Are further indirect restorations mentioned?	
2.5 Defects suitable for repair	0 (0-0.75; 0-2)
Is damage, fracture, partial loss, or partial defect of restoration discussed?	
Is secondary caries mentioned?	
Is (marginal) discoloration mentioned?	
Is ceramic chipping mentioned?	
Are marginal defects or gaps mentioned?	
2.6 Technical implementation of repair	0 (0-0; 0-2)
Is sandblasting mentioned?	
Is application of silane or universal primers mentioned?	
Are repair materials mentioned?	

[a]IQR: interquartile range.

[b]min-max: minimum and maximum score.

Table 3. Subdomains regarding generic quality and risk of bias (domain 3) were assessed using the modified DISCERN instrument [19].

Subdomain and item	Median (IQR[a]; min-max[b])
3.1 Reliability	
Are the aims clear?	0 (0-0; 0-0)
Is it[c] relevant?	0 (0-1; 0-2)
Is it clear what sources of information were used to compile the publication?	0 (0-0; 0-2)
Is it clear when the information used or reported in the publication was produced?	0 (0-1; 0-2)
Is it[c] balanced and unbiased?	0 (0-0; 0-0)
Does it[c] provide details of additional sources of support and information?	0 (0-0; 0-0)
Does it[c] refer to areas of uncertainty?	0 (0-0; 0-0)
3.2 Quality	
Does it[c] describe how each treatment works?	0 (0-0; 0-2)
Does it[c] describe the benefits of each treatment?	0 (0-0; 0-2)
Does it[c] describe the risks of each treatment?	0 (0-0; 0-0)
Does it[c] describe what would happen if no treatment is used?	0 (0-0; 0-0)
Does it[c] describe how the treatment choices affect overall quality of life?	0 (0-0; 0-0)
Is it clear that there may be more than one possible treatment choice?	0 (0-0; 0-0)
Does it[c] provide support for shared decision making?	0 (0-0; 0-0)

[a]IQR: interquartile range.

[b]min-max: minimum and maximum score.

[c]Websites' content.

Statistical Analysis

For each domain, a quality score (relative percentage: website score on all of the respective items divided by the maximum possible score sum) was calculated. Furthermore, an averaged overall quality score based on all three domains was calculated. As data were not normally distributed according to the Shapiro-Wilk test, descriptive statistical analysis contained median, quartiles, and ranges.

Differences in website scores between the three domains were analyzed using Wilcoxon signed-rank tests with a Bonferroni-Holm correction. Generalized linear modeling was used to assess the impact of practice-specific parameters on domain-related quality and the averaged overall quality score: (1) practice location (rural, town, or city); (2) practice setting (single practitioner, multiple dentists, or practice cooperation); (3) dental society membership (yes or no); and (4) year of examination or approbation. A multivariable analysis was performed and covariates entered simultaneously. Only main effects without interaction terms were tested. If no information regarding dental society membership was available, we scored this as no. If year of examination or approbation was not available, websites were treated as randomly missing and excluded from the regression analysis (n=4). Statistical analysis was performed using SPSS Statistics for Macintosh version 26.0.0.0 (IBM Corp). Statistical significance was set at $P<.05$.

Results

In total, 50 websites fulfilled the inclusion criteria. Characteristics of the included websites are shown in Table 4 (full data of all included websites are shown in Multimedia Appendix 3). Briefly, the majority of websites were from practices in Germany and situated in towns or cities. Half of the practices had single practitioners, and about half of the dentists running the websites were members of dental societies.

The median score for quality of information was 21.2% (interquartile range [IQR] 20.0%-22.3%) (Figure 1). Technical and functional aspects (55.2% [IQR 51.7%-58.6%]) showed significantly higher quality than did comprehensiveness of information (8.3% [IQR 8.3%-16.7%]) and generic quality and risk of bias (3.6% [IQR 0.0%-7.1%]; $P<.001$/Wilcoxon, Tables 1-3).

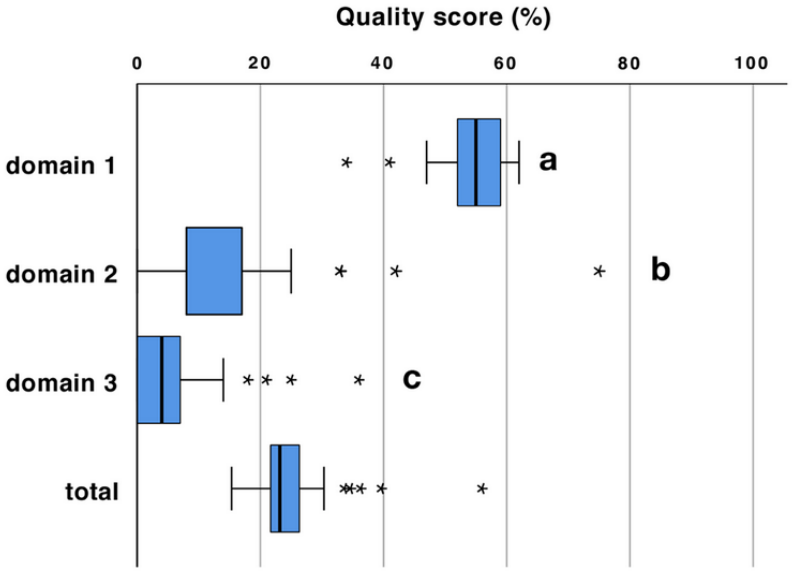
Table 4. Practice-specific parameters of the included websites (n=50).

Variable and attribute	Value
Country, n (%)	
Germany	38 (76)
Switzerland	7 (14)
Austria	4 (8)
Hungary	1 (2)
Practice location, n (%)	
Rural	10 (20)
Town	17 (34)
City	23 (46)
Practice setting, n (%)	
Single practitioner	25 (50)
Multiple dentists or practice cooperation	25 (50)
Gender, n (%)	
Female	8 (16)
Male	16 (32)
Mixed	26 (52)
Dental society membership, n (%)	
No	24 (48)
Yes	26 (52)
Year of examination or approbation[a,b], mean (SD)	1996 (9)

[a]In case of multiple dentists or practice cooperations, the years of examination or approbation were averaged, if available.

[b]There were 4 missing values.

Figure 1. Quality of different domains (relative percentage of maximum possible score sum) and total score (averaging the results from all three domains). Domain 1: technical and functional aspects; domain 2: comprehensiveness of information; domain 3: generic quality and risk of bias. Significant differences between domains are marked by different letters (*P*<.001/Wilcoxon). Outliers are marked with an asterisk (*).



Within multivariable regression analysis, none of the practice-specific parameters had a significant impact on the averaged overall quality score or domain-related quality scores (*P*>.05/generalized linear modeling, Table 5).

Table 5. Association between practice-specific parameters and website quality.

Outcome	Model fit		Practice location[a]		Practice setting (ref. single practitioner)[a]	Dental society membership (ref. no or unknown)[a]	Year of examination or approbation[a,b]
	Likelihood	*P* value	Towns (ref. rural)	Cities (ref. rural)			
Domain 1	7.88	.16	–0.45 (–4.82 to 3.93)	2.58 (–1.68 to 6.83)	1.09 (–1.87 to 4.05)	1.38 (–1.81 to 4.56)	0.12 (–0.05 to 0.29)
Domain 2	5.08	.41	6.57 (–4.69 to 17.84)	1.87 (–9.08 to 12.82)	4.51 (–3.11 to 12.13)	2.54 (–5.65 to 10.74)	–0.28 (–0.72 to 0.16)
Domain 3	3.09	.69	3.86 (–2.64 to 10.36)	2.08 (–4.23 to 8.40)	1.86 (–2.54 to 6.26)	1.18 (–3.55 to 5.91)	–0.11 (–0.36 to 0.15)
Total	3.20	.67	3.33 (–2.61 to 9.27)	2.18 (–3.60 to 7.95)	2.49 (–1.53 to 6.50)	1.70 (–2.62 to 6.02)	–0.09 (–0.32 to 0.14)

[a]Regression coefficients with 95% confidence intervals are shown.

[b]In case of multiple dentists or practice cooperations, the years of examination or approbation were averaged, if available.

Discussion

Principal Findings

In the D-A-CH countries (Germany, Austria, Switzerland), about 90% of the population has access to the internet [31]. Information on health-related aspects is increasingly assessed online, often using search engines [14,15]. Due to the broad access to the internet, operating a website has become the standard for most companies and businesses including dentists.

Search Strategy

Regarding dental health, information on dental practice websites is of special interest as patients are likely to access those websites while searching for information and an appropriate dentist. Therefore, our study focused on dentist websites only. Websites were identified using different search engines with a combined market share of more than 99% in Germany [32]. The search was performed using consumer search engines only as patients are unlikely to use scientific databases (eg, Medline).

Information Regarding Repairs

We found that only a small number of dentists included information about repairs on their websites. Dentist websites showed sufficient quality regarding technical and functional aspects but were not seen as fully trustworthy (generic quality was low, and there was a high risk of bias present). Comprehensiveness of repair-specific information was also rated low. This is in line with previous studies assessing the quality of websites regarding different dental health-related information [20,22,23,25,26,28,29]. Dental health-related information was not comprehensive and of lower quality than websites' technical and functional aspects.

A number of reasons for these findings are conceivable. First, dentists might not have enough time to create and maintain a content-comprehensive website. Dentists might also feel that informing patients online is not necessary or that it is not their task to supply patients with comprehensive health care information on their websites. The perceived lack of financial gain from providing such content online may add to this. Also, provided online content might need to be discussed with patients at the next appointment, which may be seen as a waste of time. Data from the United States demonstrated that physicians perceive appointments as less efficient and more difficult if patients have already gained information online [33]. In contrast, insufficient knowledge regarding repairs among dentists is unlikely to be a reason, as repairs are frequently taught at dental schools in Germany and all over the world [1,2]. However, dentists might regard implantology or orthodontic information to be of more importance than information on repairs, resulting in higher quality scores concerning technical aspects and generic quality and risk of bias of these websites (also measured using LIDA and DISCERN) [23,27,29].

Impact of Practice-Specific Parameters

We did not find any significant association between practice-specific parameters and website quality scores. We therefore must reject our hypothesis. This is a surprising outcome, as a range of parameters including those related to the individual practitioner and their practice seem to impact on repair behavior [1]. For example, low dentist density (ie, in a rural area), more experience or knowledge (ie, being a member of the dental associations), fewer years since dental school graduation, and working in larger group practices (ie, with multiple dentists) have been found to facilitate repairs. Also, we assumed website quality would be higher in younger dentists being more comfortable with technology and in larger group practices (with higher budgets for an online presence and marketing). Notably, a previous study on dentist websites and their quality also failed to demonstrate significant associations with most of such practice-specific parameters [22]. We mainly ascribe this to the fact that the overall quality was too poor throughout different websites, and dentists generally do not seem to prioritize providing information on repairs on their website regardless of their background or practice environment.

Limitations

This study has a number of limitations. First, the relatively small number of included websites (n=50) must be noted. Notably, the sample size was not based on a formal sample size estimation but guided by a previous study [22] and the availability of websites. Our study might have been underpowered, and the

lack of significant associations should hence be interpreted with caution. Second, we focused on German-speaking websites only. It is possible, albeit unlikely, that websites in other languages present a higher quality (eg, with regard to periodontitis, both German- and English-speaking websites showed a low quality of information) [22,34]. Last, we used established and validated criteria to assess technical aspects and generic quality and risk of bias but developed an assessment checklist for the repair-specific quality and comprehensiveness on our own. The validity of this checklist was not formally tested, and using another checklist may lead to different results.

Overcoming Observed Shortcomings

To overcome the shortcomings of dentist websites, a number of interventions are conceivable. Regulatory and legislative bodies might enforce better information standards. Professional dental bodies might assist dentists by providing high-quality information suitable for adoption on dentists' websites. Alternatively, dentists could provide links to other validated websites or organizations able to provide comprehensive information, such as dental research societies, thereby reducing the burden for the individual dentist to provide and maintain high-quality information. We did not check society websites as it can be assumed that information presented is both trustworthy and comprehensive.

Conclusion

In conclusion, only a minority of dentist websites informed patients about repair restorations. Despite sufficient technical and functional quality, the websites that did mention repairs were not comprehensive and prone to a high risk of bias. Dentists are encouraged to provide better and more trustworthy health information, including but not limited to repairs. Professional or regulatory bodies might assist dentists by providing high-quality information suitable for adoption on dentist websites. In the meantime, patients must be aware of the limitations and should seek information regarding repairs elsewhere.

Acknowledgments

The authors acknowledge support by the Open Access Publication Funds of Göttingen University. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms used in the research.
[DOC File , 56 KB - jmir_v22i4e17250_app1.doc]

Multimedia Appendix 2
Flowchart representing the search workflow.
[PDF File (Adobe PDF File), 16 KB - jmir_v22i4e17250_app2.pdf]

Multimedia Appendix 3
List of websites with information on repair restorations or restoration repair.
[DOC File , 172 KB - jmir_v22i4e17250_app3.doc]

References

1. Kanzow P, Wiegand A, Göstemeyer G, Schwendicke F. Understanding the management and teaching of dental restoration repair: systematic review and meta-analysis of surveys. J Dent 2018 Feb;69:1-21. [doi: 10.1016/j.jdent.2017.09.010] [Medline: 28943362]
2. Kanzow P, Wiegand A, Wilson NHF, Lynch CD, Blum IR. Contemporary teaching of restoration repair at dental schools in Germany: close to universality and consistency. J Dent 2018 Aug;75:121-124. [doi: 10.1016/j.jdent.2018.06.008] [Medline: 29933003]
3. Opdam NJ, Bronkhorst EM, Loomans BAC, Huysmans MCDNJM. Longevity of repaired restorations: a practice based study. J Dent 2012 Oct;40(10):829-835. [doi: 10.1016/j.jdent.2012.06.007] [Medline: 22771415]
4. van de Sande FH, Moraes RR, Elias RV, Montagner AF, Rodolpho PA, Demarco FF, et al. Is composite repair suitable for anterior restorations? A long-term practice-based clinical study. Clin Oral Investig 2019 Jun;23(6):2795-2803. [doi: 10.1007/s00784-018-2722-5] [Medline: 30368661]
5. Ruiz LF, Nicoloso GF, Franzon R, Lenzi TL, de Araujo FB, Casagrande L. Repair increases the survival of failed primary teeth restorations in high-caries risk children: a university-based retrospective study. Clin Oral Investig 2020 Jan;24(1):71-77. [doi: 10.1007/s00784-019-02899-9] [Medline: 31016542]

6. Laske M, Opdam NJM, Bronkhorst EM, Braspenning JCC, Huysmans MCDNJM. The differences between three performance measures on dental restorations, clinical success, survival and failure: a matter of perspective. Dent Mater 2019 Oct;35(10):1506-1513. [doi: 10.1016/j.dental.2019.07.010] [Medline: 31421955]
7. Martins BMC, Silva EJNLD, Ferreira DMTP, Reis KR, Fidalgo TKDS. Longevity of defective direct restorations treated by minimally invasive techniques or complete replacement in permanent teeth: a systematic review. J Dent 2018 Nov;78:22-30. [doi: 10.1016/j.jdent.2018.09.001] [Medline: 30189230]
8. Kanzow P, Wiegand A, Schwendicke F, Göstemeyer G. Same, same, but different? A systematic review of protocols for restoration repair. J Dent 2019 Jul;86:1-16. [doi: 10.1016/j.jdent.2019.05.021] [Medline: 31108118]
9. Kanzow P, Wiegand A, Schwendicke F. Cost-effectiveness of repairing versus replacing composite or amalgam restorations. J Dent 2016 Nov;54:41-47. [doi: 10.1016/j.jdent.2016.08.008] [Medline: 27575986]
10. Kanzow P, Hoffmann R, Tschammler C, Kruppa J, Rödig T, Wiegand A. Attitudes, practice, and experience of German dentists regarding repair restorations. Clin Oral Investig 2017 May;21(4):1087-1093. [doi: 10.1007/s00784-016-1859-3] [Medline: 27255959]
11. Kanzow P, Dieckmann P, Hausdörfer T, Attin T, Wiegand A, Wegehaupt FJ. Repair restorations: questionnaire survey among dentists in the Canton of Zurich, Switzerland. Swiss Dent J 2017 Apr 10;127(4):300-311 [FREE Full text] [Medline: 28480953]
12. Blum IR, Newton JT, Wilson NHF. A cohort investigation of the changes in vocational dental practitioners' views on repairing defective direct composite restorations. Br Dent J 2005 Sep;199(Suppl 5):S27-S30. [doi: 10.1038/sj.bdj.4812789] [Medline: 16192929]
13. Reineix E, Holmgren C, Muller-Bolla M, Chadelat A, Volgenant C, De Soet H, et al. Réparation des restaurations défectueuses—Connaissances et attitudes thérapeutiques en France. Le Chirurgien-Dentiste de France 2019;1842-1843:35-41 [FREE Full text]
14. Rainie L, Fox S. The online health care revolution. Washington: Pew Internet and American Life Project; 2000 Nov 26. URL: https://www.pewresearch.org/internet/2000/11/26/the-online-health-care-revolution/ [accessed 2020-03-04]
15. 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 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
16. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
17. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012;27(12):181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
18. Minervation. 2011. The LIDA Instrument: Minervation validation instrument for health care web sites, version 1.2 URL: http://www.minervation.com/wp-content/uploads/2011/04/Minervation-LIDA-instrument-v1-2.pdf [accessed 2020-03-04]
19. 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]
20. Blizniuk A, Furukawa S, Ueno M, Kawaguchi Y. Evaluation of English websites on dental caries by using consumer evaluation tools. Oral Health Prev Dent 2016;14(4):363-369. [doi: 10.3290/j.ohpd.a35746] [Medline: 26973989]
21. Aguirre PEA, Coelho MM, Rios D, Machado MAAM, Cruvinel AFP, Cruvinel T. Evaluating the dental caries-related information on Brazilian websites: qualitative study. J Med Internet Res 2017 Dec 13;19(12):e415 [FREE Full text] [doi: 10.2196/jmir.7681] [Medline: 29237585]
22. Schwendicke F, Stange J, Stange C, Graetz C. German dentists' websites on periodontitis have low quality of information. BMC Med Inform Decis Mak 2017 Aug 02;17(1):114 [FREE Full text] [doi: 10.1186/s12911-017-0511-8] [Medline: 28768511]
23. Rossi-Fedele G, Musu D, Cotti E, Dogramacı EJ. Root canal treatment versus single-tooth implant: a systematic review of internet content. J Endod 2016 Jun;42(6):846-853. [doi: 10.1016/j.joen.2016.02.011] [Medline: 27080114]
24. Aldairy T, Laverick S, McIntyre GT. Orthognathic surgery: is patient information on the Internet valid? Eur J Orthod 2012 Aug;34(4):466-469. [doi: 10.1093/ejo/cjr046] [Medline: 21459834]
25. Livas C, Delli K, Ren Y. Quality evaluation of the available Internet information regarding pain during orthodontic treatment. Angle Orthod 2013 May;83(3):500-506. [doi: 10.2319/052512-435.1] [Medline: 23025687]
26. Pithon MM, dos Santos ES. Information available on the internet about pain after orthognathic surgery: a careful review. Dental Press J Orthod 2014;19(6):86-92 [FREE Full text] [doi: 10.1590/2176-9451.19.6.086-092.oar] [Medline: 25628084]
27. Patel A, Cobourne MT. The design and content of orthodontic practise websites in the UK is suboptimal and does not correlate with search ranking. Eur J Orthod 2015 Aug;37(4):447-452. [doi: 10.1093/ejo/cju078] [Medline: 25488944]
28. Verhoef WA, Livas C, Delli K, Ren Y. Assessing the standards of online oral hygiene instructions for patients with fixed orthodontic appliances. J Am Dent Assoc 2015 May;146(5):310-317. [doi: 10.1016/j.adaj.2015.01.004] [Medline: 25925523]

29. Dogramacı EJ, Rossi-Fedele G. The quality of information on the Internet on orthodontic retainer wear: a cross-sectional study. J Orthod 2016 Mar;43(1):47-58. [doi: 10.1080/14653125.2015.1114711] [Medline: 26751763]
30. Online Broken Link Checker. URL: https://www.brokenlinkcheck.com/ [accessed 2020-03-05]
31. Miniwatts Marketing Group. 2019 Nov 30. Internet stats and Facebook usage in Europe: June 2019 statistics URL: https://www.internetworldstats.com/stats4.htm#europe [accessed 2020-03-04]
32. SEO-united. Statista. 2016 Mar. Marktanteile führender Suchmaschinen in Deutschland in den Jahren 2014 bis 2016 URL: https://de.statista.com/statistik/daten/studie/167841/umfrage/marktanteile-ausgewaehlter-suchmaschinen-in-deutschland/ [accessed 2020-03-04]
33. 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]
34. Bizzi I, Ghezzi P, Paudyal P. Health information quality of websites on periodontology. J Clin Periodontol 2017 Mar;44(3):308-314. [doi: 10.1111/jcpe.12668] [Medline: 28005268]

Abbreviations

DGZMK: German Society of Dentistry and Oral Medicine
ENTREQ: Enhancing Transparency in Reporting the Synthesis of Qualitative Research
IQR: interquartile range
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SSO: Swiss Dental Association

Edited by G Eysenbach; submitted 29.11.19; peer-reviewed by N Aly, T Cruvinel; comments to author 27.01.20; revised version received 31.01.20; accepted 16.02.20; published 15.04.20.

Please cite as:
Kanzow P, Büttcher AF, Wiegand A, Schwendicke F
Quality of Information Regarding Repair Restorations on Dentist Websites: Systematic Search and Analysis
J Med Internet Res 2020;22(4):e17250
URL: http://www.jmir.org/2020/4/e17250/
doi:10.2196/17250
PMID:32062595

Original Paper

Fiction, Falsehoods, and Few Facts: Cross-Sectional Study on the Content-Related Quality of Atopic Eczema-Related Videos on YouTube

Simon M Mueller[1], MD; Valentina N S Hongler[1], BMed; Pierre Jungo[1], MD; Lucian Cajacob[1], MD; Simon Schwegler[1], MD; Esther H Steveling[2], MD; Zita-Rose Manjaly Thomas[2], MD, PhD; Oliver Fuchs[3], MD, PhD; Alexander Navarini[1], MD, PhD; Kathrin Scherer[2], MD; Oliver Brandt[1,3], MD

[1]Department of Dermatology, University Hospital Basel, Basel, Switzerland

[2]Allergy Unit, Department of Dermatology, University Hospital Basel, Basel, Switzerland

[3]Allergy Unit, Division of Respiratory Medicine, Department of Paediatrics, Inselspital, University of Bern, Bern, Switzerland

Corresponding Author:
Simon M Mueller, MD
Department of Dermatology
University Hospital Basel
Petersgraben 4
Basel, 4031
Switzerland
Phone: 41 61 328 69 64
Email: Simon.Mueller@usb.ch

Abstract

Background: In recent years, YouTube has become a recognized source of medical information for health care consumers. Although YouTube has advantages in this context, there are potential dangers as videos may contain nonscientific, misleading, or even harmful information.

Objective: As little is known about YouTube as a source of information on atopic dermatitis (AD), we investigated the content-related quality of AD videos and their perception among YouTube users.

Methods: The quality of the 100 most viewed AD videos was assessed by using the Global Quality Scale (GQS) and the DISCERN instrument. Videos were classified as "useful," "misleading," and "potentially harmful," and the correlations of viewers' ratings (likes) with the GQS and DISCERN scores were assessed.

Results: Among the 100 videos, 68.0% (68/100) and 62.0% (62/100) were of poor and very poor scientific quality, respectively. Additionally, 32.0% (32/100) of the videos were classified as useful, 48.0% (48/100) were classified as misleading, and 34.0% (34/100) were classified as potentially harmful. Viewers' ratings did not correlate with the GQS and DISCERN scores. Overall, 50.0% (50/100) of the videos were posted by private individuals and promoters of complementary/alternative treatments, 42.0% (42/100) by therapeutical advertisers, and only 8.0% (8/100) by nonprofit organizations/universities.

Conclusions: Our study demonstrated that two-thirds of the videos analyzed were below acceptable medical quality standards and that many videos were disseminating misleading or even dangerous content. Subjective and anecdotal content was overrepresented, and viewers did not appear to be able to distinguish between high- and low-quality videos. Health promotion strategies by professional medical organizations are needed to improve their presence and visibility on YouTube.

(J Med Internet Res 2020;22(4):e15599) doi:10.2196/15599

KEYWORDS

YouTube; social media; videos; atopic eczema; atopic dermatitis; quality assessment; Global Quality Scale; DISCERN

Introduction

Atopic dermatitis (AD), also known as atopic eczema, is the most common chronic inflammatory skin disease during childhood [1] and is characterized by recurrent itchy eczematous lesions [2]. Globally, up to 20% of children and 3% of adults are affected by this condition [2]. Patients experience not only the disease itself (eg, sleep deprivation due to night-time

itching), but also the stigma associated with its visibility to others. As a result, patients with AD are often frustrated and embarrassed, which may lead to stress and perpetuation of an itch-scratch cycle, eventually worsening the condition [3]. Consequently, patients and their families frequently report a low health-related quality of life [3,4]. Considering that a high proportion of patients with skin diseases show high interest in online searches [5], patients with AD may be particularly tempted to seek information about their condition from the internet or social media for the aforementioned reasons. YouTube is such a video-based social media platform that allows users to communicate and share their disease burden and individual experience through videos, which can receive comments [6]. It is currently the second most accessed website worldwide [7], attracting approximately one-third of all internet users [8]. YouTube has an increasing number of videos containing medical information [9-11], which may disseminate inaccurate details owing to the lack of quality control and may cause severe health consequences [9,12]. Although several studies have demonstrated that YouTube is highly accessed as a source of information on dermatological topics [13-22], little is known about YouTube as a source of information on AD. This may be surprising given the high prevalence of this condition and the assumption that the populations most affected (children, adolescents, and their parents) typically belong to the age group of "digital natives" (persons born from 1980 onward) [23]. A previous publication indicated that YouTube videos are indeed highly accessed, commented, and shared in connection with AD and that many of the videos posted provide misleading guidance [24]. However, to the best of our knowledge, there have been no in-depth analyses of the topics posted, quality of medical content, upload sources, and ratings by viewers. Therefore, the objectives of this study were as follows: (1) identify the upload sources, common topics, and YouTube categories of the 100 most viewed videos; (2) investigate the content-related quality of YouTube videos as a source of information on AD by applying two different score instruments; (3) correlate viewers' ratings with our quality assessment findings; and (4) point out strategies for interventions that increase the quality of AD videos and medical content generally uploaded to YouTube and other social media platforms.

Methods

Data Collection

In this cross-sectional study, YouTube was searched on April 18, 2018, using the term "atopic eczema" with the following filter settings: "English UK" (language), "United Kingdom" (country), and "Video" (type). Thereafter, videos were sorted by view count, and the duration, upload date, title, URL, view count, uploader identity, likes/dislikes, category, license type, and origin country were recorded and analyzed. In August 2018, videos were reviewed in more detail. Although categories were re-evaluated and refined, we did not update or change numbers and video rankings. We excluded videos having poor technical quality, duplicate videos, and nonEnglish videos, unless English subtitles were displayed. Similar to many previous studies about YouTube, we limited our analysis to the 100 most viewed videos (Multimedia Appendix 1), as it has been demonstrated that videos rated lower than this have insignificant view counts with a minor impact on the analysis [22,25-28].

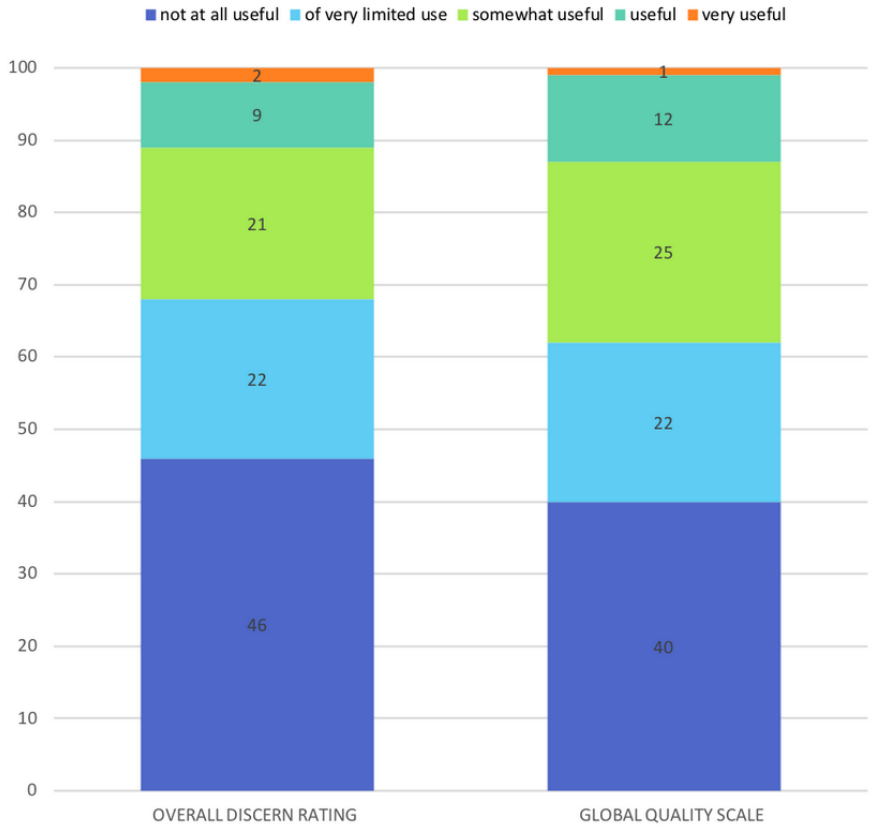
Creation of Content Topics

After gathering data, videos were categorized into themes according to the actual content presented irrespective of the titles. The following 12 categories were created: "education," "topical treatment," "systemic treatment," "complementary and alternative medicine (CAM)," "food, nutrition, and diet," "bathing/wet wrapping," "ultraviolet (UV) treatment," "irritants (clothing, sweat, heat, and allergens)," "stress prevention," "therapeutical advertisement," "unclear topic," and "other topics." Therapeutical advertising included both nonpharmaceutical and pharmaceutical advertising.

Quality Assessment

The two frequently used quality assessments DISCERN (name of the instrument) and Global Quality Scale (GQS) were applied to evaluate the medical quality of the posted videos [25,29-33]. DISCERN measures a video's quality of information about treatment choices. It includes 16 questions addressing the reliability (questions 1-8), quality of health information (questions 9-15), and overall quality (question 16) of the videos. To each of these questions 1-5 points were assigned [31,34,35] (Multimedia Appendix 2). The GQS is based on a 5-point scale measuring the content-related quality of a video, its flow, and its value as a source of information for medical laypeople [14,25,29,30,36,37] (Multimedia Appendix 3). The classification shown in Figure 1 (1=not at all useful, 2=of very limited use, 3=somewhat useful, 4=useful, and 5=very useful) was adopted from the report by Qi et al [14]. As both DISCERN and the GQS are based on a 5-point scale, the same classification was used. In addition, videos were classified as "useful" or "misleading," where useful videos had scientifically correct and accurate information about any aspect of the disease and misleading videos had scientifically unproven or inaccurate information according to currently available scientific evidence (eg, unfounded claims about pathogenesis and treatment with unproven dietary, herbal, or alternative therapy or negative portrayal of evidence-based treatment) [30,32]. If a video could not be assigned to one of these groups, it was automatically classified as "neither nor." Further, misleading videos were subdivided into "potentially harmful" and "not harmful." To assess the role of background music in terms of video popularity, we performed correlation calculations with the numbers of likes and views. Regarding all features, in case of different assessments by the analyzing dermatologists, the corresponding videos were re-evaluated and arbitrated by the principal investigator.

Figure 1. Comparison of the overall quality and usefulness of the identified videos (n=100) for patients seeking health-related advice, using the DISCERN instrument and the Global Quality Scale. The overall rating of DISCERN refers to question 16 of this tool, which is based on a 5-point scale (Multimedia Appendix 2).



Statistical Analysis

Analyses involving descriptive statistics and Spearman rank correlation coefficients for the numbers of likes/dislikes and the DISCERN and GQS scores were performed using IBM SPSS statistics, version 22.0 (IBM Corp, Armonk, New York). For evaluation of the degree of agreement among reviewers (seven experienced dermatologists listed as authors, except for ZRM, OF, and AN) regarding the videos, Cohen κ coefficients and intraclass correlation coefficients were calculated.

Results

View Count, Duration, Upload Sources, Categories, and Topics

The 100 most viewed videos garnered a total of 8,527,624 views and had a total duration of 7 hours 52 minutes (average duration of 4 minutes 44 seconds per video). All videos had a standard YouTube license, which allows the use of the videos only after obtaining permission from the author [38]. Most clips were uploaded from the United States (43.0%, 43/100), followed by unknown countries (14.0%, 14/100) and India (12.0%, 12/100) (Multimedia Appendix 4). The category most often used by the upload source was people & blogs (34.0%, 34/100), followed by education (24.0%, 24/100), how-to & style (22.0%, 22/100), science & technology (11.0%, 11/100), nonprofits & activism (7.0%, 7/100), entertainment (1.0%, 1/100), and news & politics (1.0%, 1/100) (Multimedia Appendix 5A). The most frequent topic (a video can cover more than one topic) was topical treatment (55.0%, 55/100), followed by education (38.0%, 38/100; eg, pathogenesis, risk factors, and instructions for wet-wrapping techniques), food, nutrition, and diet (28.0%, 28/100), CAM (26.0%, 26/100), bathing/wet wrapping (26.0%, 26/100), systemic treatment (15.0%, 15/100), unclear topic (15.0%, 15/100), other topics (13.0%, 13/100), irritants (12.0%, 12/100; clothing, sweat, heat, and allergens), therapeutical advertisement (10.0%, 10/100), UV treatment (5.0%, 5/100), and stress prevention (4.0%, 4/100) (Multimedia Appendix 5B).

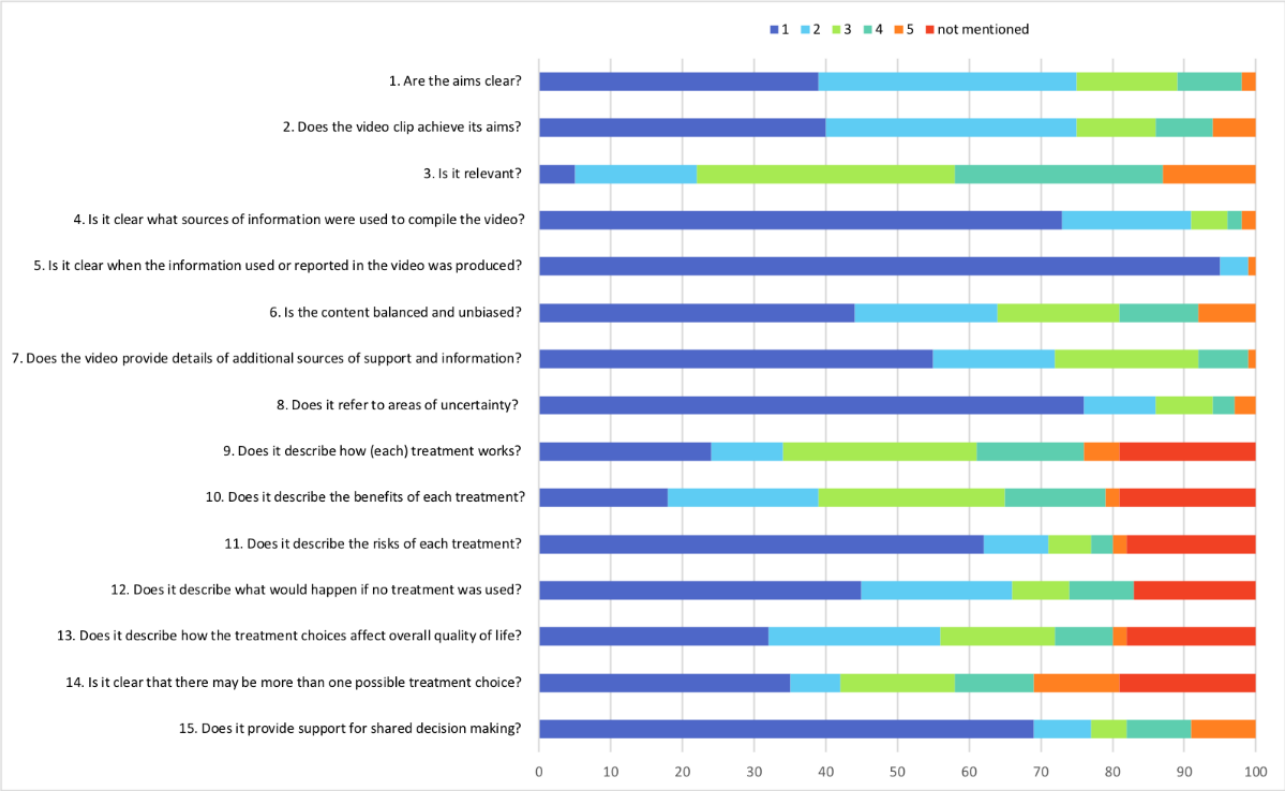
Regarding information presenters and upload sources of videos on AD (a video can cover more than one category), the most common were YouTube users with personal experiences and promoters of CAM (50.0%, 50/100), followed by alleged patients (49.0%, 49/100), therapeutical advertisers (42.0%, 42/100), dermatologists or scientists (32.0%, 32/100), health information websites and eczema associations (21.0%, 21/100), nonprofit organizations and universities (8.0%, 8/100), and television/media (6.0%, 6/100) (Multimedia Appendix 6). Videos featuring only patients had approximately 421,181 views, whereas those featuring only dermatologists/scientists had 120,736 views. Background music was used in 51.0% (51/100) of all videos, and these had 823,924 views and 3,481 likes. Background music-containing videos accounted for 53.01% (823,924/1,554,155) of the total view count and had 42.40% (3,481/8,210) of all likes. For none of the videos, YouTube statistics were accessible, whereas the comment function was disabled in only 6.0% (6/100) of the videos.

Quality Assessment and Correlation With Likes/Dislikes

Overall, 32.0% (32/100) of the videos were classified as useful, 48.0% (48/100) were classified as misleading, and 20.0% (20/100) were classified as neither nor. Among those classified as misleading, 71% (34/48) were considered potentially harmful because of possible mechanical or chemical injury or inadequate dietary recommendations and 29% (14/48) were considered not harmful (Multimedia Appendix 7A). Regarding nonscientific and potentially harmful videos, 37% (15/41) suggested unnecessary diets, 19% (8/41) discredited conventional medicine and physician advice, 15% (6/41) made unrealistic promises, 17% (7/41) did not limit the use of UV treatments, topical steroids, antibiotics, or cold packs, 7% (3/41) promoted the use of unscientific and potentially harmful procedures, and 5% (2/41) recommended topical use of potentially harmful substances (Multimedia Appendix 7B). The mean durations of useful and misleading videos were 6 minutes 2 seconds and 3 minutes 39 seconds, respectively. Regarding the view count, we excluded the most viewed video from further analysis, as it was a pharmaceutical advertisement with disabled like/dislike and comment functions, accounting for approximately 81.78% (6,973,469/8,527,624) of all views. In total, misleading videos had 870,012 views, including 789,073 views for potentially dangerous content, whereas useful videos had only 528,352 views, resulting in a misleading/useful video ratio of 1.65 (870,012/528,352).

The ratings achieved with DISCERN and the GQS were consistent, yielding the categorizations shown in Figure 1. With regard to the quality of videos, the mean overall DISCERN and GQS scores were 1.99 (SD 1.09) and 2.12 (SD 1.09), respectively, and therefore, quality was generally low, as the scores range from 1 to 5, with 5 being the highest value. Detailed analysis of the DISCERN questionnaire revealed that the major shortcomings were lack of details about the source of the presented information, unaddressed areas of uncertainty as well as risks of the presented therapy, and failure to recommend shared decision-making (Figure 2). The intraclass correlation coefficients calculated for DISCERN and the GQS were 0.97 and 0.95, respectively, indicating excellent interrater reliability when using these tools.

Figure 2. Detailed analysis of the DISCERN questionnaire (questions 1-15) used to rate the identified videos (n=100). The instrument uses a 5-point scale ranging from no to yes, where 1 means “no” (the quality criterion has not been fulfilled at all), 2-4 means “partially” (the quality criterion has been fulfilled to some extent), and 5 means “yes” (the quality criterion has been completely fulfilled).



The total numbers of likes and dislikes for the videos were 8210 and 737, respectively, yielding a like/dislike ratio of 11.14 (8210/737). In three videos, the like/dislike function was disabled. Viewers’ ratings did not correlate with the DISCERN and GQS scores (Spearman correlation ρ=0.12, *P*=.25 and ρ=0.17, *P*=.08, respectively), indicating that viewers were unable to adequately rate the quality of the videos.

Discussion

Overview

Patients with AD have a high motivation to conduct online searches, and this usually correlates with disease burden [39,40]. YouTube offers a wide range of dermatology-related videos [13], and little is known so far about the content, upload sources, topics, and scientific quality of these videos. Hitherto, no data

exist on whether viewers' ratings correlate with the quality of the medical information provided.

This study revealed that nearly half of the 100 most viewed AD-related videos on YouTube, with more than 8.5 million views, disseminated misleading information. This finding is consistent with the findings of our previously published study [22] and several other surveys investigating the content-related quality of health-related YouTube videos [12,41,42].

Comparison With Prior Work

Freemyer et al [24] thoroughly analyzed 128 videos that were gathered by screening the first two result pages of each of the five different search terms (AD, eczema, eczema tips, eczema cure, and eczema treatment) they used for AD. In contrast to our study, they distinguished between "useful," "useful-personal," "misleading," and "misleading-personal" information but not between "harmful" and "neutral" information, and therefore, they found that only 34.4% (44/128) of the videos were misleading. A reason for the markedly lower percentage of misleading videos may be the different search terms used to identify the respective videos. Although a study from 2016 showed that AD is the most commonly used term among scientific publications and studies [43], patients use the nonspecific term eczema to refer to AD [44]. As the purpose of this study was to investigate the video sample patients would come across when seeking information online, we decided to use the term atopic eczema to search YouTube. In contrast to the approach by Freemyer et al [24], we additionally sorted the search results by view count to obtain a video sample of high relevance owing to large viewership. Another aspect that makes direct comparison with our study nearly impossible is the absence of quality assessment tools in the mentioned study.

Videos of Good or Excellent Quality are Rare

According to our evaluation using DISCERN and the GQS, only 11.0% (11/100) and 13.0% (13/100) of the videos, respectively, were of good or excellent quality with unbiased evidence-based or at least science-based information. Similar results were obtained in our recently published study investigating YouTube videos on psoriasis [22]. Additionally, data by Freemyer et al [24] demonstrated that health care organizations, universities, and dermatologists are clearly underrepresented on YouTube in the context of AD.

YouTube Users Prefer Low-Quality Videos Over High-Quality Videos

Factual and informative videos are rare, and they lack popularity, as illustrated by the lower number of likes compared to those of poor-quality videos. Our previous study and many other dermatological and nondermatological studies have come to similar conclusions regarding this phenomenon [14,22,41,45]. It remains unclear why the general population tends to prefer low-quality videos over high-quality videos. Biggs et al [46] suggested that the duration of the videos might be a relevant issue. Similar to the findings in their study (mean useful video duration: 14 minutes 47 seconds; mean misleading video duration: 4 minutes 37 seconds), we found that longer videos were associated with higher quality. In our study, useful videos were nearly twice as long as misleading videos (mean useful video duration: 6 minutes 2 seconds; mean misleading video duration: 3 minutes 39 seconds), which could indeed be enough time to put off viewers looking for quick answers. Additional reasons could be that viewers specifically search for alternative content, which is usually present in videos of lower quality, and that academic videos may be less sensationalized and thus less attractive to laypersons [22,37].

Patients Prefer Advice From Fellow Patients Rather Than Physicians

Interestingly, one or more patients appeared in 49.0% (49/100) of all videos regardless of whether the video was produced by them. Of note, only 32.0% (32/100) of all videos featured a dermatologist or scientist. Analyzing the separate view count, we found that videos featuring only patients had nearly 3.5 times more views as compared with videos featuring only dermatologists/scientists (only patients: 421,181 views; only dermatologists/scientists: 120,736 views). Smith et al [47] showed that 70% of patients with chronic conditions reported experiencing one or more health care–related frustrations, such as the feeling of not being understood or taken seriously (ie, lack of empathy by the physician). Therefore, patients might follow the advice of fellow patients with the same disease rather than the instructions of health care professionals. This decision is obviously frequently made despite poor video quality.

Potentially Harmful Content

In our study, 34.0% (34/100) of the analyzed videos contained potentially harmful information. For instance, patients with AD were encouraged to not only follow unnecessary diets, such as avoid dairy or gluten, but also use topical treatments (eg, cold packs) and phototherapies without any detailed information about the duration of application or potential risks. Furthermore, conventional medicine and physician advice were discredited in various ways, while promising a fast and easy cure at the same time with the suggested therapies. Such advice was often provided by therapeutical commercials that tried to sell their products, as well as by patients who reported personal negative long-term experiences with Western conventional medicine and eventually found salvation in alternative treatments. Interestingly, these testimonials were frequently uploaded from India (accounting for 12.0% [12/100] of all videos investigated) and typically showed parents being enthusiastic about a traditional practice that healed their children who had atopic eczema.

Complementary and Alternative Medicine is a Hot Topic for YouTube Users

The confidence and high interest in CAM described above are comprehensible considering that 70%-80% of India's population is dependent on traditional systems for financing health care [48]. However, this enthusiasm about CAM appears to be shared by the rest of the world as well, because 50.0% (50/100) of the AD-related videos were uploaded by CAM promoters and YouTube users sharing their personal experiences. Reddy et al [37] recently reported the similarly dominant and controversial role of CAM in YouTube videos on food allergy. This may not be surprising considering that the question of food allergies underlying or aggravating AD is important among most affected

patients or their parents (particularly in infancy) but is often difficult to answer for clinicians.

Extensive Advertising

Another astonishing fact is that therapeutical advertisers uploaded 42.0% (42/100) of all videos investigated, although the advertised product did not necessarily appear in the video. As YouTube has become an important marketing platform over the last decade, the advertisement of products with referral links has increased in popularity [49]. Small- and medium-sized companies, in particular, can benefit from social media marketing, as a large number of potential customers can be reached at a relatively low cost [50]. This trend can be an issue for online health information seekers trying to find valid guidance, as commercial advertisements are being disguised as supposedly harmless referral links.

Along with a lack of quality recognition and an intentional search for unconventional content, a lack of entertainment may explain why fewer people are attracted to high-quality videos than to low-quality videos [22]. When examining view count and number of likes with background music as a possible clickbait factor, we found that the majority of likes were generated by videos without background music, suggesting that background music does not markedly increase the popularity of videos and is thus not a suitable strategy to attract people to high-quality videos.

A greater understanding of what kinds of measures attract a high number of viewers is pivotal for enabling health care organizations, universities, and dermatologists to appeal to more patients.

Potential Interventions to Improve the Presence and Visibility of High-Quality Videos

As the need for content-related high-quality online material is indisputable, dermatology associations, AD self-help organizations, and universities should be encouraged to produce and provide more videos containing evidence-based easy-to-understand information about pathophysiology, clinical manifestations, and therapies for patients with AD and their families, and at the same time, they need to highlight the dangers of non–evidence-based treatment options [22]. Prior to the release of videos, measures for quality assessment (eg, the DISCERN tool) should be applied to ensure high-quality content. In the long term, this approach could result in the neutralization of widely available misleading information on YouTube.

Additionally, professionals need to be aware of the fact that health information seekers mostly choose results that appear on the first page of the search engine [51] when looking for medical content online, and therefore, need to make sure that videos appear there. Investing in consulting services that improve the placement of information in search engines on a website (search engine optimization) could be more efficient than producing additional videos that are placed somewhere beyond the second results page [52]. In addition, it would be advantageous if both social media and search engine providers are encouraged to cooperate with dermatology associations and universities to position medically accurate information near the top of the results page.

Moreover, the use of a Creative Commons license instead of a standard YouTube license could lead to more visits, as has been shown in our previous study [22]. This kind of license allows individuals to use content for their own video clips [53,54]. However, it is advisable to choose a license type that forbids others to change the content in any way or to use it commercially, as misuse that can potentially result in misleading information should be prevented.

Finally, the comment section of misleading videos could be used to insert cross-links to guide viewers to trustworthy videos or websites with evidence-based information. For this approach to work, the comment function must not be disabled by the producer of the video in question, and therefore, unfortunately, this is not a feasible intervention for every video on YouTube.

Strengths and Limitations

Despite the unmistakable strengths, such as the comprehensive analyses of a high number of videos and the application of two different scoring tools (GQS and DISCERN), this study has some limitations. Although we performed comprehensive analyses, we did not evaluate the comments posted by viewers, which may contain the viewers' true opinions about a particular video and thus its popularity. Additionally, we did not evaluate the visual design of the videos using a specific tool solely developed for this purpose, making it possible that the impact of the videos on the viewers was not fully captured in this respect. Furthermore, our study is specific for an arbitrary period and therefore only includes the material available at that time. Finally, additional factors, such as instructional design and educational value, contributing to the overall quality of the videos were not assessed in this study. These aspects may certainly be worth investigating in future studies with novel assessment tools. In addition, we would like to mention that besides the assessment tools used in this study, there are other valid methods to measure the information accuracy and content quality of YouTube videos [55].

Conclusion

This study showed that two-thirds of the videos on AD analyzed were below acceptable medical quality standards and often contained misleading or even harmful information about this common disease. The fact that users tended to rate low-quality videos better than high-quality videos suggests that the majority of users are unable to distinguish between medically credible information and false information. This shows that the numbers of views and likes do not reflect the medical quality of videos. To combat this phenomenon, it is crucial that future studies investigate user motivation for such behavior in order to help medical professionals to develop approaches that contribute to improving medical information on this powerful platform.

Conflicts of Interest

None declared.

Authors' Contributions

SMM and VNSH contributed equally as first authors. OB and KS contributed equally as last authors.

Multimedia Appendix 1
Links to the YouTube videos.
[XLSX File (Microsoft Excel File), 19 KB - jmir_v22i4e15599_app1.xlsx]

Multimedia Appendix 2
DISCERN instrument.
[PDF File (Adobe PDF File), 33 KB - jmir_v22i4e15599_app2.pdf]

Multimedia Appendix 3
Global Quality Scale (GQS).
[PDF File (Adobe PDF File), 31 KB - jmir_v22i4e15599_app3.pdf]

Multimedia Appendix 4
Distribution of the videos by country of origin.
[PDF File (Adobe PDF File), 22 KB - jmir_v22i4e15599_app4.pdf]

Multimedia Appendix 5
(A) YouTube categories, in which the analyzed videos were posted (n=100); (B) Topics presented in the videos (note: a video clip can cover more than one topic).
[PDF File (Adobe PDF File), 34 KB - jmir_v22i4e15599_app5.pdf]

Multimedia Appendix 6
Distribution of information according to presenters and upload sources (multiple categories may apply to one video). CAM, complementary and alternative medicine.
[PDF File (Adobe PDF File), 25 KB - jmir_v22i4e15599_app6.pdf]

Multimedia Appendix 7
(A) Classification of the videos (n=100) into “useful,” “misleading,” and “neither nor,” with further subdivision of “misleading” into "potentially harmful” and “not harmful” (not harmful = 48 misleading videos minus potentially harmful videos [14, 29%]) shown as a bar chart (in percentage); (B) Specifications and examples of nonscientific and potentially harmful videos.
[PDF File (Adobe PDF File), 26 KB - jmir_v22i4e15599_app7.pdf]

References

1. Flohr C, Mann J. New insights into the epidemiology of childhood atopic dermatitis. Allergy 2014 Jan;69(1):3-16. [doi: 10.1111/all.12270] [Medline: 24417229]
2. Nutten S. Atopic dermatitis: global epidemiology and risk factors. Ann Nutr Metab 2015;66 Suppl 1:8-16 [FREE Full text] [doi: 10.1159/000370220] [Medline: 25925336]
3. Carroll CL, Balkrishnan R, Feldman SR, Fleischer AB, Manuel JC. The burden of atopic dermatitis: impact on the patient, family, and society. Pediatr Dermatol 2005;22(3):192-199. [doi: 10.1111/j.1525-1470.2005.22303.x] [Medline: 15916563]
4. Lewis-Jones S. Quality of life and childhood atopic dermatitis: the misery of living with childhood eczema. Int J Clin Pract 2006 Aug;60(8):984-992. [doi: 10.1111/j.1742-1241.2006.01047.x] [Medline: 16893440]
5. Whitsitt J, Karimkhani C, Boyers LN, Lott JP, Dellavalle RP. Comparing burden of dermatologic disease to search interest on google trends. Dermatol Online J 2015 Jan;21(1). [Medline: 25612125]
6. Samuel N, Alotaibi NM, Lozano AM. YouTube as a Source of Information on Neurosurgery. World Neurosurg 2017 Sep;105:394-398. [doi: 10.1016/j.wneu.2017.05.111] [Medline: 28599904]
7. Alexa. Website traffic analysis URL: https://www.alexa.com/siteinfo/youtube.com [accessed 2018-08-29]
8. Omnicore. YouTube by the Numbers: Stats, Demographics & Fun Facts URL: https://www.omnicoreagency.com/youtube- [accessed 2018-08-29]
9. Raikos A, Waidyasekara P. How useful is YouTube in learning heart anatomy? Anat Sci Educ 2014 Feb;7(1):12-18. [doi: 10.1002/ase.1361] [Medline: 23564745]

10. 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]
11. Karlsen R, Borrás Morell JE, Fernández Luque L, Traver Salcedo V. A domain-based approach for retrieving trustworthy health videos from YouTube. Stud Health Technol Inform 2013;192:1008. [Medline: 23920782]
12. Abedin T, Ahmed S, Al MM, Ahmed SW, Newaz S, Rumana N, et al. YouTube as a source of useful information on diabetes foot care. Diabetes Res Clin Pract 2015 Oct;110(1):e1-e4. [doi: 10.1016/j.diabres.2015.08.003] [Medline: 26303266]
13. Boyers LN, Quest T, Karimkhani C, Connett J, Dellavalle RP. Dermatology on YouTube. Dermatol Online J 2014 Jun 15;20(6) [FREE Full text] [Medline: 24945641]
14. 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). [Medline: 28329562]
15. Lenczowski E, Dahiya M. Psoriasis and the Digital Landscape: YouTube as an Information Source for Patients and Medical Professionals. J Clin Aesthet Dermatol 2018 Mar;11(3):36-38 [FREE Full text] [Medline: 29606999]
16. Qureshi N, Lowenstein EJ. The role of nutrition in acne pathogenesis: YouTube as a reflection of current popular thought. Skinmed 2011;9(5):279-280. [Medline: 22165041]
17. Pithadia DJ, Reynolds KA, Lee EB, Wu JJ. A cross-sectional study of YouTube videos as a source of patient information about topical psoriasis therapies. J Dermatolog Treat 2019 May 18:1-4. [doi: 10.1080/09546634.2019.1597247] [Medline: 30885022]
18. Pithadia DJ, Reynolds KA, Lee EB, Wu JJ. A cross-sectional study of YouTube videos as a source of patient information about phototherapy and excimer laser for psoriasis. J Dermatolog Treat 2019 Apr 09:1-4. [doi: 10.1080/09546634.2019.1605144] [Medline: 30963794]
19. St Claire KM, Rietcheck HR, Patel RR, Dunnick C, Dellavalle RP. Dermatology on YouTube - an update and analysis of new trends. Dermatol Online J 2018 Dec 15;24(12) [FREE Full text] [Medline: 30677792]
20. Reynolds KA, Pithadia DJ, Lee EB, Wu JJ. A cross-sectional study of YouTube videos about psoriasis biologics. Int J Dermatol 2019 May;58(3):e61-e62. [doi: 10.1111/ijd.14317] [Medline: 30478827]
21. DeBord LC, Patel V, Braun TL, Dao H. Social media in dermatology: clinical relevance, academic value, and trends across platforms. J Dermatolog Treat 2019 Aug;30(5):511-518. [doi: 10.1080/09546634.2018.1530444] [Medline: 30265614]
22. Mueller SM, Jungo P, Cajacob L, Schwegler S, Itin P, Brandt O. The Absence of Evidence is Evidence of Non-Sense: Cross-Sectional Study on the Quality of Psoriasis-Related Videos on YouTube and Their Reception by Health Seekers. J Med Internet Res 2019 Jan 16;21(1):e11935 [FREE Full text] [doi: 10.2196/11935] [Medline: 30664460]
23. Zaphiris P, Ioannou A. Learning and Collaboration Technologies. LearningTeaching. 2018 Presented at: 5th International Conference, LCT 2018, Held as Part of HCI International 2018; July 15-20, 2018; Las Vegas, NV, USA. [doi: 10.1007/978-3-319-91152-6]
24. Freemyer B, Drozd B, Suarez A. A cross-sectional study of YouTube videos about atopic dermatitis. J Am Acad Dermatol 2018 Mar;78(3):612-613. [doi: 10.1016/j.jaad.2017.09.005] [Medline: 29447681]
25. Oremule B, Patel A, Orekoya O, Advani R, Bondin D. Quality and Reliability of YouTube Videos as a Source of Patient Information on Rhinoplasty. JAMA Otolaryngol Head Neck Surg 2019 Mar 01;145(3):282-283. [doi: 10.1001/jamaoto.2018.3723] [Medline: 30605207]
26. Basch CH, Hillyer GC, Garcia P, Basch CE. Content of widely viewed YouTube videos about celiac disease. Public Health 2019 Feb;167:147-151. [doi: 10.1016/j.puhe.2018.11.004] [Medline: 30682698]
27. Ferhatoglu MF, Kartal A, Ekici U, Gurkan A. Evaluation of the Reliability, Utility, and Quality of the Information in Sleeve Gastrectomy Videos Shared on Open Access Video Sharing Platform YouTube. Obes Surg 2019 May;29(5):1477-1484. [doi: 10.1007/s11695-019-03738-2] [Medline: 30706318]
28. Basch CH, Menafro A, Mongiovi J, Hillyer GC, Basch CE. A Content Analysis of YouTube™ Videos Related to Prostate Cancer. Am J Mens Health 2017 Jan;11(1):154-157 [FREE Full text] [doi: 10.1177/1557988316671459] [Medline: 27694554]
29. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
30. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
31. 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]
32. 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]
33. Esen E, Aslan M, Sonbahar B, Kerimoğlu RS. YouTube English videos as a source of information on breast self-examination. Breast Cancer Res Treat 2019 Feb;173(3):629-635. [doi: 10.1007/s10549-018-5044-z] [Medline: 30443880]
34. Allam A, Schulz PJ, Krauthammer M. Toward automated assessment of health Web page quality using the DISCERN instrument. J Am Med Inform Assoc 2017 May 01;24(3):481-487. [doi: 10.1093/jamia/ocw140] [Medline: 27707819]

35. ReFaey K, Tripathi S, Yoon JW, Justice J, Kerezoudis P, Parney IF, et al. The reliability of YouTube videos in patients education for Glioblastoma Treatment. J Clin Neurosci 2018 Sep;55:1-4. [doi: 10.1016/j.jocn.2018.07.001] [Medline: 30075980]
36. Delli K, Livas C, Vissink A, Spijkervet FK. Is YouTube useful as a source of information for Sjögren's syndrome? Oral Dis 2016 Apr;22(3):196-201. [doi: 10.1111/odi.12404] [Medline: 26602325]
37. Reddy K, Kearns M, Alvarez-Arango S, Carrillo-Martin I, Cuervo-Pardo N, Cuervo-Pardo L, et al. YouTube and food allergy: An appraisal of the educational quality of information. Pediatr Allergy Immunol 2018 Jun;29(4):410-416. [doi: 10.1111/pai.12885] [Medline: 29512839]
38. Wondershare. Standard YouTube License vs. Creative Commons URL: https://tinyurl.com/r692rad [accessed 2019-05-05]
39. Khanna R, Shifrin N, Nektalova T, Goldenberg G. Diet and dermatology: Google search results for acne, psoriasis, and eczema. Cutis 2018 Jul;102(1):44;46;48. [Medline: 30138495]
40. Hay RJ, Johns NE, Williams HC, Bolliger IW, Dellavalle RP, Margolis DJ, et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. J Invest Dermatol 2014 Jun;134(6):1527-1534 [FREE Full text] [doi: 10.1038/jid.2013.446] [Medline: 24166134]
41. Lee JS, Seo HS, Hong TH. YouTube as a source of patient information on gallstone disease. World J Gastroenterol 2014 Apr 14;20(14):4066-4070 [FREE Full text] [doi: 10.3748/wjg.v20.i14.4066] [Medline: 24744597]
42. López-Jornet P, Pons-Fuster E, Ruiz-Roca JA. YOUTUBE videos on oral care of the organ or hematopoietic stem cell transplant patients. Support Care Cancer 2017 Apr 22;25(4):1097-1101. [doi: 10.1007/s00520-016-3497-3] [Medline: 27878369]
43. Kantor R, Thyssen JP, Paller AS, Silverberg JI. Atopic dermatitis, atopic eczema, or eczema? A systematic review, meta-analysis, and recommendation for uniform use of 'atopic dermatitis'. Allergy 2016 Oct;71(10):1480-1485 [FREE Full text] [doi: 10.1111/all.12982] [Medline: 27392131]
44. Hanifin JM. Atopic dermatitis nomenclature variants can impede harmonization. J Invest Dermatol 2012 Feb;132(2):472-3; author reply 473 [FREE Full text] [doi: 10.1038/jid.2011.292] [Medline: 21938010]
45. 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]
46. Biggs TC, Bird JH, Harries PG, Salib RJ. YouTube as a source of information on rhinosinusitis: the good, the bad and the ugly. J Laryngol Otol 2013 Aug;127(8):749-754. [doi: 10.1017/S0022215113001473] [Medline: 23866821]
47. Smith ML, Bergeron CD, Adler CH, Patel A, Ahn S, Towne SD, et al. Factors associated with healthcare-related frustrations among adults with chronic conditions. Patient Educ Couns 2017 Jun;100(6):1185-1193. [doi: 10.1016/j.pec.2016.12.033] [Medline: 28117194]
48. Gogtay NJ, Bhatt HA, Dalvi SS, Kshirsagar NA. The use and safety of non-allopathic Indian medicines. Drug Saf 2002;25(14):1005-1019. [doi: 10.2165/00002018-200225140-00003] [Medline: 12408732]
49. Schwemmer C, Ziewiecki S. Social Media Sellout: The Increasing Role of Product Promotion on YouTube. Social Media + Society 2018 Aug 14;4(3):205630511878672. [doi: 10.1177/2056305118786720]
50. Karimi S, Naghibi HS. Social media marketing (SMM) strategies for small to medium enterprises (SMEs). International Journal of Information, Business and Management 2015;7(4):86.
51. 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 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
52. Hansen DL, Derry HA, Resnick PJ, Richardson CR. Adolescents searching for health information on the Internet: an observational study. J Med Internet Res 2003 Oct 17;5(4):e25 [FREE Full text] [doi: 10.2196/jmir.5.4.e25] [Medline: 14713653]
53. Carroll MW. Creative Commons and the openness of open access. N Engl J Med 2013 Feb 28;368(9):789-791. [doi: 10.1056/NEJMp1300040] [Medline: 23445090]
54. Creative Commons. URL: https://creativecommons.org/ [accessed 2019-04-30]
55. Drozd B, Couvillon E, Suarez A. Medical YouTube Videos and Methods of Evaluation: Literature Review. JMIR Med Educ 2018 Feb 12;4(1):e3 [FREE Full text] [doi: 10.2196/mededu.8527] [Medline: 29434018]

Abbreviations

AD: atopic dermatitis
CAM: complementary and alternative medicine
GQS: Global Quality Scale
UV: ultraviolet light

Edited by G Eysenbach; submitted 25.07.19; peer-reviewed by LH Lee, T Schopf; comments to author 24.09.19; revised version received 03.10.19; accepted 17.12.19; published 24.04.20.

Please cite as:
Mueller SM, Hongler VNS, Jungo P, Cajacob L, Schwegler S, Steveling EH, Manjaly Thomas ZR, Fuchs O, Navarini A, Scherer K, Brandt O
Fiction, Falsehoods, and Few Facts: Cross-Sectional Study on the Content-Related Quality of Atopic Eczema-Related Videos on YouTube
J Med Internet Res 2020;22(4):e15599
URL: https://www.jmir.org/2020/4/e15599
doi:10.2196/15599
PMID:32329744

Original Paper

Patient Questions and Physician Responses in a Chinese Health Q&A Website: Content Analysis

Ziying Hong[1], MA; Zhaohua Deng[1,2], PhD; Richard Evans[3], PhD; Haiyan Wu[4], PhD

[1]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]School of Journalism and Information Communication, Huazhong University of Science and Technology, Wuhan, China

[3]College of Engineering, Design and Physical Sciences, Brunel University, London, United Kingdom

[4]Undergraduate School of Medical Business, Guangdong Pharmaceutical University, Zhongshan, China

Corresponding Author:
Zhaohua Deng, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
No 13 Hangkong Road, Qiaokou District
Wuhan, 430030
China
Phone: 86 19945036079
Email: zh-deng@hust.edu.cn

Abstract

Background: Since the turn of this century, the internet has become an invaluable resource for people seeking health information and answers to health-related queries. Health question and answer websites have grown in popularity in recent years as a means for patients to obtain health information from medical professionals. For patients suffering from chronic illnesses, it is vital that health care providers become better acquainted with patients' information needs and learn how they express them in text format.

Objective: The aims of this study were to: (1) explore whether patients can accurately and adequately express their information needs on health question and answer websites, (2) identify what types of problems are of most concern to those suffering from chronic illnesses, and (3) determine the relationship between question characteristics and the number of answers received.

Methods: Questions were collected from a leading Chinese health question and answer website called "All questions will be answered" in January 2018. We focused on questions relating to diabetes and hepatitis, including those that were free and those that were financially rewarded. Content analysis was completed on a total of 7068 (diabetes) and 6685 (hepatitis) textual questions. Correlations between the characteristics of questions (number of words per question, value of reward) and the number of answers received were evaluated using linear regression analysis.

Results: The majority of patients are able to accurately express their problem in text format, while some patients may require minor social support. The questions posted were related to three main topics: (1) prevention and examination, (2) diagnosis, and (3) treatment. Patients with diabetes were most concerned with the treatment received, whereas patients with hepatitis focused on the diagnosis results. The number of words per question and the value of the reward were negatively correlated with the number of answers. The number of words per question and the value of the reward were negatively correlated with the number of answers.

Conclusions: This study provides valuable insights into the ability of patients suffering from chronic illnesses to make an understandable request on health question and answer websites. Health topics relating to diabetes and hepatitis were classified to address the health information needs of chronically ill patients. Furthermore, identification of the factors affecting the number of answers received per question can help users of these websites to better frame their questions to obtain more valuable answers.

(J Med Internet Res 2020;22(4):e13071) doi:10.2196/13071

KEYWORDS

health information seeking; question classification

Introduction

Background

Rapid developments in internet capabilities have led to public changes in attitudes toward information seeking. A Pew study conducted in 2012 reported that 59% of adults in the United States searched online for health information, and half of these searchers sought information on behalf of someone else [1]. Health care consumers seek information online relating to topics such as treatment, questions before or after visiting health care providers, or advice about diet and exercise habits [2,3]. Online health information plays a significant role in patients' decision making as to how to manage a health issue or treat an illness [4,5]. During the past decade, health information has been made more readily available via the use of internet-enabled services, including social media sites [6], online communities, forums, and crowdsourcing mechanisms, to satisfy patients' information needs [7].

With developments in information technologies, health-related question and answer (Q&A) websites have grown in popularity [8]. These sites are designed to allow people to ask and obtain responses to questions on a broad range of health-related topics [9], enabling patients to exchange information by posting questions and answers, thus providing a natural form of conversation in seeking information [10]. Obtaining information from health Q&A websites is beneficial for users owing to their relatively low cost (most services are free), quick turnaround time given the broad community participation, and easy build-up of social capital [11]. Three types of online Q&A services currently exist: (1) digital reference services, (2) expert services, and (3) social Q&A services [8].

Online health Q&A websites can be characterized as providing expert-based Q&A services according to the classification proposed by Shah et al [8]. On such websites, answers are provided by a group of health care experts rather than from an open community. The medical professionals are only allowed to answer questions once they have undergone a strict registration process, which requires certification and verification. Another distinguishing feature of expert-based Q&A services is that they often include a pricing system, and are thus generally referred to as a price-based knowledge market [12], such as JustAnswer [13] and PickAnswer [14].

Many popular health Q&A websites such as the "All questions will be answered" [15] service for general patients and the "Baby Tree" service [16] for expectant mothers or those raising a baby are available to both medical professionals and the general public. Online health expert Q&A websites focus on solving the health problems experienced by patients. People raise their health-related questions and receive professional responses such as information on the prevention and treatment of diseases [17] or guidance before and after meeting with a physician [18]. As a result of developments in internet technologies, health Q&A websites have become a convenient communication platform for both health professionals and patients seeking medical information [19]. By critically reviewing the questions posted on health Q&A websites, we can assist health care information providers to better understand the literacy skills and requirements of patients and their information needs.

Numerous models and theories exist to help understand the habits of those seeking health information. Griffin and Neuwirth [20] proposed a model for risk information seeking and processing. Afifi and Weiner [21] advanced the theory of motivated information. Longo [22] proposed a conceptual model to better understand the information-seeking behaviors of patients and consumers. These models delineate the reasons motivating patients to seek information and aim to predict the underlying process. However, these studies do not outline the necessary skills required by patients to complete the information-seeking process. These skills will affect the outcome of searches and subsequent health-related decision making, as health literacy influences health information seeking [23,24]. On health Q&A websites, consumers can express their health concerns in their natural language [25]. Although a patient's ability to communicate their story and visualize the illness being experienced has been less acknowledged in research, this should be an important consideration for online health care environments [26]. In particular, there is limited knowledge as to how patients suffering from chronic illnesses are able to illustrate their situation and the problems they encounter in doing so. Further, explicit requests appear to be more successful than implicit requests [27]. In other words, stating the question explicitly, as a clear question, increases the number of responses received.

To address this issue, we explored the literacy of patients suffering from chronic illnesses in terms of their ability to describe their illness and express their question sufficiently on the online health Q&A website "All questions will be answered."

Research Question 1: Can Patients Suffering From Chronic Illnesses Describe Their Illness Accurately and Phrase Their Concerns Into a Question?

Increased understanding of patients' preferences toward health care information, including how it is presented and delivered, is important for health care providers to determine the types of topics most sought by patients [28]. A review conducted by Ramsey et al [29] indicated that information related to a specific illness or disease is the most common type of information sought by consumers. Medical treatments or procedures [30], health care professionals [31], exercise and diets [32], and symptoms [33] are also frequently searched topics [29]. Patients with different diseases have varying information needs. For patients with cancer, for instance, the most frequently requested information is related to treatment [34]. Kuske et al [35] indicated that diet, complications, exercise, medications, and pharmacological interactions are the most frequently sought topics, which led us to our second research question.

Research Question 2: What Topics Are Most Frequently Sought by Patients Suffering From Chronic Illnesses?

Many factors can potentially affect the quantity of responses received to a question. Teevan et al [36] identified that the punctuation used, number of sentences, and scope significantly

affected the response rate. With reference to "question content" and "questioner characteristics," Liu et al [37] identified 17 extrinsic factors that have a significant effect on a question's response rate, including the posting style and time period. To accelerate the time taken to receive a response from physicians, health consumers may select to pay a fee to attract specialists on health Q&A websites, thereby receiving preference. For patients, questions that receive many answers from health experts indicate highly valuable information. Rafaeli et al [38] identified that economic motivators are associated with the participation of experts on Google Answer. Similarly, the length of a question is considered to influence the quality of the response: questions with fewer sentences receive a greater number of useful responses than those with many sentences [39].

Thus, we assumed that the length of the question and the value of reward affects the number of answers received in online health Q&A websites, leading to our third research question: what is the influence of the value of reward and the length of question on the number of responses received?

Methods

Data Collection

Data were collected from a health Q&A website named "All questions will be answered" [15], one of China's most popular online question and consultation websites. A screenshot of the homepage is shown in Figure 1. We selected this platform since it is widely adopted across mainland China and attracts a broad range of users. The platform was established in 2004, and over 3,000,000 physicians are currently using it to provide services to patients online. All physicians must be certified before they are able to respond to questioners, and more than 85% of them work in tertiary hospitals with rich clinical experience.

The health Q&A website is divided into various sections, separated by department and diseases, which facilitates access to specific questions and answers. Patients can obtain information on diseases, and can access patient tools and support for self-management [40]. Such a platform enabling uncomplicated access to health information is important in the management of chronic health conditions [31]. To simplify the analysis, in this study, we focused on patients with diabetes, which is considered one of the most common chronic diseases worldwide [41]. A large percentage of people with diabetes seek health information online [40]. These patients live with a common chronic disease and need to be able to self-manage and become an expert in their treatment [42]. We also selected to focus on hepatitis because many patients with hepatitis have chronic cases [43] and seek health information, especially surrounding treatment [44]. Therefore, questions were collected under the categories of diabetes and hepatitis, thereby excluding interference from other medical departments.

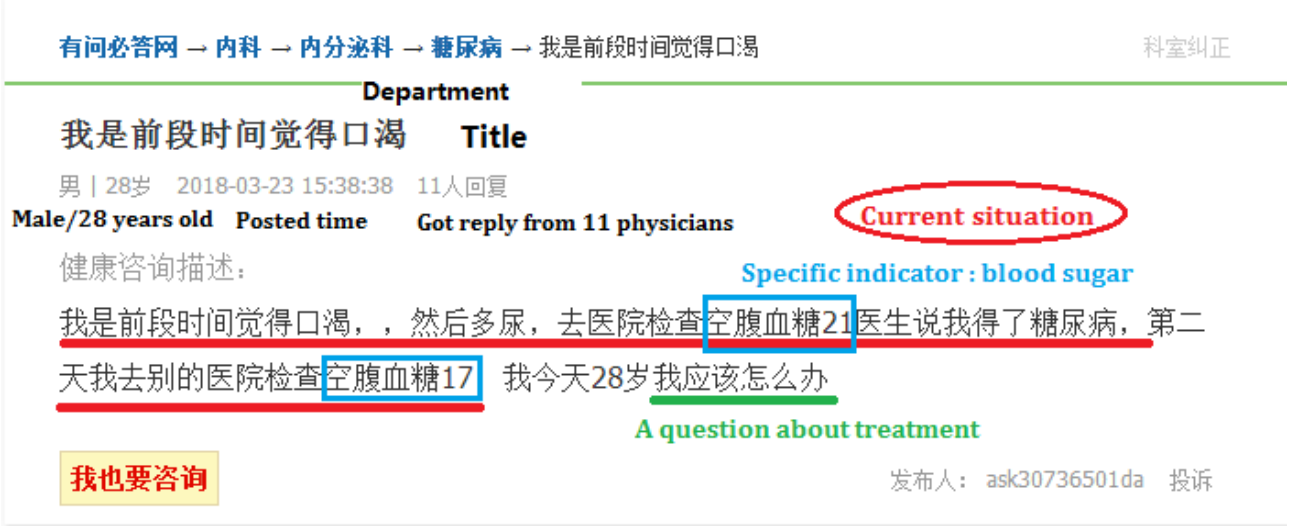
Users of the health Q&A website raise questions for themselves or for their friends and family members who are distressed by health problems. When they require an urgent reply, they may attract physicians by using "healthy coins," which are exchanged with real money at a ratio of 1:1, as shown in Figure 2. Therefore, we classified the questions into free and rewarded questions. A predesigned Java-based Web crawler was used to obtain all questions under the category of diabetes and hepatitis in February 2018. The collected data are stored in a MySQL database. The website displays the most recent 4000 (200 pages multiplied by 20 questions per page) questions of each type. First, we obtained the URLs of questions under the two categories using page parsing and information extraction, and then collected the content of questions from these URLs; an example list of questions is shown in Figure 2 and the details of the first question in Figure 2 are shown in Figure 3. We acquired 4000 free questions and 4000 rewarded question under the topics of diabetes and hepatitis in January 2018.

After filtering for duplicate and invalid data by text preprocessing, the data comprised 3618 free and 3468 rewarded diabetes-related questions, and 3695 free and 2990 rewarded hepatitis-related questions. The collected data contained the content of each question, date the question was posted, number of answers received, and the age and gender of the patient who posted the question.

Figure 1. Homepage of the website “All questions will be answered” (accessed April 18, 2018).

Figure 2. Listed questions under the category of diabetes (accessed March 23, 2018).

Figure 3. Example of a diabetes-related question (accessed March 23, 2018).



Statistical Analysis

Overall Design

The data collected were divided into two parts. The first part was the content, including the text posted by patients suffering from chronic illnesses, which was explored through content analysis. The second part involved statistical data, including age of the patient, number of responses received to each question, and value of the reward. The gender of the patient was also included in the statistical part as a categorical variable. The length of the question was calculated by the number of words in the question, which was also included in the statistical part.

To answer research question 1, content analysis was performed to explore the text element of the question. A coding framework was developed with two main parts. The first consisted of a description of the condition; if the patient provided a description about their condition using specific biological indicators, we could consider that the patient is capable of describing the condition accurately. The other part was related to the phrasing of the content of the question. The content could be phrased as a question or a statement [36]. If one or more questions were included in the content, then we could infer that the patient is capable of phrasing the content as a question. The coding framework also included the classification of question type to address research question 2.

Linear regression analysis was conducted to address research question 3, and to explore the effects of the value of the reward and the length of question on the number of answers received to each question. SPSS 22.0 software (SPSS Inc., Chicago, IL, USA) was used to conduct these analyses.

Content Analysis

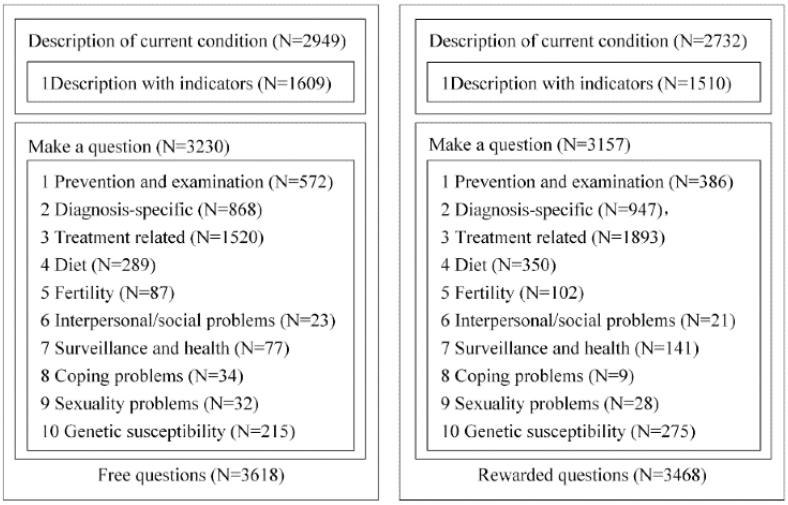
Oh et al [45] developed a coding schema for content analysis of health-related questions. The two types of information they considered were the information provided and the information sought. Participants provided a short description of their health conditions and problems and then asked question(s) about prevention, symptoms, diagnoses, or treatment of their diseases in most cases [45]. Based on this framework, and the conventional categories of questions, we developed a coding framework with two main parts: (1) description of the condition (provided information), and (2) whether or not the content was phrased as a question and the content of the question (information sought).

A sample of diabetes-related questions is provided in Figure 2. The text underlined in red is a description of a current situation, and the questions raised are underlined in green. The question providers commonly first describe the situation of the patient, and then raise the problems experienced, followed by a request for suggestions.

Coding Framework

The coding process was conducted by two research assistants following a training session. The two research assistants independently coded a random selection of 10% (362/3618, 347/3468; 370/3695, 300/2990) of the total number of questions within the pilot framework. If the result of the coding was different or they encountered concepts beyond the previous coding scheme, then the two research assistants discussed adjustment until a consensus was reached. After the independent coding was completed, the final coding framework was designed, and high intercoder reliability was demonstrated (Cohen kappa=0.80, 95% CI 0.55-1.05), indicating acceptable credibility. All questions were then coded using the newly developed framework; the coding framework is presented in Figure 4.

Figure 4. Coding results of diabetes.



Descriptions

The descriptions mainly contained information relating to the patients' current situation. A code of 1 was assigned when the question provider described the situation and a code of 0 was assigned when no description was provided. In the description group, we coded the description in the second layer. A code of 1 was assigned when the provider used specific indicators such as blood sugar and blood pressure in a question about diabetes or the value of viral DNA in a question about hepatitis.

Questions

Code 1 was assigned when the content included one or more question and code 0 was assigned for no question provided, and only a statement. The questions in group 1 were coded into 10 types after pilot coding. The classification referred to the studies of Oh [45], Rutten et al [34], and Fletcher et al [46], including: (1) prevention and examination, (2) diagnosis-specific, (3) treatment-related, (4) diet, (5) fertility, (6) interpersonal/social problem, (7) surveillance and health problem, (8) coping problem, (9) sexuality problems, and (10) genetic susceptibility problems. Code 1 was assigned when a specific content category was present, and code 0 was assigned for a question not related to these categories.

Linear Regression

The dependent variable was the number of responses received per question, which was estimated by the quantity of answers patients received from physicians after they had posted a question request on the health Q&A website. There were two independent variables: the first was the value of the reward, which was estimated by the number of "health coins," and the second was the length of the question, which was calculated by the number of words. The patients' gender and age were also included as control variables. To test our hypotheses against the value of reward and length of question, we formulated the following linear regression equation for the two diseases.

$$\text{Answer}_i = \alpha_0 + \alpha_1\text{Male}_i + \alpha_2\text{Age}_i + \alpha_3\text{Reward}_i + \alpha_4\text{Length}_i$$

Where i=1,2…,n is the index of all patients raising questions, and α_0 to α_4 are the parameters to be estimated.

Results

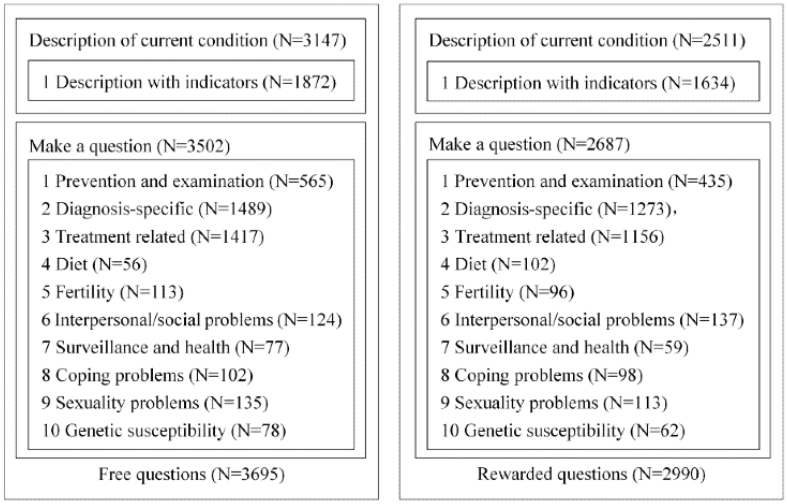
Content Analysis

The coding results are shown in Figures 4 and 5. Most patients made a specific health request in terms of medical examination. Under the category of diabetes, the large majority of the free questions and reward questions contained a description of the patient's condition. Under the category of hepatitis, the majority of free and rewarded questions included the condition of patients. Among the questions, just over half of diabetes or hepatitis question providers used one or more inspection indicators to describe the condition of patients, such as the blood sugar level. For example, in the questions shown in Figure 2, the man felt thirsty and experienced more frequent urination; he described his symptoms and reported his blood sugar level. Among all questions, 89.27% (3230/3618) and 94.77% (3502/3695) of free questions, and 85.17% (3147/3695) and 89.97% (2687/2990) of rewarded questions contained a specific request with interrogative sentences for the diabetes and hepatitis group, respectively. No specific information was provided in the remaining questions. For example, a patient raised the following question:

> *Recently, I found that I had a high blood sugar level after a medical examination. I like to eat sweet food and rice, I rarely exercise in my daily life, and I pay little attention to my diet. I am worried that I will have diabetes one day.*

The man only talks about his condition with declarative sentences, and no specific question is posed. We might infer that he wanted to know whether or not he is likely to experience diabetes in his lifetime or how to prevent it, but he did not ask a question directly.

Figure 5. Coding results of hepatitis.



Information on prevention, examination, diagnosis, and treatment was most frequently requested by patients under both disease categories. In particular, the main concern for patients with diabetes was related to treatment, whereas patients with hepatitis more frequently asked questions focused on diagnosis. Patients commonly raised more than one question in their request. For example, one patient posted the following request:

> *My mother is 61 years old, and a medical examination indicated that her fasting blood glucose is 8.5. This indicator is now regularly tested using a home glucose meter, her fasting plasma glucose is about 9.8, and the number raises to 10 in two hours after a meal. Does my mother have diabetes and how do I treat it?*

Patients with diabetes also paid attention to problems relating to diet, genetic susceptibility, fertility, surveillance, and health. Those with hepatitis cared more about sexuality, interpersonal/social issues, coping, and fertility problems.

Differences Between Free and Rewarded Questions

Like many other expert-based Q&A websites, the “All questions will be answered” health Q&A website includes a pricing system. That is, patients may pay to receive answers to the questions they ask. We compared the information of questions that were answered freely with those that were rewarded using the Mann Whitney *U* test. Female patients were seen to be more likely to post questions in the health Q&A website than males. The length of questions was calculated by the number of words in each question [39]. The distributions of age and average length of question were significantly different in the two disease groups. Free questions attracted more answers than rewarded questions. Further details are provided in Table 1.

Table 1. Descriptive data and differences between free and rewarded questions.

Variable	Diabetes		Hepatitis	
	Free question	Rewarded question	Free question	Rewarded question
Gender of patient				
Male	1673	1068	1514	996
Female	1945	2400	2181	1994
P value[a]	<.001		<.001	
Age of patient (years)				
Min	0	0	0	0
Max	98	89	92	91
Mean	42.98	30.46	31.44	18.26
SD	17.53	25.72	14.36	17.44
P value[a]	<.001		<.001	
Length of question (number of words)				
Min	12	4	10	4
Max	745	1772	430	3527
Mean	55.09	122.58	58.92	132.13
SD	34.13	132.96	32.28	190.03
P value	<.001		<.001	
Number of responses				
Min	1	1	1	1
Max	29	69	16	31
Mean	4.73	2.97	4.55	3.33
SD	2.08	3.14	1.47	2.72
P value	<.001		<.001	
Value of reward (“healthy coins”)				
Min	N/A[b]	2	N/A	2
Max	N/A	310	N/A	200
Mean	N/A	22.59	N/A	24.43

[a]*P* values are based on the Mann Whitney *U* test.

[b]N/A: not applicable.

Linear Regression

Table 2 presents the results of the linear regression analysis based on ordinary least squares. Equations are presented in hierarchical order. First, the results are shown with only control variables in Model 1, and then the independent variables were added to Model 2. The adjusted R^2 and *F* values indicated a good overall fit. The coefficients for value of the reward were significant (*P*<.001) and negative, and the length of questions was significant for diabetes (*P*<.01) but not for hepatitis (*P*=.64).

Table 2. Linear regression for number of answers.

Independent variable	Diabetes (N=3468)		Hepatitis (N=2990)	
	Model 1 B (95%CI)	Model 2 B (95%CI)	Model 1 B (95%CI)	Model 2 B (95%CI)
Constant	2.663 (2.499-2.827), *t*=31.765, *P*<.001	3.288 (3.059-3.516), *t*=28.244, *P*<.001	2.356 (2.221-2.491), *t*=34.187, *P*<.001	3.541 (3.343–3.740), *t*=34.923, *P*<.001
Gender	0.333 (0.092-0.574), *t*=2.709, *P*=.007	0.296 (0.057-0.536), *t*=2.429, *P*=.02	0.705 (0.485-0.926), *t*=6.263, *P*<.001	0.604 (0.392-0.817), *t*=5.578, *P*<.001
Age	0.007 (0.002-0.011), *t*=3.074, *P*=.002	0.007 (0.002-0.011), *t*=3.096, *P*=.002	0.040 (0.034-0.046), *t*=13.293, *P*<.001	0.029, (0.02-30.034), *t*=9.469, *P*<.001
Value of reward	—[a]	–0.016 (–0.20 to –0.011), *t*=–6.572), *P*<.001	—	–0.038 (–0.056 to –0.031), *t*=–15.413, *P*<.001
Length of questions	—	–0.001 (–0.002-0.000), *t*=–2.764, *P*=.006	—	0.000 (–0.001–0.000), *t*=–0.475, *P*=.64
Adjusted R^2	0.007	0.023	0.085	0.180
F	12.773	21.156	186.432	164.380

[a]Not applicable.

Discussion

Patients' Abilities to Describe Their Illness and Phrase the Condition as a Question

Most of the patients were found to be capable of describing their personal condition. More than 40% of patients with diabetes and more than 50% of patients with hepatitis could describe their illness accurately by using indicators in their description. More than 85% of patients phrased their condition as a question overall.

Health literacy has received substantial attention [47,48] in the field of online health care. The ability to obtain, manage, and understand health information needs will help patients make more appropriate health choices in the future. To date, most health literacy studies have focused on the evaluation of health-related content or eHealth services [48], but the patient's ability to seek appropriate information, especially when asking on a health Q&A website, is rarely investigated. Making an online request on a health Q&A website is different from a traditional offline face-to-face consultation. In a hospital setting, physicians can ask patients about their current condition and obtain crucial data for improved diagnosis and treatment. As this is a dynamic process, the physician can make adjustments during the interaction as needed. Most question providers followed this offline process and led with a description about their conditions. When they described the condition, they were also accustomed to using indicators, especially the blood sugar level as an important parameter for patients suffering from diabetes. In line with the old Chinese saying "prolonged illness made the patient a doctor," most of the patients or their caregivers are able to consciously describe their condition and raise a question about the problems they are experiencing. Nevertheless, some users cannot phrase their condition as a question, such as the man who was afraid of getting diabetes, and these patients might require a platform to express their feelings. The less formal communication element of the health Q&A website may be more suitable for these types of patients. The platform should also provide guidance for users who cannot accurately express their question, and a questioning assistant tool might be helpful in this regard.

Differences Between Patients With Diabetes and Hepatitis

Patients always seek information related to a specific illness or disease [29], and different diseases lead to various health concerns. Patients with diabetes mainly sought answers about treatment, followed by diagnosis, whereas those with hepatitis were more concerned about diagnosis, followed by treatment. Medical treatment information has always been searched most frequently [30]. The differences could be attributed to the characteristics of the two diseases: diabetes is a chronic disease, whereas some types of hepatitis are infectious. Patients with chronic diseases would be most interested in trying to find a cure or prevent disease progression, whereas those with infectious diseases may experience pressure on their social life. For the susceptible population, prevention and diagnosis are considered crucial.

For patients with diabetes, questions related to diet attracted the most attention, which is consistent with the study of Kuske et al [35]; this phenomenon might be attributed to the high sugar in foods and drinks. Patients with hepatitis also paid attention to diet but less so than patients with diabetes; their main concern with respect to diet was related to alcohol consumption, which is associated with the liver. Traditional Chinese diets are typically of ancient origin and are accepted by many Chinese people, which may explain the high number of questions about food and drink. Patients with diabetes were also concerned about other topics such as genetic susceptibility, fertility, and surveillance problems because genetic factors are the main cause of diabetes. Those with hepatitis also paid attention to sexuality, interpersonal, and fertility problems. Given that many types of hepatitis are infectious, the disease will affect the daily life and social interactions of patients. The present results enable health care website managers or physicians to understand patients' information needs on health care. To improve the usefulness of health care provider websites, managers can divide the questions into minor groups and make the questions more targeted.

Physicians can also be grouped into different types, allowing them to answer questions with which they are most familiar. Considering that treatment and diagnosis are frequently referred to, health care websites can also focus on these parts and offer educational articles about these topics.

Factors Influencing Answer Provision

When patients or caregivers put forward their questions on a health Q&A website, they want informed answers to allow them to make the right decision about their health. Patients can attract answers from physicians in two ways. One is to textually describe more aspects about their situation, and the other is to pay a fee for the response. Our results indicate that the length of question negatively affected the number of responses received only in the diabetes group, and this finding is partly consistent with the study of Morris et al [39]. The fact that the length of questions had no significant effect on the number of responses received in the hepatitis group may be attributed to the characteristics of the data, as the maximum value of the number of questions was very large (3527); therefore, these outliers may have affected the results. In the diabetes group, the length of question affected the number of questions negatively; however, the adjusted R^2 value was quite low compared with that of the hepatitis group. More significant factors in the model may lead to a smaller adjusted R^2 value, indicating that further studies are needed to verify the effect of the length of question. The coefficients were also relatively low, which may be attributed to the large sample size.

There are two possible reasons for the finding that long sentences may lead to fewer responses. First, when patients face a complicated problem, they need to use many words to make their description clear, which might make it difficult for the physicians to understand or answer the questions. Second, too many questions may be included, which require additional time and effort from physicians. Raising a question with an appropriate length of sentence is difficult. For health Q&A website users who want to get answers quickly, they should focus on refining their content and make their questions more specific. Providing a training program for patients may also be helpful for users to raise a proper, understandable question.

Some question providers pay a fee when they raise a question; however, the free questions generally received more answers than those that are rewarded. This result is in contrast to the finding of Rafaeli et al [38], who investigated economic and social motivators; hence, we must include additional factors in our further research. The amount of reward is negatively related to the number of answers. People who offer a reward are likely to make the question longer, thereby receiving fewer answers. The pricing mechanism is not clear enough for users of the website. They change their money into “health coins” at a ratio of 1:1, but whether or not these “health coins” are fairly given to the specific physicians is unknown. Therefore, the health Q&A website should provide more suitable motivation mechanisms to encourage physicians to participate on the website more frequently.

Limitations

This study has some limitations. First, the questions were only collected from one website, leading to limitations in the data source setting. The data collected might only reflect the needs of specific online users, and the conclusions may not apply to patients with other diseases or those that use other health care platforms. Future research should collect data from other platforms and make comparisons among different data sources. Second, because the proportion of free questions was higher than that of rewarded questions, the time periods were different in the two groups. Hence, the time factor may also have affected the results. Third, we may have missed some potential question types in our developed framework as some words may have more than one meaning in Chinese. Finally, the effect of descriptive variables on the willingness of physicians to answer questions should be further considered.

Conclusion

Health literacy requires many abilities in relation to health information and decision making. Our study provides insight into the ability of patients to make a suitable question request on expert health Q&A websites. The examination of patients’ health topics helps to provide a better understanding of the information needs relating to different diseases, and different health information should be made available to targeted patients based on their needs. A negative effect of the length of question and value of the reward on the number of responses received was identified. Patients do not need to pay a fee to receive a response to their questions and should aim to refine the length of their questions if they want more timely responses from specialists.

Conflicts of Interest

None declared.

References

1. Fox S, Duggan M. Health online 2013. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2019-04-22]
2. Huh J, McDonald DW, Hartzler A, Pratt W. Patient moderator interaction in online health communities. AMIA Annu Symp Proc 2013;2013:627-636 [FREE Full text] [Medline: 24551364]
3. McKinley CJ, Wright PJ. Informational social support and online health information seeking: Examining the association between factors contributing to healthy eating behavior. Comput Hum Behav 2014 Aug;37:107-116. [doi: 10.1016/j.chb.2014.04.023]

4. Fox S. Pew Internet & American Life Project. 2011. 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]
5. 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]
6. Helve O. A medical consultation service on facebook: descriptive analysis of questions answered. J Med Internet Res 2014;16(9):e202 [FREE Full text] [doi: 10.2196/jmir.3194] [Medline: 25189182]
7. Shah C. Online Question-Answering (Q&A). Social Information Seeking: Leveraging the Wisdom of the Crowd. Cham: Springer International Publishing; 2017:45-74.
8. Shah C, Oh S, Oh J. Research agenda for social Q&A. Libr Inf Sci Res 2009 Dec;31(4):205-209. [doi: 10.1016/j.lisr.2009.07.006]
9. Harper F, Raban D, Rafaeli S, Konstan J. Predictors of answer quality in online Q&A sites. 2008 Apr Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; : ACM; April 2008; Florence, Italy.
10. Oh S, Worrall A. Health answer quality evaluation by librarians, nurses, and users in social Q&A. Libr Inf Sci Res 2013 Oct;35(4):288-298. [doi: 10.1016/j.lisr.2013.04.007]
11. Kitzie V, Shah C. Faster, better, or both? Looking at both sides of online question-answering coin. Proc Am Soc Info Sci Tech 2012 Jan 11;48(1):1-4. [doi: 10.1002/meet.2011.14504801180]
12. Chen Y. Knowledge market design: A field experiment at Google Answers. J Public Econ Theory 2010 Jul 19;12(4):641-664. [doi: 10.1111/j.1467-9779.2010.01468.x]
13. JustAnswer. URL: http://www.justanswer.com/ [accessed 2019-05-16]
14. PickAnswer. URL: http://www.pickanswer.com/ [accessed 2019-05-16]
15. All question will be answered. URL: https://www.120ask.com/ [accessed 2019-05-16]
16. BabyTree. URL: http://babytree.com/ [accessed 2019-04-22]
17. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decis Support Syst 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
18. Oh S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. J Am Soc Inf Sci 2011 Nov 01;63(3):543-557. [doi: 10.1002/asi.21676]
19. Zhao J, Ha S, Widdows R. Building trusting relationships in online health communities. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):650-657. [doi: 10.1089/cyber.2012.0348] [Medline: 23786170]
20. Griffin RJ, Dunwoody S, Neuwirth K. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ Res 1999 Feb;80(2 Pt 2):S230-S245. [doi: 10.1006/enrs.1998.3940] [Medline: 10092438]
21. Afifi WA, Weiner JL. Toward a Theory of Motivated Information Management. Commun Theory 2004 May;14(2):167-190. [doi: 10.1111/j.1468-2885.2004.tb00310.x]
22. Longo DR. Understanding health information, communication, and information seeking of patients and consumers: a comprehensive and integrated model. Health Expect 2005 Sep;8(3):189-194 [FREE Full text] [doi: 10.1111/j.1369-7625.2005.00339.x] [Medline: 16098149]
23. 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]
24. 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 12;27(8):1160-1176. [doi: 10.1177/1049732316646355] [Medline: 27179023]
25. Kilicoglu H, Ben Abacha A, Mrabet Y, Shooshan SE, Rodriguez L, Masterton K, et al. Semantic annotation of consumer health questions. BMC Bioinformatics 2018 Feb 06;19(1):34 [FREE Full text] [doi: 10.1186/s12859-018-2045-1] [Medline: 29409442]
26. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
27. Francik E, Clark H. How to make requests that overcome obstacles to compliance. J Memory Lang 1985 Oct;24(5):560-568. [doi: 10.1016/0749-596x(85)90046-4]
28. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: A literature review. Health Informatics J 2016 Dec 26;22(4):992-1016. [doi: 10.1177/1460458215602939] [Medline: 26377952]
29. 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]
30. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer Use of "Dr Google": A Survey on Health Information-Seeking Behaviors and Navigational Needs. J Med Internet Res 2015 Dec 29;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]

31. Lee K, Hoti K, Hughes JD, 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;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
32. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
33. Wong C, Harrison C, Britt H, Henderson J. Patient use of the internet for health information. Aust Fam Physician 2014 Dec;43(12):875-877. [Medline: 25705739]
34. Rutten LJF, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). Patient Educ Couns 2005 Jun;57(3):250-261. [doi: 10.1016/j.pec.2004.06.006] [Medline: 15893206]
35. Kuske S, Schiereck T, Grobosch S, Paduch A, Droste S, Halbach S, et al. Diabetes-related information-seeking behaviour: a systematic review. Syst Rev 2017 Oct 24;6(1):212 [FREE Full text] [doi: 10.1186/s13643-017-0602-8] [Medline: 29065919]
36. Teevan J, Morris M, Panovich K. Factors affecting response quantity, quality, and speed for questions asked via social network status messages. 2011 Jul Presented at: Fifth International AAAI Conference on Weblogs and Social Media; 2011; Barcelona, Spain.
37. Liu Z, Jansen B. Questioner or question: Predicting the response rate in social question and answering on Sina Weibo. Inf Proc Manag 2018 Mar;54(2):159-174. [doi: 10.1016/j.ipm.2017.10.004]
38. Rafaeli S, Raban DR, Ravid G. How social motivation enhances economic activity and incentives in the Google Answers knowledge sharing market. Int J Knowl Learn 2007;3(1):1-11. [doi: 10.1504/ijkl.2007.012598]
39. Morris M, Teevan J, Panovich K. What do people ask their social networks, and why?: a survey study of status message q&a behavior Share on. 2010 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 10, 2010; Atlanta, Georgia.
40. Shaw RJ, Johnson CM. Health Information Seeking and Social Media Use on the Internet among People with Diabetes. Online J Public Health Inform 2011 Jun;3(1):ojphi.v3i1.3561 [FREE Full text] [doi: 10.5210/ojphi.v3i1.3561] [Medline: 23569602]
41. Shaw J, Sicree R, Zimmet P. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract 2010 Jan;87(1):4-14. [doi: 10.1016/j.diabres.2009.10.007] [Medline: 19896746]
42. Nordfeldt S, Johansson C, Carlsson E, Hammersjö JA. Use of the Internet to search for information in type 1 diabetes children and adolescents: a cross-sectional study. Technol Health Care 2005;13(1):67-74. [Medline: 15706065]
43. World Health Organization. Global Hepatitis Report. 2017. URL: https://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/ [accessed 2019-04-22]
44. Evon DM, Golin CE, Stoica T, Jones RE, Willis SJ, Galanko J, et al. What's Important to the Patient? Informational Needs of Patients Making Decisions About Hepatitis C Treatment. Patient 2017 Jun 23;10(3):335-344 [FREE Full text] [doi: 10.1007/s40271-016-0207-7] [Medline: 27882509]
45. Oh S, Zhang Y, Park MS. Health information needs on diseases: A coding schema development for analyzing health questions in social Q&A. Proc Am Soc Info Sci Tech 2013 Jan 24;49(1):1-4. [doi: 10.1002/meet.14504901316]
46. Fletcher C, Flight I, Chapman J, Fennell K, Wilson C. The information needs of adult cancer survivors across the cancer continuum: A scoping review. Patient Educ Couns 2017 Mar;100(3):383-410. [doi: 10.1016/j.pec.2016.10.008] [Medline: 27765377]
47. Diviani N, van den Putte B, Giani S, van Weert JC. 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]
48. Kim H, Xie B. Health literacy in the eHealth era: A systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]

Abbreviations

Q&A: question and answer

Edited by G Eysenbach; submitted 11.12.18; peer-reviewed by R Haase, L Taylor-Swanson, G Duan, W Pian; comments to author 08.04.19; revised version received 29.05.19; accepted 29.02.20; published 16.04.20.

Please cite as:
Hong Z, Deng Z, Evans R, Wu H
Patient Questions and Physician Responses in a Chinese Health Q&A Website: Content Analysis
J Med Internet Res 2020;22(4):e13071
URL: https://www.jmir.org/2020/4/e13071
doi:10.2196/13071
PMID:32297872

Original Paper

Nurse-Physician Communication Team Training in Virtual Reality Versus Live Simulations: Randomized Controlled Trial on Team Communication and Teamwork Attitudes

Sok Ying Liaw[1], PhD; Sim Win Ooi[2], BSc; Khairul Dzakirin Bin Rusli[1], BSc; Tang Ching Lau[3], MBBS, MMED, MRCP; Wilson Wai San Tam[1], PhD, MPhil; Wei Ling Chua[1], BSc, PhD

[1]Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore

[2]National University Hospital, Singapore, Singapore

[3]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Sok Ying Liaw, PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
Singapore
Singapore
Phone: 65 65167451
Email: nurliaw@nus.edu.sg

Abstract

Background: Interprofessional team training is needed to improve nurse-physician communication skills that are lacking in clinical practice. Using simulations has proven to be an effective learning approach for team training. Yet, it has logistical constraints that call for the exploration of virtual environments in delivering team training.

Objective: This study aimed to evaluate a team training program using virtual reality vs conventional live simulations on medical and nursing students' communication skill performances and teamwork attitudes.

Methods: In June 2018, the authors implemented nurse-physician communication team training using communication tools. A randomized controlled trial study was conducted with 120 undergraduate medical and nursing students who were randomly assigned to undertake team training using virtual reality or live simulations. The participants from both groups were tested on their communication performances through team-based simulation assessments. Their teamwork attitudes were evaluated using interprofessional attitude surveys that were administered before, immediately after, and 2 months after the study interventions.

Results: The team-based simulation assessment revealed no significant differences in the communication performance posttest scores (P=.29) between the virtual and simulation groups. Both groups reported significant increases in the interprofessional attitudes posttest scores from the baseline scores, with no significant differences found between the groups over the 3 time points.

Conclusions: Our study outcomes did not show an inferiority of team training using virtual reality when compared with live simulations, which supports the potential use of virtual reality to substitute conventional simulations for communication team training. Future studies can leverage the use of artificial intelligence technology in virtual reality to replace costly human-controlled facilitators to achieve better scalability and sustainability of team-based training in interprofessional education.

Trial Registration: ClinicalTrials.gov NCT04330924; https://clinicaltrials.gov/ct2/show/NCT04330924

(J Med Internet Res 2020;22(4):e17279) doi:10.2196/17279

KEYWORDS

interprofessional education; team training; nurse-physician communication; virtual reality; simulation

Introduction

The relationship between communication and patient safety has emphasized the importance of team training for health professional education [1]. Improved communication among health care teams has been associated with better quality in patient care and enhanced patient satisfaction [2]. In contrast, failures in communication have been found to be the cause of adverse patient events and negative health outcomes [3,4]. Effective communication remains a challenge between nurses

and physicians due to training differences and embedded workplace cultures [5,6]. In an earlier study on collaborative practices between nurses and junior physicians, the physicians reported insufficient information conveyed by ward nurses about changes in patients' conditions, whereas the ward nurses reported a lack of direct communication from physicians about changes in patient treatment plans. The study thereby called for interprofessional education around effective communication strategies that focus on open communication, shared information, decision making, and mutual respect [7].

There is a general consensus among health care professionals and educators that interprofessional team training needs to commence at the preregistration level and continue into workplace practice [1,3]. An evidence-based review has called for educators to incorporate tools such as Identity, Situation, Background, Assessment, and Recommendation (ISBAR) for information exchange between health care providers into interprofessional team training [8]. These tools create a shared mental model—team members' shared understanding of relevant knowledge—for teams to communicate and enable the exchange of information between team members to facilitate decision making. The integration of these tools into interprofessional education has shown to improve students' perceptions of interprofessional collaboration and confidence in communicating effectively with other team members [9]. However, these evaluations were based on self-reporting questionnaires and may not reflect actual team performances. In addition, such teaching has been primarily within uni-professional groups of medical and nursing students as there were challenges in bringing them together to learn structured communication interprofessionally [8].

Simulation, which is commonly used as a teaching method for health care team training, provides experiential opportunities in patient-care situations for teams to work together, practice their roles, and develop a shared understanding of a team and its associated tasks [9]. According to Weaver [4], team training programs are most effective with integrated learning activities that use tools to support teamwork practices. Previous studies have provided evidence on the effectiveness of incorporating simulations with communication tools for improving medical and nursing students' communication skills and teamwork attitudes [10,11]. However, these studies were limited, and few evaluated performance outcomes from the preregistration team training. In addition, logistical issues in organizing simulation training, including simulation facilities and scheduling, have proved to be challenging particularly among preregistration health care students whom are often trained in different institutions [10]. Simulations in a virtual environment can overcome time and space constraints and are gaining popularity among health care educators. Being accessible anytime and anywhere, virtual simulations provide the flexibility to be integrated into curricula and have the scalability to train large numbers of learners.

With advances in computer learning technology, virtual reality provides a viable platform for team-based simulation training that can address logistical challenges implicit in traditional simulation, such as physical location availability and scheduling for different groups of students [12]. We developed a computer-based virtual reality known as CREATIVE (Create Real-time Experience and Teamwork in Virtual Environment), where users can create physical and social presences using avatars in a 3D virtual hospital environment. CREATIVE was developed based on gaps identified from a systematic review on the use of multiuser virtual reality in health care education. The review called for the incorporation of theoretical models to inform the virtual learning process and the employment of more rigorous research including randomized controlled trials (RCT) to evaluate the impact of multiuser virtual reality on clinical performances. We applied theories including experiential learning and social constructivism to support the development and implementation of simulation learning in CREATIVE [12]. A pilot study was performed that showed the usability and feasibility of CREATIVE in supporting social interactions and collaborative practices among diverse groups of health care students [13]. Another study was conducted on CREATIVE that supported the integration of cognitive tools and virtual simulations to develop shared mental models for the delivery of optimal clinical teamwork [14]. In this study, by incorporating communication tools into a team training program, we aimed to evaluate the effectiveness of virtual reality in comparison with live simulations on medical and nursing students' communication skill performances and teamwork attitudes.

Methods

Study Designs and Participants

We conducted a prospective RCT with a pretest-posttest study design after our study was approved by the institutional review board. We used social media to recruit volunteers who were undertaking their third or fourth year of medicine or nursing courses in a local university. This group of students was targeted as they had undertaken acute care management in their respective courses. We calculated the sample size based on a medium effect size of 0.50 for an independent sample *t* test [15]. Cohen's (1992) [15] sample size table reported that a sample size of 64 participants per group (total participants=128) would be sufficient to achieve a predefined power of 0.80 at a 5% level of significance (2-sided) in detecting a medium effect size. Divided by health care courses taken and year of study, the participants were randomly assigned to a virtual or simulation group. They were grouped into interprofessional teams of 2 medical students and 2 nursing students to undertake the study intervention. Each team was facilitated by a simulation-trained faculty member and a simulated patient.

Study Interventions

The participants in the virtual and live simulation groups underwent 3 hours of team training on nurse-physician communication conducted in the university's center for health care simulation. After obtaining written informed consent from participants, they were directed to a 20-minute computer-based lesson on communication skill strategies. We adapted the strategies from the Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS) curriculum, which included ISBAR; Concerned, Uncomfortable, and Safety (CUS); feedback to acknowledge; callout; and check back [16].

The participants in the virtual group remained in the computer laboratory to undertake the virtual reality simulation. Together with their teammate, a facilitator, and a simulated patient, they logged into the 3D virtual environment using their avatar roles. Figure 1 illustrates their viewpoints when interacting with one another and the virtual environment. The virtual groups were given an orientation in which they learned to navigate between tutorial room and ward settings, talk among themselves using headsets, and perform assessments on the patient avatar. The participants in the live simulation group were brought into a simulated ward setting for their simulations and received an orientation on the ward setup, the equipment, and a standardized patient.

Table 1 describes the tasks performed by the participants in virtual reality and live simulations. After an orientation, the participants and facilitators gathered in a physical or virtual tutorial room for a briefing on the learning objectives and activities. In both groups, the participants in each team were randomly paired up (1 medicine student and 1 nursing student) and took turns role-playing and observing to participate in two simulation scenarios. The first scenario simulated a morning round situation of a postoperative patient with sepsis conditions, while the second scenario involved the same patient whose condition had deteriorated into septic shock. The participants were given time to read the case history before commencing each scenario. Each scenario began with the nursing participant performing a nursing assessment on the patient, followed by communicating to the medical participant. In these scenarios, the nursing participants were expected to use communication strategies (eg, ISBAR, CUS, check back) to communicate the assessment findings to the medical participants. The medical participants were expected to also perform physical assessments and apply communication strategies by acknowledging the nurse's concerns and communicating treatment plans using the callout strategy. The clinical findings and responses displayed by both the patient avatar and the standardized patient were similar based on the prepared scripts. Each scenario lasted about 15 to 20 minutes and was followed up by a 30-minute debriefing. The debriefing was led by a facilitator using a facilitator guide that focused on the application of communication strategies.

Figure 1. Viewpoints of different users. A: simulated patient's view; B: facilitator's view; C: medical student's view; D: nursing student's view.

Table 1. Comparison of tasks performed by the participants in virtual reality and live simulations.

Tasks	Virtual Reality	Live Simulation
Orientation	The participants navigated in the virtual environment, tested their headsets for communication, and became orientated with clickable icons and objects.	The participants were orientated to the physical ward setting, the equipment, and a simulated patient.
Briefing	The participants were briefed in real time using headsets by a facilitator avatar in the virtual tutorial room.	The participants were briefed face-to-face by a live facilitator in the physical room setting.
Reading case history	The participants read the case scenario by clicking the laptop in the virtual tutorial room and the COW[a] in the virtual ward.	The participants read the case scenario from a case file placed in the tutorial room and ward setting.
Patient assessment	The participants clicked on the body parts of the patient avatar and on monitoring machines.	The participants performed physical assessments on the live simulated patient and used monitoring machines.
Interventions and treatments	The participants clicked on the appropriate trolley or equipment.	The participants performed hands-on interventions or treatments.
Communication	The participants communicated with each other using headsets and clickable gestures.	The participants communicated with each other face-to-face.
Debriefing	The participants undertook a group debriefing in the virtual tutorial room with the facilitator avatar.	The participants undertook a face-to-face group debriefing in the physical room with the live facilitator.

[a]COW: computer on wheels.

Data Collection and Instrument

We administered the Attitudes Toward Interprofessional Health Care Team (ATHCT) and Interprofessional Socialization and Valuing Scale (ISVS) questionnaires before (baseline), immediately after (posttest), and 2 months after (follow-up) the simulation training to measure teamwork attitudes. The ATHCT, a 14-item tool using a 5-point scale validated by Kim and Ko [17], was adopted to measure the participants' attitudes toward working in interprofessional health care teams. We obtained a high internal consistency in this study, with a Cronbach alpha of .83. The ISVS, a 24-item questionnaire using a 7-point scale developed by King et al [18], was used to capture the participants' beliefs, behaviors, and attitudes in interprofessional socialization. This study reported a high Cronbach alpha of .86.

After the study intervention, the participants from both groups were brought to a room for team-based simulation assessments. They were assigned to work in pairs (medical and nursing) based on their earlier simulation teams. The teams were given a case history to read and an orientation of the simulation room with a manikin setup. The test scenario began with the nursing student performing the nursing assessment and management of the manikin, which was displaying signs and symptoms of deterioration, and calling the doctor. Upon the medical participant's arrival, they worked as a team to manage the deteriorating patient. The team-based simulation lasted about 15 minutes and the entire process was recorded. We sent the recorded videos for rating to a clinician and an academic staff member whom were blinded to the groupings. The assessors rated the team communication performances independently using a validated team communication scale. We developed the scale using a 7-item checklist with a 5-point scale and a global rating item based on observable team communication skills between the nurse and doctor. We sent the scale for content validation to an interprofessional team of 4 medicine and nursing academics and clinicians. We computed the interrater reliability across 2 raters who scored the video-recorded performances independently using the validated scale. A high intraclass correlation coefficient (ICC) of 0.96 (95% CI 0.93-0.97) was reported, indicating good interrater agreement.

Data Analysis

We applied descriptive statistics, chi-square tests, and *t* tests to analyze the demographic characteristics of the study population. We computed a paired sample *t* test to examine significant changes between the baseline and posttest performance scores and an independent sample *t* test to determine differences in the posttest scores between the groups. We performed a repeated measures analysis of variance (ANOVA) for the ATHCT and ISVS scores.

Results

Demographic Characteristics

Although we targeted 128 participants, only 120 participants completed the study. The majority were female (81, 67.5%), Chinese (105, 87.5%), and an average of 22.17 years of age (SD 2.07). There were no significant differences in the baseline characteristics, including age (P=.06), gender (P=.17), year of study (P=.85), and ethnicity (P=.94) between the virtual and simulation groups (Table 2). This supported the homogeneity of the participants between the 2 groups.

Table 2. Demographic characteristics.

Characteristics	Overall (N=120)	Virtual (N=60)	Simulation (N=60)	*P* value
Age (years), mean (SD)	22.17 (2.07)	21.82 (1.07)	22.53 (2.70)	.06
Gender, n (%)				.17
Male	39 (32.5)	16 (26.7)	23 (38.3)	
Female	81 (67.5)	44 (73.3)	37 (61.7)	
Course, n (%)				>.99
Medicine	60 (50.0)	30 (50.0)	30 (50.0)	
Nursing	60 (50.0)	30 (50.0)	30 (50.0)	
Year of study, n (%)				.85
Third year	41 (34.2)	21 (35.0)	20 (33.3)	
Fourth year	79 (65.8)	39 (65.0)	40 (66.7)	
Ethnicity, n (%)				.94
Chinese	105 (87.5)	53 (88.3)	52 (86.7)	
Indian	9 (7.5)	4 (6.7)	5 (8.3)	
Malay	6 (5.0)	3 (5.0)	3 (5.0)	

Team Communication Performance

The team-based simulation assessment revealed no significant differences in the overall communication performance posttest scores ($F_{2,58}$=1.46, *P*=.29, eta^2=0.33) between the virtual (mean 22.60, SD 5.31) and simulation groups (mean 23.97, SD 4.55). There were also no significant differences in the total checklist posttest scores ($F_{2,58}$=3.654, *P*=.29, eta^2=0.28) and global posttest scores ($F_{2,58}$=1.56, *P*=.29, eta^2=0.33) between the groups.

Teamwork Attitudes

At the baseline, there were no significant differences in the ATHCT (*P*=.33) and ISVS (*P*=.45) scores between the virtual and simulation groups, supporting the homogeneity of the participants between the groups. After the team training, there was a significant increase in the ATHCT and ISVS posttest scores from the baselines scores for both groups. There was also a significant increase in the follow-up ISVS scores from the baseline scores for the virtual group (*P*=.047) but not for the simulation group (*P*=.14). No significant differences between the baseline and follow-up ATHCT scores were found for both groups (Table 3).

Using repeated-measure ANOVA, there were no significant differences in the trend between the virtual and simulation groups for both the ATHCT (*P*-interaction=.58, eta^2=0.005) and ISVS (*P*-interaction=.61, eta^2=0.004) scores. There were also no significant differences in the ATHCT ($F_{2,118}$=0.507, *P*=.48, eta^2=0.004) and ISVS ($F_{2,118}$=0.335, *P*=.56, eta^2=0.003) scores over the 3 time points between the virtual and simulation groups.

Table 3. Comparison of teamwork attitude scores on Interprofessional Socialization and Valuing Scale and Attitudes Toward Interprofessional Health Care Team.

Group	Baseline	Posttest	*P* value	Baseline	Follow-up test	*P* value
ISVS[a] (teamwork attitude score)						
Virtual, mean (SD)	131.78 (15.81)	142.92 (14.88)	<.001	131.78 (15.81)	136.62 (6.43)	.047
Simulation, mean (SD)	134.03 (16.31)	143.95 (15.47)	<.001	134.03 (16.31)	137.45 (15.90)	.14
T test (2-tailed)	0.45	0.71	N/A[b]	0.45	0.78	N/A
ATHCT[c] (teamwork attitude score)						
Virtual, mean (SD)	58.22 (5.78)	60.05 (5.41)	.004	58.22 (5.78)	58.27 (6.57)	.95
Simulation, mean (SD)	57.12 (6.58)	59.95 (6.65)	<.001	57.12 (6.58)	57.17 (7.26)	.96
T test (2-tailed)	0.33	0.93	N/A	0.33	0.39	N/A

[a]ISVS: Interprofessional Socialization and Valuing Scale.

[b]N/A: not applicable.

[c]ATHCT: Attitudes Toward Interprofessional Health Care Team.

Discussion

Principal Findings

The outcomes in this RCT study did not show the inferiority of computer-based virtual reality on teamwork attitudes and communication skill performances when compared with live simulations. These outcomes were consistent with earlier studies [19,20], providing more evidence to support its potential use in substituting conventional team-based simulations. In this virtual reality team training, we incorporated design elements that mimicked live simulations for the experiential learning and the ability to support multiusers, including students, a simulated patient, and a facilitator to come together for collaborative learning.

Despite the different learning mediums, both offered similar learning strategies, including experiential and collaborative learning, that allowed medical and nursing students to practice the use of communication tools through roleplay exercises and engage in reflective practices through interactions with one another and their facilitators [12]. These learning activities are underpinned by the theories of experiential learning and social constructivism. Kolb's experiential learning theory supports the learning mechanism of roleplaying and debriefing, and the theory of social constructivism emphasizes social interactions underlying collaborative learning. Thus, we call upon educators and instructional designers to apply these theories to support the implementation of simulation learning in virtual reality learning environments [12].

We believe that teaching teamwork at the preregistration level is important but has limited impact on organization and patient outcomes. Our long-term evaluation of virtual reality on the retention of learning was limited to measuring changes in interprofessional attitudes 2 months after the intervention, with significant improvements in attitudes toward interprofessional socialization but not in attitudes toward interprofessional health care teams. Caution has to be taken on the validity of self-reported measures. Nonetheless, the opportunities to engage in social interactions in virtual reality could have improved students' attitudes toward interprofessional socialization. Social interactions between the medical and nursing students were also facilitated using structured communication tools, which could have promoted open communication, shared information, and decision-making [11]. One advantage of virtual reality over live simulations was its ability to allow anonymous social interactions in its environment, which may cause less social anxiety and stress for students [13,21]. We believe that positive attitudes toward interprofessional socialization are important in building future collegial working relationships between nurses and physicians.

The lack of improvement in attitudes toward interprofessional health care teams after the study interventions for both virtual and simulation groups could be due to the influence of workplace realities faced by students during their clinical practicums, wherein issues of powers, structures, and systems may limit health care teams from collaborating effectively. This highlights the importance of addressing workplace systems and structures, including the implementation of workplace team training [22]. Future studies should investigate how virtual reality can contribute to postqualification workplace-based team training. This may include the use of virtual reality and workplace simulations as a blended learning approach to optimize interprofessional learning [23].

Although we did not compare the cost and resources of the virtual and simulation interventions, we recognized the challenges for sustainability and scalability. The existing virtual reality serves as a flexible and practical platform to bring everyone together, but its scalability is constrained by human-controlled avatars. The recruitment and training of people for the avatar role of a patient were found to be less convenient and costly. In addition, as a result of unequal cohort size across different health care courses (ie, nursing students vs medical students), it is unlikely that all health care students will form interprofessional teams. We plan to develop and integrate artificial intelligence (AI) in virtual reality to replace the roles of a simulated patient and even a medical doctor. With evolving AI technology, the goal is to design AI agents that can interact with users in a natural, humanlike way such as via natural language processing [24]. The AI can cohabit virtual worlds with people and facilitate deep engagement of learning that could potentially improve learning outcomes [25]. A recent systematic review on AI in medical education highlighted the potential use of AI in virtual reality and called for more evidence to justify its effectiveness [26]. Therefore, future studies can examine the effectiveness of AI-controlled agents compared to human-controlled avatars in virtual reality simulations.

This study used the topic of sepsis and septic shock for the application of TeamSTEPPS communication strategies as the curriculum's content. However, the virtual platform was created with expandability in mind, allowing other medical conditions such as cardiac arrest and other teamwork curricula to be programmed and applied in the CREATIVE platform. We have collaborated with other local higher education institutions to develop geriatric-related cases using the CREATIVE platform to support teamwork in interprofessional rounding among diverse students of health care professions[11]. We hope to open up the possibility of international collaborative learning using this platform and to study its effectiveness.

Limitations

Although we tested the students' team performances and attitudes using simulation-based assessments and validated tools, the quality of evidence was limited by an immediate posttest on team performance and self-reported attitude questionnaires. Future studies can provide more evidence on team performance by examining the baselines and retentions of team performances. The study intervention was carried out at a single location, the university's simulation center, to capture the participants' team performances using simulation-based assessments. Future implementation can occur across different physical locations, including access from home. Future studies can be conducted to gain insights into the learning process of users and identify any potential usability issues across distant locations.

Conclusion

Interprofessional team training using simulations has proved to be logistically challenging to implement at the preregistration level due to difficulties in bringing together different groups of health care students. We implemented doctor-nurse communication team training by incorporating communication strategies from the TeamSTEPPS curriculum into computer-based virtual reality for undergraduate medical and nursing students. Our study outcomes did not show any difference between virtual and live simulations in terms of teamwork attitudes and communication skill performances, which supports the potential use of virtual reality to substitute conventional team-based simulation training. Further developments and evaluations can use AI technology to replace costly human-controlled avatars to achieve better scalability and sustainability of team-based training in interprofessional education. With its scalability and practicality, virtual reality serves as a promising learning strategy to prepare students to be part of a future collaborative workforce that can provide safe and quality patient care.

Acknowledgments

The authors would like to thank the health care students for their participation in this study. We also thank the staff of the Centre for Healthcare Simulation and National University Singapore Information Technology for supporting this study. Lastly, we would like to thank the National University Health System, Research Support Unit, for providing editing services to this manuscript. This study was supported by the Singapore Millennium Foundation Research Grant. Ethical approval has been granted by NUS-IRB, Ref No: S-17-107.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was only retrospectively registered, explained by authors with "our study is non-health related intervention and outcomes hence we did not make any clinical trial registration". The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1591 KB - jmir_v22i4e17279_app1.pdf]

References

1. Nelson S, White CF, Hodges BD, Tassone M. Interprofessional team training at the prelicensure level: a review of the literature. Academic Medicine 2017;92(5):709-716. [doi: 10.1097/acm.0000000000001435]
2. Townsend-Gervis M, Cornell P, Vardaman JM. Interdisciplinary rounds and structured communication reduce re-admissions and improve some patient outcomes. West J Nurs Res 2014 Aug;36(7):917-928. [doi: 10.1177/0193945914527521] [Medline: 24658290]
3. Allenbaugh J, Corbelli J, Rack L, Rubio D, Spagnoletti C. A brief communication curriculum improves resident and nurse communication skills and patient satisfaction. J Gen Intern Med 2019 Jul;34(7):1167-1173. [doi: 10.1007/s11606-019-04951-6] [Medline: 30997637]
4. Weaver SJ, Dy SM, Rosen MA. Team-training in healthcare: a narrative synthesis of the literature. BMJ Qual Saf 2014 May;23(5):359-372 [FREE Full text] [doi: 10.1136/bmjqs-2013-001848] [Medline: 24501181]
5. Tan T, Zhou H, Kelly M. Nurse-physician communication - an integrated review. J Clin Nurs 2017 Dec;26(23-24):3974-3989. [doi: 10.1111/jocn.13832] [Medline: 28370533]
6. Chua W, Legido-Quigley H, Jones D, Hassan N, Tee A, Liaw S. A call for better doctor-nurse collaboration: a qualitative study of the experiences of junior doctors and nurses in escalating care for deteriorating ward patients. Aust Crit Care 2020 Jan;33(1):54-61. [doi: 10.1016/j.aucc.2019.01.006] [Medline: 30772155]
7. Tang CJ, Zhou WT, Chan SW, Liaw SY. Interprofessional collaboration between junior doctors and nurses in the general ward setting: A qualitative exploratory study. J Nurs Manag 2018 Jan;26(1):11-18. [doi: 10.1111/jonm.12503] [Medline: 29178213]
8. Buckley S, Ambrose L, Anderson E, Coleman J, Hensman M, Hirsch C, et al. Tools for structured team communication in pre-registration health professions education: a Best Evidence Medical Education (BEME) review: BEME Guide No. 41. Med Teach 2016 Oct;38(10):966-980. [doi: 10.1080/0142159X.2016.1215412] [Medline: 27626840]
9. Fernandez R, Shah S, Rosenman ED, Kozlowski SWJ, Parker SH, Grand JA. Developing team cognition: a role for simulation. Simul Healthc 2017 Apr;12(2):96-103 [FREE Full text] [doi: 10.1097/SIH.0000000000000200] [Medline: 28704287]

10. Liaw SY, Zhou WT, Lau TC, Siau C, Chan SW. An interprofessional communication training using simulation to enhance safe care for a deteriorating patient. Nurse Educ Today 2014 Feb;34(2):259-264. [doi: 10.1016/j.nedt.2013.02.019] [Medline: 23518067]
11. 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]
12. 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]
13. Liaw SY, Soh SL, Tan KK, Wu LT, Yap J, Chow YL, et al. Design and evaluation of a 3D virtual environment for collaborative learning in interprofessional team care delivery. Nurse Educ Today 2019 Oct;81:64-71. [doi: 10.1016/j.nedt.2019.06.012] [Medline: 31330404]
14. Liaw SY, Wu LT, Wong LF, Soh SLH, Chow YL, Ringsted C, et al. "Getting Everyone on the Same Page": interprofessional team training to develop shared mental models on interprofessional rounds. J Gen Intern Med 2019 Dec;34(12):2912-2917. [doi: 10.1007/s11606-019-05320-z] [Medline: 31515736]
15. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
16. Ferguson SL. TeamSTEPPS: integrating teamwork principles into adult health/medical-surgical practice. Medsurg Nurs 2008 Apr;17(2):122-125. [Medline: 18517173]
17. 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]
18. 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]
19. 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]
20. Youngblood P, Harter PM, Srivastava S, Moffett S, Heinrichs WL, Dev P. Design, development, and evaluation of an online virtual emergency department for training trauma teams. Simul Healthc 2008;3(3):146-153. [doi: 10.1097/SIH.0b013e31817bedf7] [Medline: 19088658]
21. Hammick JK, Lee MJ. Do shy people feel less communication apprehension online? The effects of virtual reality on the relationship between personality characteristics and communication outcomes. Computers in Human Behavior 2014 Apr;33:302-310. [doi: 10.1016/j.chb.2013.01.046]
22. Paradis E, Whitehead CR. Beyond the lamppost: a proposal for a fourth wave of education for collaboration. Acad Med 2018 Oct;93(10):1457-1463 [FREE Full text] [doi: 10.1097/ACM.0000000000002233] [Medline: 29620672]
23. Liaw SY, Tan KK, Wu LT, Tan SC, Choo H, Yap J, et al. 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 May 28;21(5):e12537 [FREE Full text] [doi: 10.2196/12537] [Medline: 31140432]
24. Shorey S, Ang E, Yap J, Ng ED, Lau ST, Chui CK. A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study. J Med Internet Res 2019 Oct 29;21(10):e14658 [FREE Full text] [doi: 10.2196/14658] [Medline: 31663857]
25. Johnson W, Lester J. Pedagogical agents: back to the future. AIMag 2018 Jul 01;39(2):33-44 [FREE Full text] [doi: 10.1609/aimag.v39i2.2793]
26. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med Educ 2019 Jun 15;5(1):e13930 [FREE Full text] [doi: 10.2196/13930] [Medline: 31199295]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
ATHCT: Attitudes Toward Interprofessional Health Care Team
CREATIVE: Create Real-Time Experience and Teamwork in Virtual Environment
CUS: Concerned, Uncomfortable, and Safety
ISBAR: Identity, Situation, Background, Assessment, and Recommendation
ISVS: Interprofessional Socialization and Valuing Scale
RCT: randomized controlled trial
TeamSTEPPS: Team Strategies and Tools to Enhance Performance and Patient Safety.

Edited by G Eysenbach; submitted 03.12.19; peer-reviewed by S Gallagher, J Ignacio, J Thompson; comments to author 27.12.19; revised version received 02.01.20; accepted 27.01.20; published 08.04.20.

Please cite as:
Liaw SY, Ooi SW, Rusli KDB, Lau TC, Tam WWS, Chua WL
Nurse-Physician Communication Team Training in Virtual Reality Versus Live Simulations: Randomized Controlled Trial on Team Communication and Teamwork Attitudes
J Med Internet Res 2020;22(4):e17279
URL: https://www.jmir.org/2020/4/e17279
doi:10.2196/17279
PMID:32267235

Viewpoint

Multidimensional Evaluation of Virtual Reality Paradigms in Clinical Neuropsychology: Application of the VR-Check Framework

Stephan Krohn[1,2], MD, PhD; Johanne Tromp[2,3], PhD; Eva M Quinque[3,4], PhD; Julia Belger[3,4], MSc; Felix Klotzsche[2,3], MSc; Sophia Rekers[1,2], MSc; Paul Chojecki[5], Dipl-Psych; Jeroen de Mooij[3,4], BSc; Mert Akbal[2,3,6], Dip Arts; Cade McCall[7], PhD; Arno Villringer[2,3,4], MD, PhD; Michael Gaebler[2,3*], PhD; Carsten Finke[1,2*], MD, PhD; Angelika Thöne-Otto[3,4*], PhD

[1]Department of Neurology, Charité-Universitätsmedizin Berlin, Berlin, Germany
[2]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany
[3]Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[4]Clinic for Cognitive Neurology, University Hospital Leipzig, Leipzig, Germany
[5]Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany
[6]Hochschule der Bildenden Künste Saar, Saarbrücken, Germany
[7]Department of Psychology, University of York, York, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Stephan Krohn, MD, PhD
Department of Neurology
Charité-Universitätsmedizin Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 45050
Email: stephan.krohn@charite.de

Abstract

Virtual reality (VR) represents a key technology of the 21st century, attracting substantial interest from a wide range of scientific disciplines. With regard to clinical neuropsychology, a multitude of new VR applications are being developed to overcome the limitations of classical paradigms. Consequently, researchers increasingly face the challenge of systematically evaluating the characteristics and quality of VR applications to design the optimal paradigm for their specific research question and study population. However, the multifaceted character of contemporary VR is not adequately captured by the traditional quality criteria (ie, objectivity, reliability, validity), highlighting the need for an extended paradigm evaluation framework. To address this gap, we propose a multidimensional evaluation framework for VR applications in clinical neuropsychology, summarized as an easy-to-use checklist (VR-Check). This framework rests on 10 main evaluation dimensions encompassing cognitive domain specificity, ecological relevance, technical feasibility, user feasibility, user motivation, task adaptability, performance quantification, immersive capacities, training feasibility, and predictable pitfalls. We show how VR-Check enables systematic and comparative paradigm optimization by illustrating its application in an exemplary research project on the assessment of spatial cognition and executive functions with immersive VR. This application furthermore demonstrates how the framework allows researchers to identify across-domain trade-offs, makes deliberate design decisions explicit, and optimizes the allocation of study resources. Complementing recent approaches to standardize clinical VR studies, the VR-Check framework enables systematic and project-specific paradigm optimization for behavioral and cognitive research in neuropsychology.

(J Med Internet Res 2020;22(4):e16724) doi:10.2196/16724

KEYWORDS

virtual reality; neuropsychology; cognition; research design

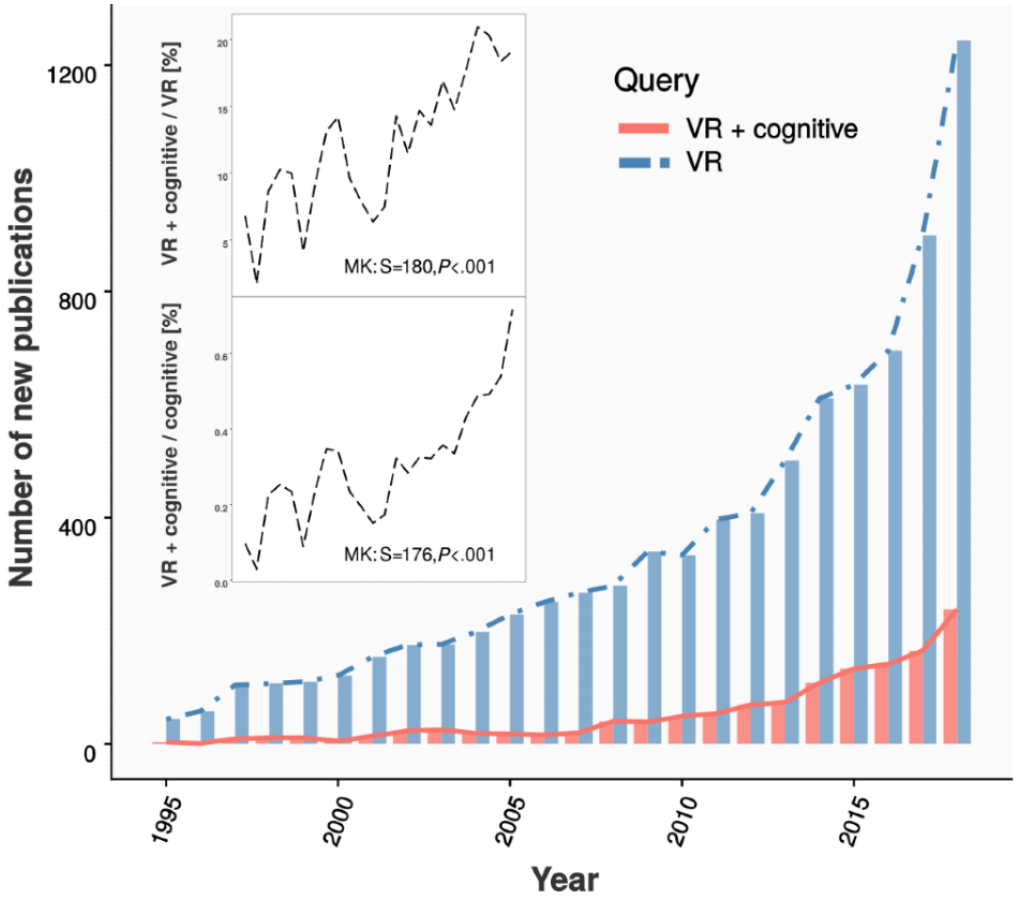
Introduction

Over the past few decades, virtual reality (VR) has emerged as one of the most rapidly advancing technologies of the 21st century, attracting substantial attention from a variety of scientific disciplines, including neuroscience. VR may be regarded as an umbrella term subsuming the real-time presentation of a computer-generated environment to a human user. Users perceive the environment through visual or multisensory stimulation and interact with it through reciprocal data exchange with the computer system, such that VR represents an advanced form of human-computer interaction [1]. VR can be broadly categorized into nonimmersive applications 2-dimensional (2D) screen presentations with interaction devices such as a keyboard or a joystick and immersive applications that are more complex and require the integration of computers with further devices such as head-mounted displays (HMDs), VR controllers, or body-tracking sensors. These immersive systems enable users to experience the virtual environment concealed from the outside world and interact with it based on head or body movements.

In the context of developing paradigms for clinical research, VR provides scientists with a unique combination of extensive design possibilities and strong experimental control. Consequently, VR-based approaches are increasingly being pursued in biomedical research and specifically with respect to investigating cognitive function with VR (Figure 1). As a result, a fast-growing number of neuropsychological VR paradigms are being developed [1-12], paralleled by decreasing costs of hardware components and the increasing availability of open-access software systems for creating new VR paradigms in a customized manner [13-16]. Although these advancements open up many opportunities to investigate the clinical potential of VR, they increasingly present researchers with the challenge of defining the optimal paradigm to answer the research question at hand and leverage the advantages of the technology. Screening the VR literature for suitable paradigms, for instance, how should one evaluate the strengths and weaknesses of a particular paradigm, weigh them against each other, and systematically compare quality across several candidate tasks? Similarly, when developing an experimental VR paradigm de novo, what task features are important to consider in the design process, which qualities should an *ideal* VR task possess, and are there trade-offs in these qualities on which a deliberate design decision must be made?

In this methodological viewpoint paper, we propose a pragmatic framework to address these questions and advance the development of VR-based research tools. To motivate our approach, we first review task evaluation based on the traditional psychometric quality criteria. We contrast these endeavors with the extensive degrees of freedom in state-of-the-art VR, illustrating that the traditional quality criteria alone are inadequate to capture the multifaceted nature of VR paradigms comprehensively. To overcome this gap, we propose a general and multidimensional evaluation framework for neuropsychological VR paradigms in the form of a checklist (VR-Check), and we illustrate the application of this framework in a concrete research project. In the following sections, we focus on VR paradigms for neuropsychological assessment, rather than rehabilitation or cognitive training paradigms. Whereas many of the VR-Check dimensions will be equally relevant to training and rehabilitation tools, we here avoid a conflation of diagnostic and therapeutic VR applications for clarity.

Figure 1. Temporal trends in the biomedical virtual reality literature. The PubMed database was searched for unique novel publications in the years 1995-2018 with the queries "Virtual Reality" (VR), "Virtual Reality" AND "cognitive" (VR + cognitive), and "cognitive." Absolute new publication numbers for the former 2 queries are displayed as bars (search: September 2019). As absolute publications rose for both the VR and cognitive query, we computed the respective ratios of publication numbers over time, as shown in the inset. The proportion of annual VR + cognitive PubMed hits over all VR PubMed hits has risen to approximately 20% over the last 20 years, and nonparametric Mann-Kendall (MK) trend analysis indicates a monotonic upward trend of this proportion (S: sample estimate; positive numbers indicate upward trend). A similar temporal trend was observed for the ratio of VR + cognitive over all cognitive PubMed hits, although this proportion remains well under 1%.



Evaluation Criteria in Classical Neuropsychological Tasks

Neuropsychological assessment tools have a long-standing history in clinical neuropsychology, with several tasks still widely in use more than half a century after their initial presentation (eg, the Wisconsin Card Sorting Test [17,18], or the Stroop Test [19,20]). Early work before the advent of neuroimaging was primarily driven by the aim to measure closely defined cognitive constructs with a clear link to specific brain areas to answer diagnostic questions not otherwise solvable at the time [21]. These early tests were predominantly evaluated according to the traditional psychometric quality criteria [22]. In brief, test results had to be independent of the experimenter (objectivity), consistently reproducible over repeated measurements (reliability), and should measure the intended construct (validity quality demands that are still widely accepted in cognitive psychology today.

With the introduction of neuroimaging into routine diagnostics, however, the mandate for clinical neuropsychologists has changed. Rather than helping to identify the neuroetiology, neuropsychologists are now faced with requests to predict and rehabilitate everyday functions, calling for a new type of paradigm tailored to do so [21]. In consequence, the need for an additional evaluation criterion, which better captures the relationship of the neuropsychological paradigm to everyday functioning, has been discussed for some time [1,21,23-26]. This relationship has been subsumed under the label *ecological relevance* [1,23]. Although in itself still subject to conceptual refinements, ecological relevance is commonly understood to posit that tests should capture the cognitive demands of daily life as closely as possible, resulting in high face validity [27], increased sensitivity to neurorehabilitation, and improved predictive power for everyday functioning [26]. One landmark publication of a test following these principles is a 1991 paper introducing the Multiple Errands Test (MET) [28] to measure multitasking. The MET comprises a list of shopping-related errands to be performed as a real-life task (ie, in a real shopping mall). Although the test features high ecological relevance, its limitations include reduced objectivity and reliability owing to unforeseeable variations in the real-world mall, high demand for resources to accompany patients in the environment, and not least safety issues and inapplicability to patients with more severe disabilities [29]. Although theoretically appealing, real-life tasks have therefore not entered routine neuropsychological assessment and are unlikely to do so due to the lack of control over the test environment. In sum, the

search for ecologically relevant, yet experimentally well-controlled tasks is still very much ongoing. In this aspect, VR has the potential to facilitate crucial progress in the field.

Overcoming the Limitations of Classical Tasks With Virtual Reality

Creating a virtual world offers many degrees of freedom: from the environment itself to the objects in that environment, and even the physics that govern the world. It is therefore possible to design environments that resemble the real world and its demands much more closely than routine paper-and-pencil tests. At the same time, VR preserves strong control over the experimental conditions (eg, the existence, type, and frequency of distractors, which are uncontrollable in real-life tasks such as the MET). Similarly, safety concerns of real-world tasks are attenuated by VR paradigms, as patients are not exposed to actual physical dangers (eg, Navarro et al [30] who used VR to test the act of crossing the street in stroke patients with neglect). Another advantage concerns the increased flexibility of the paradigm development itself: task modifications are implemented computationally, enabling a task design that specifically caters to the study population under question, the research question of interest, or an individual patient's needs. This increased flexibility also illustrates a further limitation of many classical neuropsychological tasks: the lack of parallel versions. In virtual environments, in contrast, parallel task versions are much more easily created by computational modification. Furthermore, routine neuropsychological assessment is highly personnel dependent, requiring substantial resources in terms of patient assistance and monitoring. In addition, the evaluation of behavioral performance in classical assessment tasks usually requires time-consuming processing and examination of numeric data (eg, calculating scores), which then have to be visualized in a graph or table [1]. In real-life tasks such as the MET, acquisition and evaluation of performance data are even more challenging, as a trained professional has to attend to the patient continuously. VR-based assessment, in contrast, allows for the automatic generation of standardized test scores and reduces the demand for monitoring resources during an assessment. Performance evaluation can be augmented by intuitive feedback to the user (eg, playback), which may be especially beneficial for certain age groups or patient populations [5,8,31]. Finally, the personnel dependence of traditional approaches constitutes one factor limiting the widespread availability of high-quality neuropsychological care (eg, in more rural areas or in patients with restricted mobility). In contrast, VR systems can be employed in patients' homes, offering a long-term perspective of improved ambulatory care and telerehabilitation.

Evaluation Criteria in Neuropsychological Virtual Reality Paradigms

These advantages of VR raise hopes to ameliorate some of the limitations inherent to classical neuropsychological paradigms. However, they also illustrate the multitude of features over which VR paradigms can vary. Currently, neuropsychological VR paradigms are still to be evaluated in the light of the traditional quality criteria, although the latter were initially developed for a fundamentally different kind of assessment, commonly based on paper-and-pencil tests. In general, the traditional psychometric quality criteria remain valid for newly developed tests, including VR paradigms. Nonetheless, along with the increased design possibilities of contemporary VR, new evaluation dimensions emerge above and beyond these classical criteria, highlighting the need for an extended evaluation framework to capture the multidimensional nature of VR applications more adequately. Below, we propose such an evaluation framework that allows for systematic and comparative optimization of VR paradigms in clinical neuropsychology.

VR-Check: Multidimensional Evaluation of Virtual Reality Paradigms

The framework rests on 10 evaluation dimensions, each comprising several subfeatures. These evaluation criteria are summarized in the form of a checklist (VR-Check; see below).

Domain Specificity

This evaluation dimension examines how closely the cognitive domain of interest is targeted by the candidate paradigm. This aspect is especially relevant to VR paradigms, as they differ markedly from classical tasks in both clinical and experimental paradigms: The former usually involves a paper-and-pencil test with task instruction, execution, and evaluation by a trained professional. The latter typically involves the well-controlled presentation of predefined stimuli on a 2D computer screen and the measurement of a predefined set of responses, commonly assessed by interaction devices such as a mouse or a keyboard. In both settings, stimulation is rather unisensory, and participants are limited in their ability to act outside the predefined test space. In contrast, VR allows for increased degrees of behavioral freedom, commonly including the liberty to explore the test environment. Compared with classical tasks in neuropsychology, VR furthermore permits a much higher level of self-initiated action and interactivity as well as the possibility of multisensory stimulation. Although this underscores one particular strength of the technology, these increased degrees of freedom may also recruit other cognitive domains than the one we would like to target. This can make it difficult to ascribe differences in task performance to differences in the cognitive domain under study. Therefore, a VR candidate paradigm should be evaluated on this aspect explicitly. More specifically, it is advisable to (1) consider evidence from existing literature that the candidate task will capture the cognitive domain under scrutiny (eg, are there studies relating the VR task to other assessments whose domain specificity is better established?) and (2) to vet the candidate task for potential domain confounds (eg, how strongly are visual attention or motor components implicated in solving the task?).

Ecological Relevance

VR enables researchers to simulate real-world scenarios while maintaining a high degree of experimental control. Increasing a task's similarity to the actual challenges encountered by patients in the real world may facilitate diagnostic and

rehabilitative approaches that more adequately address the patients' real-life deficits. This line of thought is commonly subsumed as the potential of VR to increase a task's *ecological validity* [1,24,32]. As noted above, there is an ongoing debate on what this umbrella term should and should not include on a conceptual level and whether a more fine-grained approach, perhaps along the axes of representativeness and generalizability [26], would be beneficial. In opting for the term *ecological relevance*, we focus on the patient perspective of everyday functional demands. We thereby deliberately scrutinize potential cognitive deficits of a patient in the domain under study that are likely to translate into real-world outcomes, such as the ability to function in the real-life environment and perform a real-life action. A candidate task is thus evaluated based on how closely it reflects these demands as encountered by the study population of interest. In consequence, a judgment is made on how relevant the paradigm is to the user's everyday life with respect to (1) the virtual environment in which the task is set, (2) the experimental stimuli to which the user is exposed, and (3) the activities performed to solve the task (ie, the user response).

Technical Feasibility

Although a candidate paradigm may possess a variety of desirable properties, one may encounter technical limitations when implementing the paradigm in VR. Technical feasibility is especially important to consider if the paradigm is designed de novo or if previously computerized versions of an existing task are not available or incompatible with state-of-the-art VR setups. We therefore evaluate whether the task can be sensibly implemented in VR in general and whether the implementation is compatible with a head-mounted display (HMD), with a 2D display device such as a tablet or a desktop computer, or both. Moreover, it is important to assess whether user interaction and navigation in the virtual world require further input devices such as VR controllers or a mouse, and if so, which input devices are technically feasible. Importantly, the technical feasibility of a candidate task may be constrained by project-specific factors such as the necessity of using a particular HMD model, a specific interaction device, or an examination room with spatial limitations.

User Feasibility

Candidate paradigms must further be evaluated in terms of feasibility for different user groups. First, is the candidate task expected to be feasible in healthy users, also considering potential differences across different age groups? Second, can one expect it to be viable in the patient population of interest, and are there possibilities to alleviate obstacles to maximize patient feasibility? Third, the task is evaluated on the complexity of the user interaction and navigation in the virtual world: How difficult is it to move and act in the virtual environment? How long will it take for healthy users and patients to learn how to carry out the task, and how intuitive are the controls? Furthermore, task duration and attentional demands might limit user feasibility. Therefore, it must be considered how long the task will take on average and whether the target user group can be expected to focus on the task sufficiently. Moreover, user feasibility may be hampered by VR-induced adverse symptoms and effects (VRISE), which are not only important for safety considerations but also because VRISE are likely to confound task performance [33]. One should therefore judge the paradigm on the likelihood of inducing VRISE such as VR-related kinetosis (*cybersickness*). Finally, it is important to evaluate any ethical concerns the task may implicate, such as the presentation of strong fear-inducing stimuli or safety considerations, as mentioned above (see also Madary and Metzinger for a detailed review of ethical considerations in VR [34]). Although these are relevant aspects to evaluate in any study population, the judgment on what is feasible in the target user group may certainly differ depending on population-specific factors such as health status or age.

In practice, the maximization of user feasibility is linked to development principles from human-computer interaction (HCI) and user experience (UX) research. This includes the application of ergonomic principles and human-centered design that maximizes accessibility and involves users and other stakeholders in an iterative development process [35,36]. Concerning virtual worlds, standard UX heuristics remain crucial [37], although some VR-specific components such as sense of control and multimodal interaction warrant additional consideration, as they have been shown to affect UX [38,39] and may be especially important in potential future multiuser scenarios and in users with neurological disorders [40].

User Motivation

Beyond mere feasibility, user motivation is crucial to ensure that participants will engage in the candidate task, especially in repeated application. To optimize user compliance, it is therefore advisable to evaluate the task with regard to factors that may facilitate user motivation. First, users may be intrinsically motivated to carry out the task due to high expected benefit or face validity of the paradigm. Second, the entertainment factor of the candidate task is judged. Next, one evaluates the possibility of a reward system, both within-session (eg, a virtual reward for successful task completion) and across-session (eg, a high-score system or advancing to higher levels). Furthermore, we examine the possibility of within- and across-session feedback on user performance. These features touch upon a gamification approach to cognitive assessment [41,42], and this represents one aspect in which VR is particularly capable of playing off its strengths against the classical neuropsychological assessments.

Task Adaptability

The ability to adapt the candidate paradigm carries important implications for both clinical and experimental settings. First, it is useful to consider how easily parallel versions of the candidate task can be created, which represents a major limitation of many classical neuropsychological tests. Not least, this aspect also constitutes a prerequisite for applying the paradigm repeatedly, for instance in a pre- and postintervention study design. Second, the task is judged on how well its difficulty can be (parametrically) adjusted. The required levels of difficulty may vary markedly between study populations (eg, patients vs healthy controls, younger vs older participants) or across multiple sessions in repeated within-participant applications. Therefore, the task is evaluated with respect to

experimental parameters that can be effectively manipulated to affect task performance systematically. In addition, it should be considered if task difficulty is adaptable enough to induce sufficient across-participant performance variance and avoid floor and ceiling effects.

Performance Quantification

A further important prerequisite for a suitable candidate paradigm concerns the ability to measure user performance in a quantitative way. One should therefore consider if outcome variables to quantify performance have been defined, or if they can be derived from the data obtained in VR. As behavior in virtual environments can be tracked digitally with high resolution in both time and space, VR offers increasingly multivariate and experimenter-independent performance parameters, facilitating more objective, data-driven, and automated analysis approaches. It is therefore evaluated to what extent the candidate paradigm allows for experimenter-independent performance evaluation.

Notably, task adaptability and automatic performance quantification in VR complement related advances in contemporary psychometrics. As VR paradigms are centered around user interactions with the virtual environment in real time, they offer the possibility of highly dynamic and individualized testing scenarios, enabling more precise and time-efficient assessments in accordance with the ideas of computerized adaptive testing [43,44]. Implementing such a reactive task design also facilitates the inclusion of large item pools with predefined difficulty of sufficient variance as well as real-time scoring (ie, immediate item evaluation) to automatically utilize the most informative items based on the participant's current performance and the assessment's goal.

Immersive Capacities

Another dimension not adequately captured by traditional test criteria concerns the capacity of VR systems to create the illusion of being located in the virtual world. There is an ongoing conceptual debate about the technical terms describing this phenomenon, specifically *immersion* and its relation to and disentanglement from the notion of *presence* [45-48]. For paradigm development, we follow Slater in the distinction that immersion describes a VR system's objective technical properties that support natural sensorimotor contingencies, whereas presence refers to the subjective illusion of *being there* in the environment as a subjective correlate of immersion [49-51]. Accordingly, one first evaluates the degree of immersion as specified by task factors and the VR system necessary to present this task. Second, the likelihood that the task (in its final implementation) will facilitate the illusion of being in the virtual environment is considered, and ideally, this judgment is informed by prior empirical evidence using presence questionnaires. This evaluation is important for two reasons: first, the degree to which participants feel present in the virtual environment may either have direct implications for the research question or represent a latent factor influencing task performance or user engagement, constituting a potential confound. Second, the state-of-the-art VR technology raises hopes that a higher degree of presence could be beneficial in diagnostic assessment, cognitive training outcome, or UX (with respect to the latter, see Brade et al [52] and Lorenz et al [53]). Indeed, there is some evidence that increased presence may have a positive impact on participants' cognitive performance, for instance, regarding fact learning [54] or memory encoding [55], although potential benefits of increased presence in clinical assessment remain to be explored.

Training Feasibility

A further consideration concerns the feasibility of the candidate paradigm to serve as the basis for a training tool. In a one-time application setting (eg, purely diagnostic assessment), the paradigm needs to fulfill fewer requirements compared with a repeated-application setting (eg, implementation of a cognitive training tool). First, one should evaluate whether there are any practical obstacles to the repeated application of the paradigm. This concerns the logistics of the VR system used for the task: Can the task be administered in multiple sites or at home, or must the user be tested in a specialized laboratory, for instance, due to the necessity of specific interaction devices such as a VR treadmill or a cave automatic virtual environment (CAVE) [56]? Potential caveats in user feasibility may yield cumulative disadvantages in the training scenario (eg, mild risk of cybersickness may be acceptable in a one-time application but could decrease compliance when repeated with high frequency). Second, one determines if the necessary prerequisites of task adaptability are met (eg, the possibility to create parallel versions, effective manipulation of difficulty). Third, it is important to consider to what extent the paradigm offers the possibility of conveying cognitive strategies for compensatory training and how these could be implemented (eg, by leveraging the extensive cueing possibilities in VR [1]). Furthermore, the likelihood of transfer effects is examined and if there is any empirical indication of their expected quality regarding near vs far transfer.

Predictable Pitfalls

Furthermore, it is advisable to vet the candidate paradigm for predictable pitfalls. As in any clinical study, implementing a VR paradigm for cognitive assessment requires time, know-how, and monetary resources that must be weighed against potential knowledge gains and patient benefit. To optimize the potential of the research endeavor, one first evaluates how well the candidate paradigm adheres to the task requirements of the current research project and if the paradigm can be modified to maximize this adherence. Second, it is considered to what extent the application of the candidate paradigm constitutes a reasonable allocation of the study resources. Not least, scrutinizing potential pitfalls early on in the development process also serves as a quality check when designing a VR paradigm de novo.

Application of the VR-Check Framework

The following sections illustrate how systematic evaluation with the VR-Check framework can guide the decision-making process in defining a neuropsychological VR paradigm for a specific research project (see Figures 2 and 3).

Figure 2. Workflow of the VR-Check evaluation. Panel A summarizes the main general steps for paradigm optimization with the virtual reality check framework. Panel B shows how this workflow applies to an exemplary research project on the assessment of spatial cognition with immersive VR (see main text). Four candidate tasks were evaluated: the Starmaze (STM), the Virtual Memory Task (VMT), the Virtual Morris Water Maze (vMWM), and the Cognitive Map Task (CMT). See Figure 3 and the main text for details on the VR-Check evaluation. Panel C visualizes the result of the optimization procedure: the immersive Virtual Memory Task (imVMT). The screenshot displays a third-person view of a user memorizing the locations of everyday objects on the table. HMD: head-mounted display; VR: virtual reality.

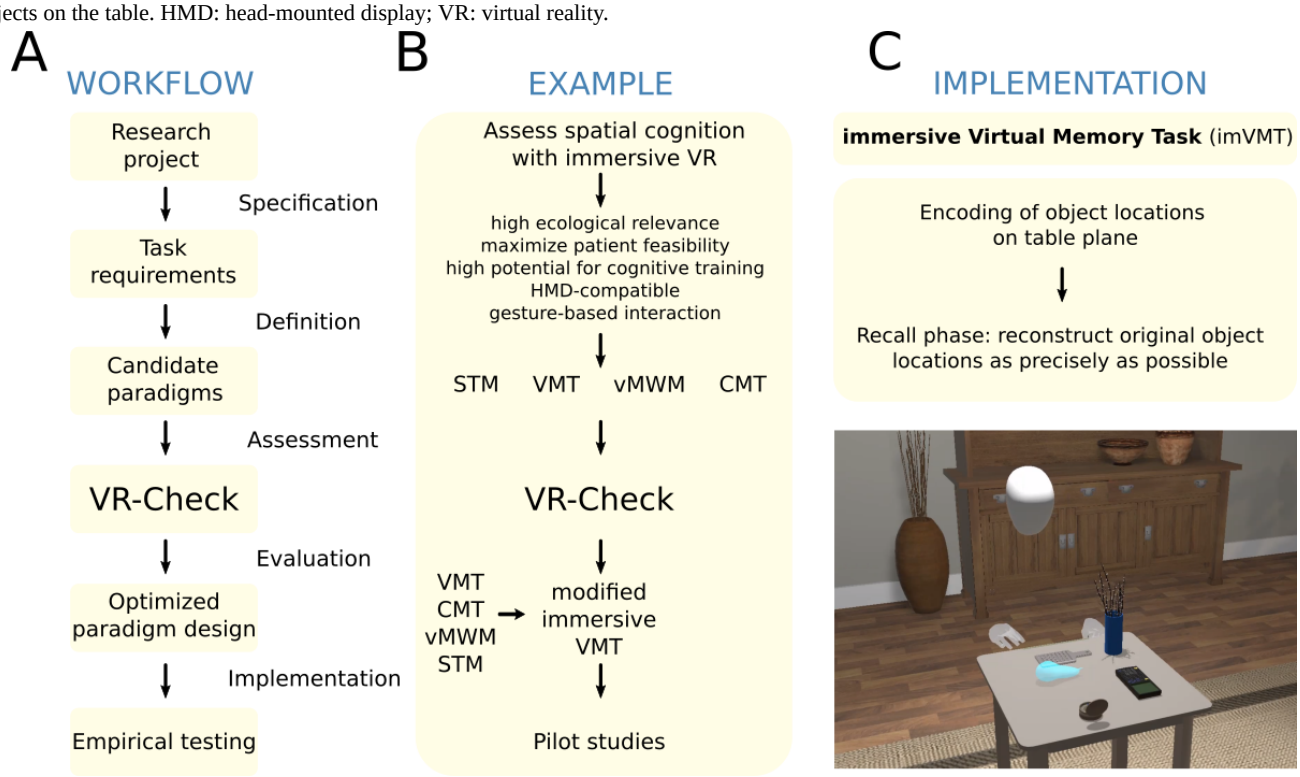
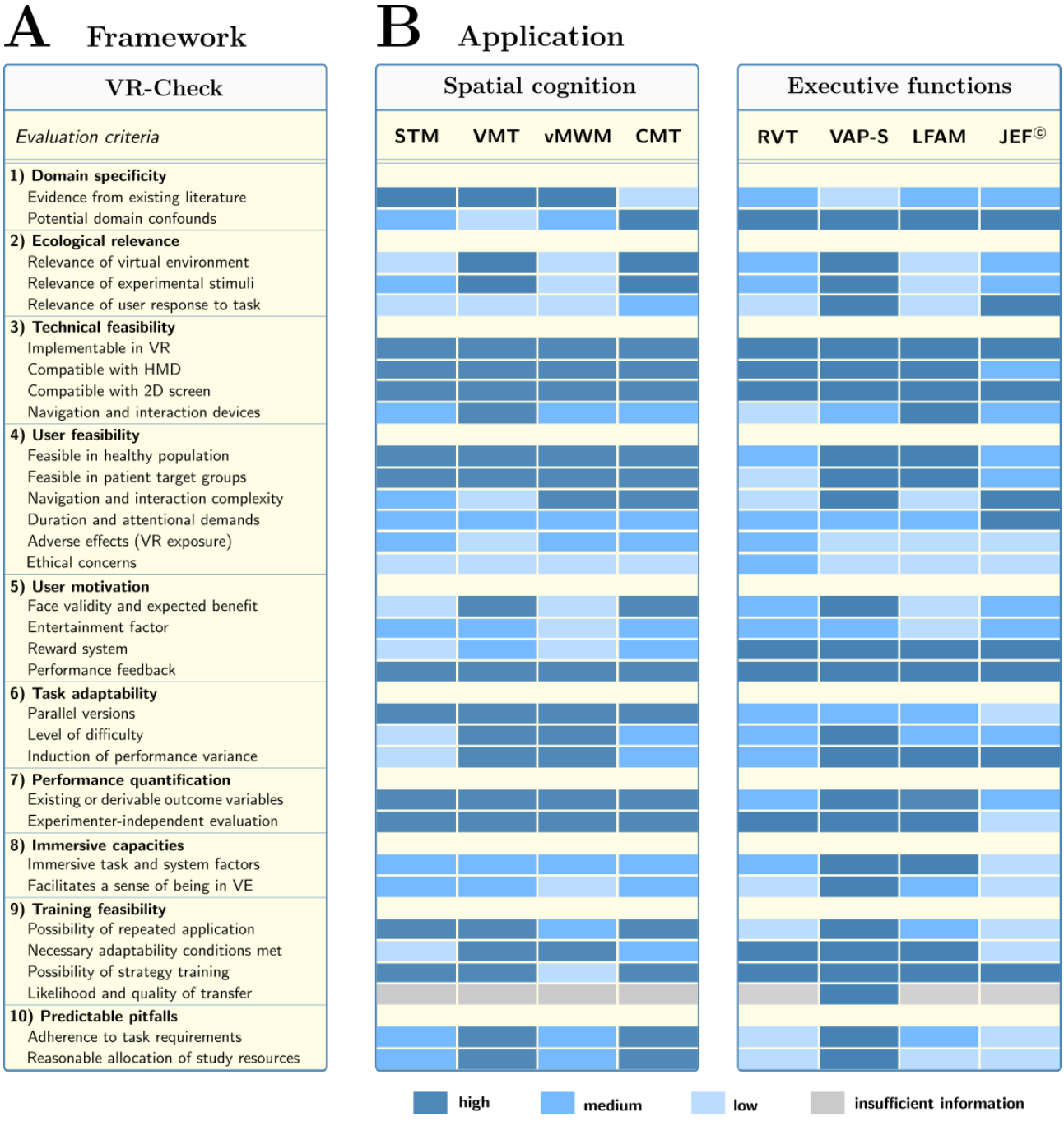


Figure 3. The VR-Check framework for virtual reality paradigms in neuropsychology. Panel A summarizes the evaluation dimensions in the form of a checklist. Panel B visualizes the application of the framework for the exemplary cases of assessing spatial cognition and executive functions with immersive VR. The color schemes display the item-wise consensus ratings on the degree to which the feature is fulfilled for a given paradigm. The evaluation procedure is illustrated for the Starmaze (STM), Virtual Memory Task (VMT), the Virtual Morris Water Maze (vMWM), and the Cognitive Map Test (CMT) for assessing spatial abilities, and the Ride in a Virtual Town (RVT), the Virtual Action Planning-Supermarket (VAP-S), the Look For A Match (LFAM) task as well as the Jansari assessment of Executive Functions (JEF) for assessing executive functions. For the given task requirements, the VMT, CMT, and the VAP-S emerged as the most suitable paradigms for the development of an immersive VR application, as detailed in the main text. 2D: 2-dimensional; HMD: head-mounted display; VE: virtual environment; VR: virtual reality.



Evaluation Workflow

First, the properties required of the VR paradigm are defined. Of note, these task requirements are necessarily project specific, such that the relative weight of the various evaluation dimensions will naturally differ across projects. Furthermore, to facilitate a comparative evaluation of tasks across the VR-Check features, a semiquantitative rating is applied to evaluate if a particular feature applies to the candidate task to a high, medium, or low degree, or if there is insufficient information to make an informed judgment (eg, asking *How high is the degree of ecological relevance of the virtual environment in paradigm X to the study population of interest?*). Researchers are thus able to systematically go through the list of features and judge each candidate paradigm according to the description above. For the de novo development of VR paradigms, the same process is applied to competing ideas or

prototypes, yielding an explicit account of which task features need to be maximized.

Example Project: Task Requirements and Candidate Paradigms

Here, the results of this evaluation procedure are presented for an exemplary research project emanating from our consortium. It is important to note that the following outcomes do not represent a judgment on the value of the paradigms *per se*, but the outcomes rather provide an illustration of the evaluation process itself and how it can inform project-specific paradigm optimization.

The goal of the exemplary research project was to apply immersive VR for the neuropsychological assessment of spatial cognition and executive functions. Suitable tasks were required to (1) be relevant to participants' everyday life, (2) be feasible in a wide range of neurological patient populations, (3) inform the development of a subsequent cognitive training tool, (4) be implemented with an HMD, and (5) allow for natural user interaction based on body-tracking devices and gesture recognition. The technical setup for implementing the task included an Oculus Rift headset (Oculus, Facebook Technologies, LLC, Menlo Park, CA), a Leap Motion controller (Leap Motion, Inc, San Francisco, CA) for hand-tracking, and a Microsoft Kinect sensor (for Windows; Microsoft Corporation, One Microsoft Way, Redmond, WA) to support body tracking. Software implementation rested on the Unity game engine (version 2019.2.11f1; Unity Technologies, San Francisco, CA) and Blender graphics suite (version 2.79; Blender, Amsterdam, the Netherlands) for VR development, a custom-built communication middleware, and a custom Web interface for data management.

Candidate paradigms were identified by literature screening of existing VR tasks and in-house paradigms from January to May 2018. All candidate tasks were assessed along the VR-Check dimensions by an interdisciplinary research consortium, including 3 cognitive neuroscientists, 2 physicians, and 4 clinical neuropsychologists. None of the team members was involved in the creation of the considered tasks or had any conflict of interest. Each paradigm was presented to the group by varying team members, followed by subsequent rating. If missing information or divergences in the individual ratings were identified, these issues were addressed in the subsequent session. A consensus was reached on all ratings through group discussion. Potential ties were to be resolved by the senior scientists, although no ties occurred for the considered paradigms. Figure 3 visualizes the ratings for a subset of four promising candidate tasks in each cognitive domain. The description below is limited to a brief account of the most decisive aspects; interested readers are referred to Multimedia Appendix 1 for a detailed point-by-point description of the systematic evaluation.

Example Project: Spatial Cognition

With respect to spatial cognition, the evaluation process is illustrated in the following candidate tasks: (1) the *Starmaze* (STM) [57-59], a VR adaptation of a rodent paradigm [60] to differentiate egocentric from allocentric navigation strategies, in which the user navigates through a point-symmetric star-shaped labyrinth to find a target; (2) the *Virtual Memory Task* (VMT) [32], a computerized spatial memory task similar to an existing real-life task [61], in which participants are required to memorize locations of everyday objects on a table; (3) the *Virtual Morris Water Maze* (vMWM) [62,63], a VR adaptation of the classical place navigation task originating from rodent research [64], in which participants learn to navigate to a concealed platform; and (4) the *Cognitive Map Task* (CMT) [65,66], a spatial learning paradigm in which participants have to construe, maintain, and retrieve a cognitive map of a virtual town by learning and finding landmarks.

On the basis of the assessment along the VR-Check dimensions, the STM and the vMWM although certainly highly appropriate paradigms for other research questions were judged to be less favorable for our purposes due to limited ecological relevance, task adaptability, and training potential. In contrast, the VMT emerged as the paradigm that most closely adhered to our task requirements, made explicit through point-by-point assessment along the VR-Check dimensions: besides high ecological relevance to our target population, favorable user feasibility, and excellent adaptability, it avoids some of the caveats of other candidate paradigms (such as high navigation complexity or the risk of adverse effects) and demands comparatively moderate implementation efforts, rendering it the optimal allocation of our study resources. Nonetheless, the VMT is limited to an assessment of spatial memory capacities due to the comparatively narrow domain target. In terms of assessing navigational abilities, the CMT was evaluated to be the most suitable starting point for the development of an immersive paradigm because of favorable ecological relevance, user feasibility and motivation, and high training potential. Notwithstanding, our evaluation process also identified potential improvements of the CMT that have to be addressed in the development process, such as a more fine-grained adaptation of difficulty.

Example Project: Executive Functions

Executive functions is an umbrella term for a multifaceted construct, including several interconnected high-level cognitive abilities that serve ongoing, goal-directed actions [67]. Subdomains include planning, problem solving, monitoring, working memory, inhibition, and task switching, and despite ongoing terminological disambiguations, there is relative agreement on the complexity and superordinate coordination role of executive functions and their importance regarding human adaptive behavior [67-70]. For a comprehensive review of executive functions paradigms in VR, see Parsons [25] and Valladares-Rodríguez et al [71]. As mentioned above, we exemplify evaluation outcomes in 4 candidate paradigms: (1) *a Ride in a Virtual Town* (RVT) [72], a prospective memory task featuring a car drive using real car components as interaction devices while completing a list of errands; (2) the *Virtual Action Planning-Supermarket* (VAP-S) [31], a grocery shopping task; (3) the *Look For A Match* (LFAM) task [73], an adaptation of the Wisconsin Card Sorting Task to a virtual beach environment; (4) the *Jansari assessment of Executive Functions* (JEF) [74], a multistep office task requiring multitasking to prepare a meeting on time.

Resulting from the VR-Check evaluation, some inconsistencies with our task requirements were identified for the LFAM (limited ecological relevance to our target populations, drawbacks in user motivation), the RVT (risk of adverse effects, incompatibilities with our interaction requirements, ecological relevance limited to drivers, ethical concerns about loss of driving capability in patient population, limited training feasibility), as well as the JEF (user feasibility limited to higher-functioning populations, ecological relevance restricted to a subgroup of our target population, incompatibilities with our immersive system factors, limited training feasibility due to caveats in task adaptability). The VAP-S, in contrast, was evaluated to be highly consistent with the project's task requirements regarding user feasibility, technical requirements, ecological relevance, and training potential, while demanding reasonable implementation efforts. The VAP-S was therefore esteemed the most favorable basis for the development of an immersive executive functions paradigm. Nonetheless, the systematic evaluation also highlighted potential caveats of the paradigm (limited domain specificity, technical solution required for large-scale multidirectional locomotion), which can thus be explicitly optimized in the implementation process.

Discussion

To leverage the potential of VR in neuropsychology, researchers are increasingly challenged with optimizing the experimental paradigm to address the study question at hand. The body of literature on biomedical VR applications is growing fast, and the importance of cognitive research within this field is steadily increasing (Figure 1), supported by the increasing availability of VR hardware and software systems. With these developments, the need arises for a new methodological framework on systematic paradigm evaluation. This gap is aggravated further by the inability of the traditional quality criteria to capture the multifaceted nature of contemporary VR. With this work, we aim to address this gap with a multidimensional evaluation protocol for VR applications in neuropsychology, summarized as an easy-to-use checklist (VR-Check, Figure 3).

Paradigm Optimization and Across-Domain Trade-Offs

The systematic evaluation approach of the VR-Check framework raises the general question of what constitutes an ideal VR paradigm for neuropsychological research. Surely, if we defined an entertaining, highly adaptable, easy-to-play, easy-to-implement, highly immersive task that is viable for any user group, targets a well-circumscribed cognitive domain, adequately captures cognitive deficits relevant to everyday functioning as measured objectively by experimenter-independent performance outcomes, and which can be applied repeatedly to induce systematic improvement in both the tested and further cognitive domains, such a paradigm would be welcomed by researchers and clinicians alike.

However, as a corollary of the multidimensional nature of VR, such an endeavor is unrealistic for two principled reasons: first, what is desired of the task is tightly linked to the research question of interest. In consequence, there is no general profile of objectively desirable properties. Although minimal requirements regarding user feasibility or technical implementation must be met by any clinical paradigm, the relative importance of the various domains will differ markedly over research applications and target populations. Indeed, the VR-Check framework serves precisely the purpose of prioritizing which domains are more important than others to address a given research question. This flexibility toward the study purpose enables researchers to weigh the different dimensions against each other and maximize the adherence to their project-specific requirements.

Second, the VR-Check framework illustrates a qualitative difference with respect to the interaction among evaluation criteria in that some are logically congruous, whereas others imply reciprocal incongruities. For instance, a paradigm featuring high training feasibility must also fulfill a variety of requirements concerning technical feasibility, user feasibility, and task adaptability and is more likely to be judged favorably in terms of user motivation because these dimensions, to some extent, inform the evaluation of training feasibility. In contrast, other comparisons yield across-domain trade-offs. Specifically, this concerns the relationship between cognitive domain specificity and ecological relevance. In the attempt to target a specific cognitive domain with high precision, the recruitment of other cognitive domains must be minimized. However, this is rarely the case in everyday functioning, when a multitude of cognitive domains are engaged simultaneously. A VR paradigm featuring high ecological relevance will therefore necessarily concede some domain specificity by recruiting other domains than the one intended. Inversely, a VR paradigm featuring high domain specificity permits only limited relevance to cognitive functioning in real life because of an artificially narrow cognitive target. As a result of this incongruousness, a deliberate decision must be made on the trade-off between domain specificity and ecological relevance.

A similar point arises with respect to the relationship between ecological relevance and experimental control. Although both task and environment are highly controllable in VR, the increased degrees of behavioral freedom can result in less controlled participant behavior as compared with classical neuropsychological assessments. This behavioral freedom comes with an increased number of error sources not related to the cognitive task itself, such as visual attention, motor control, or navigational demands. In the research context, we can increase experimental control by restricting what the participant can and cannot do in VR. However, this again entails decreased relevance to everyday functioning, as real-life behavior offers similarly many degrees of freedom and also encompasses a multitude of error sources.

In summary, the properties required of a VR paradigm are dependent on the research question at hand, and there are inevitable across-dimension trade-offs in paradigm design. These aspects necessitate deliberate design decisions to permit the project-specific optimization of the VR paradigm. The VR-Check framework guides this optimization process because it allows for a systematic account of how well a paradigm adheres to the project-specific requirements and because it makes these design decisions explicit.

Toward Improved Standardization of Clinical Virtual Reality Applications

Although the assets of VR for clinical research have been examined before, previous approaches have predominantly addressed general favorable properties of the technology [1-3,5,6] or focused on specific aspects of VR application such as avoiding VR-related adverse effects [33,75], improving UX [36,38] or ethical adversities [34]. Other studies have suggested design considerations derived from specific VR applications [76], focused on rehabilitative tools [77], or dealt with clinical study design for VR-based therapies [78].

The VR-Check framework complements these studies, as it specifically targets the project-specific optimization of the paradigm (rather than the study) design and explicitly addresses cognitive and behavioral research, and because it provides researchers with a general and easy-to-use evaluation tool. However, even though the application of the framework was highly informative in the exemplary research project, some limitations of this work deserve mentioning. First, the application of the framework was limited to the assessment of spatial cognition and executive functions, such that further research is necessary to corroborate its utility with respect to other research questions. Moreover, current evaluation outcomes are limited to semiquantitative assessment and consensus ratings, warranting further work to solidify more quantitative approaches and assess the rates of agreement across individual raters. Furthermore, it should be noted that paradigms that have been applied more often in literature might lead researchers to evaluate them more favorably simply because existing evidence makes these paradigms easier to judge. However, the fact that a paradigm may be more established in the literature does not necessarily imply that it is better suited for the study question at hand. Finally, we focused here on the design optimization of VR paradigms for neuropsychological assessment. Nonetheless, the value of therapeutic VR applications is becoming increasingly apparent [78-80], and there is an important interplay between assessment and rehabilitation, especially with regard to devising individualized therapies that cater to the patient's specific deficits (*precision medicine*). Although many of the VR-Check dimensions appear relevant to clinical VR tasks in general (eg, technical and user feasibility, adaptability, or outcome quantification), future work must investigate if the protocol is also applicable to VR tools for cognitive training and rehabilitation, or to what extent the framework must be modified to enable paradigm optimization for these applications.

Even with these limitations in mind, the VR-Check framework represents a first step toward the standardized optimization of VR paradigms in clinical neuropsychology. The potential of contemporary VR is contrasted by a relative scarcity of consensus on what should be regarded as best practice when applying VR in clinical research. With respect to paradigm development, the VR-Check framework can inform this discussion. Even with optimal paradigm design, however, proof of clinical utility ultimately requires high-quality empirical evidence such as randomized controlled trials (RCTs). In this context, the newly established Virtual Reality Committee of Outcomes Research Experts (VR-CORE) has recently suggested a framework for the development and validation of VR-based therapies [78]. This framework features 3 study phases (VR1-VR3) similar to the phase I-III model of pharmacological intervention studies. Although the authors' approach focuses explicitly on VR treatments, their systematic methodological framework is similar in spirit to our suggestions, and the two approaches complement each other (paradigm design optimization and study design optimization). For instance, the authors' call for human-centered design in early VR treatment study phases (VR1) is matched by our explicit focus on the patient perspective in the domains of technical feasibility, user feasibility, user motivation, and ecological relevance. The intermediate trial phase (VR2) is concerned with initial feasibility testing and can thus be regarded as the *study design* counterpart to the *paradigm design* feasibility dimensions of the VR-Check framework. The later VR-CORE phase (VR3) concerns RCTs to examine VR treatment efficacy vs a control condition. As such, the extension of the VR3 phase to neuropsychological assessment tasks seems natural: where a VR treatment must show intervention efficacy, a VR assessment task must show discriminatory or predictive power in empirical evaluation.

As methodological guidelines such as the VR-CORE recommendations and the VR-Check framework are further developed, they may ultimately synergize in pursuit of a more rigorous, systematic, and well-informed protocol for the development of clinical VR applications.

Acknowledgments

This work emanates from the research consortium *VReha-virtual worlds for digital diagnostics and cognitive rehabilitation*, funded by the German Federal Ministry for Education and Research (BMBF grant 13GW0206).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed description of the VR-Check application.
[DOCX File , 29 KB - jmir_v22i4e16724_app1.docx]

References

1. Rizzo A, Koenig ST. Is clinical virtual reality ready for primetime? Neuropsychology 2017 Nov;31(8):877-899. [doi: 10.1037/neu0000405] [Medline: 29376669]
2. Rizzo A, Buckwalter JG, van der Zaag C, Neumann U, Thiébaux M, Chua C, et al. Virtual Environment Applications in Clinical Neuropsychology. In: Proceedings of the 2000 IEEE Virtual Reality Conference. 2000 Presented at: VR'00; March 18-22, 2000; New Brunswick, USA p. 63-70. [doi: 10.1109/vr.2000.840364]
3. Schultheis MT, Rizzo A. The application of virtual reality technology in rehabilitation. Rehabil Psychol 2001;46(3):296-311. [doi: 10.1037/0090-5550.46.3.296]
4. Schultheis MT, Himelstein J, Rizzo A. Virtual reality and neuropsychology: upgrading the current tools. J Head Trauma Rehabil 2002 Oct;17(5):378-394. [doi: 10.1097/00001199-200210000-00002] [Medline: 12802250]
5. Rizzo A, Schultheis M, Kerns KA, Mateer C. Analysis of assets for virtual reality applications in neuropsychology. Neuropsychol Rehabil 2004;14(1-2):207-239. [doi: 10.1080/09602010343000183]
6. Rizzo A, Kim GJ. A SWOT analysis of the field of virtual reality rehabilitation and therapy. Presence 2005;14:119-146. [doi: 10.1162/1054746053967094]
7. Parsons TD, Rizzo A. Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. Cyberpsychol Behav 2008 Feb;11(1):17-25. [doi: 10.1089/cpb.2007.9934] [Medline: 18275308]
8. Rizzo A, Bowerly T, Buckwalter JG, Klimchuk D, Mitura R, Parsons TD. A virtual reality scenario for all seasons: the virtual classroom. CNS Spectr 2006 Jan;11(1):35-44. [doi: 10.1017/s1092852900024196] [Medline: 16400254]
9. Lalonde G, Henry M, Drouin-Germain A, Nolin P, Beauchamp MH. Assessment of executive function in adolescence: a comparison of traditional and virtual reality tools. J Neurosci Methods 2013 Sep 30;219(1):76-82. [doi: 10.1016/j.jneumeth.2013.07.005] [Medline: 23867080]
10. Negu A, Matu S, Sava FA, David D. Virtual reality measures in neuropsychological assessment: a meta-analytic review. Clin Neuropsychol 2016 Feb;30(2):165-184. [doi: 10.1080/13854046.2016.1144793] [Medline: 26923937]
11. Teel E, Gay M, Johnson B, Slobounov S. Determining sensitivity/specificity of virtual reality-based neuropsychological tool for detecting residual abnormalities following sport-related concussion. Neuropsychology 2016 May;30(4):474-483. [doi: 10.1037/neu0000261] [Medline: 27045961]
12. Wiener JM, Carroll D, Moeller S, Bibi I, Ivanova D, Allen P, et al. A novel virtual-reality-based route-learning test suite: assessing the effects of cognitive aging on navigation. Behav Res Methods 2019 Jun 24:- [Epub ahead of print]. [doi: 10.3758/s13428-019-01264-8] [Medline: 31236900]
13. Grübel J, Weibel R, Jiang M, Hölscher C, Hackman D, Schinazi V. EVE: A Framework for Experiments in Virtual Environments. In: Proceedings of the Biennial Conference of the German Society for Cognitive Science. 2016 Presented at: KogWis'16; September 26-30, 2016; Bremen, Germany p. 26-30. [doi: 10.1007/978-3-319-68189-4_10]
14. Howard MC. Virtual reality interventions for personal development: a meta-analysis of hardware and software. Hum–Comput Interact 2018;34(3):205-239. [doi: 10.1080/07370024.2018.1469408]
15. Zhao H, Thrash T, Wehrli S, Hölscher C, Kapadia M, Grübel J, et al. A networked desktop virtual reality setup for decision science and navigation experiments with multiple participants. J Vis Exp 2018 Aug 26(138):-. [doi: 10.3791/58155] [Medline: 30199016]
16. Brookes J, Warburton M, Alghadier M, Mon-Williams M, Mushtaq F. Studying human behavior with virtual reality: The Unity Experiment Framework. Behav Res Methods 2019 Apr 22:- [Epub ahead of print]. [doi: 10.3758/s13428-019-01242-0] [Medline: 31012061]
17. Berg EA. A simple objective technique for measuring flexibility in thinking. J Gen Psychol 1948 Jul;39:15-22. [doi: 10.1080/00221309.1948.9918159] [Medline: 18889466]
18. Eling P, Derckx K, Maes R. On the historical and conceptual background of the Wisconsin Card Sorting Test. Brain Cogn 2008 Aug;67(3):247-253. [doi: 10.1016/j.bandc.2008.01.006] [Medline: 18328609]
19. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18(6):643-662. [doi: 10.1037/h0054651]
20. MacLeod CM. Half a century of research on the Stroop effect: an integrative review. Psychol Bull 1991 Mar;109(2):163-203. [doi: 10.1037/0033-2909.109.2.163] [Medline: 2034749]
21. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. Neuropsychol Rev 2003 Dec;13(4):181-197. [doi: 10.1023/b:nerv.0000009483.91468.fb] [Medline: 15000225]
22. Moosbrugger H, Kelava A. Testtheorie und Fragebogenkonstruktion. New York: Springer; 2012.
23. Franzen MD, Wilhelm KL. Conceptual foundations of ecological validity in neuropsychological assessment. In: Sbordone RJ, Long C, editors. Ecological Validity of Neuropsychological Testing. Delray Beach, USA: CRC Press; 1996:91-112.
24. Spooner DM, Pachana NA. Ecological validity in neuropsychological assessment: a case for greater consideration in research with neurologically intact populations. Arch Clin Neuropsychol 2006 May;21(4):327-337 [FREE Full text] [doi: 10.1016/j.acn.2006.04.004] [Medline: 16769198]
25. Parsons TD. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci 2015;9:660 [FREE Full text] [doi: 10.3389/fnhum.2015.00660] [Medline: 26696869]

26. Dawson DR, Marcotte TD. Special issue on ecological validity and cognitive assessment. Neuropsychol Rehabil 2017 Jul;27(5):599-602. [doi: 10.1080/09602011.2017.1313379] [Medline: 28480821]
27. Bornstein RF. Face validity in psychological assessment: implications for a unified model of validity. Am Psychol 1996;51(9):983-984. [doi: 10.1037//0003-066x.51.9.983]
28. Shallice T, Burgess PW. Deficits in strategy application following frontal lobe damage in man. Brain 1991 Apr;114 (Pt 2):727-741. [doi: 10.1093/brain/114.2.727] [Medline: 2043945]
29. Pedroli E, Serino S, Giglioli A, Pallavcini F, Cipresso P, Riva G. The use of virtual reality tools for the assessment of executive functions and unilateral spatial neglect. In: Hu F, Lu J, Zhang T, editors. Virtual Reality Enhanced Robotic Systems for Disability Rehabilitation. Hershey, USA: IGI Global; 2016.
30. Navarro MD, Lloréns R, Noé E, Ferri J, Alcañiz M. Validation of a low-cost virtual reality system for training street-crossing. A comparative study in healthy, neglected and non-neglected stroke individuals. Neuropsychol Rehabil 2013;23(4):597-618. [doi: 10.1080/09602011.2013.806269] [Medline: 23767963]
31. Klinger E, Chemin I, Lebreton S, Marié RM. Virtual action planning in Parkinson's disease: a control study. Cyberpsychol Behav 2006 Jun;9(3):342-347. [doi: 10.1089/cpb.2006.9.342] [Medline: 16780402]
32. Koenig S, Crucian G, Dünser A, Bartneck C, Dalrymple-Alford J. Validity evaluation of a spatial memory task in virtual environments. Int J Des Innov Res 2011;6:1-13 [FREE Full text]
33. Kourtesis P, Collina S, Doumas LA, MacPherson SE. Technological competence is a pre-condition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. Front Hum Neurosci 2019;13:342 [FREE Full text] [doi: 10.3389/fnhum.2019.00342] [Medline: 31632256]
34. Madary M, Metzinger TK. Recommendations for good scientific practice and the consumers of VR-technology. Front Robot AI 2016;3:-. [doi: 10.3389/frobt.2016.00003]
35. International Organization for Standardization. 2010. 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 2020-02-07]
36. Dombrowski M, Smith P, Manero A, Sparkman J. Designing inclusive virtual reality experiences. In: Chen JY, Fragomeni G, editors. Virtual, Augmented and Mixed Reality. Multimodal Interaction. New York, USA: Springer; 2019:33-43.
37. Wang W, Cheng J, Guo JL. Usability of Virtual Reality Application Through the Lens of the User Community: A Case Study. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI EA'19; May 4 - 9, 2019; Glasgow, UK p. 1-6. [doi: 10.1145/3290607.3312816]
38. Kauhanen O, Väätäjä H, Turunen M, Keskinen T, Sirkkunen E, Uskali T, et al. Assisting Immersive Virtual Reality Development With User Experience Design Approach. In: Proceedings of the 21st International Academic Mindtrek Conference. 2017 Presented at: AcademicMindtrek'17; September 20 - 21, 2017; Tampere, Finland p. 127-136. [doi: 10.1145/3131085.3131126]
39. Chen JY, Fragomeni G. Virtual, Augmented and Mixed Reality. Applications and Case Studies. In: Proceedings of the International Conference on Human-Computer Interaction. New York, USA: Springer; 2019 Presented at: HCII'19; July 26–31, 2019; Orlando, USA. [doi: 10.1007/978-3-030-21565-1]
40. Bozgeyikli LL, Bozgeyikli E, Katkoori S, Raij A, Alqasemi R. Effects of virtual reality properties on user experience of individuals with autism. ACM Trans Access Comput 2018;11(4):1-27. [doi: 10.1145/3267340]
41. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining 'Gamification'. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek'11; September 28 - 30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
42. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
43. van der Linden WJ, Glas CA. Computerized Adaptive Testing: Theory And Practice. New York, USA: Springer; 2000.
44. van der Linden WJ, Glas CA. Elements of Adaptive Testing. New York, USA: Springer; 2010.
45. Witmer BG, Singer MJ. Measuring presence in virtual environments: A presence questionnaire. Presence 1998;7(3):225-240. [doi: 10.1162/105474698565686]
46. Slater M. Measuring presence: a response to the Witmer and Singer presence questionnaire. Presence 1999;8(5):560-565. [doi: 10.1162/105474699566477]
47. Slater M. CiteSeerX. 2003. A Note on Presence Terminology URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.800.3452&rep=rep1&type=pdf [accessed 2020-02-07]
48. Sanchez-Vives MV, Slater M. From presence to consciousness through virtual reality. Nat Rev Neurosci 2005 Apr;6(4):332-339. [doi: 10.1038/nrn1651] [Medline: 15803164]
49. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3549-3557 [FREE Full text] [doi: 10.1098/rstb.2009.0138] [Medline: 19884149]
50. Slater M, Sanchez-Vives MV. Enhancing our lives with immersive virtual reality. Front Robot AI 2016;3:-. [doi: 10.3389/frobt.2016.00074]

51. Slater M. Immersion and the illusion of presence in virtual reality. Br J Psychol 2018 Aug;109(3):431-433. [doi: 10.1111/bjop.12305] [Medline: 29781508]
52. Brade J, Lorenz M, Busch M, Hammer N, Tscheligi M, Klimant P. Being there again – Presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. Int J Hum-Comput Stud 2017;101:76-87. [doi: 10.1016/j.ijhcs.2017.01.004]
53. Lorenz M, Brade J, Diamond L, Sjölie D, Busch M, Tscheligi M, et al. Presence and user experience in a virtual environment under the influence of ethanol: an explorative study. Sci Rep 2018 Apr 23;8(1):6407 [FREE Full text] [doi: 10.1038/s41598-018-24453-5] [Medline: 29686255]
54. Cheng MT, She HC, Annetta LA. Game immersion experience: its hierarchical structure and impact on game-based science learning. J Comput Assist Learn 2015;31(3):232-253. [doi: 10.1111/jcal.12066]
55. Makowski D, Sperduti M, Nicolas S, Piolino P. 'Being there' and remembering it: presence improves memory encoding. Conscious Cogn 2017 Aug;53:194-202. [doi: 10.1016/j.concog.2017.06.015] [Medline: 28676191]
56. Gromer D, Madeira O, Gast P, Nehfischer M, Jost M, Müller M, et al. Height simulation in a virtual reality CAVE system: validity of fear responses and effects of an immersion manipulation. Front Hum Neurosci 2018;12:372 [FREE Full text] [doi: 10.3389/fnhum.2018.00372] [Medline: 30319376]
57. Iglói K, Zaoui M, Berthoz A, Rondi-Reig L. Sequential egocentric strategy is acquired as early as allocentric strategy: parallel acquisition of these two navigation strategies. Hippocampus 2009 Dec;19(12):1199-1211. [doi: 10.1002/hipo.20595] [Medline: 19360853]
58. Iglói K, Doeller CF, Berthoz A, Rondi-Reig L, Burgess N. Lateralized human hippocampal activity predicts navigation based on sequence or place memory. Proc Natl Acad Sci USA 2010 Aug 10;107(32):14466-14471 [FREE Full text] [doi: 10.1073/pnas.1004243107] [Medline: 20660746]
59. Iglói K, Doeller CF, Paradis A, Benchenane K, Berthoz A, Burgess N, et al. Interaction between hippocampus and cerebellum crus I in sequence-based but not place-based navigation. Cereb Cortex 2015 Nov;25(11):4146-4154 [FREE Full text] [doi: 10.1093/cercor/bhu132] [Medline: 24947462]
60. Rondi-Reig L, Petit GH, Tobin C, Tonegawa S, Mariani J, Berthoz A. Impaired sequential egocentric and allocentric memories in forebrain-specific-NMDA receptor knock-out mice during a new task dissociating strategies of navigation. J Neurosci 2006 Apr 12;26(15):4071-4081 [FREE Full text] [doi: 10.1523/JNEUROSCI.3408-05.2006] [Medline: 16611824]
61. Ploner CJ, Stenz U, Fassdorf K, Arnold G. Egocentric and allocentric spatial memory in idiopathic cervical dystonia. Neurology 2005 May 24;64(10):1733-1738. [doi: 10.1212/01.WNL.0000161869.83783.05] [Medline: 15911800]
62. Nedelska Z, Andel R, Laczó J, Vlcek K, Horinek D, Lisy J, et al. Spatial navigation impairment is proportional to right hippocampal volume. Proc Natl Acad Sci USA 2012 Feb 14;109(7):2590-2594 [FREE Full text] [doi: 10.1073/pnas.1121588109] [Medline: 22308496]
63. Korthauer L, Nowak N, Frahmand M, Driscoll I. Cognitive correlates of spatial navigation: associations between executive functioning and the virtual Morris Water Task. Behav Brain Res 2017 Jan 15;317:470-478. [doi: 10.1016/j.bbr.2016.10.007] [Medline: 27720743]
64. Morris R, Garrud P, Rawlins J, O'Keefe J. Place navigation impaired in rats with hippocampal lesions. Nature 1982 Jun 24;297(5868):681-683. [doi: 10.1038/297681a0] [Medline: 7088155]
65. Iaria G, Chen J, Guariglia C, Ptito A, Petrides M. Retrosplenial and hippocampal brain regions in human navigation: complementary functional contributions to the formation and use of cognitive maps. Eur J Neurosci 2007 Feb;25(3):890-899. [doi: 10.1111/j.1460-9568.2007.05371.x] [Medline: 17298595]
66. Iaria G, Palermo L, Committeri G, Barton JJ. Age differences in the formation and use of cognitive maps. Behav Brain Res 2009 Jan 23;196(2):187-191. [doi: 10.1016/j.bbr.2008.08.040] [Medline: 18817815]
67. Diamond A. Executive functions. Annu Rev Psychol 2013;64:135-168 [FREE Full text] [doi: 10.1146/annurev-psych-113011-143750] [Medline: 23020641]
68. Jurado MB, Rosselli M. The elusive nature of executive functions: a review of our current understanding. Neuropsychol Rev 2007 Sep;17(3):213-233. [doi: 10.1007/s11065-007-9040-z] [Medline: 17786559]
69. Martin J, Failows L. Executive function: theoretical concerns. In: Sokol B, Muller U, Carpendale J, Young A, Iarocci G, editors. Self- and Social-Regulation: The Development of Social Interaction, Social Understanding, and Executive Functions. Oxford, UK: Oxford University Press; 2010.
70. Hofmann W, Schmeichel BJ, Baddeley AD. Executive functions and self-regulation. Trends Cogn Sci 2012 Mar;16(3):174-180. [doi: 10.1016/j.tics.2012.01.006] [Medline: 22336729]
71. Valladares-Rodríguez S, Pérez-Rodríguez R, Anido-Rifón L, Fernández-Iglesias M. Trends on the application of serious games to neuropsychological evaluation: a scoping review. J Biomed Inform 2016 Dec;64:296-319 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.019] [Medline: 27815228]
72. Lecouvey G, Gonneaud J, Piolino P, Madeleine S, Orriols E, Fleury P, et al. Is binding decline the main source of the ageing effect on prospective memory? A ride in a virtual town. Socioaffect Neurosci Psychol 2017;7(1):1304610 [FREE Full text] [doi: 10.1080/20009011.2017.1304610] [Medline: 28567212]

73. Elkind JS, Rubin E, Rosenthal S, Skoff B, Prather P. A simulated reality scenario compared with the computerized Wisconsin card sorting test: an analysis of preliminary results. Cyberpsychol Behav 2001 Aug;4(4):489-496. [doi: 10.1089/109493101750527042] [Medline: 11708728]
74. Jansari AS, Devlin A, Agnew R, Akesson K, Murphy L, Leadbetter T. Ecological assessment of executive functions: a new virtual reality paradigm. Brain Impair 2014;15(2):71-87. [doi: 10.1017/brimp.2014.14]
75. Kourtesis P, Collina S, Doumas LA, MacPherson SE. Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. Front Hum Neurosci 2019;13:417 [FREE Full text] [doi: 10.3389/fnhum.2019.00417] [Medline: 31849627]
76. Kourtesis P, Korre D, Collina S, Doumas LA, MacPherson SE. Guidelines for the development of immersive virtual reality software for cognitive neuroscience and neuropsychology: the development of virtual reality everyday assessment lab (VR-EAL), a neuropsychological test battery in immersive virtual reality. Front Comput Sci 2020;1:-. [doi: 10.3389/fcomp.2019.00012]
77. Lange B, Koenig S, Chang C, McConnell E, Suma E, Bolas M, et al. Designing informed game-based rehabilitation tasks leveraging advances in virtual reality. Disabil Rehabil 2012;34(22):1863-1870. [doi: 10.3109/09638288.2012.670029] [Medline: 22494437]
78. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
79. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
80. García-Betances RI, Jiménez-Mixco V, Arredondo MT, Cabrera-Umpiérrez MF. Using virtual reality for cognitive training of the elderly. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):49-54. [doi: 10.1177/1533317514545866] [Medline: 25107931]

Abbreviations

CMT: Cognitive Map Task
HCI: human-computer interaction
HMD: head-mounted display
JEF: Jansari Assessment of Executive Functions
LFAM: Look for a Match
MET: Multiple Errands Test
RCT: randomized controlled trial
RVT: Ride in a Virtual Town
STM: Starmaze
UX: user experience
VAP-S: Virtual Action Planning-Supermarket
VE: virtual environment
VMT: Virtual Memory Task
vMWM: Virtual Morris Water Maze
VR: virtual reality
VRISE: VR-induced adverse symptoms and effects

Edited by G Eysenbach; submitted 18.10.19; peer-reviewed by C Parsey, M Dahl, P Wicks; comments to author 05.12.19; revised version received 07.02.20; accepted 07.02.20; published 27.04.20.

Please cite as:
Krohn S, Tromp J, Quinque EM, Belger J, Klotzsche F, Rekers S, Chojecki P, de Mooij J, Akbal M, McCall C, Villringer A, Gaebler M, Finke C, Thöne-Otto A
Multidimensional Evaluation of Virtual Reality Paradigms in Clinical Neuropsychology: Application of the VR-Check Framework
J Med Internet Res 2020;22(4):e16724
URL: https://www.jmir.org/2020/4/e16724
doi:10.2196/16724
PMID:32338614

Original Paper

Information Needs and Information-Seeking Behavior of Italian Neurologists: Exploratory Mixed Methods Study

Silvia Demergazzi[1], MSc; Luca Pastore[1], MSc; Giada Bassani[2], MSocSci; Marco Arosio[2], MSocSci; Caterina Lonati[3], MSc, PhD

[1]Medical Information Department, Teva Italia Srl, Assago, Italy

[2]Research Department, Doxa Pharma Srl, Milan, Italy

[3]Center for Preclinical Research, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

Corresponding Author:
Caterina Lonati, MSc, PhD
Center for Preclinical Research
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico
9 Pace Street
Milan
Italy
Phone: 39 392 691 2199
Email: caterina.lonati@gmail.com

Abstract

Background: Current medical professions involve an extensive knowledge of the latest validated scientific data to implement disease diagnosis, therapeutic strategies, and patient care. Although clinicians can refer to a growing number and type of information sources to keep current with new scientific achievements, there are still various concerns about medical information validity, quality, and applicability into clinical practice. Novel strategies are required to identify physicians' real-life needs with the final aim to improve modern medical information delivery.

Objective: Our research used an innovative tool to collect real-time physician queries in order to investigate information needs and seeking behavior of Italian neurologists treating patients with multiple sclerosis (MS) and migraine.

Methods: The study was designed as an exploratory mixed methods (ie, qualitative and quantitative) study involving 15 consecutive days of observation. A total of 50 neurologists (n=25 MS and n=25 migraine specialists) were recruited. Data were collected using an instant messaging mobile app designed for this research. At each information-seeking event, moderators triggered a computer-assisted personal interview including both semistructured interview and close-ended questions. Interactions and physician queries collected using the mobile app were coded into emerging themes by content analysis.

Results: Neurologist queries were relevant to the following major themes: therapy management (36/50, 71%) and drug-related information (34/50, 67%), followed by diagnostic strategies and procedures (21/50, 42%). Quantitative analysis indicated online resources were preferentially used by clinicians (48/50, 96%) compared with offline sources (24/50, 47%). A multichannel approach, in which both online and offline sources were consulted to meet the same need, was adopted in 33% (65/198) of information-seeking events. Neurologists more likely retrieved information from online relative to offline channels (F=1.7; *P*=.01). MS specialists were 53% more likely to engage in one information-seeking event compared with migraine neurologists (risk ratio 1.54; 95% CI 1.16-2.05). MS specialists tended to be more interested in patient-related content than migraine clinicians (28% [7/25] vs 10% [2/25], *P*=.06), who conversely more likely sought information concerning therapy management (85% [21/25] vs 60% [15/25], *P*=.05). Compared with MS clinicians, migraine specialists had a harder time finding the required information, either looking at online or offline channels (F=12.5; *P*=.01) and less frequently used offline channels (30% [8/25] vs 60% [15/25] of information-seeking events, *P*=.02). When multiple sources needed to be consulted to retrieve an information item, a reduced satisfaction rate was observed both among migraine and MS specialists (single source vs multiple sources *P*=.003).

Conclusions: This study provides a detailed description of real-life seeking behavior, educational needs, and information sources adopted by Italian MS and migraine neurologists. Neurologist information needs and seeking behavior reflect the specific characteristics of the specialty area in which they operate. These findings suggest identification of time- and context-specific needs of clinicians is required to design an effective medical information strategy.

(J Med Internet Res 2020;22(4):e14979) doi:10.2196/14979

KEYWORDS

information-seeking behavior; information needs; information sources; medical information delivery; neurologists; multiple sclerosis; migraine

Introduction

Modern health care professionals need to consult increasing amounts of scientific content to keep current on medical science advances [1]. It has been reported that experienced physicians use as many as 2 million pieces of information to manage their patients [2]. Moreover, the recent introduction of the precision medicine (PM) model requires a deeper understanding of properties and side effects of available drugs as treatments must be tailored to the individual patient [3]. In this landscape, physicians may feel overwhelmed by the steadily expanding flow of scientific literature [4]. The massive diffusion of online scientific resources enabling health care professionals a multichannel engagement make the selection, integration, and translation of medical information into clinical practice even more complex [5]. Paradoxically, the growth of scientific evidence and access to multiple information sources do not necessarily meet clinician needs and quality standards [2,6,7].

Whereas continuing medical education is a prominent (and often mandatory) source of medical knowledge for most physicians [8], it has been shown that such programs do not often fulfill physician needs and may fail to translate into improved clinical practice patterns [9]. Indeed, the time from educational activity to real-life information needs may not allow this information to efficiently answer questions arising directly at the point of care [6,10]. In contrast, online information sources, including the open-access resource Wikipedia and social networks, have increasingly been used by physicians to quickly retrieve medical information [6,11-15]. Therefore, effective strategies for modern medical education and information delivery should be based on extensive evaluation of physicians' real-life content needs and should likewise be prone to continuous adaptation to meet expectations in an ever-changing landscape [16].

Medical Information departments of pharmaceutical companies often deliver up-to-date, balanced, and evidence-based information on a peer-to-peer basis, answering unsolicited medical requests through different channels [16,17]. A recent survey showed that most companies in the health care sector provide some medical information [18]. Whereas reliance on industry or sponsored resources is well established in the US market, little is known on use rates for European countries and specifically for Italy.

In this study, we investigated the information needs and seeking behavior of Italian neurologists treating patients with multiple sclerosis (MS) and migraine in order to inform the content and layout of Medical Information services in our department. MS and migraine therapeutic management have evolved rather differently over the last few years. In fact, in the field of MS there were important advances, with increasing disease-modifying therapies becoming available for both progressive and relapsing-remitting MS treatment [19]. Therefore, numerous scientific educational activities and relevant online resources have been provided by pharmaceutical companies, scientific societies, and patient associations to promote MS neurologists' continuing education [19,20]. Conversely, accurate migraine diagnosis and subclassification are still challenging due to the lack of objective gold standard diagnostic criteria [21-23]. As a consequence, neurologists treating patients with migraine must cope with the lack of robust guidelines and shortage of authoritative sources of information and educational activities [24].

Information-seeking behavior is a complex phenomenon that is contextually shaped by personal needs, learning styles, available resources, and affective components among other factors [13,25,26]. While many general theories of information seeking have been proposed, we adopted a pragmatic theoretical approach to optimize knowledge gathering for the specific purpose of developing a working app for professional content delivery to clinical physicians. For this reason we refer to the sense-making approach developed since 1972 for the study of the human use of information systems; it entails the investigation of specific situations (which define the context in which a discontinuity emerges and gives rise to information needs) and information gaps (identifying the uncertainty around a specific content) so that specific instruments can be designed for content delivery [27]. The sense-making approach proposes that the moment of communication is best described by focusing on the how the actor describes the circumstances when the information gap emerges, the content of the information gap, and its attempt to bridge this gap. Therefore, at a specific moment in time and space, an individual who self-defines as facing a gap of a particular kind may use communicating tactics of a particular kind. In a different moment facing a different gap they may use a different tactic.

This research describes the real-life seeking behavior, educational needs, and information sources of Italian MS and migraine neurologists. To the best of our knowledge, these aspects have never been explored before in this context. Our data provide an initial, exploratory step in understanding the specific information needs of neurologists and give insights on the motivation, response, and gaps in the landscape of information available for this group of users. These results could be used to design novel Medical Information strategies aimed to deliver personalized, accurate, consistent, and timely information with an omnichannel approach, allowing health care professionals to make informed decisions that can improve patient care.

Methods

Study Sample and Data Collection

To evaluate physician eligibility, a screener questionnaire was administered to 72 clinicians working in different health centers and hospitals throughout the Italian territory. Based on the screener results, we enrolled 50 neurologists, of whom 25 were MS specialists and 25 migraine specialists. Recruitment was

planned to equally represent physicians from all Italian geographical macroregions in each specialty area.

The research was designed as an exploratory mixed methods (ie, qualitative and quantitative) study and involved 15 consecutive days of study. The observation was conducted through the Physician Line app, an instant messaging software app consisting in an instant messaging phone app based on WhatsApp that allowed physicians to share and describe the information needs experienced during their daily clinical practice and information sources used to retrieve the information needed. Participants were instructed on the Physician Line app functionalities by means of a kick-off video presentation. To improve response rate, respondents were rewarded with a cash incentive. Two expert researchers in the field of qualitative research (BG and MA) were able to initiate an interaction every time a physician sent a text message in the Physician Line app. Therefore, there was a direct interaction between physicians and moderators during the interview conducted by these means. In fact, each information-seeking event triggered a computer-assisted interview, during which a semistructured interview and two close-ended questions were administered to the physicians through the Physician Line app. The semistructured interview included 5 items capturing physician motivations and behavioral patterns (Multimedia Appendix 1). At the end of the semistructured interview, physicians were asked to rate how frequently they could retrieve appropriate content when needed and how satisfied they were about the information obtained. Ratings occurred on 5-point (from 0=not at all to 4=absolutely yes) and 4-point (from 0=not at all to 3=completely satisfied) Likert scales, respectively (Multimedia Appendix 1).

Ethics Approval and Consent to Participate

Due to the nature of the research, no ethics committee approval was required (Italian law Decreto 8 febbraio 2013 n. 34). In fact, the study did not involve patients or lay citizens and no health intervention had been administered to participants. This study was conducted in compliance with the European Union General Data Protection Regulation 2016/679 and in line with well-established regulatory practices and procedures governing marketing research, including the Market Research Society code of conduct (2019 revision) and the Italian Code of Professional Ethics (curated by ASSIRM 2016 revision).

Physicians actively chose to participate to the study. Records collected by Doxa Pharma Srl include data retention policies, data privacy statements, permission to take part in a data collection exercise, and agreement to the processing of personal data. The interview questions were not aimed at investigating sensitive issues like religious or political beliefs or sexual orientation. Doxa Pharma Srl ensured respect of confidentiality of collected information and pseudonymization of individual answers before primary data abstraction and analysis.

Qualitative Analysis

Physician queries and their interactions with the moderators collected with the Physician Line app were subsequently evaluated by content analysis. Following the sense-making approach, we predefined an ontology of information seeking entailing the concept of an information-seeking event described by a set of domains including situations, gaps, and tactics, namely motivation and triggers, context, information gap content, information sources, and information search strategy. An information-seeking event was defined as any action carried out by a physician in order to meet an information need (ie, consulting online sources, discussing with colleagues or sales representatives, reading a scientific article).

Each moderator abstracted relevant themes from the Physician Line app transcript with a mixed deductive-inductive content analysis of the material transcripts. Moderators precoded the transcript by highlighting codable words, sentences, or paragraphs. An initial distinction was made to discriminate motivation and triggers of information seeking and query content. The transcript was then open-coded by assigning descriptive labels to transcripts excerpts under these first two categories. Both motivation and triggers and content were further coded as follows. After consolidating redundant codes, a matrix was generated by including all codes emerging from the discussion. The codes were inductively grouped into broader categories by observing similarities of content and meaning. When disagreement occurred among coders, the item was discussed until a common taxonomy was achieved.

In a second stage, each information-seeking event was coded (for each specific motivation and content) concerning source, context, and event time (based on recording metadata). Coding of information sources was based on a predefined coding scheme (the list of codable sources is reported in Multimedia Appendix 2).

After the second stage coding, each information-seeking event was described as a vector of motivation and trigger, source, context, and event time descriptors. This information was entered in a fully codified database used for further quantitative analysis. In this context, full codification of information-seeking behavior refers to the exhaustive representation of constructs implied by the sense-making approach to describe an information-seeking event, and no additional constructs were reported.

Statistical Analysis

We computed the absolute and relative frequency for categorical variables and means and standard deviation for continuous variables. Frequency of information-seeking events was calculated considering the total number of information-seeking events over the study period, and it was expressed by person-time incidence rate (number of information-seeking events/10 person-day). We computed confidence intervals for information-seeking event rates based on the Poisson distribution. Furthermore, we used 2-way analysis of variance to evaluate differences in information retrieval and satisfaction scores across medical specialty and information channel used. Finally, differences in proportion of content type searches across specialties were assessed by Fisher exact or chi-square tests where appropriate. Analysis was conducted with SAS 9.4 (SAS Institute Inc).

Results

Participants

Participants worked in different settings, encompassing small private centers and large public organizations integrating multiple operative units aimed at providing health care for a wide catchment area. Physician characteristics are reported in Table 1.

Table 1. Sample characteristics.

Characteristics	Migraine specialists (n=25)	Multiple sclerosis specialists (n=25)	*P* value
Age in years, mean (SD)	47.5 (7.4)	50.2 (16.1)	.44
Sex, male, n (%)	12 (48.0)	9 (34.6)	.39
Geographical distribution, n (%)	—	—	**.44**
Northern regions	10 (40.0)	11 (44.0)	—
Central regions	7 (28.0)	5 (20.0)	—
Southern regions	8 (32.0)	9 (36.0)	—

Use of the instant messaging phone app Physician Line allowed real-time collection of relevant data without affecting the clinicians' daily working routine; the tool was well accepted and provided physicians the opportunity to conveniently communicate with moderators.

Qualitative Analysis

Content

Physician queries concerned 8 categories relevant to marketed and investigational drugs, clinical management, disease epidemiology and physiopathology, pharmaceutical companies and their activities, diagnostic procedures, patient-related topics, congress and educational opportunities, and other minor categories. Overall, 37 items could be consistently coded and classified in these 8 categories (Multimedia Appendix 3).

Motivations and Triggers of Information Seeking

We identified different neurologist motivations to engage in information seeking which were sorted into 2 categories. The first category included exogenous motivations (ie, triggers), external events triggering research for further information. Such events can be tentatively classified into passive and active. Specifically, passive triggers were defined as any activating content from newsletters, websites, marketing activities, or institutional or scientific communications which motivated further information seeking in the absence of a specific, preexisting information need.

> *I am reading a paper on NEJM about a phase 2 trial of <Drug_name> among MS patients. I got there because I just received the NEJM weekly newsletter and this topic was relevant to my practice.* [N1]

Conversely, active triggers were deemed to involve questions raised by patients and colleagues; emerge as noteworthy themes while the physician was actively engaged in seminars, grand round discussions, informal discussions; or were raised by clinical problems emerging in the course of a medical encounter.

> *One patient asked me about <Drug_name>. She has found the list of investigational drugs on Wikipedia and learned that the drug was under review by the FDA. She entered a secondary progressive course and she was worried about the clinical worsening of the disease.* [N2]

We also found that the need for professional growth and general scientific update rather than immediate, contingent problem-solving issues may represent a strong motivational driver for information-seeking among neurologists. In contrast to exogenous triggers, we called such experiences endogenous motivation.

> *Nowadays we, as physicians, must keep current on new drugs and scientific developments. Our field is growing in complexity with new insights into disease pathology and novel therapeutic options. For this reason, I browse PubMed on a weekly basis.* [N3]

Information Seeking Circumstances: Sources, Context, and Time

Physicians used both online and offline resources (a complete list is provided in Multimedia Appendix 2). Online resources were generally considered quick and easy to access, whereas offline resources were deemed to offer more chance for in-depth learning. Among online channels, the most cited search engines were Google and scientific literature repositories such as PubMed or Embase, which were preferred when looking for reliable, accurate, impartial, and complete information.

> *I wanted to have an overview of new drugs for MS. I did a search on PubMed by keywords and selected a few systematic reviews and editorials by reading the abstract. I will try to download the full-text in the afternoon, when I have more time.* [N5]

Online resources, like PubMed, were considered more convenient when accessible through mobile apps. One limitation in the use of professional scientific literature was the lack of institutional subscription to professional scientific journals since most published research requires the payment of access fees. Nonprofessional search engines such as Google were exploited for initial exploratory search or when quick answer to simple questions were needed. This was particularly true when active triggers motivated the initiation of information-seeking behavior and small pieces of information were quickly required to complement clinical decision-making or make sense of a

question raised in the course of medical interactions or educational events.

> *During a visit I needed some info about dosing regimens for a patient with liver disease. I searched Google and easily found the leaflet of the drug online.* [N6]

Whereas the use of PubMed or other indexed repositories of scientific literature inherently leads to consultation of authoritative scholarly articles, the use of general public search engines such as Google requires an additional selection process on the part of physicians. Physicians reported use of institutional and noninstitutional websites, portals specializing in scientific dissemination, medical content websites, social media and blogs, and professional or patient discussion forums depending on the type and content of the piece of information searched for. Despite this extensive and multifaceted use, responders pointed out the difficulty of evaluating validity and reliability of online contents retrieved by such means. For this reason, in some cases portals of governmental institutions and scientific societies were used to find guidelines and specific authoritative grey literature material.

> *I received notice of definitive approval of <Drug_name>. I consulted ECTRIMS library to learn more about the drug.* [N7]

Offline resources included books, seminars, roundtables, workshops, educational events, and practical training. By their very nature such offline materials and events were considered authoritative sources and were used to satisfy needs related to professional development, gain an in-depth understanding of a disease, and learn new complex skills. A list of offline resources mentioned is reported in Multimedia Appendix 2.

> *Congresses are great opportunities to expand my network, have a grasp of current developments in the field, and improve my understanding of the disease and new drugs.* [N7]

Quantitative Analysis

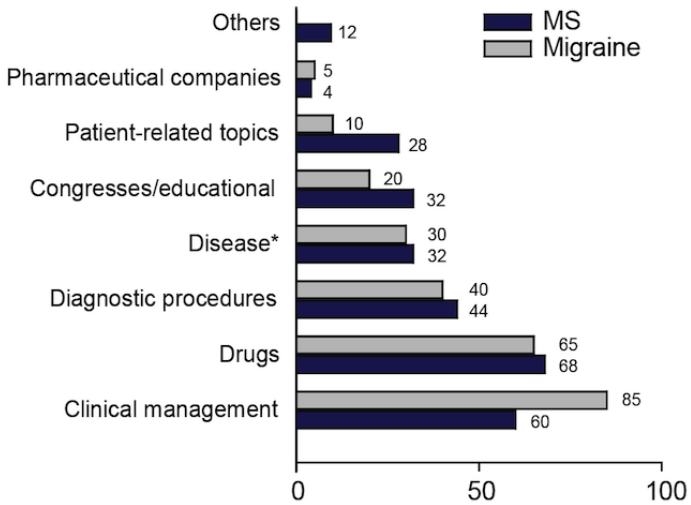
Frequency and Distribution of Information-Seeking Events

Over the 15 days of study, a total of 198 information-seeking events were collected corresponding to 2.64 information-seeking events/10 person-days (95% CI 2.29-3.28). More specifically, 120 information-seeking events (61%; 3.69 events/10 person-days; 95% CI 3.06-4.40) were sent by MS specialists, while 78 (39%; 2.4 events/10 person-days; 95% CI 1.91-2.98) were sent by migraine neurologists. Hence, MS specialists were 53% more likely to engage in one information-seeking event compared with migraine specialists (risk ratio 1.54; 95% CI 1.16-2.05). Each information-seeking event included an average of 1.98 different information searches (95% CI 1.79-2.18), for a total of 392 needs recorded.

Distribution of the Expressed Information Need Categories

Overall, the majority of physician expressed at least one need concerning therapy management (36/50, 71%), followed by drug-related content (34/50, 67%), diagnostic strategies and procedures (21/50, 42%), disease-related content (16/50, 31%), congresses and educational opportunities (14/50, 27%), patient-related content (10/50, 20%) and other/miscellaneous including administrative issues, pharmacoeconomics (2/50, 4%), pharmaceutical companies (2/50, 4%), topical issues (1/50, 2%), and unclassified content (5/50, 10%). Distribution of major information needs was slightly different across specialties (Figure 1). Migraine specialists tended to seek information concerning therapy management more often than MS specialists (85% [21/25] vs 60% [15/25], P=.05); on the other hand, MS specialists tended to be more interested in patient-related content compared with migraine specialists (28% [7/25] vs 8% [2/25], P=.06).

Figure 1. Distribution of neurologists' major information needs. Information needs expressed by neurologists treating multiple sclerosis (n=25) and migraine (n=25) were grouped into 9 major categories by content analysis. Bars denote the share of physicians (%) reporting information seeking for each major category. *Disease epidemiology and physiopathology. MS: multiple sclerosis.



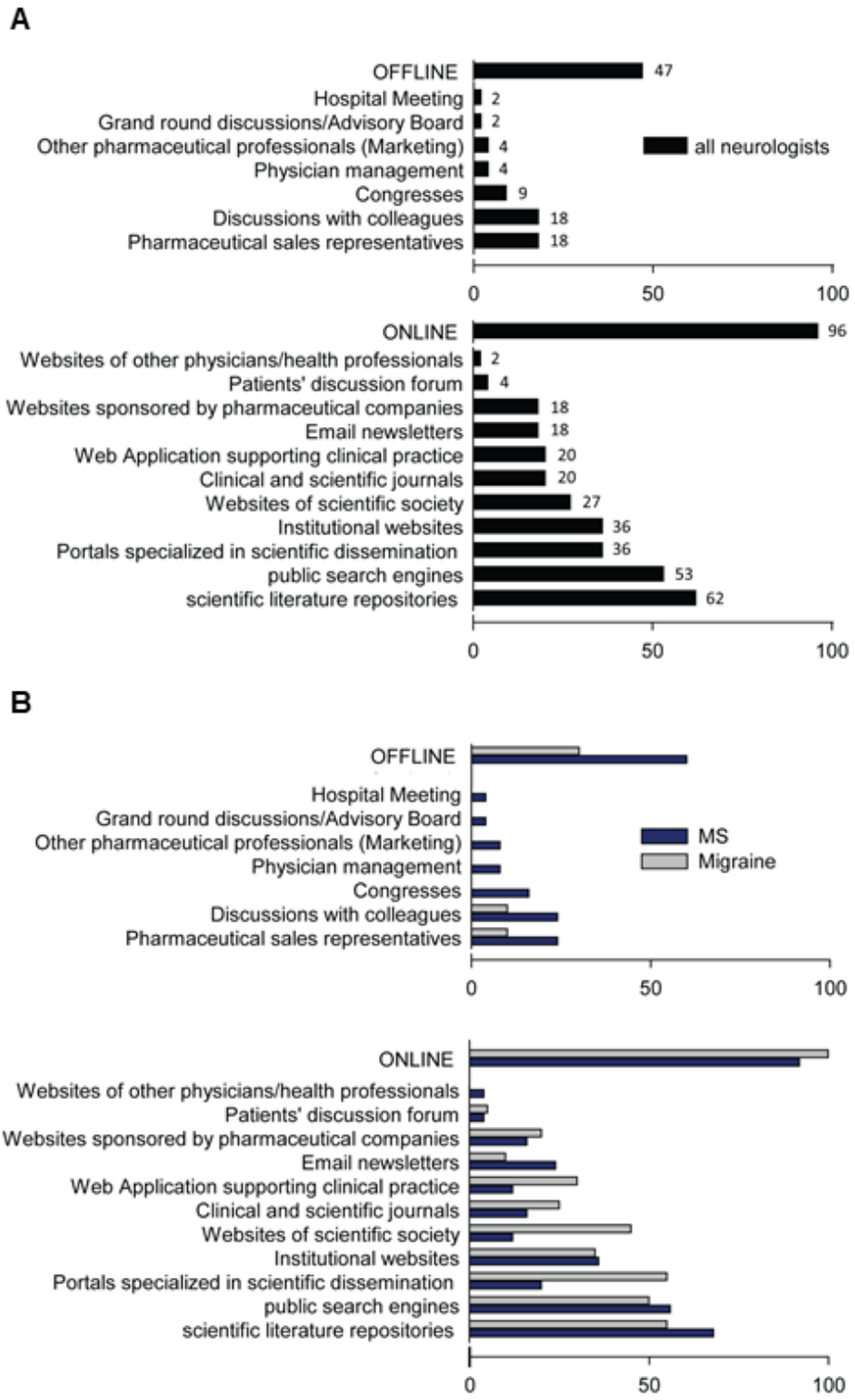
Information Sources

As shown in Figure 2A, online resources were used by the majority of clinicians (48/50, 96%), while offline resources were less often consulted (24/50, 47%). Figure 2B shows details on sources used across specialties. Migraine specialists used offline channels less frequently compared with MS neurologists (30% [8/25] vs 60% [15/25], *P*=.02).

Overall, an initial adoption of a Web search query did not exclude subsequent use of offline channels and vice versa; indeed, online and offline channels were frequently used in combination (33% [65/198] of the information seeking events), adopting a multichannel approach. Of interest, migraine specialists more likely engaged a multichannel search than MS specialists (43% [34/78] vs 27% [32/120] of the information-seeking events, *P*=.03). An explanation of this behavior can be found in the analysis of Physician Line app data, which revealed that multichannel search was adopted for complex issues needing articulated responses or when the desired information was not available in a single authoritative source. Especially in the migraine field, this implied adopting an iterative tree-like search strategy, in which answers to the original questions raised further information needs that triggered further information-seeking behaviors.

We found tentative evidence that different sources of information were selected to satisfy different information needs (Multimedia Appendix 4). For example, PubMed and other professional repositories of the scientific literature were used more often than public search engines when physicians were looking for information about disease epidemiology and physiopathology (39% [18/46] vs 11% [5x/46], *P*=.02) and diagnostic procedures (21% [7/34] vs 12% [4/34], *P*=.16). Conversely, public search engines tended to be used more frequently than professional scientific repositories to find out about congresses (9% [2/23] vs 0 [0/23], *P*=.07) and patient-related topics (31% [5/16] vs 19% [3/16], *P*=.07). Finally, professional and general search engines were equally used to look for information about drugs (18% [17/96] vs 19% [18/96]) and “clinical management” (20% [34/169] vs 19% [32/169]). Websites of scientific societies or institutions were more likely searched to retrieve information about congresses (8/23, 35%), disease epidemiology and physiopathology (6/46, 13%), and diagnostic procedures (6/34, 18%) compared with other themes (2% [3/169] for “clinical management”; 1% [1/96] for “drugs,” *P*=.01).

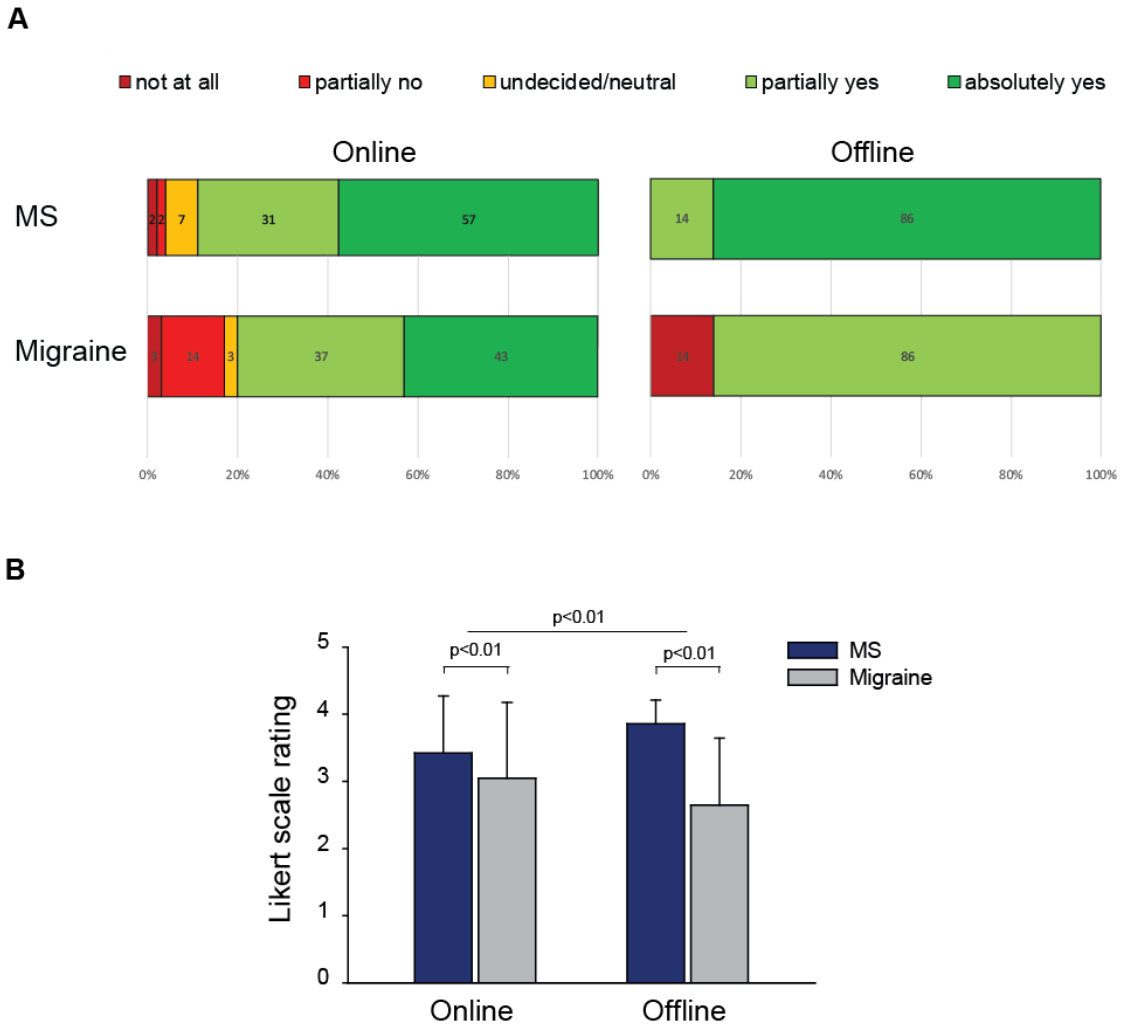
Figure 2. Distribution of information sources used by neurologists for information seeking. A. Channels used by neurologists (n=50) for information seeking. B. Source utilization across specialties (MS specialists, N=25, and migraine specialists, n=25). Bars denote the share of physicians (%) using each specific source for information seeking. MS: multiple sclerosis.



Information Retrieval Rating

Data collected with the Physician Line app indicate that the information needed was available in at least one resource for most physicians. Average information retrieval rating was 3.27 (SD 0.99) indicating that physicians most often retrieved the information they needed. In very few instances, clinicians reported they did not find the information needed (7% [26/392] of information needs, Figure 3A). Additionally, physicians more likely retrieved information from online versus offline channels (Figure 3B, F=1.7; P=.01). Migraine specialists had a harder time finding answers to their questions compared with MS neurologists, either looking at online or offline channels (Figure 3B, omnibus test, F=12.5; P=.01).

Figure 3. Information retrieval. A. Distribution of information retrieval ratings by information channel type and medical specialty (MS specialists, n=25, and migraine specialists, n=25). Physicians were asked to rate if they could retrieve the needed information. Diagrams show percentage of information-seeking events for each rating score of the 5-point Likert scale used (from 0=not at all to 4=absolutely yes); B. Differences in average information retrieval scores across specialty area and information channel and medical specialty. *P* values refer to 2-way analysis of variance. MS: multiple sclerosis.



Satisfaction About Information Quality

Physicians reported they were generally satisfied with the information retrieved (Figure 4A and 4B). Differences across channels and specialties in satisfaction ratings were similar to those observed for information retrieval ratings (Figure 4B, omnibus test, F=49.0; P=.01). In particular, migraine specialists were significantly less satisfied about offline channels compared with MS neurologists (P=.02). Conversely, satisfaction about information retrieved through online channels was similar in both specialty groups (P=.69). Finally, we observed reduced satisfaction when multiple sources were interrogated for a given information need (Figure 5, omnibus test: F=7.18, P=.001; single source vs multiple sources: F=12.02, P=.003). Such difference was observed among MS (P=.006) and migraine specialists (P=.005).

Figure 4. Neurologist satisfaction about information quality. A. Distribution of satisfaction ratings for retrieved information by information channel type and medical specialty (MS specialists, n=25, and migraine specialists, n=25). Physicians were asked to rate how satisfied they were about the retrieved information. Diagrams show percentage of information-seeking events for each score of the 4-point Likert scale used (from 0=not at all to 4=very satisfied); B. Differences in average satisfaction scores by information channel and medical specialty. *P* values refer to 2-way analysis of variance. MS: multiple sclerosis.

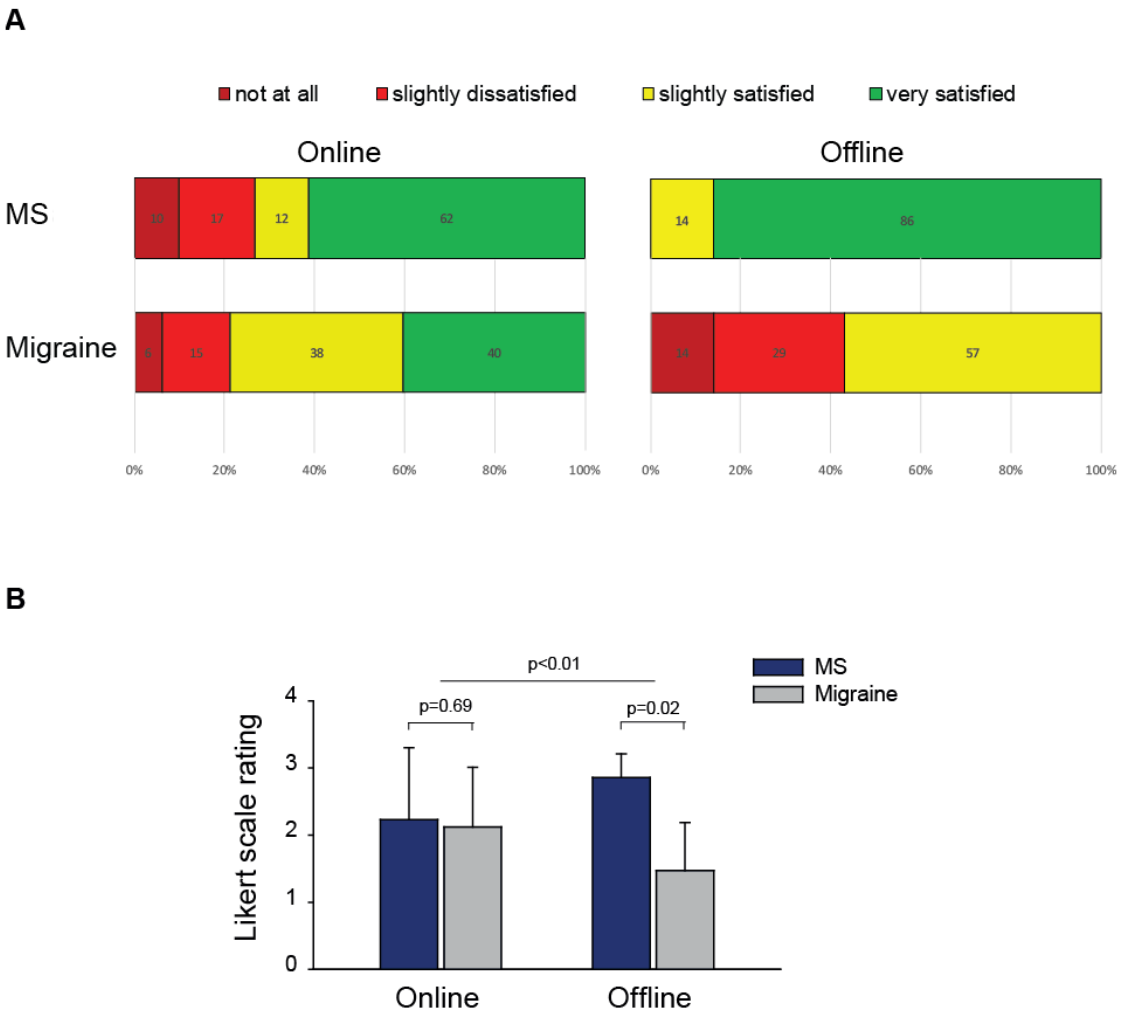
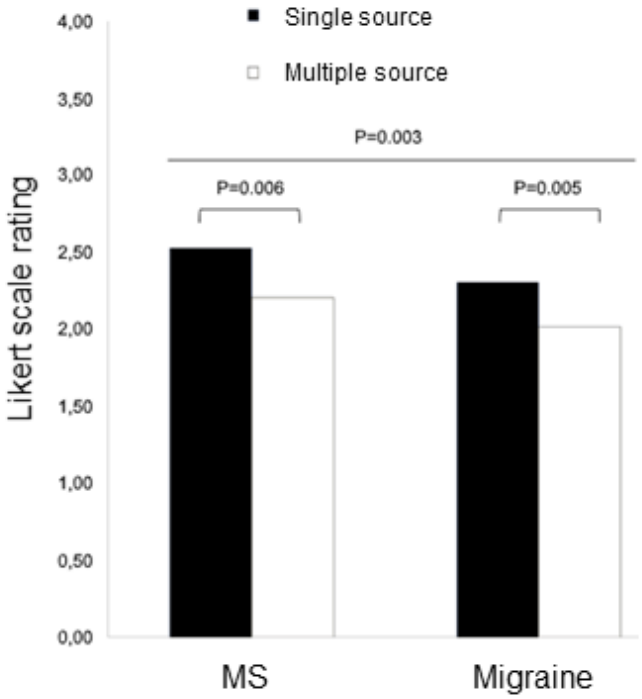


Figure 5. Neurologist satisfaction when using single or multiple sources. Average satisfaction ratings by medical specialty (MS specialists, n=25, and migraine specialists, n=25) and information sources adopted for each information-seeking event. *P* values refer to 2-way analysis of variance. MS: multiple sclerosis.



Discussion

Principal Findings

This research used an instant messaging phone app to investigate real-life needs, learning triggers, and information sources of Italian neurologists treating patients with either MS or migraine. This exploratory study provides a first time preliminary profiling of neurologists that could be useful for subsequent development of personalized approaches to medical education.

An interesting finding of this study is that seeking behavior is different between MS and migraine specialists, suggesting that the specific clinical scenario of specialization determines a distinctive pattern of interests, information needs, issues to be addressed, and preferred information sources. There are significant differences between current therapeutic management of MS and migraine, mainly due to the recent advances in the treatment strategies for MS. Our data on information needs and seeking behavior of neurologists seem to reflect the differences occurring in the therapeutic area in which physicians operate. Indeed, compared with MS specialists, migraine neurologists were less prone to engage in information-seeking behavior and, when they did look for medical information, they more frequently asked questions about therapy management rather than other topics. Furthermore, migraine physicians often needed to consult multiple information channels to find answers and more frequently relied on the use of public search engines rather than professional scientific repositories. Finally, satisfaction of migraine neurologists with offline channels (eg, including seminars, discussion with pharmaceutical sales representatives, congresses) was significantly lower than that of MS specialists.

A deeper analysis of resource use according to information needs indicates that neurologists have developed need- and time-specific research patterns that depend on the type and content of the piece of information searched for. Indeed, when starting new information seeking, physicians choose the most effective search process as well as the most suitable information source to answer that particular question. This requires neurologists to estimate the properties, reliability, weaknesses, and strength of each available channel. Moreover, the choice to use online or offline channels or a combination also depends on the environment in which the information need occurs (ie, during a patient encounter, at home, during a discussion with colleagues). For example, consistent with previous research [10,12,28], in our study group public search engines were exploited for initial exploratory researches or to rapidly meet simple information needs. This was particularly true when small pieces of information were needed to quickly answer simple questions raised in the course of medical interactions or educational events. In other words, it appears that these online contents are used when physicians need a prompt transfer of medical information into clinical practice. However, due to the difficulty in assessing the validity of the overwhelming amount of general public search engines or free online encyclopedia, physicians subsequently referred to more authoritative scientific resources, including PubMed or other indexed repositories of scientific literature and offline channels. On the other hand, to address complex issues needing articulated responses, a multichannel tree-like search process was generally adopted, in which both online and offline sources were used. The integrative use of online and offline channels often occurred when a single channel failed to meet a specific information need, as in the case of migraine specialists.

The complexity around information seeking and learning modalities recently gave rise to the principles of personalized education (PE). Like the emerging framework of PM [3,29,30], PE posits that the offer of educational services should reflect the specific content, timing, conciseness, ergonomics, and use channel needs at the point of use rather than the one-size-fits-all approach adopted so far [13,31,32]. In other words, PE involves identification of specific context-by-user information needs, optimization of educational activities, and design of user-centered learning [33,34]. In this context, medical information departments have the possibility to provide reliable medical content in a timely manner answering specific inquiries from specialists. However, in contrast with evidence from North American data [18], our findings show that the use of websites sponsored by pharmaceutical companies was marginal. On the other hand, recourse to pharmaceutical sales representatives was a prominent information source among offline channels, representing an important credibility asset for medical information services. In light of these data, customization of Medical Information services on the basis of specific needs, information gaps, and information-seeking behaviors for different therapeutic areas is both a challenge and an opportunity for pharmaceutical companies.

Strengths and Limitations

This study was based on the use of the Physician Line app, a novel data collection tool that enables real-life observation and sharing of the information needs experienced by physicians. This unique feature allows granular, simultaneous data collection, thus minimizing recall bias. Therefore, the Physician Line app offers the opportunity to obtain results similar to those provided by ethnographic research, in which moderators interact with participants in their real-life environment. Moreover, this tool allows the identification of time- and context-specific information needs. Taken together, these observations indicate that the Physician Line app may represent a valuable strategy for identifying the educational needs and information search strategies of health care professional.

Some limitation of this research also needs to be acknowledged. First, the exploratory design involves a small sample size and a short observation period. Further, we lacked potentially important personal and contextual data that could bring further insight into information-seeking profiles among MS and migraine specialists.

Conclusion

Our research shows that Italian neurologists practicing in different therapeutic areas experience different information needs, adopt different seeking behaviors, and refer to different information sources during their clinical practice and/or professional development. These findings suggest that an effective information delivery approach requires customization of both the information to be provided and the communication methods to be adopted, which must be tailored to the specific situations, needs, and requirements of health care professionals.

Based on this, identification of time- and context-specific needs as well as physician' profiling appear to be essential steps to design personalized information delivery and medical education strategies for neurologists involved in different subspecialty therapeutic areas. This requires novel instruments enabling us to thoroughly assess physicians' real-life information needs. This proof-of-concept study investigated the suitability of the Physician Line app as a tool to collect relevant data without affecting clinicians' daily working routine and providing, at the same time, the opportunity for interaction between physicians and moderators. Based on these preliminary results, the Physician Line app appears to be a valuable tool enabling identification of specific context-by-user information needs, optimization of educational activities, and design of user-centered learning, in accordance with PE principles. Further research should be designed to evaluate the potential of this novel instrument for data collection in a larger sample population. Delivery of personalized, accurate, consistent, and timely information to physicians is essential to improve patient care.

Acknowledgments

This research was funded by Teva Italy Srl. The authors thank Mario Cepparulo, medical director in Teva Italy Srl, for his support of this project. The views and opinions expressed in this publication are those of the individual authors and do not necessarily reflect the official position of the authors' companies. The authors thank Margherita Racano for her contribution to data analysis.

Authors' Contributions

SD designed the research and drafted the manuscript; LP contributed to manuscript drafting; GB designed the research, performed the qualitative analysis, and analyzed the data; MA designed the research, performed the quali-quantitative analysis, and analyzed the data; and CL performed the quantitative data analysis and wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Physician Line app items of the semistructured interview and Likert-based questions.
[DOCX File , 16 KB - jmir_v22i4e14979_app1.docx]

Multimedia Appendix 2
Information sources mentioned during the observation period.
[DOCX File , 18 KB - jmir_v22i4e14979_app2.docx]

Multimedia Appendix 3
Themes emerging from content analysis.
[DOCX File , 21 KB - jmir_v22i4e14979_app3.docx]

Multimedia Appendix 4
Use of information sources according to information needs.
[DOCX File , 19 KB - jmir_v22i4e14979_app4.docx]

References

1. Bornmann L, Mutz R. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J Assn Inf Sci Tec 2015 Apr 29;66(11):2215-2222. [doi: 10.1002/asi.23329]
2. Smith R. What clinical information do doctors need? BMJ 1996 Oct 26;313(7064):1062-1068 [FREE Full text] [doi: 10.1136/bmj.313.7064.1062] [Medline: 8898602]
3. The AOMS. Acad Med Sci. 2015 May 12. Stratified, personalised or P4 medicine: a new direction for placing the patient at the centre of healthcare and health education URL: https://acmedsci.ac.uk/file-download/38266-56e6d483e1d21.pdf [accessed 2020-02-18]
4. Avorn J. New York Times. 2013 Jun 12. Healing the overwhelmed physician URL: https://www.nytimes.com/2013/06/12/opinion/healing-the-overwhelmed-physician.html?_r=0 [accessed 2020-02-18]
5. 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]
6. Formoso G, Rizzini P, Bassi M, Bonfanti P, Rizzardini G, Campomori A, et al. Knowledge transfer: what drug information would specialist doctors need to support their clinical practice? Results of a survey and of three focus groups in Italy. BMC Med Inform Decis Mak 2016 Sep 01;16:1-9 [FREE Full text] [doi: 10.1186/s12911-016-0355-7] [Medline: 27581665]
7. Gray JA. Where's the chief knowledge officer? To manage the most precious resource of all. BMJ 1998 Sep 26;317(7162):832-840 [FREE Full text] [doi: 10.1136/bmj.317.7162.832] [Medline: 9748173]
8. Ahmed K, Ashrafian H. Life-long learning for physicians. Science 2009 Oct 09;326(5950):227. [doi: 10.1126/science.326_227a] [Medline: 19815754]
9. Haines ST. It's time to dismantle CE and build a CPI system. Am J Pharm Educ 2018 Jun;82(5):382-384 [FREE Full text] [doi: 10.5688/ajpe6985] [Medline: 30013249]
10. Cook DA, Sorensen KJ, Wilkinson JM, Berger RA. Barriers and decisions when answering clinical questions at the point of care: a grounded theory study. JAMA Intern Med 2013 Nov 25;173(21):1962-1969. [doi: 10.1001/jamainternmed.2013.10103] [Medline: 23979118]
11. Masters K. For what purpose and reasons do doctors use the Internet: a systematic review. Int J Med Inform 2008 Jan;77(1):4-16. [doi: 10.1016/j.ijmedinf.2006.10.002] [Medline: 17137833]
12. Bernard E, Arnould M, Saint-Lary O, Duhot D, Hebbrecht G. Internet use for information seeking in clinical practice: a cross-sectional survey among French general practitioners. Int J Med Inform 2012 Jul;81(7):493-499. [doi: 10.1016/j.ijmedinf.2012.02.001] [Medline: 22425281]
13. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]
14. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. J Am Med Inform Assoc 2012;19(5):777-781 [FREE Full text] [doi: 10.1136/amiajnl-2012-000990] [Medline: 22759618]
15. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
16. TenBarge AM, Riggins JL. Responding to unsolicited medical requests from health care professionals on pharmaceutical industry-owned social media sites: three pilot studies. J Med Internet Res 2018 Oct 29;20(10):e285 [FREE Full text] [doi: 10.2196/jmir.9643] [Medline: 30373730]
17. Fung SM, Sud C, Suchodolski M. Survey of customers requesting medical information: preferences and information needs of patients and health care professionals to support treatment decisions. Drug Inf J 2019 Jan 02:216847901881545. [doi: 10.1177/2168479018815455] [Medline: 30602300]
18. Giffin S, Shah R, Soloff A, Vaysman A, Oreper J, Gažo A, et al. Pharma Collaboration for Transparent Medical Information (phactMI) Benchmark Study: trends, drivers, success factors, and value of globalization in medical information. Ther Innov Regul Sci 2019 May;53(3):332-339. [doi: 10.1177/2168479018779920] [Medline: 29916261]

19. Ziemssen T, Derfuss T, de Stefano N, Giovannoni G, Palavra F, Tomic D, et al. Optimizing treatment success in multiple sclerosis. J Neurol 2016 Jun;263(6):1053-1065 [FREE Full text] [doi: 10.1007/s00415-015-7986-y] [Medline: 26705122]
20. Río J, Comabella M, Montalban X. Multiple sclerosis: current treatment algorithms. Curr Opin Neurol 2011 Jun;24(3):230-237. [doi: 10.1097/WCO.0b013e328346bf66] [Medline: 21499098]
21. Weatherall MW. The diagnosis and treatment of chronic migraine. Ther Adv Chronic Dis 2015 May;6(3):115-123 [FREE Full text] [doi: 10.1177/2040622315579627] [Medline: 25954496]
22. Bartleson J, Cutrer F. Migraine update: diagnosis and treatment. Minn Med 2010 May;93(5):36-41. [Medline: 20572569]
23. Gilmore B, Michael M. Treatment of acute migraine headache. Am Fam Physician 2011 Feb 01;83(3):271-280 [FREE Full text] [Medline: 21302868]
24. Fumal A, Schoenen J. Current migraine management: patient acceptability and future approaches. Neuropsychiatr Dis Treat 2008 Dec;4(6):1043-1057 [FREE Full text] [doi: 10.2147/ndt.s3045] [Medline: 19337450]
25. Wilson T. Models in information behaviour research. J Doc 1999 Aug;55(3):249-270 [FREE Full text] [doi: 10.1108/EUM0000000007145]
26. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J 2013 Sep;30(3):178-190. [doi: 10.1111/hir.12036] [Medline: 23981019]
27. Dervin B. From the mind's eye of the "user": the sense-making qualitative-quantitative methodology. Qual Res Inf Manag 1992;9:61-84. [Medline: 25246403]
28. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007 Jun;24(2):78-94 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00713.x] [Medline: 17584211]
29. Miles A, Loughlin M, Polychronis A. Evidence-based healthcare, clinical knowledge and the rise of personalised medicine. J Eval Clin Pract 2008 Oct;14(5):621-649. [doi: 10.1111/j.1365-2753.2008.01094.x] [Medline: 19018885]
30. Samani NJ, Tomaszewski M, Schunkert H. The personal genome—the future of personalised medicine? Lancet 2010 May 01;375(9725):1497-1498. [doi: 10.1016/S0140-6736(10)60598-3] [Medline: 20435212]
31. Kluijtmans M, de Haan E, Akkerman S, van Tartwijk J. Professional identity in clinician-scientists: brokers between care and science. Med Educ 2017 Jun;51(6):645-655 [FREE Full text] [doi: 10.1111/medu.13241] [Medline: 28247420]
32. Karlikova M, Polivka J, Strojil J, Topolcan O. A road toward better education in personalized medicine at universities and beyond. Per Med 2015 Jun;12(3):259-267. [doi: 10.2217/pme.14.84] [Medline: 29771646]
33. Grace ES, Wenghofer EF, Korinek EJ. Predictors of physician performance on competence assessment: findings from CPEP, the Center for Personalized Education for Physicians. Acad Med 2014 Jun;89(6):912-919 [FREE Full text] [doi: 10.1097/ACM.0000000000000248] [Medline: 24871243]
34. DeLuca GC, Ovseiko PV, Buchan AM. Personalized medical education: Reappraising clinician-scientist training. Sci Transl Med 2016 Jan 13;8(321):321fs2. [doi: 10.1126/scitranslmed.aad0689] [Medline: 26764155]

Abbreviations

MS: multiple sclerosis
PE: personalized education
PM: precision medicine

Edited by G Eysenbach; submitted 11.06.19; peer-reviewed by C Knoepke, K Hameen-Anttila; comments to author 29.09.19; revised version received 22.12.19; accepted 03.02.20; published 08.04.20.

Please cite as:
Demergazzi S, Pastore L, Bassani G, Arosio M, Lonati C
Information Needs and Information-Seeking Behavior of Italian Neurologists: Exploratory Mixed Methods Study
J Med Internet Res 2020;22(4):e14979
URL: https://www.jmir.org/2020/4/e14979
doi:10.2196/14979
PMID:32181742

Original Paper

Online Health Information–Seeking Among Older Women With Chronic Illness: Analysis of the Women's Health Initiative

Mina S Sedrak[1*], MS, MD; Enrique Soto-Perez-De-Celis[2*], MD; Rebecca A Nelson[1], PhD; Jennifer Liu[1], BS; Molly E Waring[3], PhD; Dorothy S Lane[4], MPH, MD; Electra D Paskett[5], PhD; Rowan T Chlebowski[6], MD, PhD

[1]City of Hope National Medical Center, Duarte, CA, United States

[2]Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran, Mexico City, Mexico

[3]University of Connecticut, Storrs, CT, United States

[4]Stony Brook University School of Medicine, Stony Brook, NY, United States

[5]Ohio State University Comprehensive Cancer Center, Columbus, OH, United States

[6]The Lundquist Institute, Torrance, CA, United States

[*]these authors contributed equally

Corresponding Author:
Mina S Sedrak, MS, MD
City of Hope National Medical Center
1500 E Duarte Rd
Duarte, CA, 91010
United States
Phone: 1 626 256 4673 ext 86635
Email: msedrak@coh.org

Abstract

Background: Understanding how older patients with chronic illnesses use the internet to obtain health information is relevant for the design of digital interventions aimed at improving the health and well-being of adults aged 65 years and older; this cohort represents the sickest, most expensive, and fastest-growing segment of the US population.

Objective: The objective of our study was to describe online health information–seeking behavior among older patients with chronic illnesses and to compare the characteristics of patients who report using the internet to obtain health information with those who do not.

Methods: The study population included 72,806 women aged 65 years and older enrolled in the Women's Health Initiative (WHI), a national cohort study, who completed a 2014 supplemental questionnaire assessing everyday technology use and internet use for researching health conditions. Comparisons were made between participants with and without a history of chronic illness and between users and nonusers of online sources for health information. Multivariate logistic regression was used to estimate odds ratios (ORs) and 95% CIs.

Results: Of the total, 59% (42,887/72,806) of older women used the internet for health information. Compared with women who did not use the internet to obtain health information, those who used the internet were younger (median age: 76 vs 81 years), more likely to be non-Hispanic white (90% [38,481/42,887] vs 87% [26,017/29,919]), earned a higher income (over $US 50,000: 55% [23,410/42,887] vs 33% [9991/29,919]), achieved a higher educational level (more than high school: 87% [37,493/42,887] vs 75% [22,377/29,919]), and were more likely to live with a partner (52% [22,457/42,887] vs 36% [10,759/29,919]) (all P<.001). Women with Alzheimer disease were least likely to report online health information–seeking compared to those without the disease (OR 0.41, 95% CI 0.38-0.43). In contrast, women with a recent diagnosis of cancer, within the previous 2 years (OR 1.23, 95% CI 1.11-1.36) or 2-5 years ago (OR 1.09, 95% CI 1.00-1.19), were most likely to use the internet for health information.

Conclusions: Nearly 6 in 10 older women participating in the WHI reported using the internet to obtain health information. Patients recently diagnosed with cancer are more likely to be looking for health information online, even after adjustment for age, suggesting that these patients may have a greater need for digital health resources.

(J Med Internet Res 2020;22(4):e15906) doi:10.2196/15906

KEYWORDS

online health information–seeking; digital health; technology; chronic disease; internet

Introduction

Digital health technology has been proposed as a potential tool to improve the quality, cost, and safety of health care for older patients [1]. Despite the widespread availability of health information on the internet and recent increases in internet use among older adults in the United States [1,2], there is limited data regarding internet usage among older adults with chronic illnesses for seeking health information [3]. Understanding how older patients with chronic illnesses obtain information is relevant for the design of educational materials and strategies aimed at improving awareness and empowering older adults.

Most of what we know about older adults' use of technology comes from prior research that did not examine online health information–seeking behavior by the presence of particular health conditions. Research by the Pew Research Center shows that older adults remain largely disconnected from the digital world, with one-third of adults aged ≥65 years never having used the internet, and roughly half (49%) without internet services at home [2]. Moreover, a recently published analysis of the National Health and Aging Trends Study [1] showed that the use of everyday technology in older adults (aged ≥65 years) was below that of the general population. For instance, only 16% of older adults obtained health information using health technology, compared to up to 60% in younger populations. However, most patients in this cohort considered themselves to be in excellent or very good health. Interestingly, older adults with more comorbidities and those taking more medications were more likely to use digital health technologies [1]. In another study that evaluated patients hospitalized with acute coronary syndrome, 31% of which included older adults (aged ≥65 years), 57% of patients looked online for health information, with no difference by history of type 2 diabetes, hyperlipidemia, hypertension, cardiovascular disease, or cancer [3]. There was also no difference by age groups. In addition, few qualitative studies [4,5] have been conducted to explore the challenges patients with multiple chronic conditions face when using technology for health-related purposes. These prior studies have limited generalizability, and it remains unclear whether existing technological tools are meeting the needs of patients with high levels of illness burden.

Our objective was to assess the frequency of online health information–seeking among patients with chronic illnesses compared to patients without chronic conditions and to examine characteristics associated with online health information–seeking.

Methods

Study Design, Data Collection, and Study Population

We examined online health information–seeking among a subset of older participants in the Women's Health Initiative (WHI). The WHI study included 61,808 postmenopausal women, aged 50-79 years, enrolled at 40 US clinical centers between 1993 and 1998. Women participated in randomized clinical trials with three overlapping components (n=68,132) or an observational study (n=93,676) [6,7]. Follow-up of participants is ongoing, and the study was approved by the institutional review boards of all participating institutions. All participants provided written informed consent. Research study staff involved in data collection were trained, certified, and recertified annually to carry out specific data collection procedures.

For this study, we included 72,806 postmenopausal women (representing 92% of all active participants) aged ≥65 years, who participated in the WHI cohort study and completed a 2014 supplemental survey on technology and internet use for researching health conditions. This study examined deidentified data. As the current analysis does not meet the criteria for research on human subjects, it did not require approval from an institutional review board.

Measures

In 2014, as part of the WHI Extension Study Supplemental Questionnaire (Form 156), participants were asked about their use of mobile phones, other mobile devices, and computers to access the internet (Yes or No). Additionally, they were asked whether they used the internet to search for health information ("Do you use the Internet to look for health information?"; responses were Yes or No). Multimedia Appendix 1 presents a complete list of technology questions included in the WHI survey [8]. Age, race or ethnicity, annual household income, and smoking status, and medical conditions were self-reported at baseline.

Statistical Analysis

We examined the characteristics of participants who reported using the internet to obtain health information compared to those who did not using chi-square tests for categorical variables and the Wilcoxon signed-rank test for continuous variables. We considered 2-sided *P* values <.05 to be significant. Multivariate logistic regression was used to estimate odds ratios (ORs) and 95% CIs of online health information–seeking among women with and without chronic illness. We performed all analyses using SAS software (SAS Institute Inc).

Results

Of the total participants, 59% (42,887/72,806) reported using the internet to obtain health information. Compared with women who did not use the internet to obtain health information, those who used the internet were younger (median age: 76 vs 81 years); more likely to be non-Hispanic white (90% [38,481/42,887] vs 87% [26,017/29,919]); earned a higher income (over $US 50,000: 55% [23,410/42,887] vs 33% [9991/29,919]); achieved a higher educational level (more than high school: 87% [37,493/42,887] vs 75% [22,377/29,919]); were more likely to be non-smokers (94% [40,203/42,997] vs 91% [27,108/29,919]); and were more likely to use other technology including mobile phones (93% [39,670/42,887] vs 76% [22,662/29,919]), computers (96% [41,042/42,887] vs 47% [14,097/29,919]), text messaging (47% [20,343/42,887] vs 22% [6442/29,919]), email (94% [40,485/42,887] vs 41% [12,153/29,919]), and smartphones (46% [19,843/42,887] vs 15% [4470/29,919]) (all *P*<.001; Table 1). Women who used the internet to obtain health information were more likely to live with a partner (52% [22,457/42,887] vs 36% [10,759/29,919]) but were less likely to live alone (34%

[14,415/42,997] vs 41% [12,383/29,919]) or with others (4% [1890/42,887] vs 6% [1646/29,919]). Internet users were less likely to report frequent utilization of special services (11% [4550/42,887] vs 16% [4863/29,919]), assisted living (2% [715/42,887] vs 5% [1515/29,919]), and nursing home stays within the last year (2% [691/42,887] vs 3% [830/29,919]) (all $P<.001$).

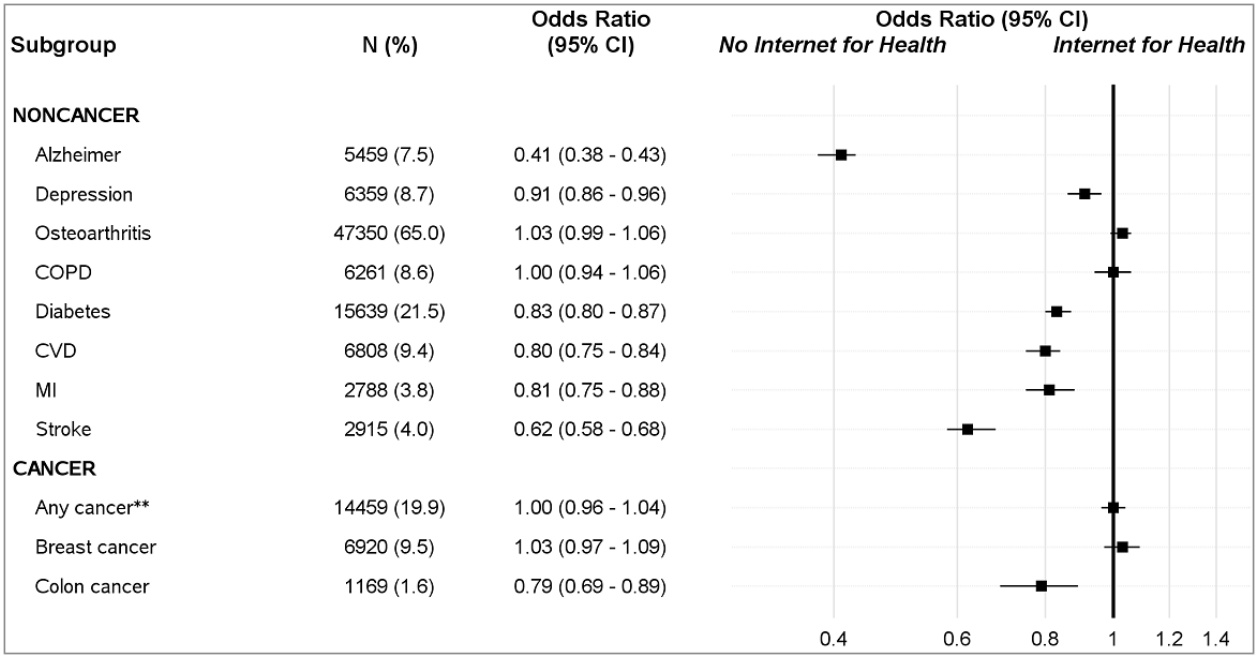
Table 1. Characteristics of older women participating in the Women's Health Initiative in relation to online health information–seeking.

Demographic characteristics	Used the internet for health information (n=42,887)	Did not use the internet for health information (n=29,919)	*P* value
Age at enrollment (years), median (IQR)	59 (55-63)	64 (59-68)	<.001
Age at time of survey (years), median (IQR)	76 (73-80)	81 (76-85)	<.001
Age group at time of survey (years), n (%)			<.001
65-74 years	17,495 (41)	5461 (18)	
75-84 years	21,034 (49)	15,633 (52)	
≥85 years	4358 (10)	8825 (30)	
Race/ethnicity, n (%)			<.001
Non-Hispanic white	38,481 (90)	26,017 (87)	
Annual household income (US $), n (%)			<.001
<20,000	2228 (5)	4324 (15)	
20,000-50,000	15,260 (36)	13,898 (47)	
>50,000	23,410 (55)	9991 (33)	
Education, n (%)			<.001
High school or less	5103 (12)	7357 (25)	
More than high school	37,493 (87)	22,377 (75)	
Living situation, n (%)			
Lives with partner	22,457 (52)	10,759 (36)	<.001
Lives alone	14,415 (34)	12,383 (41)	<.001
Lives with other (child, relative, etc)	1890 (4)	1646 (6)	<.001
Receives special services	4550 (11)	4863 (16)	<.001
Resides in a place with special services	715 (2)	1515 (5)	<.001
Stayed in a nursing home in the past year	691 (2)	830 (3)	<.001
Smoking status, n (%)			<.001
Nonsmoker	40,203 (94)	27,108 (91)	
Smoker	746 (2)	723 (2)	
Media/internet use characteristics, n (%)			
Owns a mobile phone	39,670 (93)	22,662 (76)	<.001
Uses a computer	41,042 (96)	14,097 (47)	<.001
Receives text messages on a mobile phone	20,343 (47)	6442 (22)	<.001
Uses email	40,485 (94)	12,153 (41)	<.001
Uses the internet (for any purpose)	40,934 (95)	10,974 (37)	<.001
Uses a smartphone	19,843 (46)	4470 (15)	<.001

Women with a history of many specific health conditions were less likely to report online health information–seeking (Figure 1). For example, compared to those without a disease, patients with a disease were less likely to use the internet for health information if they had a diagnosis of Alzheimer disease (OR 0.41, 95% CI 0.38-0.43), stroke (OR 0.62, 95% CI 0.58-0.68), colon cancer (OR 0.79, 95% CI 0.69-0.89), cardiovascular disease (OR 0.80, 95% CI 0.75-0.84), myocardial infarction (OR 0.81, 95% CI 0.75-0.88), diabetes (OR 0.83, 95% CI 0.80-0.87), and depression (OR 0.91, 95% CI 0.86-0.96). The only health condition associated with a higher likelihood of online health information–seeking was a recent cancer diagnosis; women diagnosed with cancer within the previous 2 years or 2-5 years were more likely to seek health information online

than women without a history of cancer (OR 1.23, 95% CI 1.11-1.36 and OR 1.09, 95% CI 1.00-1.19, respectively). There were no differences in online health information–seeking among those with osteoarthritis, chronic obstructive pulmonary disease, and breast cancer.

Figure 1. Online health information–seeking behavior in relation to chronic illness status among older women. Diagnosis might have happened at any time prior to the survey. COPD: chronic obstructive pulmonary disease; CVD: cardiovascular disease; MI: myocardial infarction. *All factors were adjusted for the current age group, race, income, and education. **Cancer sites included in the survey: anal, adrenal, appendix, the base of the tongue, biliary, bladder, bone, brain, breast, cerebrospinal, cervical, colon, endocrine, esophagus, eye, gallbladder, genital, gum, hypopharynx, kidney, larynx, leukemia, liver, lung, lymphoma, myeloma, mouth floor, nasopharynx oropharynx, other digestive, ovary, palate, pancreas, parotid, peritoneum, renal pelvis, respiratory, salivary, sinus, stomach, thymus, thyroid, tongue, tonsil, trachea, ureter, urinary, vagina, vulva, and other cancers.



Discussion

Findings

Nearly 6 in 10 postmenopausal women participating in the WHI reported using the internet to obtain health information. Variables associated with less internet usage for health information were older age, nonwhite race, high school education or less, and an income of US $50,000 or less. Although women with several specific health conditions were less likely to engage in online health information–seeking, those who were diagnosed with cancer in the past 5 years were more likely to look to the internet for health information [9].

There are several important implications of this study. First, a large proportion of older adults are using the internet to seek health information. Our findings add to the existing literature [3] by expanding our understanding of technology use in older adults. A recently published analysis of the National Health and Aging Trends Study (NHATS) showed that the use of everyday technology in older adults was below that of the general population, where only 16% of older adults obtained health information using health technology compared to up to 60% in younger populations [1,10]. However, most NHATS patients considered themselves to be in excellent or very good health. Interestingly, older adults with more comorbidities and those taking more medications were more likely to use digital health technologies in the current study. Our findings differ likely because most of the patients who participated in the WHI have chronic health conditions, unlike prior studies of the older adult population.

Second, disparities in online health information–seeking exist among patients who are ethnic minorities and of a lower socioeconomic status. Consistent with prior studies, our findings add to the evidence that digital health is not reaching all seniors equally [1,3,11-14]. Although recent studies have shown that the gap between those who have access to digital technology and those who do not has become increasingly narrow over time [15], digital health interventions are not reaching a proportion of older patients, and this disparity in access and seeking behavior of online health information may contribute to worsening disparities in health outcomes.

Third, the timing and type of chronic illness may play an important role in the online health information–seeking behavior of older patients. Specifically, patients recently diagnosed with cancer within the last 5 years are more likely to be looking for health information online, even after adjustment for age, suggesting that these patients may have a greater need for digital health resources. This is consistent with findings in the literature. One study has shown that information seeking was reported most frequently by cancer survivors than by the general population [16]. In contrast to previous studies that comprise most of the current literature on the use of digital health technology among older adults, the high proportion of patients with a cancer diagnosis in the WHI provides a unique opportunity to evaluate the use of online health

information–seeking among older adults with cancer. Our findings suggest that time since diagnosis of cancer is an important factor in the use of the internet for health among older adults. We also found a relationship between cancer type and online information–seeking behaviors, with patients with colorectal cancer being significantly less likely to use the internet for health information. This mirrors findings from previous studies [16] and may be explained by the fact that the median age at diagnosis of colon cancer in women is 72 years [17], compared with 62 years for women with breast cancer [18].

Limitations

There are limitations to this study. The WHI participants included only postmenopausal women, thus findings may not be applicable to older men. Further studies are needed to better understand online health information–seeking behaviors in relation to chronic illness among older men. Moreover, the experiences of women in the WHI may not be representative of all older women with chronic health conditions. A prior study has shown that women who participated in the WHI observational study may be healthier than same-age women in the United States [19]. Nevertheless, the diversity of participants' backgrounds suggests that our findings may be generalizable to a wider population of older women with chronic health conditions. A further limitation is that our findings are based on responses in 2014. Access to and use of digital tools may have become more widespread since participants were surveyed, as a result of continuous technological advancement. Furthermore, we were not able to report the specific reasons why participants did not use the internet to look for health information. It is possible that some women may be interested in looking for health information online but cannot do so due to a lack of access to the internet. It also may be possible that some women benefit from health information available on the internet through their family members or friends [20,21]. Finally, we are only able to report on older adults' access to and use of the internet to search for health information. We acknowledge that this only reflects one aspect of health information–seeking behavior, and further studies are needed to better understand behaviors such as uses of online portals, online prescribing, and telemedicine.

Conclusions

Understanding the use of digital health technology among older adults with chronic illnesses and whether they obtain health-related information utilizing the internet are essential pieces of information for designing effective and widespread digital health interventions. Our results show that a significant proportion of older women may not be adequately reached by online information, and thus, there is still a need for more traditional forms of media for the dissemination of health information. This is one of the first studies, to our knowledge, to describe the prevalence of online health information–seeking in older adults with chronic illnesses. Our results provide important information regarding online health information–seeking among older women, particularly those with chronic conditions, and could inform the development of health messaging tailored for this population.

Acknowledgments

MSS has received funding support from NIH K12 (K12CA001727), NIA R03 (R03AG064377), the Waisman Innovation Fund, Circle 1500, and the HOPE Foundation. ESPDC received support from a 2016 Long Term International Fellowship from the Conquer Cancer Foundation of the American Society of Clinical Oncology. Preliminary data from this study were presented as a poster abstract at the International Society of Geriatric Oncology 2018 Annual Meeting in Amsterdam, Netherlands.

Authors' Contributions

MSS had full access to all of the study data and takes responsibility for the integrity of the data and the accuracy of the data analysis. MSS, ESPDC, and RTC were involved in study conceptualization and design; MSS, ESPDC, RAN, JL, DSL, MEW, EDP, and RTC were involved in data acquisition, analysis, or interpretation; MSS, RAN, and JL drafted the manuscript; MSS, ESPDC, RAN, JL, DSL, MEW, EDP, and RTC performed critical revision of the manuscript for important intellectual content; RAN completed statistical analysis; and MSS and RTC supervised the study.

Conflicts of Interest

MSS reports research funding from the National Institute of Aging (NIA R03AG064377), the National Cancer Institute (NCI K12CA001727), the Waisman Innovation Fund, Circle 1500, the HOPE Foundation, Seattle Genetics, Novartis, and Eli Lilly. EDP was a stakeholder for Pfizer, and reports grant funding from Pfizer, the Merck Foundation, the Breast Cancer Research Foundation, Foxconn Technology Group, Susan G Komen, and the One Family Foundation. All other authors have no conflicts of interest to declare.

Multimedia Appendix 1
A complete list of technology questions included in the Women's Health Initiative survey.
[DOCX File , 15 KB - jmir_v22i4e15906_app1.docx]

References

1. 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]
2. Anderson M, Perrin A. Tech adoption climbs among older adults. Washington, DC: Pew Research Center; 2017 May. URL: https://tinyurl.com/sjnmfns [accessed 2019-05-09]
3. Waring ME, McManus DD, Amante DJ, Darling CE, Kiefe CI. Online health information seeking by adults hospitalized for acute coronary syndromes: Who looks for information, and who discusses it with healthcare providers? Patient Educ Couns 2018 Nov;101(11):1973-1981 [FREE Full text] [doi: 10.1016/j.pec.2018.06.016] [Medline: 30305253]
4. Kim BY, Lee J. Smart Devices for Older Adults Managing Chronic Disease: A Scoping Review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
5. Alwashmi MF, Fitzpatrick B, Davis E, Gamble J, Farrell J, Hawboldt J. Perceptions of Health Care Providers Regarding a Mobile Health Intervention to Manage Chronic Obstructive Pulmonary Disease: Qualitative Study. JMIR Mhealth Uhealth 2019 Jun 10;7(6):e13950 [FREE Full text] [doi: 10.2196/13950] [Medline: 31199330]
6. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 2002 Jul 17;288(3):321-333. [doi: 10.1001/jama.288.3.321] [Medline: 12117397]
7. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials 1998 Feb;19(1):61-109. [doi: 10.1016/s0197-2456(97)00078-0] [Medline: 9492970]
8. Women's Health Initiative. Form 156 - Supplemental questionnaire. Seattle: Women's Health Initiative; 2016. URL: https://www.whi.org/researchers/data/WHIDataDict/f156_ctos_inv.pdf [accessed 2019-11-01]
9. Mattsson S, Olsson EMG, Johansson B, Carlsson M. Health-Related Internet Use in People With Cancer: Results From a Cross-Sectional Study in Two Outpatient Clinics in Sweden. J Med Internet Res 2017 May 15;19(5):e163 [FREE Full text] [doi: 10.2196/jmir.6830] [Medline: 28506959]
10. Fox S, Duggan M. Health online 2013. 2013. Pew Research Center URL: http://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2018-12-22] [WebCite Cache ID 74sbsnhYG]
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 Jul 04;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]
12. Raven MC, Kaplan LM, Rosenberg M, Tieu L, Guzman D, Kushel M. Mobile Phone, Computer, and Internet Use Among Older Homeless Adults: Results from the HOPE HOME Cohort Study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e10049 [FREE Full text] [doi: 10.2196/10049] [Medline: 30530464]
13. 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]
14. 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;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
15. Smith A. Pew Research Center. 2015. U.S. smartphone use in 2015 URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2015/03/PI_Smartphones_0401151.pdf [accessed 2019-08-16]
16. 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]
17. American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019. Atlanta: American Cancer Society, Inc; 2017. URL: https://tinyurl.com/qoueoua [accessed 2018-03-08] [WebCite Cache ID 6xliPrZTN]
18. American Cancer Society. Breast Cancer Facts & Figures 2017-2018. Atlanta: American Cancer Society, Inc; 2018. URL: https://tinyurl.com/y9dfnpu4 [accessed 2019-01-11] [WebCite Cache ID 75LyQ9TL6]
19. Langer RD, White E, Lewis CE, Kotchen JM, Hendrix SL, Trevisan M. The Women's Health Initiative Observational Study: baseline characteristics of participants and reliability of baseline measures. Ann Epidemiol 2003 Oct;13(9 Suppl):S107-S121. [doi: 10.1016/s1047-2797(03)00047-4] [Medline: 14575943]
20. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney RLJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
21. Luger TM, Hogan TP, Richardson LM, Cioffari-Bailiff L, Harvey K, Houston TK. Older Veteran Digital Disparities: Examining the Potential for Solutions Within Social Networks. J Med Internet Res 2016 Nov 23;18(11):e296 [FREE Full text] [doi: 10.2196/jmir.6385] [Medline: 27881361]

Abbreviations

NHATS: National Health and Aging Trends Study
OR: odds ratio

WHI: Women's Health Initiative

Edited by G Eysenbach; submitted 16.08.19; peer-reviewed by J McCullough, C Hao, H Lee; comments to author 21.10.19; revised version received 04.11.19; accepted 10.02.20; published 09.04.20.

Please cite as:
Sedrak MS, Soto-Perez-De-Celis E, Nelson RA, Liu J, Waring ME, Lane DS, Paskett ED, Chlebowski RT
Online Health Information–Seeking Among Older Women With Chronic Illness: Analysis of the Women's Health Initiative
J Med Internet Res 2020;22(4):e15906
URL: http://www.jmir.org/2020/4/e15906/
doi:10.2196/15906
PMID:32271152

Original Paper

Comparing Medical Term Usage Patterns of Professionals and Search Engine and Community Question Answering Service Users in Japan: Log Analysis

Kazuya Taira[1], MHS; Taichi Murayama[2], MA; Sumio Fujita[3]; Mikiko Ito[1], PhD; Kei Kamide[4], PhD; Eiji Aramaki[2], PhD

[1]Department of Public Health Nursing, Shiga University of Medical Science, Otsu, Shiga, Japan

[2]Social Computing Laboratory, Nara Institute of Science and Technology, Ikoma, Japan

[3]Yahoo Japan Corporation, Tokyo, Japan

[4]Division of Health Science, Osaka University, Osaka, Japan

Corresponding Author:
Kazuya Taira, MHS
Department of Public Health Nursing
Shiga University of Medical Science
Seta Tsukinowa-cho
Otsu, Shiga, 520-2192
Japan
Phone: 81 +81775482398
Fax: 81 77 548 2398
Email: tairak@belle.shiga-med.ac.jp

Abstract

Background: Despite increasing opportunities for acquiring health information online, discussion of the specific words used in searches has been limited.

Objective: The aim of this study was to clarify the medical information gap between medical professionals and the general public in Japan through health information–seeking activities on the internet.

Methods: Search and posting data were analyzed from one of the most popular domestic search engines in Japan (Yahoo! JAPAN Search) and the most popular Japanese community question answering service (Yahoo! Chiebukuro). We compared the frequency of 100 clinical words appearing in the clinical case reports of medical professionals (clinical frequency) with their frequency in Yahoo! JAPAN Search (search frequency) logs and questions posted to Yahoo! Chiebukuro (question frequency). The Spearman correlation coefficient was used to quantify association patterns among the three information sources. Additionally, user information (gender and age) in the search frequency associated with each registered user was extracted.

Results: Significant correlations were observed between clinical and search frequencies (r=0.29, P=.003), clinical and question frequencies (r=0.34, P=.001), and search and question frequencies (r=0.57, P<.001). Low-frequency words in clinical frequency (eg, “hypothyroidism,” “ulcerative colitis”) highly ranked in search frequency. Similarly, “pain,” “slight fever,” and “numbness” were highly ranked only in question frequency. The weighted average of ages was 34.5 (SD 2.7) years, and the weighted average of gender (man –1, woman +1) was 0.1 (SD 0.1) in search frequency. Some words were specifically extracted from the search frequency of certain age groups, including “abdominal pain” (10-20 years), “plasma cells” and “inflammatory findings” (20-30 years), “DM” (diabetes mellitus; 30-40 years), “abnormal shadow” and “inflammatory findings” (40-50 years), “hypertension” and “abnormal shadow” (50-60 years), and “lung cancer” and “gastric cancer” (60-70 years).

Conclusions: Search and question frequencies showed similar tendencies, whereas search and clinical frequencies showed discrepancy. Low-clinical frequency words related to diseases such as “hypothyroidism” and “ulcerative colitis” had high search frequencies, whereas those related to symptoms such as “pain,” “slight fever,” and “numbness” had high question frequencies. Moreover, high search frequency words included designated intractable diseases such as “ulcerative colitis,” which has an incidence of less than 0.1% in the Japanese population. Therefore, it is generally worthwhile to pay attention not only to major diseases but also to minor diseases that users frequently seek information on, and more words will need to be analyzed in the future. Some characteristic words for certain age groups were observed (eg, 20-40 years: “cancer”; 40-60 years: diagnoses and diseases identified in health examinations; 60-70 years: diseases with late adulthood onset and “death”). Overall, this analysis demonstrates that

medical professionals as information providers should be aware of clinical frequency, and medical information gaps between professionals and the general public should be bridged.

(J Med Internet Res 2020;22(4):e13369) doi:10.2196/13369

KEYWORDS

health knowledge; internet; search engine; community question answering service; information-seeking behavior

Introduction

Since the 1990s, the popularization of the internet and personal information devices such as personal computers, smartphones, and tablets has become widespread. Further, opportunities to acquire health information from the web have been increasing. A study examining changes in the basic attributes of information users on the internet in Sweden in 2010 and 2013 showed that the range of users is widening to include younger generations and women [1]. In the United States, 59% of adults were reported to acquire their health information from the internet in 2013 [2]. A survey conducted by the Ministry of Internal Affairs and Communications in Japan in 2015 showed that approximately 80% of people acquired health information via the internet, regardless of gender or age [3]. Moreover, there are many difficulties faced by general users in the process of seeking health information and determining its reliability. In fact, misleading information from the internet could result in serious health hazards [4].

Despite these limitations, health information services have been reported to reduce costs such as medical expenses, to improve production efficiency, and to bring benefits beyond investment [5-7]. The World Health Organization recommended the promotion of such services to member countries from the viewpoint of the quality and safety of medical care, and the possibility of improving access to medical information [8,9]. Taking measures such as the proper management of health information on the web, which is more widely used by the general population, as well as the improvement of services in hospitals and facilities can help promote the management of chronic diseases [10], improve patients' self-efficacy [11], and support treatment decisions [12]; further, the cost effectiveness for improving the health condition of citizens will be high.

In Japan, eHealth Net developed by the Ministry of Health, Labour and Welfare [13], and the websites of Japanese local governments, national research institutes, and academic societies provide reliable health information. Because professionals create the content of these websites, people believe in the validity and current relevance of the information provided and deem the information to have a reliable level of quality. However, it is difficult to find the necessary information from a wide variety of websites where information is segmented according to expertise. In some countries, the government (or its subsidiary organizations) provides comprehensive services, such as the National Health Service in the United Kingdom and MedlinePlus in the United States; however, there is currently no such health service counterpart in Japan.

Conversely, community question answering (CQA) services can directly answer users' questions, and websites with user-generated content such as NAVER [14] are very convenient because they can provide information (gathered for easy skimming) on a given topic; however, the trustworthiness of the provided information cannot be guaranteed due to the lack of expertise of the content providers. To address this issue, there is a movement to ensure the reliability of health information provided on such websites created by the private sector through certification systems such as the Health on the Net Foundation Code of Conduct [15].

It is important to evaluate websites that provide medical information, as the general public tends to be confused by technical terms in their search for reliable websites and information with regard to not only understanding available information but also formulating search queries. For example, searching for "my skin itches" may not lead one to the correct website, which might instead be found by searching for "pruritus cutaneous." It is also necessary for experts who disseminate information to understand these differences when organizing information on websites.

A related study reported that there are not many medical or health-related search queries on the Web, and that the total number of such queries is decreasing compared to the rise in e-commerce search terms [11]. However, another report indicated that health queries account for 4.5% of all searches on two search engines, and that at least 6.75 million health searches are conducted daily [16], indicating that people access a significant amount of health information online.

Further, studies on the quality of health-information websites analyzing search engine rankings and page view statistics demonstrated that English-language Wikipedia is a useful online resource among websites that provide health information [9]. The tools developed to evaluate health information on the internet have also been examined; however, it is still unclear whether or not they are useful because the reliability and validity of many existing tools have not been determined [10].

In addition, studies analyzing people's search process and search logs have reported that they are unable to find the health or medical information they were seeking and have difficulty in formulating effective queries [17,18]. One study that analyzed logs on the Japanese CQA service Yahoo! Chiebukuro reported that people find it difficult to ask questions, and that they are more likely to be interested in information on various health or disease stages and in the content posted by people with similar experiences [12]. In the search process, symptoms and disease names are searched for alternately, and users tend to experience a sense of anxiety during this process [19-21]. Moreover, adverse health effects may result from self-diagnosis and self-treatment [13,22]; therefore, it is crucial to develop a web

environment that does not cause excessive anxiety and health hazards.

Studies comparing search queries between medical professionals and the general public have shown that health care expertise affects users' query selection and assessment of website quality [23]; furthermore, medical professionals use search engines more often, and spend more time in searching and formulating longer queries [24,25]. Thus, leveraging the search behavior of medical professionals to reformulate queries by general users can help improve their search results [26].

Although the above-mentioned studies attempted to improve users' search results to help the general public access more relevant medical information [23,26,27], it is also important for medical professionals to understand users' medical information needs in order for them to provide not only reliable but also accessible medical information on the internet.

Therefore, the aim of the current study was to clarify the difference in the frequently used words by medical professionals and by general internet users; for this purpose, we used data from one of the largest Japanese search engines and a CQA service primarily involving Japanese-speaking users. The results of this study will be useful in identifying the unmet needs of general internet users and in helping health professionals provide medical information tailored to general users.

Methods

We analyzed the query/question logs of the search engine "Yahoo! JAPAN Search" and the CQA service "Yahoo! Chiebukuro" (the Japanese counterpart of Yahoo! Answers), which were provided by the Yahoo Japan Corporation. Data were acquired from the earliest available date for each service up to August 2018: September 2013 to August 2018 for "Yahoo! JAPAN Search" and November 2005 to August 2018 for "Yahoo! Chiebukuro."

We compared the frequency of 100 clinical words appearing in a clinical case search log by medical professionals (clinical frequency) with their frequency in the search logs of Yahoo! JAPAN Search (search frequency) and in the posted questions of Yahoo! Chiebukuro (question frequency). The top 100 words were used for each word frequency category, as this is frequently used by medical staff [28]. The frequency of words in the ~45,000 case reports of the Japan Science Foundation was used for calculating the word use frequency of medical professionals. We analyzed electronic medical records and discharge summaries written by medical professionals from cooperating hospitals, and extracted 100 frequently used words from the MANBYO Dictionary [29], containing not only disease names specified by international standards such as the ICD 10 Standard Disease Name Master (V 4.04 Revised April 1, 2018) but also all symptoms, including abbreviations and the English names of diseases. Approximately 1.6 million words related to symptoms and diseases were extracted from about 290,000 documents, including approximately 363,000 frequently occurring words related to symptoms and diseases merged with words in the ICD 10. These words are hereafter referred to as "clinical frequency 100 words" [28].

To assess the validity of limiting the study to 100 clinical words, we employed three medical and nursing professionals to assist in determining the 100 clinical words. The 100 most frequent clinical words included both generally common words such as "fever," "diabetes," and "hypertension," and medical specialty words such as "multiple myeloma" and "dermatomyositis." Therefore, in clarifying the medical information gap between medical professionals and the general public, we concluded that the use of 100 words was appropriate from the viewpoint of visibility.

We counted the frequency of the clinical frequency 100 words in the query logs of "Yahoo! JAPAN Search" (hereafter referred to as "search frequency 100 words") and about 16 million questions posted in the category of "Health, Beauty, and Fashion" of "Yahoo! Chiebukuro" (hereafter referred to as "question frequency 100 words"). In calculating the question frequency 100 words, a morphological analysis was performed on the character information data with MeCab using the Mecab-ipadic-Neologd dictionary [30], and the occurrence frequency was counted. Thereafter, the Spearman correlation coefficient was used to analyze the association patterns among the three frequencies.

We also performed a qualitative analysis of words with a low ranking in the clinical frequency 100 words (particularly the top 10 words) but a high ranking in the search and question frequency categories. That is, words that were not frequently used by medical staff but are often searched on the web.

In addition, for the search frequency 100 words, the retrieval history associated with registered user information was extracted, and descriptive statistics by gender and age were confirmed. For aggregation by gender and age, the weighted average was calculated, and words that were out of the range of an average value with 2 SD were extracted as unique words for each age and gender:

Gender: $\Sigma_{(k=1}^{100}$ [–1 × Man_{Nk} + 1 × $Woman_{Nk}$]/(Man_{Nk} + $Woman_{Nk}$)/100, where N is the number of searches.

Age: $\Sigma_{k=1}^{100}$ ([10s_{Nk} × 10 + 20s_{Nk} × 20 + 30s_{Nk} × 30 + 40s_{Nk} × 40 + 50s_{Nk} × 50 + 60s_{Nk} × 60]/[10s_{Nk} + 30s_{Nk} + 40s_{Nk} + 50s_{Nk} + 60s_{Nk}])/100, where Xs_N is the number of searches according to X decade of age.

The data analyzed in this study were based on the sum of search results for each of the following devices: personal computer, tablet, and smartphone. The analysis was performed using Python ver. 3.6.5 (Python Software Foundation [31]).

Results

According to StatCounter Global Stats [32], "Yahoo! JAPAN Search" accounted for 28.06% (range 15.82%-40.49%) of the Japanese search engine share in all devices (personal computer, tablet, and smartphone) from 2009 to 2018. Globally, this share is 3.28% (range 1.61%-6.01%) on average. Therefore, most users of this search engine are Japanese, and it was thus considered to be a useful source for understanding the general search situation in Japan.

The results of the correlation analysis of the clinical frequency 100 words, search frequency 100 words, and question frequency 100 words are shown in Table 1, and the respective plot diagrams are shown in Figures 1-3 (also see Multimedia Appendix 1). Note that in Figures 1-3, only words in all uppercase letters were searched for using the Latin alphabet, not in Japanese, and boldface words indicate the top 10 words with regard to clinical frequency.

Figure 1. Rank plot of clinical and search frequencies in the search engine.

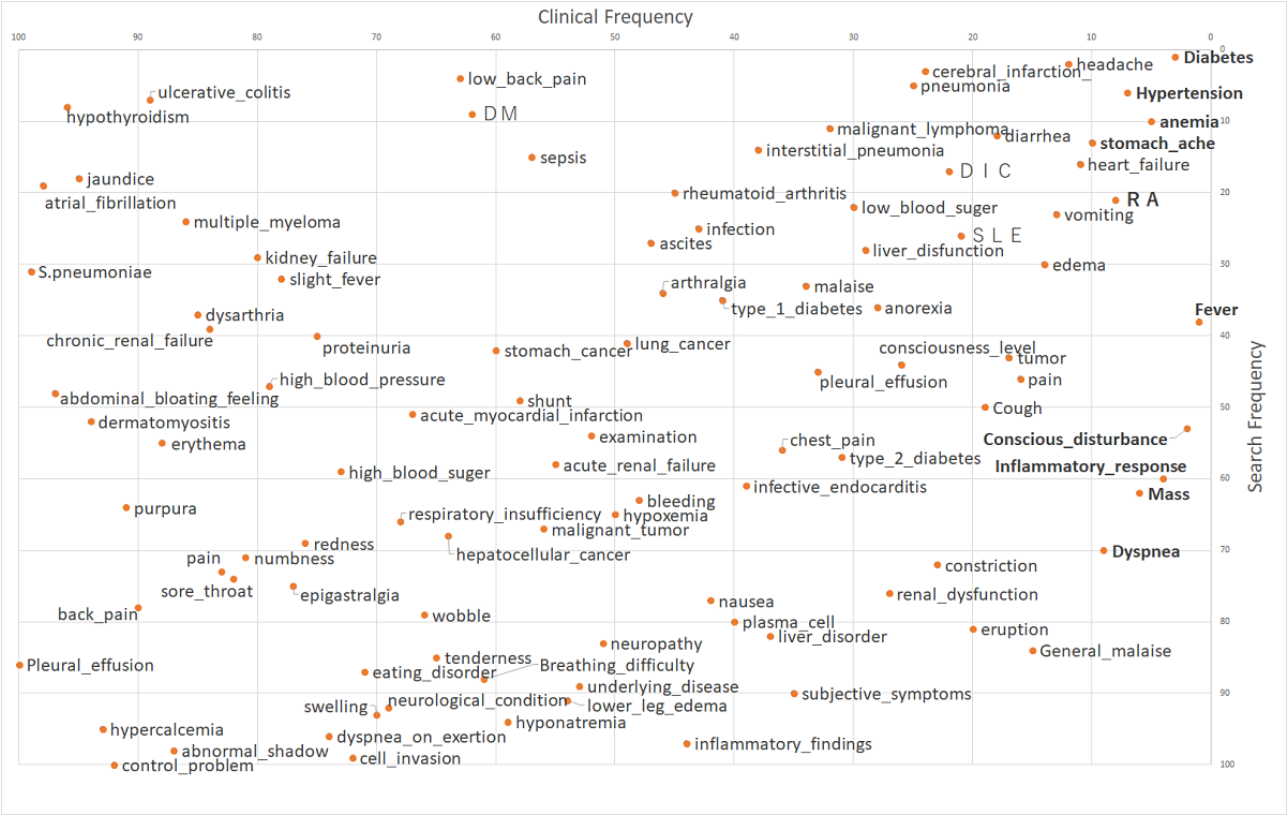


Figure 2. Rank plot of clinical and question frequencies in the community question answering service.

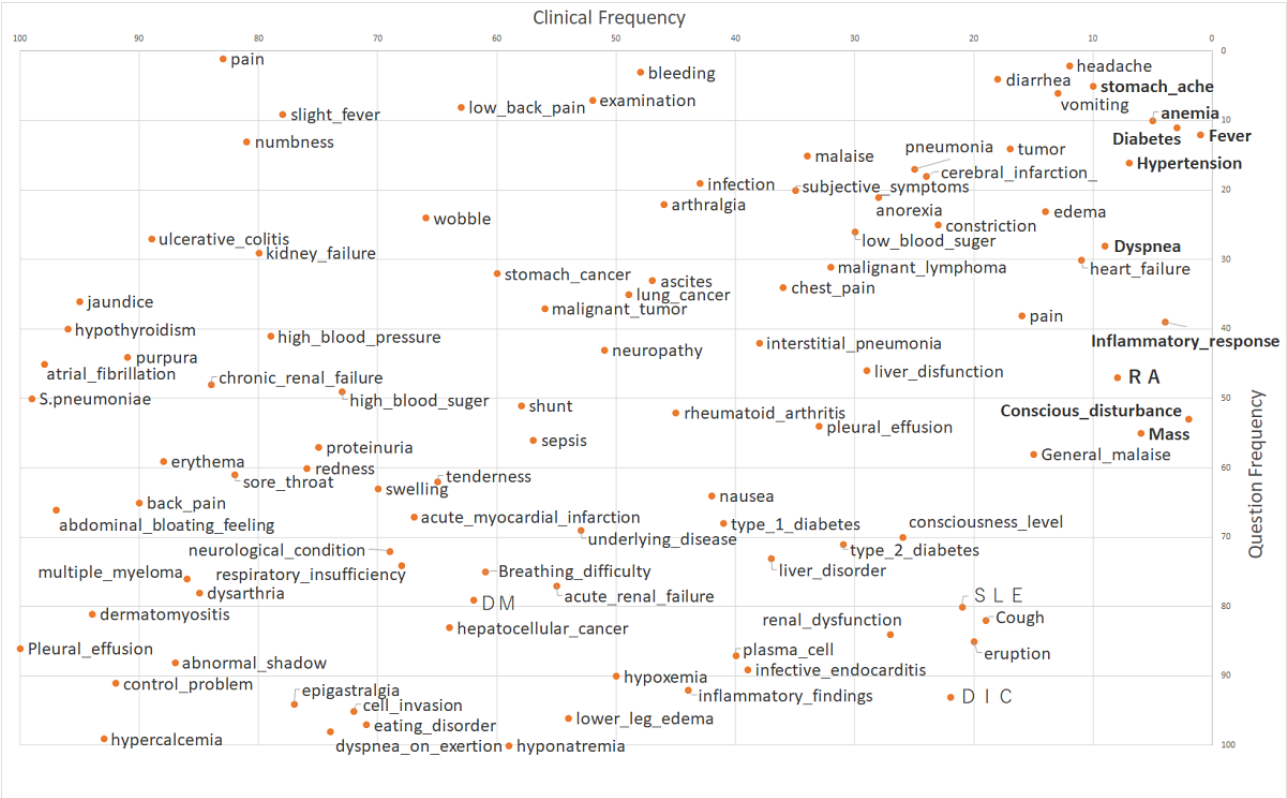


Figure 3. Rank plot of question frequency in the community question answering service and search frequency in the search engine.

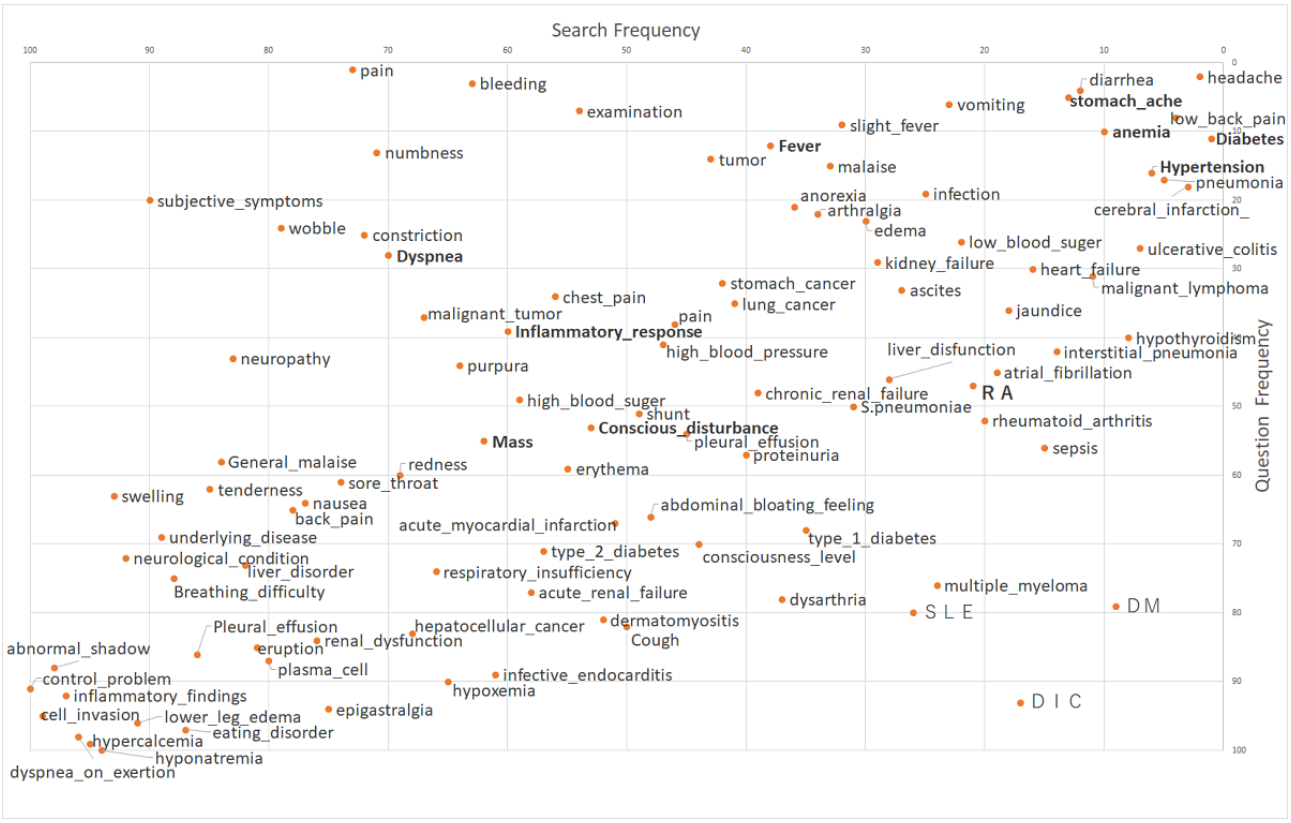


Table 1. Correlation coefficients of the search rank of each platform.

Frequency type	Clinical Frequency	Search Frequency	Question Frequency
Clinical Frequency	1.000	0.290, *P*=.003	0.337, *P*=.001
Search Frequency (Search Engine)	—[a]	1.000	0.569, *P*<.001
Question Frequency (CQA[b] service)	—	—	1.000

[a]Not applicable.

[b]CQA: community question answering.

There was a weak correlation between clinical frequency 100 words and search frequency 100 words (Table 1). Qualitative analysis showed that diseases that are the main cause of death in Japan and the lifestyle diseases that cause them such as diabetes, hypertension, headache, anemia, abdominal pain, heart failure, and cerebral infarction ranked highly for both frequencies. By contrast, words for the evaluation of a patient's condition that are often used by medical professionals, such as ascites retention, hypercalcemia, poor control, and back pain, had a low ranking in both clinical and search frequency 100 words. Words with a low ranking in clinical frequency 100 words and high ranking in search frequency 100 words—that is, the frequency of use by medical staff is relatively low but that of the general public in searches is high—included thyroid function degeneration, ulcerative colitis, jaundice, atrial fibrillation, multiple myeloma, and renal failure.

A weak correlation was also found between clinical frequency 100 words and question frequency 100 words (Table 1). Qualitative analysis showed that words related to symptoms such as headache, abdominal pain, diarrhea, vomiting, anemia, diabetes, and fever were among the top-ranked words for both categories, whereas words such as ascites retention, hypercalcemia, poor control, abnormal shadow, and dermatomyositis ranked lower. Words that were ranked lower in clinical frequency 100 words and higher in question frequency 100 words—that is, the frequency of use by medical staff is relatively low but the question frequency in "Yahoo! Chiebukuro" is high—included pain, slight fever, numbness, ulcerative colitis, and renal failure.

Finally, a moderate correlation was found between search frequency 100 words and question frequency 100 words (Table 1). The top words in both frequencies were headaches, lower back pain, diarrhea, abdominal pain, anemia, and diabetes, whereas the words that ranked lower in both frequencies were hyponatremia, cell invasion, dyspnea on exertion, high anemia, lower edema, inflammatory findings, and poor control. Furthermore, among the words with a large divergence between search and question frequency 100 words, those with a high ranking in question frequency 100 words were subjective symptoms, pain, numbness, and wandering, while those with a high ranking in search frequency 100 words were DM (diabetes

mellitus), DIC (disseminated intravascular coagulation), and SLE (systemic lupus erythematosus).

Considering the top 10 words in bold type in Figures 1-3, the same trend was observed as found for the top 100 words. The words used by medical professionals when managing patients' medical conditions, such as conscious disturbance, inflammatory response, and dyspnea, were found less frequently in both search and question frequencies.

The search ratio by gender for search frequency 100 words was 44.4% (SD 8.7%) for men and 54.6% (SD 9.2%) for women; thus, the search rate was higher for women. Likewise, the weighted average according to gender (men coded as –1 and women coded as +1) was 0.1 (SD 0.1), showing a tendency toward greater searching by women (Table 2). Moreover, the search ratios by age were 2.9% (SD 1.5%) in the 10-20 years age group, 24.3% (SD 8.6%) in the 20-30 age group, 26.0% (SD 4.9%) in the 30-40 age group, 24.6% (SD 5.2%) in the 40-50 age group, 13.9% (SD 4.1%) in the 50-60 age group, and 7.3% (SD 3.4%) in the 60-70 age group; thus, those in the 20-50 age group had the highest search rates. The weighted average was 34.5 (SD 2.7) years (Table 2).

To extract words specific to each gender or age, we searched for words whose weighted average value was larger or smaller than the mean (2 SD), which are summarized in Table 3 and Table 4, respectively.

Table 2. The number of search words by weighted average (WA) by gender and age.

WA	Number of search words
WA by gender[a]	
WA<–0.3	2
–0.3≤WA<–0.2	2
–0.2≤WA<–0.1	5
–0.1≤WA<0	14
0≤WA<0.1	22
0.1≤WA<0.2	28
0.2≤WA<0.3	18
0.3≤WA<0.4	7
WA≥0.4	1
WA by age (years)	
WA<28	1
28≤WA<30	2
30≤WA<32	19
32≤WA<34	23
34≤WA<36	26
36≤WA<38	14
38≤WA<40	13
WA≥40	1

[a]Man, –1; Woman, +1.

Table 3. Search rate by gender and gender-specific queries.

Gender	Mean (SD)	High Frequency	Low Frequency
Men	44.4 (8.7)	RA[a], DM[b]	None
Women	54.6 (9.2)	None	RA, DM

[a]RA: rheumatoid arthritis.

[b]DM: diabetes mellitus.

Table 4. Search rate by age and age-specific queries.

Age (years)	Mean (SD)	High Frequency	Low Frequency
10-20	2.9 (1.5)	dyspnea, stomachache, subjective symptoms, plasma cell	None
20-30	24.3 (8.6)	plasma cell, inflammatory findings, hypoxemia, cell invasion, dyspnea on exertion	None
30-40	26.0 (4.9)	diabetes mellitus, jaundice	high blood pressure
40-50	24.6 (5.2)	abnormal shadow	plasma cell, inflammatory findings, hyponatremia, cell invasion, hypercalcemia
50-60	13.9 (4.1)	rheumatoid arthritis, high blood pressure, abnormal shadow	inflammatory findings
60-70	7.3 (3.4)	interstitial pneumonia, lung cancer, stomach cancer, high blood pressure, atrial fibrillation, *Streptococcus pneumoniae* infection	None

Discussion

Principal Findings

We found a moderate correlation between search and question frequency 100 words, whereas clinical frequency 100 words was only weakly correlated with the other two frequencies. Therefore, the words that are frequently used by medical professionals may differ from words used by general users in search engines or when consulting CQA services. In addition, although the content searched in the search engine and in the CQA service was similar, they also showed unique characteristics.

In qualitative analysis, diabetes and hypertension were recognized as words with a high ranking in all frequency categories. Regarding differences, search frequency 100 words showed a slightly higher frequency for disease names such as heart failure and cerebral infarction, and question frequency 100 words showed a slightly higher frequency for more symptomatic names such as headache, diarrhea, and vomiting.

The characteristics of words that are less frequently used by medical professionals but are frequently searched on the internet by general users differed in both search frequency 100 words and question frequency 100 words. For example, in search frequency 100 words, diseases such as hypothyroidism (rank 8) and ulcerative colitis (rank 7) were common, whereas in question frequency 100 words, symptoms such as pain (rank 1), slight fever (rank 9), and numbness (rank 13) were more typical. Related to this finding, Zhang [12] claimed that questions in the CQA service correspond to the disease stage of the user. The results of the present study suggest that, based on the characteristics of the CQA service, people with specific concrete worries and consultation needs more frequently use the CQA service than searching by queries. Therefore, we consider that the top words in search frequency 100 words are conceptual, whereas the top words used in question frequency 100 words are related to more specific symptoms. Studies on the query logs of both experts and the general public have reported conflicting results, with some indicating that the main focus of searches is on symptoms rather than diseases [20], while others suggesting that the main focus is on diseases rather than symptoms [25]; however, differences may arise depending on whether the search is done on a search engine or through a question on a CQA service.

In addition, words with high search frequencies and question frequencies included designated intractable diseases such as “ulcerative colitis”, with an incidence of less than 0.1% in Japan. Thus, information on diseases that affect a large population of patients is not necessarily high, and factors such as the age at which the disease is likely to develop, severity of the symptoms, prognosis in terms of survival, and presence or absence of a treatment method may also be relevant. Despite the fact that ulcerative colitis is designated as an intractable disease, it does not directly influence survival, and even if remission is achieved, a complete cure is not possible. In fact, the specific questions posted in Japanese on Yahoo! Chiebukuro included the following: “Can you play soccer with ulcerative colitis?,” “Can I become a firefighter or a police officer even if I have ulcerative colitis?,” “Can’t ulcerative colitis be cured completely? It is not a remission but a complete cure,” “Ulcerative colitis doesn’t cause death, does it?? My parent has the disease. I’m very worried. Someone please answer m (* _ _) m! [conventional notation for a bowing gesture].”

General users seek to obtain health information from the internet for various purposes, such as interacting with people who have the same experience [33], looking for advice [34], and understanding the diagnosis [20]. As represented by the words “ulcerative colitis” and “hypothyroidism,” people who are not completely cured but are in remission may be more likely to have such purposes, but because of the low risk to life, such diseases and purposes may be given a lower priority by medical professionals as medical information providers. Since there is generally less information available for minor diseases, it is important to value user-oriented information needs and not only information based on major diseases or professional judgment.

Regarding gender in search frequency 100 words, abbreviations used by medical professionals such as “RA” and “DM” ranked

at the top of gender-specific words. However, when the details of the search related to these words and the words that were searched together were examined, other words with similar spelling such as “JRA” (Japan Racing Associations), “ZARA” (a fashion brand name), and “DMM” (a company name) were also found. Therefore, careful interpretation of alphabetical abbreviations is necessary. Nevertheless, there were no specific queries according to gender, but the overall proportion of females searching was high. These results are consistent with previous studies [35]. In Japan, the proportion of men participating in child rearing and nursing care has also increased, but as women still carry out many of these traditional roles, it was predicted that women are not only searching for themselves but also for their family’s health-related problems.

Regarding age in search frequency 100 words, for people in their 40s and 50s who are expected to have an increasing number of diagnoses of diseases by medical examination, words such as abnormal shadow and hypertension were highly ranked, whereas words such as stomach and lung cancer, atrial fibrillation, interstitial pneumonia, and pneumococcus ranked higher for those in their 60s. Therefore, there may be a connection between diseases whose prevalence increases with age and health information needs. However, it was somewhat difficult to interpret words such as plasma cells, cell wetness, hypoxemia, and exercise dyspnea, which were extracted as specific words by users in their 20s and 30s. More detailed examination of the searches related to these words and the words that were searched together showed that “nursing” was searched together with these words or long sentences, such as case examples used as tasks for nursing students, which suggested that many searches by nursing students were likely included in this category. However, these words are all related to “cancer” as well as cancer treatment of the adolescent and young adult generation in Japan in recent years. Therefore, we believe that further analysis is necessary to determine the detailed trends of the search tendency of each age group.

Strengths and Limitations

This study has several limitations. First, since the analysis was limited to the top 100 clinical frequency words, infrequent words used in search engines and CQA services were also included, which may have confounded the correlation analysis between the three word frequencies. Clinical frequency 100 words are centered on general medical terms and do not include words used by medical personnel in psychiatry, dermatology, obstetrics, and gynecology. Second, the log data analyzed in this study are biased toward Japanese users and cannot be generalized globally, and the relationship between the incidence of disease and the information needs of the general population has not been fully explained; thus, further research is needed.

Nevertheless, this study is significant in that the findings reveal that the frequency of some words differed between clinical and search/question frequencies. Further, compared to search engines, CQA data contained more words about symptoms than diseases. Although no clear causal relationship can be established, the number of diseases may be high, at least depending on the nature of the disease. In addition, more women were found to search for medical information on the internet. Lastly, our analysis highlights that the medical information needs differ according to age group.

Conclusion

In conclusion, when providing medical information on the internet in Japan, medical professionals as information providers should be aware of clinical frequency, and medical information gaps between professionals and the general public should be bridged. Moreover, such information should take into account users’ age and gender, as well as the delivery format of the website and CQA service.

Acknowledgments

We would like to thank Editage (www.editage.jp) for English language editing. This work was supported by the Japan Society for the Promotion of Science KAKENHI Grant Number 18K17637 and the President Discretionary Research Expenses of Shiga University of Medical Science.

Authors' Contributions

KT, conceptualization, formal analysis, writing, and funding acquisition; TM, formal analysis, methodology; SF, data curation, validation, and supervision; MI, validation and methodology; KK, validation and supervision; EA, investigation, methodology, and supervision.

Conflicts of Interest

SF is an employee of Yahoo Japan Corporation, Japanese Internet service company that provides "Yahoo! JAPAN Search" and "Yahoo! Chiebukuro" services analyzed in the paper.

Multimedia Appendix 1
Number of searches and rank of Top 100 words with three types of frequency.
[PNG File , 509 KB - jmir_v22i4e13369_app1.png]

References

1. Wahlstedt E, Ekman B. Patient choice, Internet based information sources, and perceptions of health care: Evidence from Sweden using survey data from 2010 and 2013. BMC Health Serv Res 2016 Dec 01;16:325 [FREE Full text] [doi: 10.1186/s12913-016-1581-5] [Medline: 27480937]
2. Fox S, Duggan M. Health Online 2013.: Pew Research Center: Internet & Technology; 2013 Jan 15. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2020-02-21]
3. Ministry of Public Management. Information Communication White Paper Version; Part 1 Chapter 2 Section 2. 2015. Diffusion of utilization of ICT beyond age URL: http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/pdf/index.html [accessed 2020-02-21]
4. Ministry of Health, Labour and Welfare of Japan. 2017. Death from infant botulism caused by honey [press release] URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-11130500-Shokuhinanzenbu/0000161263.pdf [accessed 2018-12-12]
5. World Health Organization. Executive summary. In: WHO recommendations on home-based records for maternal, newborn and child health: vii?viii. Geneva, Switzerland: World Health Organization; 2018:978-992.
6. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
7. Toms EG, Latter C. How consumers search for health information. Health Informatics J 2007 Sep;13(3):223-235. [doi: 10.1177/1460458207079901] [Medline: 17711883]
8. Fox S. The social life of health information.: Pew Research Center: Internet & Technology; 2009 Jun 11. URL: http://www.pewinternet.org/2009/06/11/the-social-life-of-health-information/ [accessed 2020-02-21]
9. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479. [doi: 10.1197/jamia.M3059] [Medline: 19390105]
10. Jadad AR, Gagliardi A. Rating health information on the Internet: navigating to knowledge or to Babel? JAMA 1998 Feb 25;279(8):611-614. [doi: 10.1001/jama.279.8.611] [Medline: 9486757]
11. Spink A, Yang Y, Jansen J, Nykanen P, Lorence DP, Ozmutlu S, et al. A study of medical and health queries to web search engines. Health Info Libr J 2004 Mar;21(1):44-451. [doi: 10.1111/j.1471-1842.2004.00481.x] [Medline: 15023208]
12. Zhang Y. Contextualizing consumer health information searching: An analysis of questions in a social Q&A community. 2010 Presented at: 1st ACM International Health Informatics Symposium; Nov 11-12, 2010; Arlington, Virginia, USA p. 11-12. [doi: 10.1145/1882992.1883023]
13. Ministry of Health, Labor and Welfare Health information site for lifestyle-related disease prevention. e-health net URL: https://www.e-healthnet.mhlw.go.jp [accessed 2018-12-12]
14. NAVER. URL: https://matome.naver.jp/ [accessed 2020-02-20]
15. The Health On the Net Foundation. Health on the Net. URL: https://www.hon.ch/en/ [accessed 2020-02-21]
16. Eysenbach G, Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. AMIA Annu Symp Proc 2003:225-229 [FREE Full text] [Medline: 14728167]
17. Zeng QT, Kogan S, Plovnick RM, Crowell J, Lacroix E, Greenes RA. Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. Int J Med Inform 2004 Feb;73(1):45-55. [doi: 10.1016/j.ijmedinf.2003.12.015] [Medline: 15036078]
18. 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] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
19. White RW, Horvitz E. Cyberchondria. In: ACM Trans. Inf. Syst. 2009 Nov 01 Presented at: Studies of the onset and persistence of medical concerns in search logs. SIGIR'12 Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval; Aug 12-16, 2012; Portland, Oregon, USA p. 1-37. [doi: 10.1145/1629096.1629101]
20. Cartright M, White R, Horvitz E. Intentions and attention in exploratory health search. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval Jul 24-28; Beijing, China. ACM New York. 2011 Presented at: The 34th International ACM SIGIR Conference on Research and Development in Information Retrieval; Jul 24-28, 2011; Beijing, China p. 65-74. [doi: 10.1145/2348283.2348322]
21. Ryen WW, Eric H. SIGIR'12 Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. Presented at: 35th International ACM SIGIR Conference on Research and Development in Information Retrieval; August 12-16, 2012; Portland, Oregon, USA p. 12-16. [doi: 10.1145/1629096.1629101]
22. Zuccon G, Koopman B, Palotti J. Diagnose this if you can. In: Hanbury A, Kazai G, Rauber A, Fuhr N, editors. Advances in Information Retrieval. ECIR 2015. Lecture Notes in Computer Science, vol 9022. Cham: Springer; 2015:562-567.
23. Tamine L, Chouquet C. On the impact of domain expertise on query formulation, relevance assessment and retrieval performance in clinical settings. Inform Process Manag 2017 Mar;53(2):332-350. [doi: 10.1016/j.ipm.2016.11.004]
24. Tamine L, Chouquet C, Palmer T. Analysis of biomedical and health queries: Lessons learned from TREC and CLEF evaluation benchmarks. J Assn Inf Sci Tec 2015 Jun 01;66(12):2626-2642. [doi: 10.1002/asi.23351]
25. Palotti J, Hanbury A, Müller H, Kahn CE. How users search and what they search for in the medical domain. Inf Retrieval J 2015 Oct 24;19(1-2):189-224. [doi: 10.1007/s10791-015-9269-8]

26. Soldaini L. The Knowledge and Language Gap in Medical Information Seeking. SIGIR Forum 2019 Jan 17;52(1):178-179. [doi: 10.1145/3308774.3308807]
27. Schwarz J, Morris M. Augmenting web pages and search results to support credibility assessment. 2011 Presented at: the SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC, Canada p. 7-12. [doi: 10.1145/1978942.1979127]
28. Aramaki E, Wakamiya S, Yano K, Nagai H, Okahisa T, Ito K. Development of the Clinical Corpus with Disease Name Annotation. Journal of Natural Language Processing 2018;25(1):119-152. [doi: 10.5715/jnlp.25.119]
29. MANBYO Dictionary. Large-scale disease name dictionary for tabulating and analyzing disease names actually used in clinical settings URL: http://sociocom.jp/~data/2018-manbyo/index.html [accessed 2020-02-20]
30. GitHub. Mecab-ipadic-Neologd URL: https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md [accessed 2020-02-20]
31. Python Software Foundation. Python: A dynamic, open source programming language URL: https://www.python.org [accessed 2020-02-18]
32. StatCounter Global Stats. StatCounter. URL: http://gs.statcounter.com/ [accessed 2020-02-21]
33. 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. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
34. Skeels MM, Unruh KT, Powell C, Pratt W. Catalyzing Social Support for Breast Cancer Patients. Proceedings of the Sigchi Conference on Human Factors in Computing systems. CHI Conference, 31 March 2010 2010:173-182 [FREE Full text] [doi: 10.1145/1753326.1753353] [Medline: 21654894]
35. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. Soc Sci Med 2015 Mar;128:151-158. [doi: 10.1016/j.socscimed.2015.01.028] [Medline: 25618604]

Abbreviations

CQA: community question answering
DM: diabetes mellitus
DIC: disseminated intravascular coagulation
RA: rheumatoid arthritis
SLE: systemic lupus erythematosus

Edited by G Eysenbach; submitted 11.01.19; peer-reviewed by J Paparrizos, E Rasmussen; comments to author 02.10.19; revised version received 12.11.19; accepted 04.02.20; published 13.04.20.

Please cite as:
Taira K, Murayama T, Fujita S, Ito M, Kamide K, Aramaki E
Comparing Medical Term Usage Patterns of Professionals and Search Engine and Community Question Answering Service Users in Japan: Log Analysis
J Med Internet Res 2020;22(4):e13369
URL: https://www.jmir.org/2020/4/e13369
doi:10.2196/13369
PMID:32281938

Original Paper

Using Nonexpert Online Reports to Enhance Expert Knowledge About Causes of Death in Dental Offices Reported in Scientific Publications: Qualitative and Quantitative Content Analysis and Search Engine Analysis

Meike Gaiser[1*]; Joachim Kirsch[1*], MD; Till Sebastian Mutzbauer[1,2*], MD, DDS

[1]Institute for Anatomy and Cell Biology, University of Heidelberg, Heidelberg, Germany

[2]Maxillofacial Surgery and Dental Anesthesiology, Mutzbauer & Partner, Zürich, Switzerland

[*]all authors contributed equally

Corresponding Author:
Till Sebastian Mutzbauer, MD, DDS
Maxillofacial Surgery and Dental Anesthesiology
Mutzbauer & Partner
Tiefenhoefe 11
Zürich, 8001
Switzerland
Phone: 41 44211 ext 1465
Email: tsmutzbauer@gmail.com

Abstract

Background: Fatalities rarely occur in dental offices. Implications for clinicians may be deduced from scientific publications and internet reports about deaths in dental offices.

Objective: Data involving deaths in dental facilities were analyzed using Google as well as the PubMed database. By comparing both sources, we examined how internet data may enhance knowledge about deaths in dental offices obtained from scientific medical publications, which causes of death are published online, and how associated life-threatening emergencies may be prevented.

Methods: To retrieve relevant information, we searched Google for country-specific incidents of death in dental practices using the following keywords: “death at the dentist,” “death in dental practice,” and “dying at the dentist.” For PubMed searches, the following keywords were used: “dentistry and mortality,” “death and dental treatment,” “dentistry and fatal outcome,” and “death and dentistry.” Deaths associated with dental treatment in a dental facility, attributable causes of death, and documented ages of the deceased were included in our analysis. Deaths occurring in maxillofacial surgery or pre-existing diseases involved in the death (eg, cancer and abscesses) were excluded. A total of 128 cases from online publications and 71 cases from PubMed publications that met the inclusion criteria were analyzed using chi-square statistics after exclusion of duplicates.

Results: The comparison between the fatalities from internet (n=117) and PubMed (n=71) publications revealed that more casualties affecting minors appeared online than in PubMed literature (online 68/117, 58.1%; PubMed 20/71, 28%; $P<.001$). In PubMed articles, 10 fatalities in patients older than 70 years of age were described, while online sources published 5 fatalities ($P=.02$). Most deaths, both from internet publications and PubMed literature, could be assigned to the category *anesthesia, medication, or sedation* (online 80/117, 68.4%; PubMed 25/71, 35%; $P<.001$). Deaths assigned to the categories *infection* and *cardiovascular system* appeared more often in the PubMed literature (*infection*: online 10/117, 8.5%; PubMed 15/71, 21%; $P=.01$; *cardiovascular system*: online 5/117, 4.3%; PubMed 15/71, 21%; $P<.001$). Furthermore, sedative drugs were involved in a larger proportion of fatal incidents listed online compared to in PubMed (online 41/117, 35.0%; PubMed: 14/71, 20%, $P=.03$). In the United States, more deaths occurred under sedation (44/96, 46%) compared to those in the other countries (Germany and Austria 1/17, 6%, $P=.002$; United Kingdom 1/14, 7%, $P=.006$).

Conclusions: Online and PubMed databases may increase awareness of life-threatening risks for patients during dental treatment. Negative aspects of anesthesia and sedation, as well as the number of deaths of young patients, were underestimated when reviewing PubMed literature only. Medical history of patients, medication dosages, and vital function monitoring are significant issues for practitioners. A high-impact finding from online reports was the underestimation of risks when performing sedation

and even general anesthesia. Detailed knowledge of the definition and understanding of *deep sedation* and *general anesthesia* by dentists is of major concern. By avoiding potentially hazardous procedures, such as sedation-aided treatments performed solely by dentists, the risk of treatment-induced life-threatening emergencies may be reduced.

(J Med Internet Res 2020;22(4):e15304) doi:10.2196/15304

KEYWORDS

dental death; dental practice; dental sedation; risk; internet search engine

Introduction

Fortunately, complications with a fatal outcome rarely occur in dental offices. In the United Kingdom, for example, one dental-related death occurs per 464-758 dentist years [1]. Fatalities may occur not only in elderly and multimorbid patients but also in younger patients, especially in connection to more extensive surgical procedures [2]. The probability that one will experience a medical emergency in a dental setting is remarkably higher; in the United Kingdom, 9 to 11 emergencies occur over a period of 40 years as a dentist [1]. In Saxony, Germany, a study was conducted to find out how often medical emergencies occur in the practice of dentistry; more than 50% of the respondents reported 1 to 3 medical emergencies having occurred within a year and 36% reported the occurrence of up to 10 emergencies [3]. However, it is not possible to determine the fatality risk of life-threatening emergencies happening in dental facilities.

Furthermore, it cannot be determined if fatalities occurred coincidentally or if the dentists' treatments contributed to the fatal events. Despite the low incidence of fatalities in dental facilities, it is important to analyze the causes of death in order to gain knowledge about how these life-threatening emergencies could be prevented. Publications in scientific medical journals may contribute to awareness about medical emergencies occurring in dental facilities. However, it is assumed that additional information could be extracted from internet reports, which could enhance the knowledge about the background and pathogenesis of fatal events. Additional management strategies could be developed when considering information being published on the internet about medical emergencies having resulted in deaths. At best, dentists would avoid pursuing a life-threatening course of treatment, for example, by not performing interventions defined as at-risk procedures.

In 2017, Reuter et al [4] performed a systematic review examining deaths that were related to dental procedures. They reviewed various specialized literature libraries without date restrictions and extracted data from North America, Europe, Asia, and South America about causes and affected patients. A total of 148 fatalities, with an average patient age of 34.6 years, were investigated. In scientific medical literature, as well as on the internet, deaths in dental facilities associated with sedation or general anesthesia have been reported; however, other causes have also been reported [4-32].

The Google Trends database can display the search volume of a keyword, for instance, "dental." Some differences in search volumes can be found for the keywords "dental" and "death" when comparing searches of these words in relation to the total counts of global searches using this database in the United States, the United Kingdom, and in German-speaking countries. However, because of a lack of data, probably due to a low number of searches, there are no statistics available regarding the search term "death dental office" in these respective countries. The term "dental" is searched for twice as often as the term "death" in the United States, the United Kingdom, and in Switzerland; in Germany and Austria, the term "death" is searched for even more seldomly in relation to the term "dental" (see Multimedia Appendix 1).

The importance of the internet in daily life is increasing. Patients use the internet not only to obtain information but also to exchange experiences and to rate doctors. In a study from 2017, it was demonstrated that the number of doctor ratings were increasing [5]. For clinicians, the internet offers additional resources and scientific insights into problems and trends that might be helpful in their daily work as well as for the prevention of diseases. Forums, for example, may deal with patient fears and could be helpful for clinicians as well; they may benefit from patients' and laypersons' points of view and adjust their treatments accordingly.

To the best of our knowledge, there is no overview of deaths in dental offices that has been published on the internet. Hence, we performed a qualitative and quantitative analysis of online publications and medical literature covering fatalities in dental offices in German-speaking countries, France, the United Kingdom, and the United States. In addition to the causes of death, patient groups were analyzed.

This study aims to examine content regarding death in dental treatment facilities that has been published on the internet and retrieved by the Google search engine and compare these findings with content extracted from the scientific medical literature.

The specific research topics of the study are as follows:

1. Which causes of death in dental offices are published on the internet?
2. How can the internet as a database enhance knowledge from scientific medical publications regarding deaths in dental offices?
3. What implications for clinicians can be deduced from internet reports and scientific publications about the reasons for deaths in dental offices?

Methods

Searches Using the Google Search Engine

Between December 21, 2016, and May 3, 2017, we performed a country-specific Google search for deaths in dental facilities.

For the German language, where we used the URL top-level domains .de, .at, and .ch, German equivalents of the keywords "death at the dentist," "death in dental practice," and "dying at the dentist" were used. The terms "death at dentist," "death in dental office," and "dying at the dentist" were used for the English language, where the URL top-level domains .uk and .com were used. The term "death at dentist" was searched in the French language, where the URL top-level domain .fr was used.

An article related to a fatality was included when it had not been published in a medical specialist journal. Additionally, internet pages containing a collection of deaths in a dentist´s or a physician´s office were included. The internet was additionally searched for related content on May 31, 2018, and on October 31, 2018.

Selection of Websites

Articles were included if a patient´s death had been associated with a dental treatment in a dental facility. Furthermore, the age of the casualty and the cause of death had to be indicated, as well as information about the type of anesthesia if the fatality was anesthesia related. In case of a severe pre-existing disease related to the oral cavity (eg, existing abscesses, infections, or tumor diseases), the case was excluded. Deaths related to oral and maxillofacial surgery were also excluded.

Systematization of Articles Found Using Google

The systematization regarding the use of anesthesia was done as follows. Anesthesia was assigned to the category *general anesthesia* when "general anesthesia" or an intubation had been mentioned or when only the term "anesthesia" had been used by the author, even if local anesthesia had been used additionally. The category *sedation* was chosen when the terms "sedation," "sedating drugs," "sedate," "tranquilization," or "nitrous oxide" had been used.

Where there was a lack of information regarding anesthesia, an article only qualified for inclusion if the type of anesthesia had had no influence on the death or the information was unnecessary for the allocation of the cause of death. Cases with general anesthesia or sedation were evaluated with relation to the presence of an anesthesiologist. Furthermore, the anesthesia used was examined with relation to the administration of nitrous oxide. Table 1 lists the categories of the causes of death.

The patients were divided into the following age groups: 0-5 years, 6-17 years, 18-30 years, 31-55 years, 56-70 years, and >70 years. Imprecise age information was rounded; for instance, *early 30s* was rounded to 32 and *mid-70s* to 75.

Table 1. Categories of causes of death.

Categories	Examples
Anesthesia, medication, or sedation	Overdose, anaphylaxis, hypoxia, interactions, or any other recognizable cause
Airway or respiratory system	Suffocation, aspiration, false intubation, or laryngeal edema
Bleeding or coagulation	High blood loss, bleeding, or disseminated intravascular coagulation
Cardiovascular system	Heart attack, cardiac arrest, aneurysm, or air embolism
Infection	Bacterial, viral, or fungal infection; endocarditis, pneumonia, fasciitis, or mucormycosis
Others	Causative, but not temporary; temporary, but not causative

Searches in PubMed

On May 11, 2017, a systematic literature search was performed in PubMed. We searched for the following Medical Subject Headings keywords: Dentistry AND Mortality OR Death AND Dental Treatment OR Dentistry AND Fatal Outcome OR Death AND Dentistry. Through filters, the search was limited to a period of the last 30 years and the term "humans." PubMed was additionally searched on May 31, 2018, and on October 31, 2018.

Selection and Systematization of Relevant Articles From PubMed

The inclusion criteria described in the above sections applied to online literature as well. Editorials, letters, and duplicates were excluded. Articles from Europe, North America, and Australia were included. The systematization was completed analogously to the online database.

Comparison of Selected Publications by Chi-Square Statistics

Eventually, all included deaths from online publications were summarized and compared to the included deaths found in PubMed publications. After carefully reviewing all the cases, duplicates or possible duplicates were excluded before the evaluation of the results using chi-square statistics; the statistical software XLSTAT, version 2018 (Addinsoft), was used for analysis. For further statistical analysis and to definitively exclude duplicate cases, the deaths were assessed by year of occurrence and additionally divided into two groups: cases that occurred before 1995 and cases that occurred after 1994.

Results

Online Publications

Online Results in Total

A total of 18 casualties were found in Germany and 2 were found in Austria. Groups of deaths in the two countries were summarized as deaths in German-speaking countries. In Switzerland, no deaths were reported on the internet. A total of

120 deaths were reported from the United States, 15 from the United Kingdom, and 1 from France. After excluding 28 cases, there were a total of 128 casualties reported from online publications (see Figure 1).

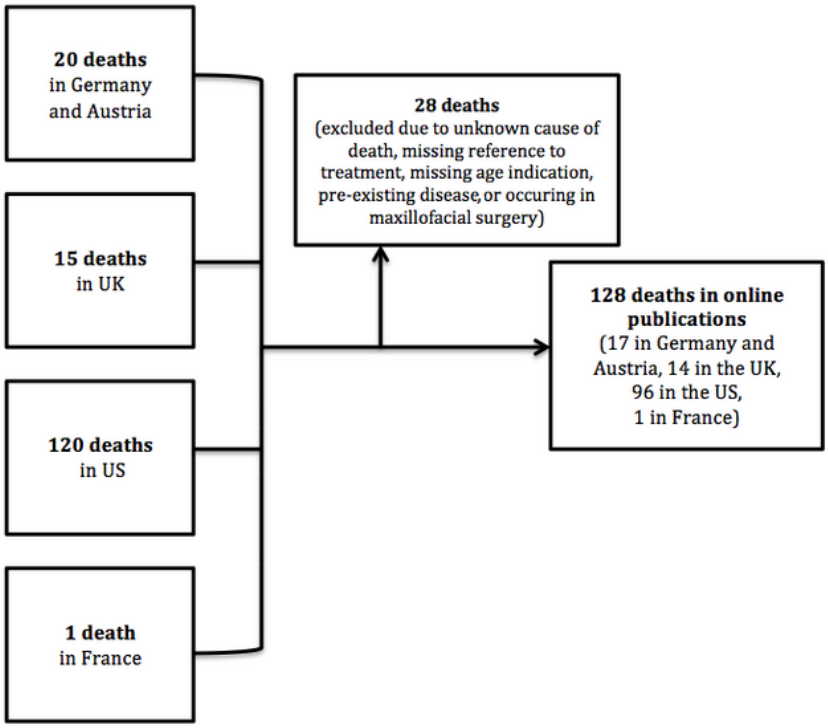
Online Results for Germany and Austria

A total of 3 casualties were excluded from the 20 found in Germany and Austria. The mean age of the 17 remaining casualties was 19.3 years (SD 21.9; median 10 years). The dominating age group was 0-5 years and included 8 fatalities (8/17, 47%). A total of 2 of the deceased (2/17, 12%) were in the 6-17-year age group, whereas the age groups 18-30 years and 31-55 years had 3 patients in each (3/17, 18%). Of the 17 deceased, 1 (6%) was older than 70 years of age and the age group 56-70 years was not represented. A total of 11 patients out of 17 were female (65%) and 6 were male (35%).

The most common cause of death could be attributed to the category *anesthesia, medication, or sedation* (14/17, 82%), followed by *bleeding or coagulation* (2/17, 12%) and *infection* (1/17, 6%). For a breakdown of the number of casualties due to *anesthesia, medication, or sedation* by age, see Multimedia Appendix 2.

Prior to their death, 13 out of 17 casualties (76%) had been treated under general anesthesia. In 1 case each out of 17 (6%), sedation or local anesthesia had been administered, and in 2 cases (12%) no details were given.

Figure 1. Flowchart for article selection from Google results on the basis of article title. UK: United Kingdom; US: United States.



Online Results for the United Kingdom

The mean age of the 14 deceased patients was 30.5 years (SD 22.1; median 28.5 years). The dominating age groups were 6-17 years and 31-55 years, with 4 cases out of 14 (29%) each, whereas 2 patients (14%) in the age group 0-5 years died. The age group 18-30 years was represented with 1 case out of 14 (7%). Of the 14 deceased, 3 (21%) were in the 56-70-year age group, and the over-70-year age group was not affected at all. Half of the deceased were male (7/14, 50%) and half were female (7/14, 50%).

In the United Kingdom, the cause of death was attributed to the *anesthesia, medication, or sedation* category in 9 of the 14 cases (64%), to *bleeding or coagulation* in 3 cases (21%), and to *infection* and *others* in 1 (7%) case each. In the *others* category, a case of a small girl who had been starving to death after a traumatic visit to the dentist was described. For a breakdown of the number of deaths due to *anesthesia, medication, or sedation* by age, see Multimedia Appendix 3.

General anesthesia had been administered in 6 of the 14 cases (43%), sedation in 1 case (7%), and local anesthesia in 5 cases (36%). In addition, 2 patients (14%) did not receive any anesthesia at all.

Online Results for the United States

The mean age of the deceased patients was 22.6 years (SD 21.6; median 17 years). More than every second deceased patient was a minor: 25 out of 96 (26%) were in the 0-5-year age group, 29 (30%) were in the 6-17-year age group, 18 (19%) were in the 18-30-year age group, and 12 (13%) were in the 31-55-year age group. In the age group 56-70 years, 7 patients (7%) died, and 5 (5%) of the deceased were older than 70 years. Out of 96 deceased patients, 54 (56%) were male and 42 (44%) were female.

In the United States, 61 of 96 casualties (64%) were assigned to the *anesthesia, medication, or sedation* category. Out of 96 cases, 51 (53%) were related to anesthesia (ie, general anesthesia, sedation, or local anesthesia) and the other 10 cases (10%) were attributed to medication given before or after

treatment (eg, pain medication and antibiotics). For a breakdown of the number of deaths due to *anesthesia, medication, or sedation* by age, see Multimedia Appendix 4. A total of 15 out of 96 (16%) deaths were assigned to the *airway or respiratory system* category, 8 cases (8%) to the *cardiovascular system*, 9 (9%) to *infection*, and 2 (2%) to *others*, which included 1 that was defined as a "natural cause of death" and 1 (1%) that was attributed to an asthma attack. Out of 96 cases, 1 (1%) was assigned to the *bleeding or coagulation* category.

In 29 out of 96 cases (30%), no information about the administered anesthesia had been reported. Most of the casualties had undergone sedation (44/96, 46%), and 16 patients (17%) had been treated under general anesthesia. In 7 cases (7%), only local anesthesia had been used.

From the United States, papers including 11 cases that had received nitrous oxide had been published. In 10 of these cases, the cause of death was assigned to *anesthesia, medication, or sedation*; these are subdivided in Table 2.

Table 2. Dental practice casualties involving nitrous oxide assigned to the anesthesia, medication, or sedation category from online publications in the United States.

Sedative agents	Casualties (n=10), n (%)
Nitrous oxide with local anesthesia	
Total	3 (30)
Nitrous oxide reported to be causative	1 (10)
Nitrous oxide with sedative drugs or general anesthesia	
Total	7 (70)
Nitrous oxide reported to be causative	2 (20)

In the 11th case where nitrous oxide was used, the cause of death was assigned to the *airway or respiratory system* category. This case had also been reported in the PubMed literature; the little girl had died after having aspirated a cotton roll. In only 1 of the 11 cases (9%), an anesthesiologist had been on site, and in 1 case (9%) there was no information regarding an anesthesiologist standing by.

In Table 3, all fatalities under general anesthesia and sedation were listed regarding the presence of an anesthesiologist or anesthetist. In the cases without a supervising anesthesiologist, a second dentist or even the treating dentist himself or herself had acted as *anesthesiologist*. It is unclear whether the respective dentists had completed an advanced training course prior to the procedures.

Table 3. Fatalities in dental practices under general anesthesia and sedation with respect to the presence of an anesthesiologist or anesthetist, based on data from online publications in the United States.

Setting of fatalities	Anesthesiologist present, n (%)	Anesthetist present, n (%)	No anesthesiologist or anesthetist present, n (%)	No information, n (%)
Fatalities where general anesthesia or sedation were used (n=60)	6 (10)	1 (2)	33 (55)	20 (33)
Fatalities where general anesthesia or sedation were used and cause of death was assigned to *anesthesia, medication, or sedation* category (n=47)	6 (13)	1 (2)	26 (55)	14 (30)

Online Results for France

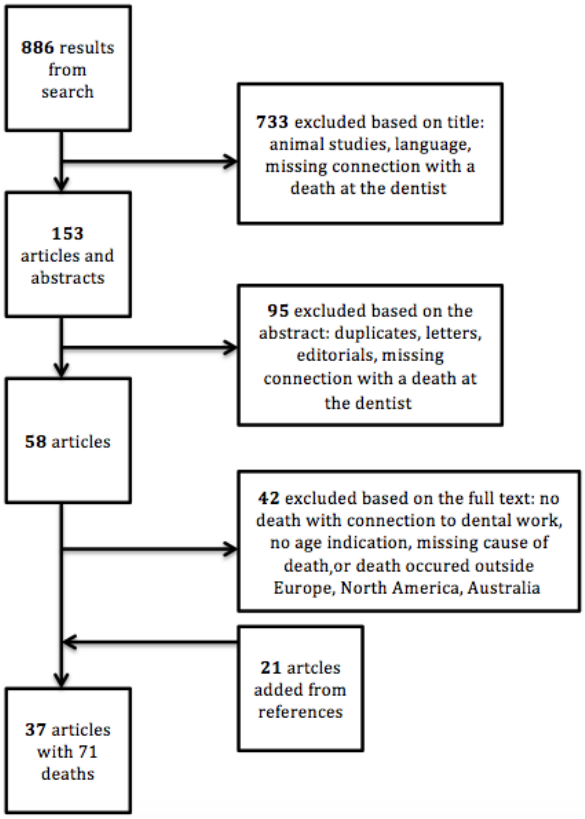
Only 1 casualty occurred in France. The deceased was a 7-year-old boy who had aspirated a protective cover for the x-ray film and had suffocated in the dental office. Consequently, the death was classified as *airway or respiratory system*. No anesthesia had been administered to the patient.

PubMed Publications

PubMed Publications in Total

We found 886 publications from our PubMed search. The flowchart in Figure 2 describes the process of exclusion and inclusion of publications. A total of 733 publications were excluded based on the title, 95 were excluded based on the abstract, and 58 publications underwent a full-text review. Finally, 37 publications reporting a total of 71 fatalities were included in the study. A total of 20 publications were case reports [9,10,12,13,18,20,22,24-26,28-37], 5 were retrospective studies [6,14,21], and 12 were case studies [7,8,11,17,19,23,26,38-42].

Figure 2. Flowchart for article selection from PubMed results.



PubMed Results

The mean age of the 71 deceased from the PubMed literature was 37.9 years (SD 26.4; median 32 years) and included 43 (61%) male and 27 (38%) female patients. In 1 case (1%), the author had not specified a gender. The dominating age groups were 56-70 years and minors up to 5 years of age, with 15 cases (21%) each. The age group between 6 and 17 years was represented by 5 cases (7%), and the age group between 18 and 30 years was represented by 14 cases (20%). Of the 71 deceased, 12 (17%) died between 31 and 55 years of age and 10 (14%) were older than 70 years of age.

According to the PubMed literature, 25 cases out of 71 (35%) were assigned to the *anesthesia, medication, or sedation* category. A total of 24 cases (34%) were related to anesthesia (ie, general anesthesia, sedation, or local anesthesia), while the 25th patient died due to an anaphylactic shock originating from the dental impression material. The *infection* and the *cardiovascular system* categories were represented by 15 cases (21%) each, the *airway or respiratory system* category by 12 cases (17%), and the *bleeding or coagulation* category by 4 cases (6%). For a breakdown of the number of deaths due to *anesthesia, medication, or sedation* by age, see Multimedia Appendix 5.

The most common cause of death was from general anesthesia (22/71, 31%), followed by local anesthesia and sedation with 14 cases each (20%). Only 1 casualty (1%) had occurred completely without anesthesia. In the remaining 20 cases (28%), the author had not provided any information about the anesthesia administered.

Out of 71 fatalities, the 36 cases (51%) where general anesthesia or sedation had been used were examined with respect to the presence of an anesthesiologist or anesthetist (see Table 4).

Table 4. Fatalities in dental practices with general anesthesia and sedation with respect to the presence of an anesthesiologist or anesthetist, based on data from PubMed publications.

Setting of fatalities	Anesthesiologist present, n (%)	Anesthetist present, n (%)	No anesthesiologist or anesthetist present, n (%)	No information, n (%)
Fatalities where general anesthesia or sedation were used (n=36)	3 (8)	6 (17)	5 (14)	22 (61)
Fatalities where general anesthesia or sedation were used and cause of death was assigned to *anesthesia, medication, or sedation* category (n=19)	3 (16)	2 (10)	4 (21)	10 (53)

In PubMed, 11 fatalities where nitrous oxide had been used were described. In 8 of these cases (73%), the cause of death was classified as *anesthesia, medication, or sedation*. However, it was not obvious that nitrous oxide was the cause of death in any of the cases; only the involvement of nitrous oxide was mentioned. The other 3 cases (27%) involving nitrous oxide

were assigned to *the airway or respiratory system* category. Those patients had died because of aspiration of a cotton roll, bronchospasm, or laryngeal edema. In Table 5, the PubMed data are compared with online results.

Table 5. Age and sex of casualties, cause of death, and anesthesia used in dental facilities found in online publications from three country groups compared with the PubMed literature.

Variable	Germany and Austria (n=17), n (%)	United Kingdom (n=14), n (%)	United States (n=96), n (%)	PubMed (N=71), n (%)
Age (years)				
0-5	8 (47)	2 (14)	25 (26)	15 (21)
6-17	2 (12)	4 (29)	29 (30)	5 (7)
18-30	3 (18)	1 (7)	18 (19)	14 (20)
31-55	3 (18)	4 (29)	12 (13)	12 (17)
56-70	0 (0)	3 (21)	7 (7)	15 (21)
>70	1 (6)	0 (0)	5 (5)	10 (14)
Sex				
Female	11 (65)	7 (50)	42 (44)	27 (38)
Male	6 (35)	7 (50)	54 (56)	43 (61)
Not specified	0 (0)	0 (0)	0 (0)	1 (1)
Cause of death				
Anesthesia, medication, or sedation	14 (82)	9 (64)	61 (64)	25 (35)
Bleeding or coagulation	2 (12)	3 (21)	1 (1)	4 (6)
Infection	1 (6)	1 (7)	9 (9)	15 (21)
Others	0 (0)	1 (7)	2 (2)	0 (0)
Cardiovascular system	0 (0)	0 (0)	8 (8)	15 (21)
Airway or respiratory system	0 (0)	0 (0)	15 (16)	12 (17)
Anesthesia				
General anesthesia	13 (76)	6 (43)	16 (17)	22 (31)
Sedation	1 (6)	1 (7)	44 (46)	14 (20)
Local anesthesia	1 (6)	5 (36)	7 (7)	14 (20)
None	0 (0)	2 (14)	0 (0)	1 (1)
No information	2 (12)	0 (0)	29 (30)	20 (28)

Comparison of PubMed Literature With Summarized Online Publications

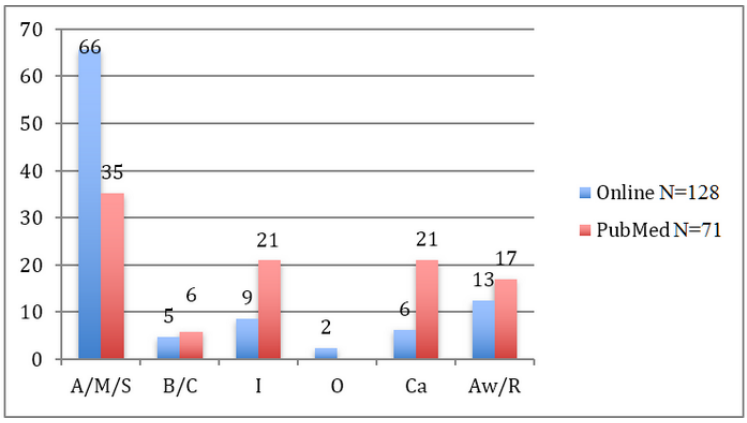
Age

The mean age of the deceased patients in online publications was 22.9 years (SD 21.6; median 17 years) compared with 37.9 years (SD 26.4; median 32 years) in PubMed articles. For a compilation of the age distribution, see Multimedia Appendix 6.

Cause of Death

The *anesthesia, medication, or sedation* category dominated the online and the PubMed publications (see Table 5); however, the proportions in PubMed were only about half as high compared to online publications. Whereas *bleeding or coagulation* problems did not show a large difference, *infection*, *cardiovascular system*, and *airway or respiratory system* problems were overly represented in PubMed (see Figure 3).

Figure 3. Percentage of casualties for each cause of death in dental facilities using data from online publications and PubMed. Possible duplicates were not excluded. A/M/S: anesthesia, medication, or sedation; B/C: bleeding or coagulation; I: infection; O: others; Ca: cardiovascular system; Aw/R: airway or respiratory system.



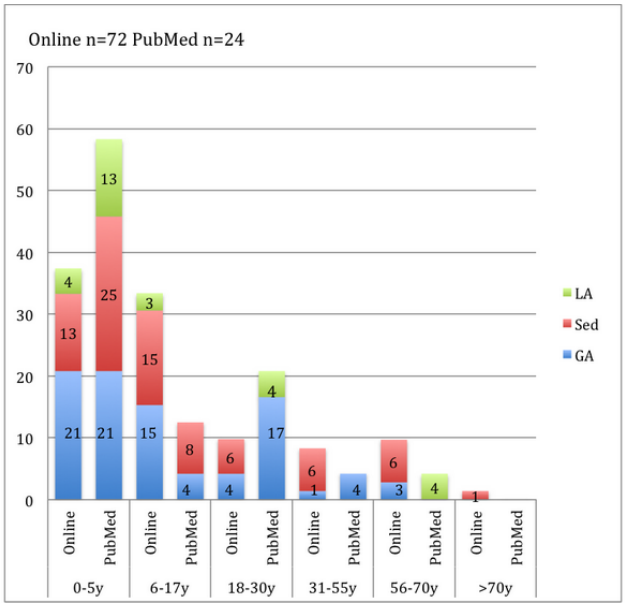
Anesthesia

Whereas local anesthesia as the reason for fatal incidents showed a higher proportion in PubMed publications, sedation was more strongly represented in online reports. No large differences in general anesthesia fatality rates were found. For a breakdown of the types of anesthesia used, see Multimedia Appendix 7.

Anesthesia and Age

The analysis of PubMed articles showed higher proportions of both sedation and local anesthesia leading to fatalities among patients 5 years of age and younger compared to online publications. However, the proportion of sedation leading to fatalities dominated the 6-17-year-old and 31-years-and-older age groups in online media. General anesthesia showed higher percentages in online publications among the 6-17-year-old and 56-70-year-old age groups and higher percentages in PubMed articles among the 18-30-year-old and 31-55-year-old age groups (see Figure 4).

Figure 4. Percentage of deaths in dental facilities reported in online and PubMed publications categorized as *anesthesia, medication, or sedation* with the use of general anesthesia (GA), sedation (Sed), or local anesthesia (LA) in each age group (years, y). Possible duplicates were not excluded.



Comparison by Chi-Square Statistics

Selection of Publications to Compare

Out of 128 deaths found in online publications, 11 cases (8.6%) were excluded for statistical analysis, as these were also found in the PubMed literature. Therefore, 117 cases from online publications were compared with 71 cases from the PubMed literature using chi-square statistics. When comparing the three country groups—Germany and Austria, the United Kingdom, and the United States—all cases were included.

Age

A higher proportion of fatalities in the 17-years-and-younger age group was found in online publications (68/117, 58.1%) compared to PubMed articles (20/71, 28%). However, among the group of patients older than 70 years, fatal outcomes were found in 14% (10/71) of PubMed publications compared to 4.3% (5/117) of online publications (see Table 6).

Table 6. Comparison of fatalities reported in online and PubMed publications with respect to the age of deceased patients.

Patient age group	Number of fatalities, n (%)		*P* value
	Online (N=117)	PubMed (N=71)	
Over and under 18 years			<.001
<18 years	68 (58.1)	20 (28)	
≥18 years	49 (41.9)	51 (72)	
Over and under 70 years			.02
≤70 years	112 (95.7)	61 (86)	
>70 years	5 (4.3)	10 (14)	

Causes of Death

Fatalities in the *anesthesia, medication, or sedation* category dominated online, whereas *infection* and *cardiovascular system* rates were higher in PubMed. No differences between online and PubMed publications were found when *airway or respiratory system* and *bleeding or coagulation* categories were compared (see Table 7).

Table 7. Comparison of fatalities reported in online and PubMed publications with respect to causes of death.

Cause of death	Number of fatalities, n (%)		*P* value
	Online (N=117)	PubMed (N=71)	
Anesthesia, medication, or sedation			<.001
Yes	80 (68.4)	25 (35)	
No	37 (31.6)	46 (65)	
Infection			.01
Yes	10 (8.5)	15 (21)	
No	107 (91.5)	56 (79)	
Cardiovascular system			<.001
Yes	5 (4.3)	15 (21)	
No	112 (95.7)	56 (79)	
Airway or respiratory system			.26
Yes	13 (11.1)	12 (17)	
No	104 (88.9)	59 (83)	
Bleeding or coagulation			.88
Yes	6 (5.1)	4 (6)	
No	111 (94.9)	67 (94)	

Comparison Regarding the Use of Sedation

The number of fatalities attributed to sedation were higher in online reports covering the US population compared to data from other countries (see Table 8). The proportion of sedation as the reason for fatal outcomes was higher in the online reports (41/117, 35.0%) compared to articles from PubMed (14/71, 20%) (see Table 9).

Table 8. Comparison of fatalities in dental facilities as a result of sedation reported in online publications with respect to country.

Country	Fatalities resulting from sedation, n (%)	Fatalities not resulting from sedation, n (%)	*P* value
Germany and Austria (n=17)	1 (6)	16 (94)	N/A[a]
United Kingdom (n=14)	1 (7)	13 (93)	N/A
United States (n=96)	44 (46)	52 (54)	N/A
Germany and Austria vs United Kingdom	N/A	N/A	.89
Germany and Austria vs United States	N/A	N/A	.002
United Kingdom vs United States	N/A	N/A	.006

[a]N/A: not applicable.

Table 9. Comparison of fatalities in dental facilities as a result of sedation in online and PubMed publications.

Publication source	Fatalities resulting from sedation, n (%)	Fatalities not resulting from sedation, n (%)	*P* value
Online (N=117)	41 (35.0)	76 (65.0)	.03
PubMed (N=71)	14 (20)	57 (80)	

Comparison Before 1995 and After 1994

To fully exclude duplicates, the online cases after 1994 were compared with the PubMed cases before 1995. As shown in Table 10, there is no difference between the respective time frames regarding PubMed or when the numbers in the categories leading to fatalities were compared.

Whereas data identifying *anesthesia, medication, or sedation* as contributing factors dominated in online reports after 1994 compared to publications listed in PubMed before 1995, *sedation*-related casualty proportions in both of the examined databases were not different (see Table 11).

Table 10. Comparison of fatalities in dental facilities described in PubMed before 1995 and after 1994 with respect to deaths assigned to the anesthesia, medication, or sedation and sedation-only categories.

Cause of death	Fatalities reported in PubMed, n (%)		*P* value
	Before 1995 (n=23)	After 1994 (n=48)	
Anesthesia, medication, or sedation			.63
Yes	9 (39)	16 (33)	
No	14 (61)	32 (67)	
Sedation only			.77
Yes	5 (22)	9 (19)	
No	18 (78)	39 (81)	

Table 11. Comparison of fatalities in dental facilities described in PubMed before 1995 and online after 1994 with respect to deaths assigned to the anesthesia, medication, or sedation and sedation-only categories.

Cause of death	Fatalities reported, n (%)		*P* value
	PubMed before 1995 (n=23)	Online after 1994 (n=108)	
Anesthesia, medication, or sedation			.008
Yes	9 (39)	74 (68.5)	
No	14 (61)	34 (31.5)	
Sedation only			.24
Yes	5 (22)	37 (34.3)	
No	18 (78)	71 (65.7)	

Discussion

Principal Findings and Implications

Overview

The main finding of this study was that reports regarding causes of death in dental offices published on the internet enhance knowledge about the respective fatalities compared to extracting data from PubMed only. Fatalities associated with medication, general anesthesia, and sedation would have been underrepresented if only examining fatalities published in PubMed without considering cases reported on the internet. The same applies to the fatality rates of minors having died related to a dental treatment. In comparison to Reuter et al [4], the PubMed literature review conducted in this study was restricted to the last 30 years and to publications from Europe, North America, and Australia. Therefore, a lower number of death cases were obtained in this study's PubMed search. This procedure was applied in order to use the same criteria in both the PubMed and online searches and to enable a reasonable comparison of the results. Reuter et al's publication [4], as well as this study, revealed a fatality value of 21% for the category *cardiovascular system*; in addition, as a main result, it can be concluded that in both works most fatalities were assigned to the category *anesthesia, medication, or sedation*.

Age

The differing mean age in online publications (mean 22.9 years, SD 21.6; median 17 years) and in the PubMed literature (mean 37.9 years, SD 26.4; median 32 years) can possibly be explained by an increased awareness by the public when younger people die as a result of dental treatment.

On the other hand, the authors of the PubMed articles evaluated available data using a scientific approach. This may explain the higher average age in scientific literature as reflecting a more even age distribution.

Cause of Death

Overview

Death categories in PubMed articles were more evenly distributed compared to online publications. In both the online publications and in the PubMed literature, the most common cause of death could be assigned to the *anesthesia, medication, or sedation* category; the percentage in online publications was 65.6% (84/128), which is almost twice as high as in the PubMed articles (25/71, 35%).

Presumably, articles describing deaths associated with anesthesia or sedation (ie, *anesthesia, medication, or sedation* category) may induce a compassionate effect in a normal reader and are, therefore, more suitable for headlines, especially when the group of patients is under 18 years of age, which was overly represented.

Furthermore, it is noticeable that only 11 out of 128 (8.6%) of the fatalities reported in online publications occurred after infection, whereas 21% (15/71) of infection-related fatalities were found in PubMed articles. This could be explained by symptoms of infection having developed with delay after dental treatment. Compared with a death directly related to the administration of anesthesia or sedation for dental treatment, delayed infection may be less suitable as an online headline.

In addition, a clear difference between online publications (8/128, 6.0%) and PubMed articles (15/71, 21%) in the category *cardiovascular system* was found. This difference may have resulted from the higher average age of the patients in the PubMed articles. Furthermore, it is assumed that deaths in the *cardiovascular system* category were not associated with dental treatments, especially by medical laypersons.

Anesthesia, Medication, or Sedation

In the case of deaths attributed to the category *anesthesia, medication, or sedation*, omissions in the use of adequate equipment or medication overdose had predominantly been responsible for the death of the patient. Furthermore, medication side effects or undiscovered diseases had sometimes been attributed to the fatal outcome. Regarding these causes of death, dentists' adequate knowledge of the equipment, medication, and respective complications is rated as a major issue. Whether anesthesiologists have been underinformed or have been ignoring related health or safety issues in the respective cases remains unclear.

Data from the 1960s until now already show an impact of anesthesia, medication, or sedation procedures on fatalities in dental facilities. However, some bias may arise regarding substances historically available that are no longer in use in many countries, such as halothane [4]. Enhanced knowledge about fatalities in dental facilities covering the last 30 years has been obtained from additional information derived from internet reports compared to PubMed literature alone within this time frame. Although some historical medication identified as a cause of death in previous publications is no longer available, performing sedation or even general anesthesia in certain dental facilities without the assistance of an anesthesiologist is still considered dangerous.

Frequently, online publication analysis revealed that both dentists and dental staff had had no or insufficient knowledge of lifesaving procedural management. Regular emergency management and cardiopulmonary resuscitation training is advisable [43]. Emergency management of the respective complications would not be necessary if at-risk interventions, such as deep sedation or even general anesthesia, by dentists could be avoided, since these complications would not even occur. Dentists should probably be discouraged from performing these procedures. Awareness of the actual status of these procedures in the dental office as well as knowledge of the definition of *deep sedation* and *general anesthesia* may be of major concern.

Bleeding or Coagulation

Deaths attributed to the category *bleeding or coagulation* may have occurred because of underestimation of the bleeding risk. Although a coagulation disorder or the use of anticoagulants was known, surgical procedures had often been performed without consulting the patient's physician or without using adequate hemostyptic medical devices. In online publications, cases had also been reported in which patients had not returned

to the dental office despite recurrent bleeding and had then bled to death, for example, during sleep. It is quite possible that patients may underestimate a rebleeding episode and, therefore, avoid another consultation with the dentist. In particular, because of an increasing number of patients taking anticoagulant medication, the importance of an accurate survey of the patient's history and consultation with the patient's physician, as well as sufficient instructions for the patient after a surgical procedure, is highlighted [44]. Furthermore, an accurate assessment of a patient's bleeding risk enables the dentist to be prepared for intra- or postoperative bleeding (eg, with topical hemostatic agents or mechanical techniques) [45].

Other causes of death in this category are unknown diseases, for example, liver-related diseases, which may also influence the coagulation system. Unfortunately, these risks cannot be completely ruled out, even if the patient is carefully examined and the surgery performed properly.

Infection

In the PubMed literature, in particular, a considerable number of deaths that were categorized as *infection* and deaths because of endocarditis were described. While 7 cases were found in PubMed, only 1 case was reported in online publications, where a patient had died because of endocarditis. In 4 of these cases, it was stated that the treating dentist had not administered antibiotic prophylaxis before the treatment. In the other 3 cases, this information could not be clearly tracked from the publication. With an incidence of 3-7 cases per 100,000 people, infectious endocarditis is a rare but serious disease with a mortality rate of 20%-25% [46-50].

Dentists need to be aware of the indications for administering antibiotic prophylaxis. The recommendations to administer antibiotic prophylaxis to patients at risk before dental treatment have recently undergone certain modifications [51]. To prevent infectious endocarditis, dentists should consult the patient's physician or cardiologist when treatment of at-risk patients is planned.

Cardiovascular System

In both the online publications and the PubMed literature, deaths assigned to the *cardiovascular system* category were described. As reported earlier [4], in this study, cases were occasionally found in which a relationship to the anesthesia or medication used could not be completely ruled out. Nevertheless, as in other categories, a detailed survey of the patient's history is essential and, if necessary, a consultation with the patient's physician or cardiologist would be recommended. A fatal cardiovascular incident can occur anywhere; in a dental office, a distinction is made between cases that would have occurred independently from dental treatment and cases triggered by the treatment (eg, stress and local anesthesia) [43]. Air embolism is a very rare complication but has been described during drilling procedures for dental implant placement. Therefore, air-operated handpieces are obsolete for this procedure, and the equipment recommended by the manufacturer of the implants should be used [52].

Airway or Respiratory System

In online publications, about half of the deaths in the *airway or respiratory system* category had been caused by aspiration of foreign bodies (eg, tooth, cotton roll, swabs, and instruments) or stomach contents. However, swelling (eg, abscess and hematoma) with airway occlusion, as well as 1 case of suffocation having resulted from the patient's position during treatment, has been described as well. In the PubMed publications, mainly hereditary angioedema had been responsible for the described fatalities. As it can be triggered even by oral manipulation during dental treatment, a patient's lack of knowledge even after having been informed about the diagnosis years ago may have contributed to the occurrence. Since the onset of swelling may be delayed more than 24 hours after a dental procedure, the dentist may not even be involved in emergency management. Therefore, a survey of the patient's history should involve questions about these respective issues. In cases with a positive history, dental treatment should only be performed after consultation with the primary care physician [23,53].

Aspiration of foreign bodies may be prevented using protective membranes, such as rubber dams, especially in endodontic treatments [54]. However, this does not protect against aspiration of extracted teeth, cotton rolls, or gauze positioned in the throat in cases of general anesthesia and endotracheal intubation.

Anesthesia and Sedation

Overview

Anesthesia and sedation are considered priority topics in both this and other studies [4], therefore, they are discussed in more detail in this section. While the proportions of general anesthesia related to the respective deaths were almost equal in online publications (35/128, 27.3%) and in the PubMed literature (22/71, 31%), differences were noted in the use of sedation and local anesthetics. Under sedation, the percentage of deaths in online publications (46/128, 35.9%) was almost twice as high as those in the PubMed articles (14/71, 20%). Compared with general anesthesia, sedation of dental patients might be perceived as a harmless procedure.

The primary concern of the authors of online publications had probably not been to draw attention to the underestimation of risks by dental sedation. However, the additional knowledge extracted from the internet reports in this study may contribute significantly to the awareness of these risks in the dental community.

Local Anesthesia

In online publications, 10.2% (13/128) of the described deaths had been attributed to local anesthetic treatment compared with 20% (14/71) in the PubMed literature. It is unclear whether the deaths caused by this type of anesthesia are underreported on the internet.

The importance for including cases with supposedly less-harmful local anesthetic for the professional world is emphasized to draw attention to the risks and to define measures of care. It is noticeable that fatalities under local anesthesia in the PubMed articles covered mainly the group of children up to 5 years of age, where the dentists may have underestimated the risk of overdose.

Sedation and General Anesthesia

In online publications and in the PubMed literature, cases were described in which nitrous oxide had been used. In the specialist literature, however, no direct involvement of nitrous oxide in the death of the patients could be detected. In the fatalities described in online publications from the United States, nitrous oxide had been reported to be causative in 3 out of 10 cases (30%). A limitation could be the perception of the author of the respective internet post, who may have reported speculation as fact. In addition, a coroner's report had not been cited.

Nitrous oxide is still widely used in dentistry. In online publications, deaths involving nitrous oxide are reported exclusively from the United States. This may be due to the perception of maximum safety of nitrous oxide. In the last 30 years, the number of states where no further approval is needed for its administration has been decreasing [55].

Only 1 sedation case could be retrieved online from Germany and Austria. Younger fatalities, especially below 5 years of age, had occurred under general anesthesia.

As there were no respective fatality reports from online publications covering the United Kingdom, the German-speaking countries, as well as France, nitrous oxide involvement can only be analyzed from the US data. Nitrous oxide had contributed to the fatal course of treatment in only 10 cases. Because of the low number, no statement can be made about whether additional sedative or anesthetic drugs involved may have enhanced the medical problem leading to death. Only 1 fatality caused by foreign body aspiration in connection to nitrous oxide was published on the internet. However, as extracted from the PubMed database, the problem of aspiration and related issues, such as bronchospasm and laryngeal edema, seems to be more relevant during nitrous oxide use. This should increase the perception of the respective risk under adjunctive nitrous oxide or additional sedation during dental treatment under local anesthesia. It is expected that in most situations where nitrous oxide is used in dental treatment facilities the dentist has no additional anesthesiologist support. However, in the data obtained here, additional sedative medication had been used in 7 fatalities and in only 1 case an anesthesiologist was standing by.

In Germany and Austria, the predominant adjunctive measures had been general anesthesia performed by a certified anesthesiologist, whereas sedation procedures or even general anesthesia had been performed without a certified anesthesiologist in a significant number of cases in the United States. Sedation-related problems leading to the death of a patient seem to be more relevant in the United States compared with the other areas examined (see Table 8).

A significant association of deaths due to anesthesia or sedation with the presence or absence of an anesthesiologist has been demonstrated [4]. In more than 50% of fatalities reported in online publications covering the United States, where adjunctive general anesthesia or sedation had been used, it became obvious that no anesthesiologist had been on site during either general anesthesia or sedation. Our PubMed search found only 5 cases out of 36 (14%) where no anesthesiologist had been on site during adjunctive general anesthesia or sedation procedures. However, in 22 out of 36 cases (61%), no information about a specialist standing by had been provided. Nevertheless, this difference is remarkable, as the internet reports regarding this category almost exclusively cover the United States database. Even if an overlap between the internet and PubMed databases cannot fully be excluded, the high percentage in the internet reports may increase awareness among clinicians that the factor *anesthesiologist on site* is significantly influencing the course of respective incidents. This factor would have been underreported without the internet analysis.

It is also remarkable that in more than one-third of deaths reported in the PubMed literature and in almost two-thirds of deaths found in the internet publications covering the United States and the United Kingdom, an *anesthesia, medication, or sedation*-related problem was at least partly involved in the fatal outcome (see Table 5). Data may focus the awareness of dental practitioners—not only those in the United States—on anesthesiologist support, especially when sedative medication is used. Thereby, patient safety would be enhanced and preventable deaths of patients at risk may significantly be reduced.

PubMed data have not shown an increase of fatalities where sedation, as such, was involved when publications from the era before internet data availability were compared with those from the internet age beginning around 1995. Furthermore, when comparing the internet data after the beginning of 1995 (ie, after 1994) with PubMed data before 1995, no differences can be found regarding sedation-associated fatality rates in relation to fatality rates not associated with sedation. However, a much higher proportion of fatalities in the *anesthesia, medication, or sedation* category can be extracted from the internet reports. This category would have been underestimated if the internet reports had not been included in this study.

In 2001, a recommendation from the Department of Health in the United Kingdom was published and restricted general anesthesia for dental treatment to the hospital setting [56].

Limitations and Weaknesses of the Applied Method

First, a standardized evaluation is difficult since some of the online reports were written by laypersons without scientific backgrounds. The authors may have reported incorrect medical data or may have used incorrect terminology, possibly resulting in the assignment of a death to an inappropriate category. Second, the assignment of the cases from the internet and PubMed publications to the cause of death categories was difficult and was based on the authors' impressions. Some fatalities could have been allocated to more than one category. Third, terms that referred to deaths in dental offices were used, but there might be other terms that could be related to dental deaths that were not searched for when browsing the internet. Despite a precise search using Google, there might be cases that may not have been taken into account. Fourth, care was taken to exclude possible duplicate cases. By comparing fatalities before 1995 and after 1994, we have tried to fully separate the populations examined. Nevertheless, duplicates cannot be ruled out completely. Fifth, the search was restricted to six countries, which might limit the generalizability. Further research might

be needed to identify differences or overlaps in dental deaths in other countries.

The Internet as a Data Source Evaluating Casualties in Dental Facilities

"Death in the dental office" may be a phrase that comes into the view of a regular online visitor, for example, on a news page. However, it is not a term of interest for the general public. The number of internet searchers was so low that Google Trends database analysis has not created statistics since 2004. However, dental clinicians or researchers may benefit to a large extent from the data obtained from such searches.

The implication of scientific publications indexed by PubMed may be significantly enhanced by adding data obtained from Google database searches, such as the topic in this paper.

Conclusions

In this study, deaths having occurred in dental offices published on the internet and in PubMed were examined. It is advisable to be aware of the risks, no matter how low; the risk factors; and the diseases associated with dental offices in order to prevent possibly life-threatening situations in relation to dental treatment. The additional searches on the internet revealed relevant data regarding fatalities, which would have remained hidden if only extracting data from PubMed. Fatalities associated with general anesthesia or sedation would have been underrepresented, as well as the number of minors having died related to dental treatments, without the analysis of internet-based reports.

Regarding anesthesia, especially when used with children and even with local anesthetics, the risk of overdose should not be underestimated. The type and dose of medication should be of major concern. The definitions and understanding of deep sedation and general anesthesia should be a focus in the education of dentists willing to perform adjunctive sedation during treatment, in order to be aware of associated life-threatening problems. By avoiding potentially hazardous procedures, such as sedation-aided treatments performed exclusively by the dentist without the assistance of an experienced physician or anesthesiologist, the risk of treatment-induced life-threatening emergencies may be reduced.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Searches for the words "death" and "dental" in the United States using Google Trends.
[PNG File , 220 KB - jmir_v22i4e15304_app1.png]

Multimedia Appendix 2
Number of deaths due to *anesthesia, medication, or sedation* by age group using data from online publications in Germany and Austria.
[PNG File , 41 KB - jmir_v22i4e15304_app2.png]

Multimedia Appendix 3
Number of deaths due to *anesthesia, medication, or sedation* in each age group using data from online publications in the United Kingdom.
[PNG File , 53 KB - jmir_v22i4e15304_app3.png]

Multimedia Appendix 4
Number of deaths due to *anesthesia, medication, or sedation* in each age group using data from online publications in the United States.
[PNG File , 81 KB - jmir_v22i4e15304_app4.png]

Multimedia Appendix 5
Number of deaths in the *anesthesia, medication, or sedation* category due to general anesthesia, sedation, or local anesthesia in each age group from the PubMed literature.
[PNG File , 60 KB - jmir_v22i4e15304_app5.png]

Multimedia Appendix 6
Percentage of casualties in dental facilities for each patient age group, based on data from online publications and PubMed. Possible duplicates were not excluded.
[PNG File , 96 KB - jmir_v22i4e15304_app6.png]

Multimedia Appendix 7

Percentage of anesthesia types used in cases of casualties in dental facilities, based on data from online publications and PubMed. [PNG File , 68 KB - jmir_v22i4e15304_app7.png]

References

1. Atherton GJ, McCaul JA, Williams SA. Medical emergencies in general dental practice in Great Britain. Part 1: Their prevalence over a 10-year period. Br Dent J 1999 Jan 23;186(2):72-79 [FREE Full text] [doi: 10.1038/sj.bdj.4800023] [Medline: 10079576]
2. Helm M, de Buhr K, Lampl L, Schramm A. Notfallmedizinische ausbildungskonzepte für studierende der zahnmedizin. Notfall Rettungsmed 2010 Oct 30;14(5):389-395. [doi: 10.1007/s10049-010-1382-z]
3. Müller MP, Hänsel M, Stehr SN, Weber S, Koch T. A state-wide survey of medical emergency management in dental practices: Incidence of emergencies and training experience. Emerg Med J 2008 May;25(5):296-300. [doi: 10.1136/emj.2007.052936] [Medline: 18434471]
4. Reuter NG, Westgate PM, Ingram M, Miller CS. Death related to dental treatment: A systematic review. Oral Surg Oral Med Oral Pathol Oral Radiol 2017 Feb;123(2):194-204.e10. [doi: 10.1016/j.oooo.2016.10.015] [Medline: 27989710]
5. McLennan S, Strech D, Reimann S. Developments in the frequency of ratings and evaluation tendencies: A review of German physician rating websites. J Med Internet Res 2017 Aug 25;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]
6. D'eramo EM, Bookless SJ, Howard JB. Adverse events with outpatient anesthesia in Massachusetts. J Oral Maxillofac Surg 2003 Jul;61(7):793-800; discussion 800. [doi: 10.1016/s0278-2391(03)00238-6] [Medline: 12856252]
7. D'Eramo EM. Morbidity and mortality with outpatient anesthesia: The Massachusetts experience. J Oral Maxillofac Surg 1992 Jul;50(7):700-704. [doi: 10.1016/0278-2391(92)90101-5] [Medline: 1607998]
8. D'Eramo EM, Bontempi WJ, Howard JB. Anesthesia morbidity and mortality experience among Massachusetts oral and maxillofacial surgeons. J Oral Maxillofac Surg 2008 Dec;66(12):2421-2433. [doi: 10.1016/j.joms.2008.06.095] [Medline: 19022119]
9. Davies JM, Campbell LA. Fatal air embolism during dental implant surgery: A report of three cases. Can J Anaesth 1990 Jan;37(1):112-121. [doi: 10.1007/BF03007491] [Medline: 2295094]
10. Forrest A, Milne N, Soon A. Hereditary angioedema: Death after a dental extraction. Aust Dent J 2017 Mar;62(1):107-110 [FREE Full text] [doi: 10.1111/adj.12447] [Medline: 27491932]
11. Nkansah PJ, Haas DA, Saso MA. Mortality incidence in outpatient anesthesia for dentistry in Ontario. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1997 Jun;83(6):646-651. [doi: 10.1016/s1079-2104(97)90312-7] [Medline: 9195616]
12. Sterzik V, Tatschner T, Roewer N, Barrera D, Bohnert M. Fatal visit to the dentist. Int J Legal Med 2015 Jan;129(1):219-222. [doi: 10.1007/s00414-013-0930-8] [Medline: 24202697]
13. Gangemi S, Spagnolo EV, Cardia G, Minciullo PL. Fatal anaphylactic shock due to a dental impression material. Int J Prosthodont 2009;22(1):33-34. [Medline: 19260424]
14. Chicka MC, Dembo JB, Mathu-Muju KR, Nash DA, Bush HM. Adverse events during pediatric dental anesthesia and sedation: A review of closed malpractice insurance claims. Pediatr Dent 2012;34(3):231-238. [Medline: 22795157]
15. Smith MM, Barbara DW, Mauermann WJ, Viozzi CF, Dearani JA, Grim KJ. Morbidity and mortality associated with dental extraction before cardiac operation. Ann Thorac Surg 2014 Mar;97(3):838-844. [doi: 10.1016/j.athoracsur.2013.10.034] [Medline: 24360092]
16. Flick WG, Katsnelson A, Alstrom H. Illinois dental anesthesia and sedation survey for 2006. Anesth Prog 2007;54(2):52-58 [FREE Full text] [doi: 10.2344/0003-3006(2007)54[52:IDAASS]2.0.CO;2] [Medline: 17579504]
17. Holmes SM, Udey DK. What are the lessons we can glean from a review of recent closed malpractice cases involving oral and maxillofacial infections? Oral Maxillofac Surg Clin North Am 2011 Nov;23(4):601-607, vii. [doi: 10.1016/j.coms.2011.07.010] [Medline: 21982612]
18. Okada Y, Suzuki H, Ishiyama I. Fatal subarachnoid haemorrhage associated with dental local anaesthesia. Aust Dent J 1989 Aug;34(4):323-325. [doi: 10.1111/j.1834-7819.1989.tb04640.x] [Medline: 2775019]
19. Lytle JJ, Stamper EP. The 1988 anesthesia survey of the Southern California Society of Oral and Maxillofacial Surgeons. J Oral Maxillofac Surg 1989 Aug;47(8):834-842. [doi: 10.1016/s0278-2391(89)80043-6] [Medline: 2746393]
20. Marshall DA, Berry C, Brewer A. Fatal disseminated intravascular coagulation complicating dental extraction. Br J Oral Maxillofac Surg 1993 Jun;31(3):178-179. [doi: 10.1016/0266-4356(93)90120-l] [Medline: 8512913]
21. Coplans MP, Curson I. Deaths associated with dentistry and dental disease 1980-1989. Anaesthesia 1993 May;48(5):435-438 [FREE Full text] [doi: 10.1111/j.1365-2044.1993.tb07022.x] [Medline: 8317656]
22. McKechnie J. Prostatic carcinoma presenting as a haemorrhagic diathesis after dental extraction. Br Dent J 1989 Apr 22;166(8):295-296. [doi: 10.1038/sj.bdj.4806813] [Medline: 2785811]
23. Bork K, Barnstedt S. Laryngeal edema and death from asphyxiation after tooth extraction in four patients with hereditary angioedema. J Am Dent Assoc 2003 Aug;134(8):1088-1094. [doi: 10.14219/jada.archive.2003.0323] [Medline: 12956349]
24. Ely EW, Stump TE, Hudspeth AS, Haponik EF. Thoracic complications of dental surgical procedures: Hazards of the dental drill. Am J Med 1993 Nov;95(5):456-465. [doi: 10.1016/0002-9343(93)90327-l] [Medline: 8238061]

25. Fogarty C, Regennitter F, Viozzi CF. Invasive fungal infection of the maxilla following dental extractions in a patient with chronic obstructive pulmonary disease. J Can Dent Assoc 2006 Mar;72(2):149-152 [FREE Full text] [Medline: 16545177]
26. Kim J, Fortson JK, Cook HE. A fatal outcome from rhinocerebral mucormycosis after dental extractions: A case report. J Oral Maxillofac Surg 2001 Jun;59(6):693-697. [doi: 10.1053/joms.2001.23407] [Medline: 11381399]
27. Kunkel M, Kleis W, Morbach T, Wagner W. Severe third molar complications including death-lessons from 100 cases requiring hospitalization. J Oral Maxillofac Surg 2007 Sep;65(9):1700-1706. [doi: 10.1016/j.joms.2007.04.014] [Medline: 17719386]
28. Levine M. Understanding how a dental infection may spread to the brain: Case report. J Can Dent Assoc 2013;79:d9 [FREE Full text] [Medline: 23522149]
29. Lopez R, Flavell S, Thomas C. A not very NICE case of endocarditis. BMJ Case Rep 2013 Jan 29;2013:1-2 [FREE Full text] [doi: 10.1136/bcr-2012-007918] [Medline: 23365162]
30. Moore C, Addison D, Wilson JM, Zeluff B. First case of Fusobacterium necrophorum endocarditis to have presented after the 2nd decade of life. Tex Heart Inst J 2013;40(4):449-452 [FREE Full text] [Medline: 24082377]
31. Ricci ML, Fontana S, Pinci F, Fiumana E, Pedna MF, Farolfi P, et al. Pneumonia associated with a dental unit waterline. Lancet 2012 Feb 18;379(9816):684. [doi: 10.1016/S0140-6736(12)60074-9] [Medline: 22340301]
32. Schmitt M, Puri S, Dalal NR. Aortic valve endocarditis causing fatal myocardial infarction caused by ostial coronary artery obliteration. Heart 2004 Mar;90(3):303 [FREE Full text] [doi: 10.1136/hrt.2003.026534] [Medline: 14966053]
33. Wochna K, Jurczyk AP, Krajewski W, Berent J. Sudden death due to malignant hyperthermia during general anesthesia. Arch Med Sadowej Kryminol 2013;63(1):11-14, 17. [Medline: 23879013]
34. Hersh EV, Helpin ML, Evans OB. Local anesthetic mortality: Report of case. ASDC J Dent Child 1991;58(6):489-491. [Medline: 1783701]
35. No authors listed. Katie Dougal. A legacy for improvement. SAAD Dig 1998 Apr;15(2):5-7. [Medline: 11833436]
36. Auroy Y, Benhamou D, Péquignot F, Jougla E, Lienhart A. Hyponatraemia-related death after paediatric surgery still exists in France. Br J Anaesth 2008 Nov;101(5):741 [FREE Full text] [doi: 10.1093/bja/aen282] [Medline: 18854387]
37. Kupiec TC, Kemp P, Raj V, Kemp J. A fatality due to an accidental methadone substitution in a dental cocktail. J Anal Toxicol 2011 Sep;35(7):512-515. [doi: 10.1093/anatox/35.7.512] [Medline: 21871161]
38. Dhoble A, Vedre A, Abdelmoneim SS, Sudini SR, Ghose A, Abela GS, et al. Prophylaxis to prevent infective endocarditis: To use or not to use? Clin Cardiol 2009 Aug;32(8):429-433 [FREE Full text] [doi: 10.1002/clc.20583] [Medline: 19685514]
39. Malamed SF. Morbidity, mortality and local anaesthesia. Prim Dent Care 1999 Jan;6(1):11-15. [Medline: 10752458]
40. Robinson EM. Death in the dental chair. J Forensic Sci 1989 Mar;34(2):377-380. [Medline: 2651555]
41. Schulman NJ, Owens B. Medical complications following successful pediatric dental treatment. J Clin Pediatr Dent 1996;20(4):273-275. [Medline: 9151618]
42. Lee H, Milgrom P, Huebner CE, Weinstein P, Burke W, Blacksher E, et al. Ethics rounds: Death after pediatric dental anesthesia: An avoidable tragedy? Pediatrics 2017 Dec;140(6):1-6 [FREE Full text] [doi: 10.1542/peds.2017-2370] [Medline: 29114060]
43. Kaeppler G, Daubländer M, Hinkelbein R, Lipp M. Quality of cardiopulmonary resuscitation by dentists in dental emergency care [Article in German]. Mund Kiefer Gesichtschir 1998 Mar;2(2):71-77. [doi: 10.1007/s100060050033] [Medline: 9567061]
44. Schmelzeisen R. Zahnärztliche Mitteilungen. 2002. Zahnärztliche chirurgie bei patienten mit antikoagulantientherapie URL: https://www.zm-online.de/archiv/2002/07/zahnmedizin/zahnaerztliche-chirurgie-bei-patienten-mit-antikoagulantientherapie/ [accessed 2020-02-09]
45. Kamoh A, Swantek J. Hemostasis in oral surgery. Dent Clin North Am 2012 Jan;56(1):17-23, vii. [doi: 10.1016/j.cden.2011.06.004] [Medline: 22117940]
46. Hoen B, Alla F, Selton-Suty C, Béguinot I, Bouvet A, Briançon S, Association pour l'Etude et la Prévention de l'Endocardite Infectieuse (AEPEI) Study Group. Changing profile of infective endocarditis: Results of a 1-year survey in France. JAMA 2002 Jul 03;288(1):75-81. [doi: 10.1001/jama.288.1.75] [Medline: 12090865]
47. Selton-Suty C, Célard M, Le Moing V, Doco-Lecompte T, Chirouze C, Iung B, AEPEI Study Group. Preeminence of Staphylococcus aureus in infective endocarditis: A 1-year population-based survey. Clin Infect Dis 2012 May;54(9):1230-1239. [doi: 10.1093/cid/cis199] [Medline: 22492317]
48. Duval X, Delahaye F, Alla F, Tattevin P, Obadia J, Le Moing V, AEPEI Study Group. Temporal trends in infective endocarditis in the context of prophylaxis guideline modifications: Three successive population-based surveys. J Am Coll Cardiol 2012 May 29;59(22):1968-1976 [FREE Full text] [doi: 10.1016/j.jacc.2012.02.029] [Medline: 22624837]
49. Tleyjeh IM, Steckelberg JM, Murad HS, Anavekar NS, Ghomrawi HM, Mirzoyev Z, et al. Temporal trends in infective endocarditis: A population-based study in Olmsted County, Minnesota. JAMA 2005 Jun 22;293(24):3022-3028. [doi: 10.1001/jama.293.24.3022] [Medline: 15972564]
50. Mylonakis E, Calderwood SB. Infective endocarditis in adults. N Engl J Med 2001 Nov 01;345(18):1318-1330. [doi: 10.1056/NEJMra010082] [Medline: 11794152]
51. Habib G, Lancellotti P, Antunes MJ, Bongiorni MG, Casalta J, Del Zotti F, ESC Scientific Document Group. 2015 ESC Guidelines for the management of infective endocarditis: The Task Force for the Management of Infective Endocarditis of

the European Society of Cardiology (ESC). Endorsed by: European Association for Cardio-Thoracic Surgery (EACTS), the European Association of Nuclear Medicine (EANM). Eur Heart J 2015 Nov 21;36(44):3075-3128. [doi: 10.1093/eurheartj/ehv319] [Medline: 26320109]
52. Girdler NM. Fatal sequel to dental implant surgery. J Oral Rehabil 1994 Nov;21(6):721-722. [doi: 10.1111/j.1365-2842.1994.tb01188.x] [Medline: 7830208]
53. Tse K, Zuraw BL. Recognizing and managing hereditary angioedema. Cleve Clin J Med 2013 May;80(5):297-308 [FREE Full text] [doi: 10.3949/ccjm.80a.12073] [Medline: 23636922]
54. Cohen S, Schwartz S. Endodontic complications and the law. J Endod 1987 Apr;13(4):191-197. [doi: 10.1016/S0099-2399(87)80139-5] [Medline: 3471843]
55. Boynes SG, editor. Dental Anesthesiology: A Guide to the Rules and Regulations of the United States of America. 6th edition. Pittsburgh, PA: No-No Orchard Publishing; 2018.
56. Pike D. A conscious decision. A review of the use of general anaesthesia and conscious sedation in primary dental care. SAAD Dig 2000 Jul;17(3):13-14. [Medline: 11404865]

Edited by G Eysenbach; submitted 30.06.19; peer-reviewed by M Kurrer, A Lau; comments to author 29.09.19; revised version received 29.12.19; accepted 03.02.20; published 17.04.20.

Please cite as:
Gaiser M, Kirsch J, Mutzbauer TS
Using Nonexpert Online Reports to Enhance Expert Knowledge About Causes of Death in Dental Offices Reported in Scientific Publications: Qualitative and Quantitative Content Analysis and Search Engine Analysis
J Med Internet Res 2020;22(4):e15304
URL: http://www.jmir.org/2020/4/e15304/
doi:10.2196/15304
PMID:32038029

Original Paper

Longitudinal Study of the Variation in Patient Turnover and Patient-to-Nurse Ratio: Descriptive Analysis of a Swiss University Hospital

Sarah N Musy[1,2], MSc; Olga Endrich[3,4], MD; Alexander B Leichtle[4,5], MD; Peter Griffiths[6,7,8], PhD; Christos T Nakas[5,9], PhD; Michael Simon[1,2], PhD

[1]Institute of Nursing Science, University of Basel, Basel, Switzerland

[2]Nursing and Midwifery Research Unit, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[3]Medical Directorate, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[4]Insel Data Science Center, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[5]University Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[6]Health Sciences, University of Southampton, Southampton, United Kingdom

[7]National Institute for Health Research Applied Research Collaboration (Wessex), Southampton, United Kingdom

[8]LIME Karolinska Institutet, Stockholm, Sweden

[9]Laboratory of Biometry, University of Thessaly, Volos, Greece

Corresponding Author:
Michael Simon, PhD
Institute of Nursing Science
University of Basel
Bernoullistrasse 28
Basel, 4057
Switzerland
Phone: 41 61 267 09 12
Email: m.simon@unibas.ch

Abstract

Background: Variations in patient demand increase the challenge of balancing high-quality nursing skill mixes against budgetary constraints. Developing staffing guidelines that allow high-quality care at minimal cost requires first exploring the dynamic changes in nursing workload over the course of a day.

Objective: Accordingly, this longitudinal study analyzed nursing care supply and demand in 30-minute increments over a period of 3 years. We assessed 5 care factors: patient count (care demand), nurse count (care supply), the patient-to-nurse ratio for each nurse group, extreme supply-demand mismatches, and patient turnover (ie, number of admissions, discharges, and transfers).

Methods: Our retrospective analysis of data from the Inselspital University Hospital Bern, Switzerland included all inpatients and nurses working in their units from January 1, 2015 to December 31, 2017. Two data sources were used. The nurse staffing system (tacs) provided information about nurses and all the care they provided to patients, their working time, and admission, discharge, and transfer dates and times. The medical discharge data included patient demographics, further admission and discharge details, and diagnoses. Based on several identifiers, these two data sources were linked.

Results: Our final dataset included more than 58 million data points for 128,484 patients and 4633 nurses across 70 units. Compared with patient turnover, fluctuations in the number of nurses were less pronounced. The differences mainly coincided with shifts (night, morning, evening). While the percentage of shifts with extreme staffing fluctuations ranged from fewer than 3% (mornings) to 30% (evenings and nights), the percentage within “normal” ranges ranged from fewer than 50% to more than 80%. Patient turnover occurred throughout the measurement period but was lowest at night.

Conclusions: Based on measurements of patient-to-nurse ratio and patient turnover at 30-minute intervals, our findings indicate that the patient count, which varies considerably throughout the day, is the key driver of changes in the patient-to-nurse ratio. This demand-side variability challenges the supply-side mandate to provide safe and reliable care. Detecting and describing patterns in variability such as these are key to appropriate staffing planning. This descriptive analysis was a first step towards identifying time-related variables to be considered for a predictive nurse staffing model.

(J Med Internet Res 2020;22(4):e15554) doi:10.2196/15554

KEYWORDS

patient safety; electronic health records; nurse staffing; workload; routine data

Introduction

Determining appropriate nursing staff numbers and skill mixes in hospital units is vital to both ensure quality of care [1-4] and maintain health care budgets [5]. Understaffing or poor skill mixes can lead to adverse patient outcomes, while overstaffing can lead to budgetary overruns and ultimately close hospitals. In economic terms, the relationship between patients and nurses is one of supply and demand, respectively, representing the amount of care required by the patients versus the nursing staff's capacity to provide that care.

Undoubtedly, care demands and staffing requirements vary widely across departments, units, seasons, months, days of the week, shifts (morning, evening, or night), and even hours [6-8]. Each unit's patient count fluctuates with patient turnover resulting from admissions, discharges, and transfers within and between units [9-12]. In turn, turnover affects the volume of nurses' clinical and administrative duties [13,14].

Another notable factor is patient acuity, or the amount of time each patient requires. Newly admitted or transferred patients tend to have high levels of acuity because they require baseline assessments and treatments [15,16]. Specific patient characteristics, including demographics (eg, age, gender, family support, socioeconomic factors), personal background, diagnoses, and treatment regimes, can also increase acuity. For example, patients who are older [17-19], lack family support [18], or have limited knowledge of their health condition(s) [20] have more complex care needs, with strong implications for nurse staffing. Depending on each shift's patient numbers and combined acuity, nursing workloads can vary across and within wards. Meeting their specific needs requires an appropriate number and skill mix of nurses: registered nurses (RNs), licensed practical nurses (LPNs), and unlicensed personnel. Some studies have used nursing staff with or without qualification in their analyses [21,22]. In Switzerland, RNs typically represent the major proportion of hospital nursing staff; however, direct and indirect patient care typically involves collaborations between nurses with a broad range of qualifications.

To date, the majority of research about nurse staffing has fit into two categories: longitudinal studies conducted over relatively long periods (eg, years) or across locations and cross-sectional studies often conducted across multiple locations. However, in both cases, large-scale views fail to capture shift-level or daily variations in either supply or demand. This lack of detail limits the understanding of the association between staffing, patient turnover, and relevant human and economic outcomes [7]. Noting that these limitations severely limit the value of research findings for staffing guidelines [22], studies have begun to highlight the advantages of both focusing on unit-level dynamics and using hospital record data longitudinally rather than cross-sectionally [21,23,24].

Every nursing unit manager's job includes assuring patient safety and quality of care every hour of the day. However, due to the noted limitations and considering the principle that nurse-patient relationships occur on the individual level [25], previous studies have offered a limited view of the small-scale supply and demand dynamic of nursing workload. Nurse staffing planners are particularly challenged by demand-side variability, which occurs over very short periods. Therefore, to both optimize staffing levels (ie, maintain levels that will allow safe patient care while minimizing personnel costs) and develop reliable staffing guidelines, it is necessary to record and explore fluctuations in nursing workload throughout the day rather than simply considering daily or shift averages.

Therefore, this study's overall aims were to describe the supply and demand dynamics of nursing care and identify mismatches between supply and demand from a longitudinal perspective. Specifically, in 30-minute increments, across a range of units in a Swiss University Hospital (Inselspital, Bern), we describe every recorded change in patient numbers (ie, care demand); nurse numbers (ie, supply); patient-to-nurse ratios for the various nurse groups; extreme supply and demand mismatches; and patient turnover (ie, numbers of admissions, discharges, and transfers).

Methods

Study Design and Setting

This retrospective, descriptive, observational study used routinely collected patient data from the Inselspital University Hospital, Bern, Switzerland. As one of Switzerland's five University Hospitals, the Inselspital treats approximately 48,000 inpatients annually [26]. Only inpatient units with data for the full 3 years were included. Our data were drawn from 10 departments: (1) Internal Medicine; (2) Cardiology & Cardiovascular Surgery; (3) Orthopedics & Plastic Surgery; (4) Neurology, Neurosurgery, Otolaryngology, Head and Neck Surgery, & Ophthalmology; (5) Visceral Surgery and Medicine, Gastroenterology, Thoracic Surgery, & Pulmonology; (6) Dermatology, Urology, Rheumatology, & Nephrology; (7) Hematology & Oncology; (8) Maternity & Gynecology; (9) Pediatrics; and (10) Intensive Care. Full data (2015–2017) were available for all these departments for the full study period.

Participants

Patients

All inpatients were included. No further specific eligibility criteria were applied.

Nurses

All staff providing direct or indirect nursing care were considered in the analysis, independent of educational background. We divided nursing staff into five groups: RNs, including nurses in supervisory positions (group 1); LPNs (group

2); others, including unlicensed and administrative personnel (group 3); students (group 4); and external nurses (agency staff; group 5). In Switzerland, RNs complete a 3-4–year tertiary professional or university-based education (group 1). Unlike in other countries, Switzerland also offers 3 years of secondary-level professional training for nursing assistants (group 2). Group 3 included unlicensed personnel, including nursing aides with minimal education or training, and administrative staff. Group 4 included both nursing and medical students working as nursing aides.

Data Sources and Variables

We extracted our data from two sources: the tacs nurse staffing system (ie, datasheets organized in a relational database) and medical discharges. From tacs, we extracted four care-relevant factors: nurses, patients, activities, and care-related working hours. The tacs system records time allocations provided by every nurse at the end of every shift. In addition to administrative work, teaching duties, and continuous education, each record specifies the time devoted to each patient's direct and indirect care. No further details about the type of activity such as medication, mobility, or respiratory therapy are currently available. Nurses' absences such as holidays, illnesses, or accidents are also recorded. Patient turnover information is provided with the nursing unit, date, and time, as well as whether inpatient or outpatient services were provided. Finally, medical discharge data include patient demographics, admission and discharge details, and diagnoses. Each data record identifies the relevant unit and includes identifiers for the nurse (and her or his contract) and each case (patient) cared for during that shift. Based on these identifiers, the 5 datasheets were linked at the patient, nurse, and unit levels in a single dataset. Then, all patient, nurse, and unit identifiers were deidentified, leaving only department names. Table 1 describes the 17 variables used in the analysis.

Table 1. Description of the 17 variables used for the current analysis, listed in alphabetical order.

Variable and short description	Source[a]
Type of activity performed by the nurse	
Indirect and direct care	A[b]
Administrative work	A
Teaching assignments	A
Continuous education	A
Absences (ie, holidays, illnesses, accidents)	A
Admission date	
Patient's hospital admission date	P[c]
Admission time	
Patient's hospital admission time	P
Age	
Patient's age at admission	M[d]
Case identifier	
Unique code for the patient's case (deidentified to "Patient1", "Patient2", etc)	A, P, M
Departments of the Inselspital, Bern University Hospital	
Cardiology & Cardiovascular Surgery	A, M
Neurology, Neurosurgery, Otolaryngology, Head and Neck Surgery, & Ophthalmology	A, M
Intensive Care	A, M
Pediatrics	A, M
Dermatology, Urology, Rheumatology, & Nephrology	A, M
Visceral Surgery and Medicine, Gastroenterology, Thoracic Surgery & Pulmonology	A, M
Internal Medicine	A, M
Maternity & Gynecology	A, M
Orthopedics & Plastic Surgery	A, M
Hematology & Oncology	A, M
Contract identifier	
Unique code for each nursing position/contract (a nurse can have multiple contracts within the hospital involving various qualifications or working units), which was deleted after merging	A, N[e], W[f], M
Date	
Working date of the nurse	A, W
Discharge date	
Patient's hospital discharge date	P, M
Discharge time	
Patient's hospital discharge time	P
End time	
Time at which the nurse stopped work for the shift or started a break	W
Group (classifications of nurse qualifications)	

Variable and short description	Source[a]
Registered nurses	N
Licensed practical nurses	N
Others (eg, unlicensed and administrative personnel)	N
Students	N
External nurses	N
Main diagnosis	
ICD-10-GM[g] codes for the patient's main diagnosis	M
Nurse identifier	
Unique code for a nurse (deidentified to "Nurse1", "Nurse2", etc)	A, N, W
Start time	
Time at which the nurse began work or returned from a break	W
Transfer date	
Transfer date of the patient within and between departments	P
Transfer time	
Transfer time of the patient within and between departments	P
Unit identifier	
Unique code for the unit (deidentified to "Unit1", "Unit2", etc within each department)	A, N, P

[a]Source: nurse staffing system (tacs) or medical discharge data.

[b]A: nurse staffing system activity data.

[c]P: nurse staffing system patient data.

[d]M: medical discharge data.

[e]N: nurse staffing system nurse data.

[f]W: nurse staffing system working hours data.

[g]ICD-10-GM: 10th revision of the International Statistical Classification of Diseases and Related Health Problems, German Modification.

Ethical Considerations

Our acquisition of data from the Inselspital (University Hospital of Bern) was outside the purview of the Cantonal Ethic Commission of Bern based on the Swiss legislation on research with humans (Req-2016-00618). All data involving patients, nurses, and units were deidentified.

Statistical Analyses

All statistical analyses were conducted using R, version 3.5.1 for Mac OS and Linux [27]. To handle and manipulate the data, we used the purrr [28], dplyr [29], tidyr [30], and data.table [31] packages. To manipulate time and date, we used the lubridate [32], chron [33], and padr [34] packages. To create plots, we used the ggplot2 [35] and scales [36] packages.

Linking Procedure

Data were merged based on 6 key variables: patient identifier, nurse identifier, contract identifier, unit identifier, time, and date. First, a subset of activity data was created for data on inpatient units and direct and indirect care. This subset was then divided into nurse and patient activity fields, and any duplicate records were deleted. Each nurse's activities were merged first with her or his other data, then with the data she or he supplied regarding time use, contract identifier, and date. Contract identifiers were deleted, and nurses and units were deidentified. Similarly, each patient's activity data were merged first with their other data, then with medical controlling data by case identifier. To maintain consistent patient counts, healthy newborn babies in the Maternity & Gynecology department were excluded. Patients and units were then deidentified. Finally, the merged nurse and patient data were expanded to allow assessment of the number of nurses, patients, admissions, discharges, and transfers in 30-minute increments.

Descriptive Overview

For each department, the total numbers of units and patients were recorded. Mean (SD) and median (interquartile range [IQR]) were calculated for patient age. Length of stay (LOS) in days was computed by subtracting the discharge date from the admission date. Median (IQR) was calculated for LOS and for the number of patients per day per unit. Finally, for each department, we identified the two most common diagnoses by incidence (%).

Number of Patients

To keep computational complexity to a reasonable limit, patient numbers (ie, demand) were calculated at 30-minute intervals. Alternative increments (20, 40, and 60 minutes) had no relevant effect on the results. However, as patients rarely arrive or leave at shift divisions and some do not stay on the unit for one full shift, a short interval length ensures precise patient numbers.

Unit-level calculations correspond to every 30-minute interval (ie, 48 data points per day) totaling 153,792 points per unit during the study period.

Number of Nurses

As with the number of patients, the number of nurses in each group (ie, supply) was calculated for each 30-minute increment. As nurses may work only half shifts or overlapping shifts, this increment length ensured precise numbers. For each of the 48 daily data points for the 3 years covered by the study, the numbers of RNs, LPNs, and other staff were calculated. Unfortunately, as external nurses and students are not classified as typical employees, our datasets included no breakdown of their time allocation. As only daily information was available for these groups, their total numbers over the 3-year study period as well as their daily means and medians were calculated to provide an image of their effects across each unit.

Patient-to-Nurse Ratio

The patient-to-nurse ratio was computed by dividing the number of patients by the number of nurses at every data point. Along with numbers of patients and nurses, patient-to-nurse ratios were plotted separately for each day's 48 data points and for each day of the week using each unit's and each department's mean with CIs.

Extreme Mismatch Between Supply and Demand

Additionally, for three key time points of each day, namely at 2 am, 10 am, and 6 pm, the unit-level median (IQR) patient-to-nurse ratio was calculated for weekdays and weekends and divided by department. Further, we calculated when 50% more or less work was required per nurse for every data point and unit based on the median patient-to-nurse ratio. Two variables were created:

extreme lower threshold = median - (median/2)

extreme higher threshold = median + (median/2)

These arbitrary cut-offs were set to illustrate extreme staffing situations. Extreme staffing situations are important to identify times where supply and demand do not match (eg, the demand is too high for the given supply or vice versa). This is an indicator of whether supply and demand are staying within the "normal" range throughout the year. For the 3-year study period, the percentages of data points falling far below or far above the thresholds were calculated. Medians (IQRs) and extreme higher and lower thresholds (% of data points) were plotted with bar charts to highlight variations in patient-to-nurse ratios. Graphics and calculations were computed separately for nurse groups 1 (RNs), 2 (LPNs), and 3 (others).

Patient Turnover

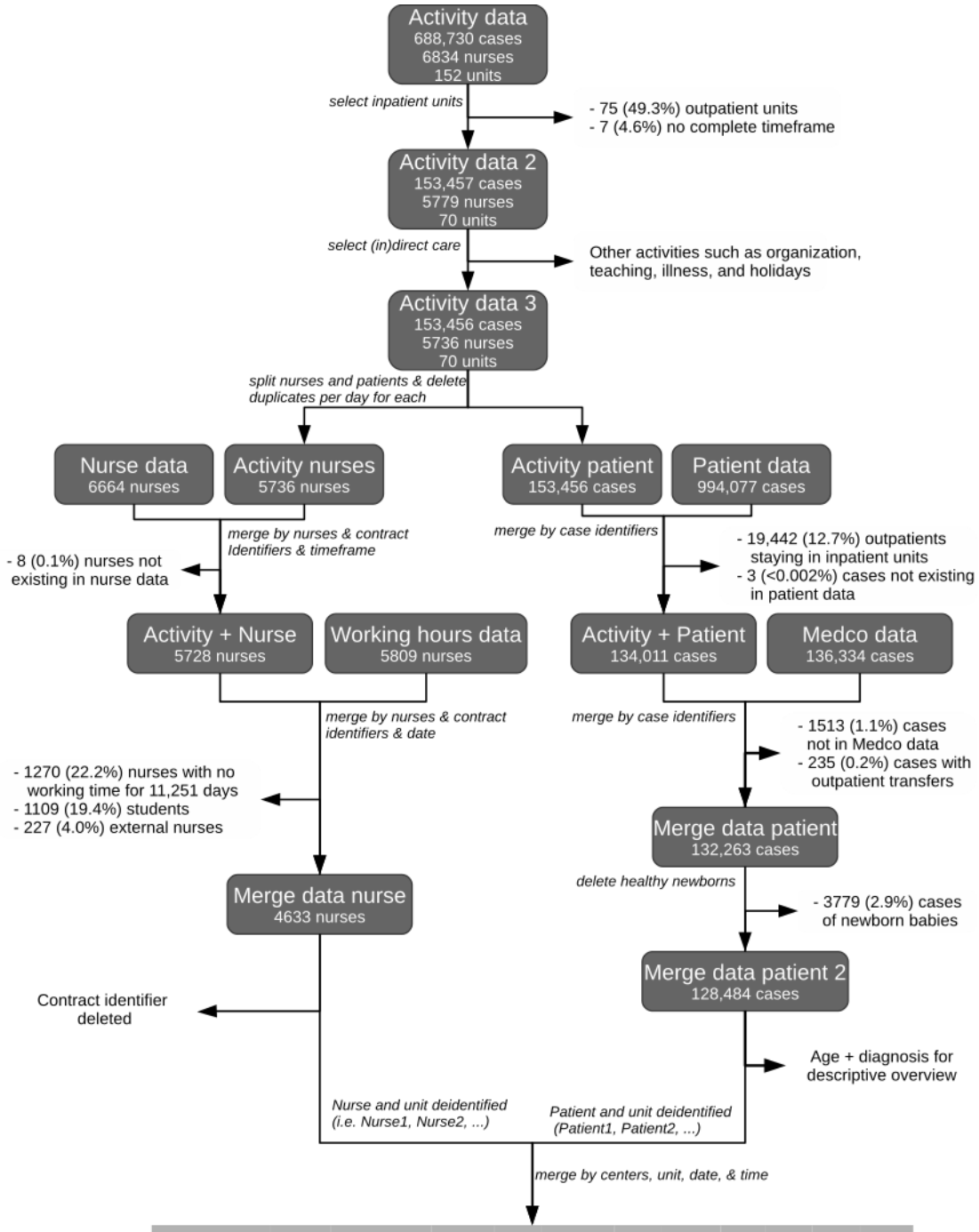
For every unit, the numbers of admissions, discharges, and transfers were computed for every 30-minute data point during the 3-year study period. Admission corresponds to any entry of a patient to the hospital and discharge to any exit from the hospital. Transfers, corresponding to movement of admitted patients from one unit to another, were divided into "Transfers in" and "Transfers out" of each relevant unit. As the units were of different sizes, the numbers of admissions, discharges, and transfers in and out were divided by the number of patients present at each specific data point. This allowed us to obtain a ratio for patient turnover that could be plotted on the same scale for all the units. Finally, vertical bar charts were created for weekdays and weekends by calculating the mean of the units for the departments. The left side of the vertical bar charts represents patients leaving the unit (ie, discharges and transfers out), while the bars on the right side represent incoming patients (ie, admissions and transfers in).

Results

Linking Procedure

The main activity data were drawn from 688,730 cases and 6834 nurses in 152 units. After the exclusion of outpatient units and noncare activities (whether direct or indirect), the activity data reflected 153,456 cases (153,456/688,730, 22.2%) and 5736 nurses (5736/6834, 83.9%) in 70 units (70/152, 46.1%). Of the remaining 5736 nurses, data from 4633 (80.8%) were usable for the final analyses. A number of nurses (1270) were excluded for specific days only, as they had recorded no working time data. Those exclusions correspond to 11,251 (1.5%) person-days. Another main reason for exclusion was that 1109 students and 227 external nurses did not have exact working hours. However, the data from both groups were usable for our descriptive analyses. Numerous students and external nurses became RNs over the 3 years of the study period. This largely explains why the final number of nurses was higher than 5736. Regarding the number of cases, a total of 128,484 (124,484/153,456, 83.7%) cases were used. Two main factors explain this reduction: outpatients (19,442/153,456, 12.7%) and healthy newborn babies (3779/153,456, 2.5%) from the Maternity & Gynecology department. Further details of the linking procedure are shown in Figure 1. For patient-to-nurse ratios and patient turnover analyses, we included 10 departments, including 70 inpatient units in which 4633 nurses (>22 million data points) provided care to 128,484 cases (>35 million data points).

Figure 1. Process to link the two datasets and variables used for the analysis. Nn: number of nurses; Np: number of patients; PNR: patient-to-nurse ratio; Na: number of admissions; Nd: number of discharges; Nti: number of transfers in; Nto: number of transfers out; RN: registered nurse; LPN: licensed practical nurse.



Department	Unit	Date	Time	Nn	Group	Np	PNR	Na	Nd	Nti	Nto
Intensive care	Unit1	01.01.2015	00:00	4	RN	6	1.5	0	0	0	0
Intensive care	Unit1	01.01.2015	00:00	2	LPN	6	3	0	0	0	0
Intensive care	Unit1	01.01.2015	00:00	1	Other	6	6	0	0	0	0
Intensive care	Unit1	01.01.2015	00:30	4	RN	6	1.5	1	0	0	0

Descriptive Overview

The number of patient cases in the included units in each department over the study period (2015-2017) ranged from 5007 for Hematology & Oncology to 28,377 for Cardiology & Cardiovascular Surgery. In almost all departments, mean and median patient ages were >54 years. The exceptions were patients in Maternity & Gynecology, who had a mean age of 36.5 years (SD 15.4 years) and median age of 33 years (IQR 28-40 years), and in Pediatrics, who had mean and median ages of 3.8 years (SD 5 years) and 1 year (IQR 0-7 years), respectively. The Hematology & Oncology, Internal Medicine, and Orthopedics & Plastic Surgery departments had the highest median LOS, at 7 days (IQR 4-14 days), 6 days (IQR 3-10 days), and 5 days (IQR 3-9 days), respectively. Cardiology &

Cardiovascular Surgery and Intensive Care had the lowest median LOS, at 2 days (IQR 1-7 days) and 2 days (IQR 2-3 days), respectively. The most common diagnoses were tumors, predominantly in the Hematology & Oncology department (4161/5007, 83.1%); circulatory system diseases, mainly in the Cardiology & Cardiovascular Surgery department (24,206/28,377, 85.3%); and traumatic injuries, poisonings, and other consequences of external causes, which were highest in the Orthopedics & Surgery department (5213/10,489, 49.7%). Further details are provided in Table 2.

Table 2. Descriptive overview of each department classified by the overall number of patients for 2015-2017.

Department	Number of patients	Number of units	Age (years), mean (SD)	Age (years), median (IQR[a])	LOS[b], median (IQR)	Patients/day/unit, median (IQR)	Top 2 diagnoses, n/N (%)	
							First	Second
Cardiology & Cardiovascular Surgery	28,377	12	67.3 (14.4)	70 (59-78)	2 (2-5)	10 (7-12)	Circulatory system diseases, 24,206/28,377 (85.3)	Traumatic injuries, poisonings, and other consequences of external causes, 1220/28,377 (4.3)
Neurology, Neurosurgery, Otolaryngology, Head & Neck Surgery, & Ophthalmology	27,916	10	58.5 (18.5)	61 (47-73)	4 (3-6)	13 (11-17)	Circulatory system diseases, 6421/27,916 (23)	Nervous system diseases, 4327/27,916 (15.5)
Intensive Care	21,359	8	61.6 (16.6)	64 (52-74)	2 (2-3)	8 (7-10)	Circulatory system diseases, 8095/21,359 (37.9)	Tumors, 3503/21,359 (16.4)
Pediatrics	19,543	10	3.8 (5)	1 (0-7)	3 (2-7)	11 (8-14)	Some conditions whose origin is the perinatal period, 3987/19,543 (20.4)	Respiratory system diseases, 3479/19,543 (17.8)
Dermatology, Urology, Rheumatology, & Nephrology	16,381	7	59.6 (17.3)	62 (48-73)	4 (3-7)	12 (6-17)	Genitourinary system diseases, 5160/16,381 (31.5)	Tumors, 3473/16,381 (21.2)
Visceral Surgery and Medicine, Gastroenterology, Thoracic Surgery, & Pulmonology	14,250	5	58.2 (17.1)	61 (48-71)	4 (3-7)	17 (12-21)	Digestive system diseases, 5073/14,250 (35.6)	Tumors, 4190/14,250 (29.4)
Internal Medicine	12,506	6	66 (18.3)	70 (54-80)	6 (3-10)	15 (13-18)	Circulatory system diseases, 2389/12,506 (19.1)	Infectious and parasitic diseases, 1163/12,506 (9.3)
Maternity & Gynecology	11,894	3	36.5 (15.4)	33 (28-40)	4 (3-5)	18 (14-21)	Pregnancy, childbirth, and the puerperium, 7172/11,894 (60.3)	Tumors, 1998/11,894 (16.8)
Orthopedics & Plastic Surgery	10,489	5	52.9 (19.6)	54 (37-68)	5 (3-9)	14 (12-16)	Traumatic injuries, poisoning, and some other consequences of external causes, 5213/10,489 (49.7)	Diseases of the osteo-articular system, muscles and connective tissue, 3346/10,489 (31.9)
Hematology & Oncology	5007	4	59.2 (15.5)	61 (51-70)	7 (4-14)	11 (7-18)	Tumors, 4161/5007 (83.1)	Endocrine, nutritional, and metabolic diseases, 225/5007 (4.5)

[a]IQR: interquartile range.

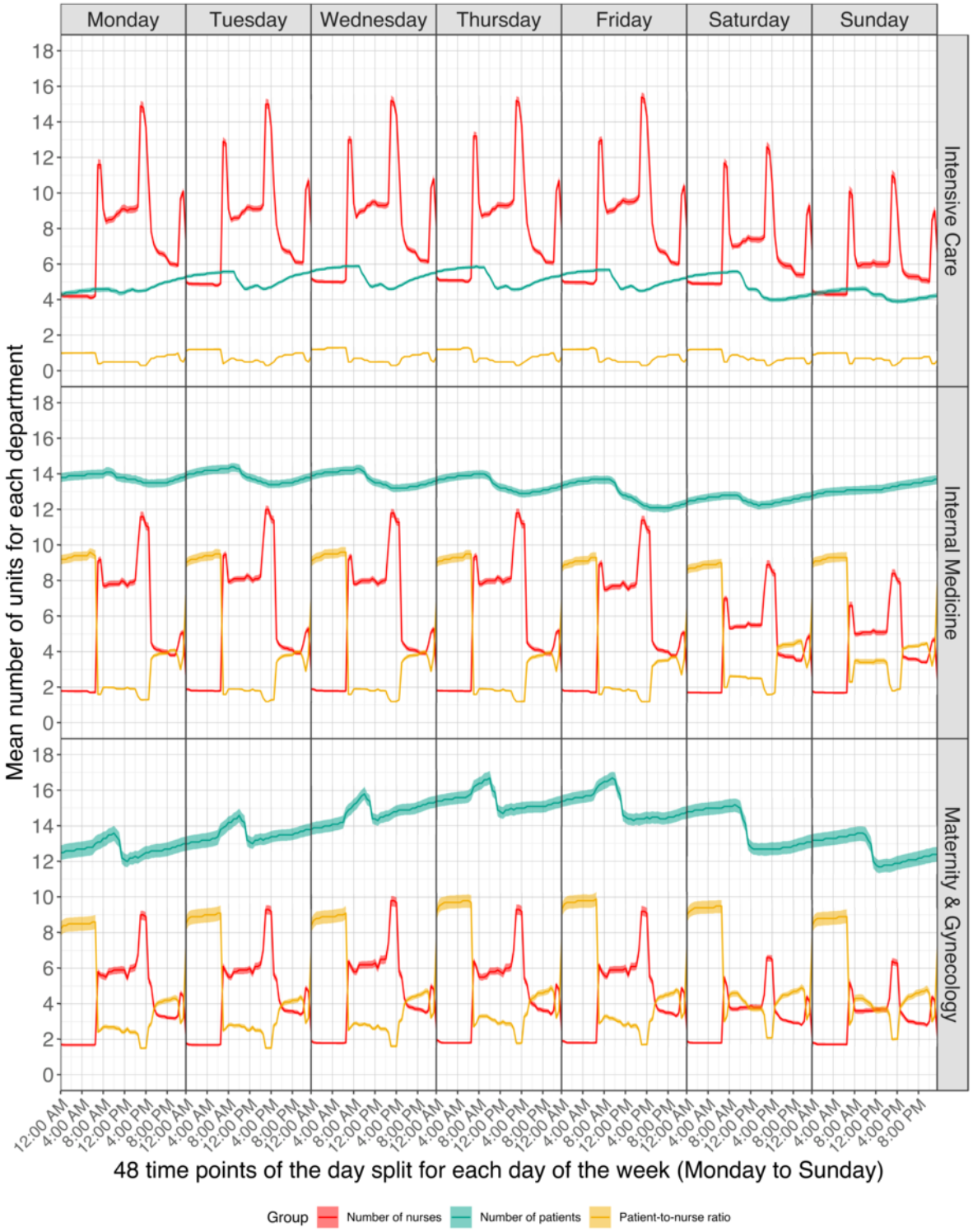
[b]LOS: length of stay.

Number of Patients

Numbers of patients and nurses and patient-to-nurse ratios were plotted against the 48 data points per day for each day of the week, each of the 10 departments, and each of the 3 nurse groups. Considering the large number of plots this generated, we show only the 3 plots that show the key characteristic patterns: the RN group (group 1) for the Intensive Care,

Maternity & Gynecology, and Internal Medicine departments (see Figure 2). All plots can be found in Multimedia Appendix 1 (group 1, RNs), Multimedia Appendix 2 (group 2, LPNs), and Multimedia Appendix 3 (group 3, others).

Figure 2. Plots of the number of patients, number of registered nurses, and patient-to-nurse ratio with the CIs.



On the demand side, a number of broad patterns emerged, several of which occurred across departments. Demand fluctuated not only throughout the day, with various clear peaks, but also through the week, as shown for the Maternity & Gynecology department, where patient numbers peaked on Thursday and Friday. Overall, patient numbers increased for 6 departments (6/10, 60%) from Monday to Thursday or Friday, and patient numbers peaked daily between 8:00 am and 10:00 am, then either stabilized or decreased. On Friday mornings, patient numbers decreased in preparation for the weekend. The exception was in the Internal Medicine department, where the

number of patients increased continuously from Friday evening until Monday morning (see Figure 2).

Number of Nurses

From the care supply perspective, variation was far less pronounced than on the demand side. Three main variations were apparent. First, fewer nurses were present through the weekends and occasionally on Friday. Second, Sundays generally had fewer nurses than on Saturdays. Third, the numbers of nurses were highest in the morning, then dropped for the afternoon shift and again for the night shifts. These patterns where quite stable throughout the week. In 6 (6/10, 60%) of the departments, gaps of 1 or 2 nurses were clearly discernible between 11:00 am and 1:00 pm.

Except for 3 (30%) of the 10 departments with almost no staff on the weekends, 1-2 LPNs were mainly present between 06:00 am and 5:00 pm. The pattern for nurse group 3 (ie, others) was similar to that of LPNs, although generally with roughly 1 more care staff. All departments increased their staff by 1-3 nurses for all or several of the following times: 7:00 am to 8:00 am, 2:00 pm to 4:30 pm, and 10:00 pm to 12:00 am (see Figure 2). As mentioned, apart from daily information, no records of working time were available for either students or external nurses. For external nurses and students, the median daily number of nurses was 0, except in the Internal Medicine department, where there was a median of 1 student. The daily number of external nurses ranged from 0 to 9 and of students from 0 to 12. For both, the maxima occurred only once, on a Sunday, during the 3 years.

Patient-to-nurse Ratio

Figure 2 shows the plots, and Figure 3 shows the median (IQR) of the patient-to-RN ratio for 3 key time points on weekdays and weekends. For RNs, the ratio was highest at night and lowest in the morning. During the night, the median ratio was 4-11 patients per RN, except in the Intensive Care department, which had a ratio of 1.1 patients per RN. In the morning, the ratio ranged from 0.6 (IQR 0.4-0.8) on weekdays to 0.8 (IQR 0.6-1) on weekends in the Intensive Care department and ranged from 2.8 (IQR 2-3.5) on weekdays to 4 (IQR 3-5) on weekends in the Maternity & Gynecology department. For LPNs, the median number was always 0 (IQR 0-0) at night, while the median number in the morning ranged from 3 to 8. For nurse group 3 (ie, others), a median number of 0 (IQR 0-0) staff members was generally present during the night shift. In the morning shifts, 7 departments' (7/10, 70%) median patient-to-nurse ratios increased for group 3 (ie, others) from weekdays (4.3 to 8) to weekends (6 to 12). In the afternoon shifts, all medians decreased (see Multimedia Appendix 4).

Figure 3. Median (interquartile range [IQR]) patient-to-registered nurse ratios for key time points, with the percentages of shifts with an extremely high threshold (EHT) or an extremely low threshold (ELT). Three departments are displayed split by weekdays and weekends.

Extreme Mismatch Between Supply and Demand

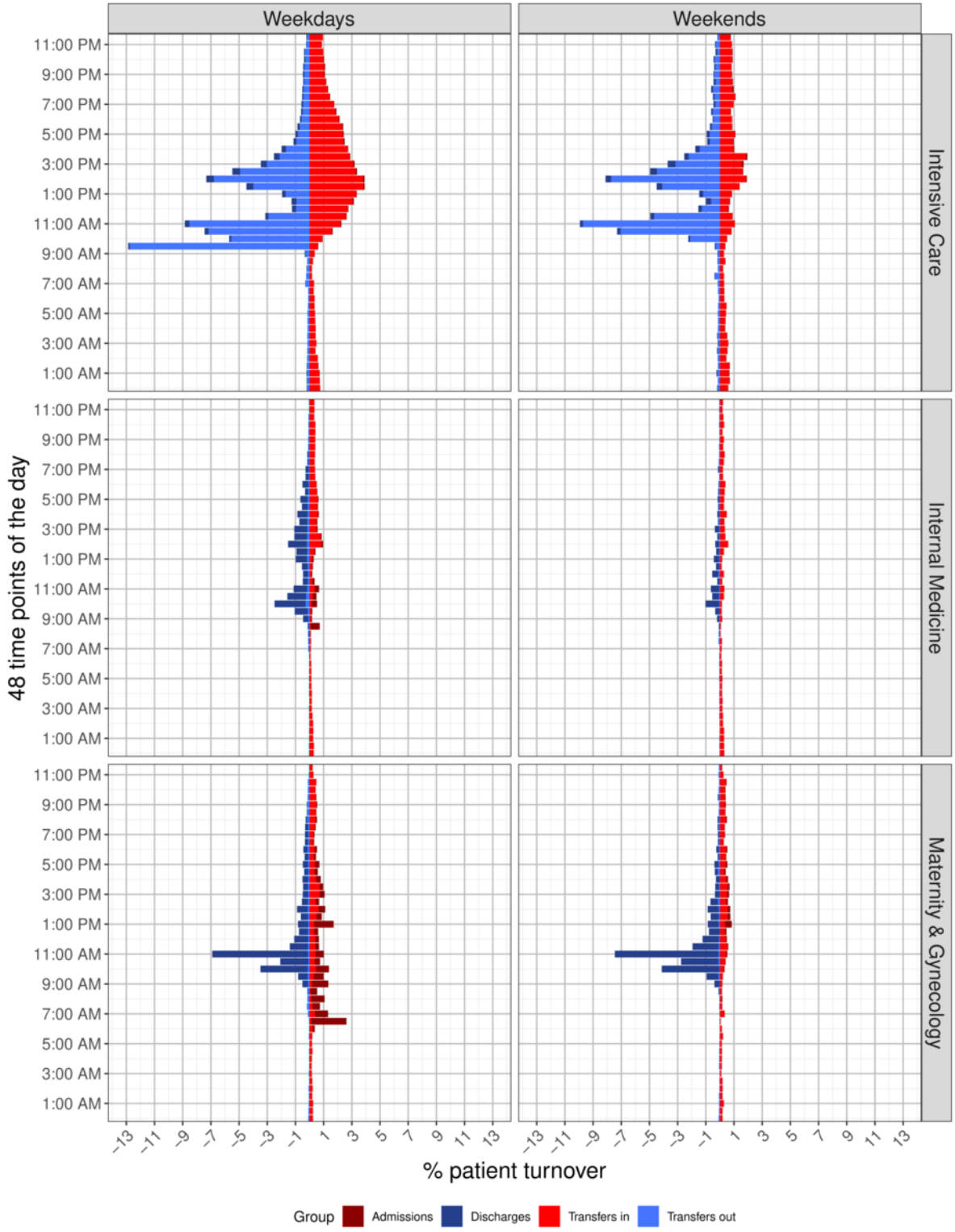
Figure 3 shows not only the weekday and weekend medians (IQRs) of the patient-to-RN ratios for 3 key time points (namely 2 am, 10 am, and 6 pm) but also the threshold values and percentages of shifts with extreme supply-and-demand mismatches. Complete results can be found in Multimedia Appendix 4. For 7 departments (7/10, 70%), the percentages of shifts with extreme staffing increased from morning to night and from weekdays to weekends. For RNs, the lowest percentages of extreme understaffing and overstaffing happened on 2.5% and 0.1% of weekday and weekend mornings, respectively. For both extremely high and extremely low patient-to-nurse ratios, the highest incidence (around 30% of shifts) occurred in the evening and weekend nights. These ratios occurred in the Cardiology & Cardiovascular Surgery (lower);

Dermatology, Urology, Rheumatology, & Nephrology (lower); and Hematology & Oncology (higher) departments. The same 3 departments had the lowest incidence of shifts with "normal" staffing levels (below 55%) for weekend nights. On the other hand, more than 80% of all shifts in the Orthopedics & Plastic Surgery department fell within "normal" staffing levels. For nurse groups 2 (LPNs) and 3 (others), the incidence of extreme staffing ranged from very high (49.5%) to very low (1.5%) to none (0%). Possibly because of these groups' low nurse numbers, no clear patterns were apparent.

Patient Turnover

Similar to the patient-to-nurse ratio, bar charts for patient turnover are displayed for only the Intensive Care, Internal Medicine, and Maternity & Gynecology departments (see Figure 4). Bar charts for all 10 departments are displayed in Multimedia Appendix 5. All departments showed reductions in patient turnover during the weekends. Entries (admissions and transfers in) and exits (discharges and transfers out) of patients occurred at very similar times for all departments: 09:00 am to 11:00 am and 1:00 pm to 3:00 pm. The Intensive Care department had the highest percentage of transfers (peaking at almost 13% at 9:30 am). As shown in Figure 3, the numbers of extreme staffing mismatches also fluctuated as a result of the number of either nurses or patients present. Figure 4 shows the variation in the patterns throughout the day that influenced the patient-to-nurse ratio. For example, if on a given day a peak of discharges occurs at 10:00 am with no or few admissions or transfers in, the patient-to-nurse ratio will decrease, potentially leading to extreme overstaffing. The same is true for the inverse. A peak in admissions or transfers in can increase the patient-to-nurse ratio, leading to extreme understaffing.

Figure 4. Percentages of patient turnover for the 48 data points by weekdays and weekends.



Discussion

Main Results

For the first time, we longitudinally analyzed demand for and supply of care at 30-minute intervals for 24 hours a day over a period of 3 years in a large university hospital. Data from the nurse staffing system (tacs) and medical discharge records were used to explore patient-to-nurse ratios and patient turnover. The 10 departments belonged to the Inselspital (Bern University Hospital) and varied by purpose, size, number of units, and patient population.

From the demand side, continuous turnover meant patient numbers fluctuated across each day and varied across units. Less variation was seen in the supply side, where the change in the number of nurses occurred mainly for each shift (night, morning, evening). RNs accounted for roughly three-quarters of the nurse workforce, making them the largest staff subgroup

involved with patient care. The remaining quarter was comprised of others (including unlicensed personnel and administrative staff) and LPNs. These smaller groups were mainly present during day shifts on weekdays.

Simultaneous longitudinal data on patient and nurse numbers allowed us to determine which had the greater influence on patient-to-nurse ratios (ie, the effects the variations in the numbers of nurses and patients had on nurse workload). Most published studies have shown that during weekends, patient-to-nurse ratios tend to increase [6,7]. We confirmed this observation, not only because the nursing staff was reduced over weekends but also because in many departments (eg, Internal Medicine), patient numbers tended to increase on Saturday and Sunday. In fact, we found that the supply side remained quite constant; it was mainly the demand side that drove patient-to-nurse ratios.

The ramifications of a demand-driven workload are particularly clear regarding weekend staff planning. Reducing staffing for reduced weekend demand might be justified on Friday nights, as a surge in discharges Thursday and Friday decreases patient numbers. However, patient numbers increased in several departments over the weekends, as patient entries outnumbered exits.

To set the reported patient-to-nurse ratios in context, the European cross-sectional study RN4CAST, which was conducted in 2009-2011 with 35 Swiss acute care hospitals, reported an average of 7.9 patients per nurse in medical and surgical units [37,38]. However, the reported ratios were aggregated at the hospital level, ignoring the ratios per shift. In 2015, the Swiss cross-sectional Match^{RN} study followed up the same hospitals that participated in RN4CAST [39]. The overall average patient-to-nurse ratio of the 23 participating hospitals was 7.8, and the shift averages were 5.9 for the morning shift, 7.3 for the evening shift, and 14.2 for the night shift [40,41]. Both cross-sectional studies surveyed only RNs. From our study, the median patient-to-nurse ratio for RNs was 2-3 for the morning shift, 3-5 for the evening shift, and 4-10 for the night shift, excluding the Pediatrics, Maternity & Gynecology, and Intensive Care departments. The ratios are difficult to compare with the overall averages from the RN4CAST and Match^{RN} studies. At the shift level, our patient-to-nurse ratios were lower than the Match^{RN} ratios, suggesting an above-average staffed hospital. High patient-to-nurse ratios have been associated with worse patient outcomes; however, due to conflicting results, the relationship remains unclear [42,43]. One main reason for the lack of conclusive evidence is researchers' tendency to seek associations between mean patient-to-nurse ratios and patient outcomes [44]. However, this means obscuring sharp changes in supply and demand, thereby concealing periods when staffing levels are low. As no consensus exists concerning the definition of an extreme staffing situation, we chose arbitrary cut-offs of double or half of the median work per nurse to define extremely high or low patient-to-nurse ratios, respectively. These cut-offs showed that our extreme thresholds were commonly crossed during certain shifts and in certain departments.

This observation underscores the potential volatility of nurse workload, even over a single shift, and the need for longitudinal approaches to staffing research to help identify and counteract this volatility [45]. The distribution of extreme shifts also indicated that individual departments (eg, the Orthopedics & Surgery department, with more than 80% of units and shifts staffed within the "normal" range) can maintain their patient-to-nurse ratio quite effectively. Identifying the most meaningful thresholds to define extreme staffing will require further research.

As illustrated in Figure 4, while patient turnover was continuous, it was concentrated at various times throughout the day. Consistent with previous findings, entries and exits were both rare during the night [46,47]. Moving patients to the units where they can receive the most appropriate care is essential for their recovery. Also, as demand is independent of available resources, patient turnover occurs when units are short on either staff or beds. For the former, the unit is closed, and the patient is moved to a similar unit; for the latter, the patient is placed in an intermediate unit until a bed becomes available [48]. However, both cases lead to increased administrative work, and even where beds are available and staff sufficient, transfers, admissions, and discharges all entail higher volumes of administrative requirements and patient care needs. Therefore, our analyses confirmed that patient turnover is one factor for nursing workload [5,47,49].

The impact of turnover can be greater when entries and exits occur at the same time, as illustrated in the first row of Figure 4 for the Intensive Care department. Between 1.6% and 32.3% of nursing time is spent on patient turnover-related activities [50,51]. Associations between patient turnover and nursing care quality have been documented, where higher turnover led to higher nursing workload, possibly compromising nursing care quality [13,38,52]. High patient turnover is associated with more adverse events, including mortality [53], nosocomial infections [54], and medication errors [52] as well as more readmissions [55,56]. The current approach is somewhat unrealistic. Where each patient case or event is rooted in a unique context, much of the current literature treats all patient turnover and patients as the same [47,57]. In contrast, Tierney et al [58] used weighted patient acuity and patient turnover variables to account for intercase variation. Studies also showed that death was a more likely outcome in contexts featuring high patient-to-nurse ratios and patient turnover [13,59,60].

Potential Implications

Because of the granularity of the analysis, patient-to-nurse ratios were analyzed with patient turnover, as even in cases where the patient-to-nurse ratio might appear normal, both patient entries and exits increase nurse workload. During periods of high turnover, the time available for patient care can be severely reduced. Hospitals or departments that fail to account for this additional burden commonly operate with suboptimal nursing staff levels [51-53].

Previous research has suggested a relationship between higher patient-to-nurse ratios and worse patient outcomes [61-66]. Mandatory minimum patient-to-nurse ratios are often suggested as an approach to ensure safe staffing levels. As the nursing

supply is quite constant at the level of individual shifts, the question may arise as to whether that supply can realistically be changed in response to midshift fluctuations in demand. The patient turnover variability illustrated in this study showed that where entries matched exits, patient numbers remained reasonably constant and where imbalances occurred between entries and exits, patient numbers fluctuated.

However, neither of these cases adequately reflects nurse workload. In the first, even while a balanced turnover resulted in a constant patient-to-nurse ratio, if both sides increased, the additional work required for each incoming and outgoing case would represent a considerable burden. In the second, records of patient numbers alone give no indication as to whether the supply of nurses was adjusted accordingly. These two examples illustrate the necessity of considering both supply and demand data for staffing purposes.

Certain patterns were clearly associated with routines that applied to specific days and times of the week. Defining and clarifying those periods for each unit would help improve assessment of staffing levels. Given that some hospital departments do not operate on the weekends, further detailed analysis of weekend work for nurses is needed to determine how the workload increases. Current research on hospital staffing is predominantly based on cross-sectional data, which cannot show fluctuations in patient turnover [67]. Lacking longitudinal data, it is extremely difficult to match the rather constant nurse supply to the varying patterns in patient demand.

To our knowledge, only one previous study examined the longitudinal associations between nurse staffing and patient turnover. Its findings indicated large variations in patient turnover [68]. As the demand side is quite volatile, to anticipate when the nurse supply should be changed to match changes in workload, it is important to identify the times of day, days of the week, and even months during which specific entry and exit patterns occur. Armed with this information, staffing levels might differ across not only units and hospitals but also countries. Thus, unit-level analysis offers the best chance to detect patterns of supply and demand. Identifying the complex relationships involved and then building more efficient predictive models that capture all meaningful variations will require further studies examining longitudinal nurse and patient data.

Limitations

Certain limitations were encountered during our analyses. One of these concerned the tacs nurse staffing system, as this was the first time that routine data were used for research and linking purposes. During the process, we found that a small minority of nurses (~ 1%) were not using the system consistently. Also, due to issues with merging data, a number of nurses and patients were excluded from the analysis. To maximize the data quality for this and future studies, these issues were discussed with the nurse staffing system software developers.

Also, while we selected persons providing direct or indirect care for the analysis, it was impossible to know whether those people also performed tasks not associated with patient care, such as organizational tasks or teaching. Patient-to-nurse ratios were calculated for all persons present in each unit studied. Although we measured variation in nurses' workload with a high level of granularity, the significance of the short peaks in demand relative to the supply over short periods is hard to judge because nurses' work involves multitasking and they can prioritize urgent tasks and delay others without necessarily harming patients [69]. Minutes of care were also available from the data; however, due to data quality concerns regarding the time allocated to each patient, these data were excluded, and metadata were included in their place. This may have solved the patient care time data limitation by providing the exact time invested for each patient during working hours.

Further, the results were limited by the absence of accurate working time data for external nurses and students. Even if these groups had a daily median presence of 0, their assistance might have been crucial when they were present, as for night shift support. This type of task shifting between individuals and departments to compensate for staffing shortfalls is a key tool to handle demand and avoid gaps in supply but was not recorded in the available data.

Our study looked only at nurses, but the hospital environment is multidisciplinary. All health care providers play an important role, and their collaboration is crucial for patients [70]. For example, studies showed a positive impact on patients' outcomes by incorporating or improving nurse-physician or pharmacist-physician collaborative practices [71-73].

Finally, the study was undertaken in a single hospital, and we explored many sources of variation, but not patient acuity and severity. Nursing workload depends on not only the amount of direct and indirect care, patient turnover, and patient-to-nurse ratio but also patient acuity and severity [74,75]. Further investigation is thus needed.

Conclusions

To our knowledge, this is the first detailed study to use data on patient-to-nurse ratios and patient turnover in time increments as low as 30 minutes. The goal was to illustrate fluctuations in these two variables between and within departments and days of the week. The choice of 30 minutes was subjective and based on available computational resources. While the literature includes references to the fluctuations we observed, no study to date has illustrated those fluctuations in such fine detail. The key driver for care was clearly patient demand, which showed high variability even during individual shifts. This volatility challenges health care suppliers to provide safe and reliable care when demand is high while avoiding overstaffing when it is low. Detecting patterns of variation will help optimize staffing. This descriptive analysis was a first step towards detecting fluid variables to be considered in developing a predictive model on which to base health care staff planning, possibly including the introduction of innovative working/shift schemes in this sensitive sector.

Acknowledgments

MS and SNM developed the idea for the study. MS, ABL, PG, and CTN contributed to the concept, design, and supervision of the project. SNM conducted the analysis and contributed to the drafting of the manuscript. All authors contributed to the interpretation of the data and the critical revision of the manuscript. All authors approved the final version.

We would like to thank Barbara Ammann, Wilco Laan, and Elias Panizza from Wigasoft for helping us to understand the nurse staffing system and for delivering the additional variables necessary to conduct the analysis. Additionally, we would also like to thank Chris Shultis for professional editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of patients and nurses with patient-to-nurse ratios plotted (with confidence intervals) for RN group. The x-axis shows the 48 time points of the day split for each day of the week (Monday to Sunday); the y-axis represents the mean number of units for each of the ten departments.
[PDF File (Adobe PDF File), 179 KB - jmir_v22i4e15554_app1.pdf]

Multimedia Appendix 2
Number of patients and nurses with patient-to-nurse ratios plotted (with confidence intervals) for LPN group. The x-axis shows the 48 time points of the day split for each day of the week (Monday to Sunday); the y-axis represents the mean number of units for each of the ten departments.
[PDF File (Adobe PDF File), 173 KB - jmir_v22i4e15554_app2.pdf]

Multimedia Appendix 3
Number of patients and nurses, with patient-to-nurse ratios plotted (with confidence intervals) for Others group. The x-axis shows the 48 time points of the day split for each day of the week (Monday to Sunday); the y-axis represents the mean number of units for each of the ten departments.
[PDF File (Adobe PDF File), 175 KB - jmir_v22i4e15554_app3.pdf]

Multimedia Appendix 4
Median of patient−to−nurse ratio for key time points, split by weekdays/weekends for each group of nurses, together with percentages of shifts with extreme patient−to−nurse ratios.
[PDF File (Adobe PDF File), 54 KB - jmir_v22i4e15554_app4.pdf]

Multimedia Appendix 5
Patient turnover in percentages for the 48 data points, split for weekdays and weekends.
[PDF File (Adobe PDF File), 65 KB - jmir_v22i4e15554_app5.pdf]

References

1. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA 2002;288(16):1987-1993. [doi: 10.1001/jama.288.16.1987] [Medline: 12387650]
2. Mark BA, Hughes LC, Belyea M, Chang Y, Hofmann D, Jones CB, et al. Does safety climate moderate the influence of staffing adequacy and work conditions on nurse injuries? J Safety Res 2007;38(4):431-446 [FREE Full text] [doi: 10.1016/j.jsr.2007.04.004] [Medline: 17884430]
3. Manojlovich M, Sidani S, Covell CL, Antonakos CL. Nurse dose: linking staffing variables to adverse patient outcomes. Nurs Res 2011;60(4):214-220. [doi: 10.1097/NNR.0b013e31822228dc] [Medline: 21691239]
4. Fasoli DR, Haddock KS. Results of an integrative review of patient classification systems. Annu Rev Nurs Res 2010;28:295-316. [doi: 10.1891/0739-6686.28.295] [Medline: 21639031]
5. Swiger PA, Vance DE, Patrician PA. Nursing workload in the acute-care setting: A concept analysis of nursing workload. Nurs Outlook 2016;64(3):244-254. [doi: 10.1016/j.outlook.2016.01.003] [Medline: 26944266]
6. Palese A, Petean M, Cerne D. Unexpected deaths in medical wards during night shifts: a narrative analysis of nursing experiences. J Clin Nurs 2014 Sep;23(17-18):2599-2608. [doi: 10.1111/jocn.12477] [Medline: 24354422]
7. Patrician PA, Loan L, McCarthy M, Fridman M, Donaldson N, Bingham M, et al. The association of shift-level nurse staffing with adverse patient events. J Nurs Adm 2011 Feb;41(2):64-70. [doi: 10.1097/NNA.0b013e31820594bf] [Medline: 21266884]

8. Aspden P, Wolcott J, Bootman J, Cronenwett L. Preventing medication errors: quality chasm series. In: Committee on Identifying and Preventing Medication Errors. Washington (DC): The National Academies Press; 2007.
9. Blay N, Duffield CM, Gallagher R. Patient transfers in Australia: implications for nursing workload and patient outcomes. J Nurs Manag 2012 Apr;20(3):302-310. [doi: 10.1111/j.1365-2834.2011.01279.x] [Medline: 22519607]
10. Blay N, Roche MA, Duffield C, Gallagher R. Intrahospital transfers and the impact on nursing workload. J Clin Nurs 2017 Dec;26(23-24):4822-4829. [doi: 10.1111/jocn.13838] [Medline: 28382638]
11. Duffield C, Diers D, Aisbett C, Roche M. Churn: patient turnover and case mix. Nurs Econ 2009;27(3):185-191. [Medline: 19558079]
12. McGillis Hall L, Kiesners D. A narrative approach to understanding the nursing work environment in Canada. Soc Sci Med 2005 Dec;61(12):2482-2491. [doi: 10.1016/j.socscimed.2005.05.002] [Medline: 15964672]
13. Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR, Harris M. Nurse staffing and inpatient hospital mortality. N Engl J Med 2011 Mar 17;364(11):1037-1045. [doi: 10.1056/NEJMsa1001025] [Medline: 21410372]
14. Griffiths P, Maruotti A, Recio Saucedo A, Redfern OC, Ball JE, Briggs J, Missed Care Study Group. Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study. BMJ Qual Saf 2019 Aug;28(8):609-617 [FREE Full text] [doi: 10.1136/bmjqs-2018-008043] [Medline: 30514780]
15. Huber E, Kleinknecht-Dolf M, Müller M, Kugler C, Spirig R. Mixed-method research protocol: defining and operationalizing patient-related complexity of nursing care in acute care hospitals. J Adv Nurs 2017 Jun;73(6):1491-1501. [doi: 10.1111/jan.13218] [Medline: 27878843]
16. Welton JM. Measuring Patient Acuity: Implications for Nurse Staffing and Assignment. J Nurs Adm 2017 Oct;47(10):471. [doi: 10.1097/NNA.0000000000000516] [Medline: 28957863]
17. Halm M, Peterson M, Kandels M, Sabo J, Blalock M, Braden R, et al. Hospital nurse staffing and patient mortality, emotional exhaustion, and job dissatisfaction. Clin Nurse Spec 2005;19(5):241-51; quiz 252. [doi: 10.1097/00002800-200509000-00007] [Medline: 16179855]
18. Titler M, Dochterman J, Xie X, Kanak M, Fei Q, Picone DM, et al. Nursing interventions and other factors associated with discharge disposition in older patients after hip fractures. Nurs Res 2006;55(4):231-242. [doi: 10.1097/00006199-200607000-00003] [Medline: 16849975]
19. Weiss ME, Piacentine LB, Lokken L, Ancona J, Archer J, Gresser S, et al. Perceived readiness for hospital discharge in adult medical-surgical patients. Clin Nurse Spec 2007;21(1):31-42. [doi: 10.1097/00002800-200701000-00008] [Medline: 17213738]
20. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 1995 May 01;122(9):681-688. [doi: 10.7326/0003-4819-122-9-199505010-00007] [Medline: 7702231]
21. Diya L, Van den Heede K, Sermeus W, Lesaffre E. The relationship between in-hospital mortality, readmission into the intensive care nursing unit and/or operating theatre and nurse staffing levels. J Adv Nurs 2012 May;68(5):1073-1081. [doi: 10.1111/j.1365-2648.2011.05812.x] [Medline: 21883408]
22. Griffiths P, Ball J, Drennan J, Dall'Ora C, Jones J, Maruotti A, et al. Nurse staffing and patient outcomes: Strengths and limitations of the evidence to inform policy and practice. A review and discussion paper based on evidence reviewed for the National Institute for Health and Care Excellence Safe Staffing guideline development. Int J Nurs Stud 2016 Nov;63:213-225. [doi: 10.1016/j.ijnurstu.2016.03.012] [Medline: 27130150]
23. Welton JM. Nurse staffing and patient outcomes: Are we asking the right research question? Int J Nurs Stud 2016 Nov;63:A1-A2. [doi: 10.1016/j.ijnurstu.2016.08.015] [Medline: 27599693]
24. Clarke S, Donaldson N. Nurse staffing and patient care quality and safety. In: Hughes RG, editor. Patient safety and quality: An evidence-based handbook for nurses. Rockville (MD): Patient Safety and Quality: An Evidence-Based Handbook for Nurses; 2008.
25. Yakusheva O, Lindrooth R, Weiss M. Economic evaluation of the 80% baccalaureate nurse workforce recommendation: a patient-level analysis. Med Care 2014 Oct;52(10):864-869. [doi: 10.1097/MLR.0000000000000189] [Medline: 25215646]
26. Insel Gruppe AG. Leistung: Unsere Spitäler stehen allen offen 2018 URL: https://jahresbericht.inselgruppe.ch/#leistung; [accessed 2019-05-09]
27. R Core T. Development. 2013. R: A language and Environment for Statistical Computing URL: https://www.r-project.org/ [accessed 2020-02-20]
28. Henry L, Wickham H. R package version 0.2. 2018. purrr: Functional Programming Tools URL: https://cran.r-project.org/web/packages/purrr/ [accessed 2020-02-21]
29. Wickham H, François R, Henry L, Müller K. R package version 0.7.6. 2018. dplyr: A Grammar of Data Manipulation URL: https://cran.r-project.org/web/packages/dplyr/ [accessed 2020-02-21]
30. Wickham H. RDocumentation. 2019. tidyr v0.8.3: Easily Tidy Data with 'spread()' and 'gather()' Functions URL: https://www.rdocumentation.org/packages/tidyr/versions/0.8.3 [accessed 2020-02-23]
31. Dowle M, Srinivasan A. R package version 1.10. 2017. data. table: Extension of ‘data. frame’ URL: https://cran.r-project.org/web/packages/data.table/index.html [accessed 2020-02-21]

32. Grolemund G, Wickham H. Dates and Times Made Easy with {lubridate}. J Stat Soft 2011;40(3):1-25. [doi: 10.18637/jss.v040.i03]
33. James D, Hornik K. R package version 23. 2018. chron: Chronological Objects which Can Handle Dates and Times URL: https://cran.r-project.org/web/packages/chron/ [accessed 2020-02-21]
34. Thoen E. R package version 0.4.1. 2018. padr: Quickly Get Datetime Data Ready for Analysis URL: https://cran.r-project.org/web/packages/padr/index.html [accessed 2020-02-21]
35. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2016.
36. Wickham H. R package version 1.0.0. 2018. scales: Scale Functions for Visualization URL: https://cran.r-project.org/web/packages/scales/index.html [accessed 2020-02-21]
37. Sermeus W, Aiken LH, Van den Heede K, Rafferty AM, Griffiths P, Moreno-Casbas MT, RN4CAST consortium. Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. BMC Nurs 2011 Apr 18;10:6 [FREE Full text] [doi: 10.1186/1472-6955-10-6] [Medline: 21501487]
38. Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Sermeus W, RN4CAST Consortium. Nurses' reports of working conditions and hospital quality of care in 12 countries in Europe. Int J Nurs Stud 2013 Feb;50(2):143-153. [doi: 10.1016/j.ijnurstu.2012.11.009] [Medline: 23254247]
39. Bachnick S, Ausserhofer D, Januel J, Schubert M, Schwendimann R, De Geest S, et al. Matching Registered Nurse services with changing care demands (Match): study protocol of a natural experiment multi-centre study. J Adv Nurs 2017 Jul;73(7):1735-1746. [doi: 10.1111/jan.13287] [Medline: 28251690]
40. Bachnick S, Ausserhofer D, Baernholdt M, Simon M, Match RN study group. Patient-centered care, nurse work environment and implicit rationing of nursing care in Swiss acute care hospitals: A cross-sectional multi-center study. Int J Nurs Stud 2018 May;81:98-106. [doi: 10.1016/j.ijnurstu.2017.11.007] [Medline: 29554590]
41. Dhaini SR, Denhaerynck K, Bachnick S, Schwendimann R, Schubert M, De Geest S, Match RN study group. Work schedule flexibility is associated with emotional exhaustion among registered nurses in Swiss hospitals: A cross-sectional study. Int J Nurs Stud 2018 Jun;82:99-105. [doi: 10.1016/j.ijnurstu.2018.03.019] [Medline: 29626703]
42. Sales A, Sharp N, Li Y, Lowy E, Greiner G, Liu C, et al. The association between nursing factors and patient mortality in the Veterans Health Administration: the view from the nursing unit level. Med Care 2008 Sep;46(9):938-945. [doi: 10.1097/MLR.0b013e3181791a0a] [Medline: 18725848]
43. Numata Y, Schulzer M, van der Wal R, Globerman J, Semeniuk P, Balka E, et al. Nurse staffing levels and hospital mortality in critical care settings: literature review and meta-analysis. J Adv Nurs 2006 Aug;55(4):435-448. [doi: 10.1111/j.1365-2648.2006.03941.x] [Medline: 16866839]
44. Lee A, Cheung YSL, Joynt GM, Leung CCH, Wong W, Gomersall CD. Are high nurse workload/staffing ratios associated with decreased survival in critically ill patients? A cohort study. Ann Intensive Care 2017 Dec;7(1):46 [FREE Full text] [doi: 10.1186/s13613-017-0269-2] [Medline: 28466462]
45. Schubert M, Glass TR, Clarke SP, Aiken LH, Schaffert-Witvliet B, Sloane DM, et al. Rationing of nursing care and its relationship to patient outcomes: the Swiss extension of the International Hospital Outcomes Study. Int J Qual Health Care 2008 Aug;20(4):227-237 [FREE Full text] [doi: 10.1093/intqhc/mzn017] [Medline: 18436556]
46. Baernholdt M, Cox K, Scully K. Using clinical data to capture nurse workload: implications for staffing and safety. Comput Inform Nurs 2010;28(4):229-234. [doi: 10.1097/NCN.0b013e3181e1e57d] [Medline: 20571375]
47. Jennings BM, Sandelowski M, Higgins MK. Turning over patient turnover: an ethnographic study of admissions, discharges, and transfers. Res Nurs Health 2013 Dec;36(6):554-566 [FREE Full text] [doi: 10.1002/nur.21565] [Medline: 24242196]
48. VanFosson CA, Yoder LH, Jones TL. Patient Turnover: A Concept Analysis. ANS Adv Nurs Sci 2017;40(3):298-310. [doi: 10.1097/ANS.0000000000000171] [Medline: 28266962]
49. Myny D, Van Hecke A, De Bacquer D, Verhaeghe S, Gobert M, Defloor T, et al. Determining a set of measurable and relevant factors affecting nursing workload in the acute care hospital setting: a cross-sectional study. Int J Nurs Stud 2012 Apr;49(4):427-436. [doi: 10.1016/j.ijnurstu.2011.10.005] [Medline: 22030021]
50. Abbey M, Chaboyer W, Mitchell M. Understanding the work of intensive care nurses: a time and motion study. Aust Crit Care 2012 Feb;25(1):13-22. [doi: 10.1016/j.aucc.2011.08.002] [Medline: 21937236]
51. Cornell P, Herrin-Griffith D, Keim C, Petschonek S, Sanders AM, D'Mello S, et al. Transforming nursing workflow, part 1: the chaotic nature of nurse activities. J Nurs Adm 2010 Sep;40(9):366-373. [doi: 10.1097/NNA.0b013e3181ee4261] [Medline: 20798619]
52. Duffield C, Diers D, O'Brien-Pallas L, Aisbett C, Roche M, King M, et al. Nursing staffing, nursing workload, the work environment and patient outcomes. Appl Nurs Res 2011 Nov;24(4):244-255. [doi: 10.1016/j.apnr.2009.12.004] [Medline: 20974086]
53. Wagner J, Gabler NB, Ratcliffe SJ, Brown SES, Strom BL, Halpern SD. Outcomes among patients discharged from busy intensive care units. Ann Intern Med 2013 Oct 01;159(7):447-455 [FREE Full text] [doi: 10.7326/0003-4819-159-7-201310010-00004] [Medline: 24081285]
54. Cunningham JB, Kernohan WG, Rush T. Bed occupancy, turnover intervals and MRSA rates in English hospitals. Br J Nurs 2006;15(12):656-660. [doi: 10.12968/bjon.2006.15.12.21398] [Medline: 16835541]

55. Heggestad T. Operating conditions of psychiatric hospitals and early readmission--effects of high patient turnover. Acta Psychiatr Scand 2001 Mar;103(3):196-202. [doi: 10.1034/j.1600-0447.2001.00166.x] [Medline: 11240576]
56. Wagner C, Budreau G, Everett LQ. Analyzing fluctuating unit census for timely staffing intervention. Nurs Econ 2005;23(2):85-90, 55. [Medline: 15881494]
57. Fieldston ES, Ragavan M, Jayaraman B, Allebach K, Pati S, Metlay JP. Scheduled admissions and high occupancy at a children's hospital. J Hosp Med 2011 Feb;6(2):81-87. [doi: 10.1002/jhm.819] [Medline: 21290580]
58. Tierney SJ, Seymour-Route P, Crawford S. Weighted staffing plans for better prediction of staffing needs. J Nurs Adm 2013 Sep;43(9):461-467. [doi: 10.1097/NNA.0b013e3182a23e41] [Medline: 23979036]
59. Neuraz A, Guérin C, Payet C, Polazzi S, Aubrun F, Dailler F, et al. Patient Mortality Is Associated With Staff Resources and Workload in the ICU: A Multicenter Observational Study. Crit Care Med 2015 Aug;43(8):1587-1594. [doi: 10.1097/CCM.0000000000001015] [Medline: 25867907]
60. Griffiths P, Ball J, Bloor K, Böhning D, Briggs J, Dall'Ora C, et al. Nurse staffing levels, missed vital signs observations and mortality in hospital wards: retrospective longitudinal observational study. Health Services and Delivery Research 2018;6(38):1-120. [doi: 10.3310/hsdr06380] [Medline: 30516947]
61. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. N Engl J Med 2002 May 30;346(22):1715-1722. [doi: 10.1056/NEJMsa012247] [Medline: 12037152]
62. Blegen MA, Goode CJ, Reed L. Nurse staffing and patient outcomes. Nurs Res 1998;47(1):43-50. [doi: 10.1097/00006199-199801000-00008] [Medline: 9478183]
63. Kovner C, Gergen PJ. Nurse staffing levels and adverse events following surgery in U.S. hospitals. Image J Nurs Sch 1998;30(4):315-321. [Medline: 9866290]
64. Rafferty AM, Clarke SP, Coles J, Ball J, James P, McKee M, et al. Outcomes of variation in hospital nurse staffing in English hospitals: cross-sectional analysis of survey data and discharge records. Int J Nurs Stud 2007 Feb;44(2):175-182 [FREE Full text] [doi: 10.1016/j.ijnurstu.2006.08.003] [Medline: 17064706]
65. McCloskey JM. Nurse staffing and patient outcomes. Nurs Outlook 1998;46(5):199-200. [doi: 10.1016/s0029-6554(98)90047-1] [Medline: 9805336]
66. Weissman JS, Rothschild JM, Bendavid E, Sprivulis P, Cook EF, Evans RS, et al. Hospital workload and adverse events. Med Care 2007 May;45(5):448-455. [doi: 10.1097/01.mlr.0000257231.86368.09] [Medline: 17446831]
67. Driscoll A, Grant MJ, Carroll D, Dalton S, Deaton C, Jones I, et al. The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units: a systematic review and meta-analysis. European Journal of Cardiovascular Nursing 2018;17(1):a. [doi: 10.1177/1474515117721561] [Medline: 28718658]
68. He J, Staggs VS, Bergquist-Beringer S, Dunton N. Nurse staffing and patient outcomes: a longitudinal study on trend and seasonality. BMC Nurs 2016;15:60 [FREE Full text] [doi: 10.1186/s12912-016-0181-3] [Medline: 27757068]
69. Patterson ES, Ebright PR, Saleem JJ. Investigating stacking: How do registered nurses prioritize their activities in real-time? International Journal of Industrial Ergonomics 2011 Jul;41(4):389-393. [doi: 10.1016/j.ergon.2011.01.012]
70. Morley L, Cashell A. Collaboration in Health Care. J Med Imaging Radiat Sci 2017 Jun;48(2):207-216. [doi: 10.1016/j.jmir.2017.02.071] [Medline: 31047370]
71. Brown SS, Lindell DF, Dolansky MA, Garber JS. Nurses' professional values and attitudes toward collaboration with physicians. Nurs Ethics 2015 Mar;22(2):205-216. [doi: 10.1177/0969733014533233] [Medline: 24878573]
72. Matzke GR, Moczygemba LR, Williams KJ, Czar MJ, Lee WT. Impact of a pharmacist-physician collaborative care model on patient outcomes and health services utilization. Am J Health Syst Pharm 2018 Jul 15;75(14):1039-1047. [doi: 10.2146/ajhp170789] [Medline: 29789318]
73. Sabone M, Mazonde P, Cainelli F, Maitshoko M, Joseph R, Shayo J, et al. Everyday ethical challenges of nurse-physician collaboration. Nurs Ethics 2020 Feb 23;27(1):206-220. [doi: 10.1177/0969733019840753] [Medline: 31014168]
74. Acar I, Butt SE. Modeling nurse-patient assignments considering patient acuity and travel distance metrics. J Biomed Inform 2016 Dec;64:192-206 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.006] [Medline: 27742350]
75. Alghamdi MG. Nursing workload: a concept analysis. J Nurs Manag 2016 May;24(4):449-457. [doi: 10.1111/jonm.12354] [Medline: 26749124]

Abbreviations

IQR: interquartile range.
LOS: length of stay.
LPNs: licensed practical nurses.
RNs: registered nurses.

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by M Smith, JH Lee; comments to author 12.10.19; revised version received 28.10.19; accepted 03.02.20; published 02.04.20.

Please cite as:
Musy SN, Endrich O, Leichtle AB, Griffiths P, Nakas CT, Simon M
Longitudinal Study of the Variation in Patient Turnover and Patient-to-Nurse Ratio: Descriptive Analysis of a Swiss University Hospital
J Med Internet Res 2020;22(4):e15554
URL: https://www.jmir.org/2020/4/e15554
doi:10.2196/15554
PMID:32238331

Original Paper

Leveraging Eye Tracking to Prioritize Relevant Medical Record Data: Comparative Machine Learning Study

Andrew J King[1,2], PhD; Gregory F Cooper[1,3], MD, PhD; Gilles Clermont[2], MSc, MD; Harry Hochheiser[1,3], PhD; Milos Hauskrecht[4], PhD; Dean F Sittig[5], PhD; Shyam Visweswaran[1,3], MD, PhD

[1]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States

[2]Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[3]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, United States

[4]Department of Computer Science, University of Pittsburgh, Pittsburgh, PA, United States

[5]Department of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Shyam Visweswaran, MD, PhD
Department of Biomedical Informatics
University of Pittsburgh
The Offices at Baum
5607 Baum Blvd., Suite 523
Pittsburgh, PA
United States
Phone: 1 412 648 7119
Email: shv3@pitt.edu

Abstract

Background: Electronic medical record (EMR) systems capture large amounts of data per patient and present that data to physicians with little prioritization. Without prioritization, physicians must mentally identify and collate relevant data, an activity that can lead to cognitive overload. To mitigate cognitive overload, a Learning EMR (LEMR) system prioritizes the display of relevant medical record data. Relevant data are those that are pertinent to a context—defined as the combination of the user, clinical task, and patient case. To determine which data are relevant in a specific context, a LEMR system uses supervised machine learning models of physician information-seeking behavior. Since obtaining information-seeking behavior data via manual annotation is slow and expensive, automatic methods for capturing such data are needed.

Objective: The goal of the research was to propose and evaluate eye tracking as a high-throughput method to automatically acquire physician information-seeking behavior useful for training models for a LEMR system.

Methods: Critical care medicine physicians reviewed intensive care unit patient cases in an EMR interface developed for the study. Participants manually identified patient data that were relevant in the context of a clinical task: preparing a patient summary to present at morning rounds. We used eye tracking to capture each physician's gaze dwell time on each data item (eg, blood glucose measurements). Manual annotations and gaze dwell times were used to define target variables for developing supervised machine learning models of physician information-seeking behavior. We compared the performance of manual selection and gaze-derived models on an independent set of patient cases.

Results: A total of 68 pairs of manual selection and gaze-derived machine learning models were developed from training data and evaluated on an independent evaluation data set. A paired Wilcoxon signed-rank test showed similar performance of manual selection and gaze-derived models on area under the receiver operating characteristic curve (P=.40).

Conclusions: We used eye tracking to automatically capture physician information-seeking behavior and used it to train models for a LEMR system. The models that were trained using eye tracking performed like models that were trained using manual annotations. These results support further development of eye tracking as a high-throughput method for training clinical decision support systems that prioritize the display of relevant medical record data.

(J Med Internet Res 2020;22(4):e15876) doi:10.2196/15876

KEYWORDS

electronic medical record system; eye tracking; machine learning; intensive care unit; information-seeking behavior

Introduction

Background

Electronic medical record (EMR) systems capture considerable amounts of patient data, especially in data-rich care settings like intensive care units (ICUs) [1]. The large amount of data per patient necessitates that the user interfaces of EMR systems help users rapidly understand the medical state of each patient [2]. However, patient data in EMR systems are typically presented to physicians with little prioritization, leading to the risk of cognitive overload, which in turn may lead to reductions in physician ability to identify important patient data for clinical assessment and management and increases in risks of medical error [2]. One approach to overcoming this drawback envisions an adaptive EMR system that draws the physician's attention to the right data, for the right patient, at the right time, and in the right way [3]. As a step toward this goal, we developed a Learning EMR (LEMR, pronounced lemur) system that uses physician information-seeking behavior to prioritize the display of relevant medical record data [4]. Our system relies on supervised machine learning models that predict which patient data are likely to be useful in a specific context, where the context refers to a particular combination of clinical user, clinical task, and patient case [5]. The models will be trained on thousands of observations of patient data items that different physicians have sought as relevant across a wide range of patient cases and clinical tasks.

The acquisition of training data is a critical barrier to the development of a LEMR system. Some observations of physician information-seeking behavior are recorded in commercial EMR systems through the capture of mouse clicks, but such data on user behavior are often captured with insufficient resolution for training a LEMR system. Another method for collecting behavior data is through manual annotation of relevant data by clinical users. In a prior study, we collected manual annotations from physicians who reviewed patient cases and identified which data were relevant for a simulated clinical task [5]. With this data, we developed machine learning models that enable the LEMR system to identify likely relevant patient data in similar clinical contexts. However, manual annotation is disruptive and time-consuming, and thus, it limits the amount of training data that can be collected. To have a broad coverage of clinical contexts, larger amounts of training data are needed for model construction. Moreover, while manual annotation is possible in a research setting, it is infeasible in routine clinical practice where it would dramatically increase the clinical workload. Eye tracking offers a method that can potentially capture information-seeking behavior unobtrusively as a byproduct of care delivery [6]. In this paper, we investigated the use of eye tracking as an automated high-throughput method to capture physician information-seeking behavior.

Related Work

The use of eye tracking in studies of EMR systems has focused on understanding users and their interactions with systems. Investigators have used eye tracking to understand clinical reasoning [7], discern information search patterns [8], measure cognitive loads while performing tasks in the EMR system [9], evaluate usability [10], and measure time use [11]. Moacdieh et al [12] used eye tracking to demonstrate that display of irrelevant data increases cognitive workload, and Gold et al [13] showed that a commercial eye-tracking device can be used to assess an EMR system's usability.

We are not aware of any studies that have used eye tracking to train machine learning models of information-seeking behavior in an EMR system. In previous work, we established the feasibility of using a low-cost eye tracker to identify items of interest on an EMR display [14]. We conjectured that the resulting eye-tracking data can be used to infer which medical record data are viewed by physicians. Such data can then be used to train machine learning models that the LEMR system would use to identify medical record data in a future patient case that are likely to be relevant to the user. Moreover, the availability of inexpensive, portable eye-tracking devices could make broad deployment of such systems feasible.

In this paper, we propose and evaluate a novel approach to using eye tracking to capture information-seeking behavior that can subsequently be used to build models that identify relevant patient data in a LEMR system. We compare the performance of supervised machine learning models trained on eye tracking–derived annotations with that of models trained on manually obtained annotations. We hypothesize that eye tracking is an effective, high-throughput alternative to manually annotating training data for the LEMR system.

Learning Electronic Medical Record System

A LEMR system prioritizes the display of medical record data based on predicted relevance [4]. To learn relevance, our prototype LEMR system enables the collection of physician information-seeking behavior. We used two collection methods in this study: (1) manual selection, where the user annotates relevant medical record data by clicking checkboxes displayed on the user interface, and (2) gaze-derived, where an eye-tracking device captures eye gaze patterns while the user is reviewing a patient's medical record. Figure 1 shows the LEMR system's user interface on an eye tracking–equipped computer monitor. The software for the LEMR user interface is available on GitHub [15].

Figure 1. A computer monitor displaying the Learning Electronic Medical Record user interface with an eye-tracking device mounted at the bottom edge of the monitor. The interface temporally displays patient medical record data in four scrollable panels: (from left to right) panel 1 contains vital signs, ventilator settings, and intake and output; panel 2 contains medication administrations; panel 3 contains laboratory test results; and panel 4 contains free-text notes and reports. The remote eye-tracking device is magnetically attached to the bottom edge of the monitor and connected to the computer via a universal serial bus cable (screenshot is of a deidentified patient case).

Methods

Eye Tracking

We used the Tobii EyeX (Tobii Gaming) eye-tracking device, an inexpensive portable eye-tracking device and software package marketed for computer gaming and virtual reality applications [16] (as of publication, EyeX's successor product is called Tobii Eye Tracker 4C). The EyeX device samples eye gaze coordinates on the monitor at approximately 60 Hz. To operate the device and read the stream of gaze data, we used open-source Python bindings for the Tobii Gaze Software Development Kit [17].

A recent evaluation of the Tobii EyeX found that for many research applications temporal and spatial resolutions are modest, precision is moderate, and sampling frequency is low [18]. Despite the EyeX's limitations, the authors noted that it is adequate for applications that do not require more than monitoring of simple eye movements. For example, the EyeX would not be suitable for measuring every saccade and fixation of a person reading a paragraph of text but is suitable for determining if a region of the monitor was gazed upon. In a prior study, we found the EyeX to be comparable in performance to a more expensive, research-grade eye-tracking device for capturing information-seeking behavior while using the LEMR user interface (the difference in errors of the two devices was less than a predefined noninferior margin of 11 pixels at the 95% confidence interval) [14].

The eye-tracking device provides time-stamped gaze position coordinates on the computer monitor but does not provide concomitant information on data displayed on the monitor. Therefore, to use eye tracking to automatically capture physician information-seeking behavior, the stream of gaze position coordinates must be mapped onto elements of the user interface. Because the LEMR user interface is dynamic (it changes as users scroll through the medical record data), we separately record a stream of interface element locations on the monitor that can be mapped to gaze position coordinates obtained from eye tracking. Each interface element contains one medical record data item (eg, the time series of blood glucose measurements). Using time-stamps in the two data streams, gaze position coordinates are translated to medical record data items. A data item such as blood glucose measurements with a gaze dwell time that is longer than a set threshold (eg, 250 milliseconds) is designated as a positive training sample for that data item. Positive and negative training samples with respect to a particular data item denote that the data item was viewed and was not viewed by a user, respectively. We developed and

implemented an algorithm that rapidly processed concomitantly acquired gaze position coordinates and interface element locations. This algorithm is described in our prior work [14]. The scripts for eye tracking are available on GitHub [19].

Data Collection

Overview

For machine learning, we created two training data sets of physician information-seeking behavior. One data set consisted of targets obtained from manual user selection; the second consisted of targets obtained from gaze position coordinates captured by eye tracking. We created an independent evaluation data set using manual annotation. Reviewers were presented with both structured EMR data (eg, demographics, diagnosis, vital sign measurements, ventilator settings, laboratory test results, medication administrations, procedures, microbiology culture results, and intake and output) and free text data (eg, admission notes, radiology reports) when reviewing a case. However, only structured EMR data were used for machine learning.

Reviewers

We recruited critical care medicine physicians from the University of Pittsburgh's Department of Critical Care Medicine to review and annotate patient cases: 11 physicians participated in the training phase and 12 physicians participated in the evaluation phase. Five reviewers who participated in the training phase also participated in the evaluation phase. Details of characteristics of the reviewers are shown in Table 1. Every reviewer used the LEMR system (shown in Figure 1). All reviewers were able to familiarize themselves with the interface by reviewing several practice cases before data collection began.

Table 1. Characteristics of reviewers.

Phase of study	Number of reviewers	Time in years since medical school graduation, mean (SD)	Time in years spent in ICU[a], mean (SD)	Weeks per year spent rounding in the ICU, mean (SD)
Training	11	5.3 (3.0-10.0)	1.8 (0.3-7.0)	34 (26-42)
Evaluation	12	5.4 (3.0-11.0)	1.7 (0.6-4.0)	36 (28-44)

[a]ICU: intensive care unit.

Simulated Clinical Task

Reviewers were asked to use the LEMR interface to conduct prerounding, which involves the review of a patient's medical record and identification of relevant data for a summary presentation at morning rounds. Each patient case was loaded into the LEMR system and shown to a physician reviewer (an example patient case is shown in Figure 1). The physician reviewed the case and completed three tasks: familiarization, preparation, and selection.

In the familiarization task, the physician was shown the patient's medical record data from hospital admission until 8:00 am on a random ICU day during the ICU stay between day 2 and the day before discharge from the ICU. The physician was instructed to review the data and become familiar with the case as if it were one of their patients.

In the preparation task, the physician was shown an additional 24 hours of the patient's medical record data and instructed to review the data and prepare to present the case during morning rounds. During this task, we used eye tracking to record the physician's gaze position coordinates from which we inferred the physician's information-seeking behavior.

In the selection task, a checkbox was added to each user interface element containing a medical record data item (such as blood glucose measurements). The physician manually annotated (by clicking on checkboxes) data items they considered to be relevant for the task of prerounding.

Patient Cases

For the creation of training data sets, 178 patient cases were randomly selected from patients who were admitted to an ICU between June 1, 2010, and April 30, 2012, at the University of Pittsburgh Medical Center (PA). For the creation of the evaluation data set, 18 patient cases were randomly selected from patients admitted to an ICU between June 1, 2012, and December 31, 2012. Each selected patient had a diagnosis of either acute kidney failure (AKF; ICD-9 584.9 or 584.5) or acute respiratory failure (ARF; ICD-9 518.81). The number of patients with each diagnosis was roughly equal in the training and evaluation data sets. EMR data for the patients were extracted from a research database and deidentified to remove all unique identifiers except for dates and times related to events.

Annotations

We created two training data sets that contained the same patient cases but differed in the construction of the targets. In the manual selection training data set, targets were derived from the checkbox clicks recorded during the selection task; in the gaze-derived training data set, targets were derived from eye tracking during the preparation task. From the 178 patient cases reviewed, 44 cases were discarded either because the eye-tracking data were incomplete (which occurred when a reviewer's head was not within the trackable range of the eye-tracking device) or the selection task was skipped inadvertently. Thus, the two training data sets consisted of the same 134 patient cases. Note that each patient case was reviewed by a single reviewer to maximize the number of patient cases that could be reviewed with a limited number of physicians. For the evaluation data set, 18 patient cases were each reviewed by 4 physicians for a total of 72 patient cases. Four cases were discarded because a reviewer inadvertently skipped the annotation task, and thus the evaluation data set contained 68 patient cases with manually annotated targets.

Machine Learning

Problem Description

The LEMR system prioritizes the display of relevant medical record data. To determine which data are relevant in a specific context (a combination of user, task, and patient), the LEMR system uses supervised machine learning models. For this study, we focused on the information-seeking differences among patient cases when the user and task were constant (all users were critical care physicians who performed the same task of prerounding). The training data sets consisted of a large number of predictor and target variables.

Predictor Variables and Feature Construction

The patient data we used as predictor variables included simple atemporal variables such as demographics (6) and diagnosis (1) and complex variables representing time series, including vital sign measurements (14), ventilator settings (9), laboratory test results (814), medication administrations (1207), procedures (394), microbiology culture results (10), and fluid intake and output (7). We derived a fixed set of features from each complex variable so that standard machine learning methods could be applied readily. For example, for a time-stamped sequence of serum glucose levels, we generated 35 features that include the first glucose value during the ICU stay, most recent value, highest and lowest values until the current time, difference between the most recent two values, and 30 other features [20]. Thus, we generated 35 features for each laboratory test and each vital sign, 31 features for each ventilator setting, 9 for each medication administration, 4 for each culture and each procedure, and 2 for each intake and output variable. In summary, a patient case was represented by a fixed-size vector of 13,596 features that summarized the clinical evolution of the patient's condition from the time of admission to the ICU to the current time.

Target Variables

We represent physician information-seeking behavior with a set of binary target variables. A target variable is created for any medical record data item that a physician could seek as relevant. For a specific patient case, a data item's corresponding target variable indicates if the item was sought as relevant or not (eg, the glucose target variable is positive if the physician sought the serum glucose levels). We derived two sets of targets; one set of targets was derived from manual selections and another from gaze position coordinates. In the manual selection training data set, a target variable, such as glucose target, was assigned the value positive if the physician selected the checkbox that was displayed with the data item; and a target was assigned the value negative if the associated checkbox was not selected. In the gaze-derived training data set, a target was assigned the value positive if the physician gazed at that item for at least 250 milliseconds and was assigned the value negative if the physician gazed at it for less than 250 milliseconds. (This threshold corresponds to the average fixation time while a person is reading [21]). In a patient case where a data item was not available (eg, serum glucose levels were not measured), the corresponding target was deemed to be absent, and the case was excluded from the training data used to train a model to predict the relevance of that data item.

Preprocessing of Training Data

We applied several preprocessing steps to the training data sets. A feature was removed if it had the same value in all cases. If two or more features had identical values for every case, only one of those features was retained. For example, the binary feature "ever measured" often had the same values across all cases for one or more laboratory tests that are part of a single panel; this occurred because such tests are ordered together. Missing predictor values were imputed using two different methods. In the first method, features were imputed with the median (nominal features were imputed with the mode). In the second method, continuous features were imputed with linear regression, and discrete features were imputed with logistic regression. The two imputation methods resulted in two distinct data sets for the manual selection and gaze-derived targets, respectively (a median-imputed data set and a regression-imputed data set).

To reduce feature dimensionality, we applied a feature selection algorithm that independently considered each predictor variable's set of constructed features. If a predictor variable's constructed feature set was predictive of a target (ie, cross-validated area under the receiver operating characteristic curve [AUROC] was greater than 0.55), the feature set was kept in the training set for the model corresponding to that target variable. Otherwise, the feature set was discarded from the training set for the model corresponding to that target variable.

Model Training

We applied three classification methods (scikit-learn implementations of L2-penalized logistic regression, support vector classifier, and random forest classifier [22]) to the manual selection and gaze-derived training data sets to develop two categories of models. Within each category, we used cross-validation to select the best performing combination of imputation method and classification method for each target. The best performing combination was then used to train a model from the full training data set. The scripts for feature construction [23], imputation of missing values [24], feature selection [25], and for training and evaluating models [26] are available on GitHub.

Results

Table 2 provides a summary of the data sets, models trained, and models evaluated. The gaze-derived data set had more positive targets on average, resulting in more targets (ie, data items) having enough training data (>3 positive samples) for model construction. In total, 87 manual selection and 115 gaze-derived models were trained. For 68 targets, models were derived from both data sets. These 68 pairs of models were used in the evaluation, and by coincidence the evaluation data set consisted of 68 patient cases.

Table 2. Summary of data sets, models trained, and models evaluated. Each model predicts if a single electronic medical record data item is relevant.

Data set	Counts		
	Number of patient cases	Number of models trained	Number of models evaluated (using the manual-selection evaluation data set)
Manual selection training	134	87	68
Gaze-derived training	134	115	68
Manual selection evaluation	68	—	—

Table 3 provides a summary of the models selected for evaluation. Based on the distribution of models (ie, the number of models columns), random forests and median imputation tended to have higher cross-fold AUROC values on the training set than the other machine learning and imputation methods, respectively.

The AUROCs of the 68 pairs of models applied to the evaluation data set are plotted in Figure 2 and shown in Table 4. In Figure 2, points below the diagonal line indicate that manual selection models perform better, and points above the diagonal line indicate that gaze-derived models perform better. The best performing pair of models was for the target alanine aminotransferase (AUROCs of 0.97 for manual selection and 0.90 for gaze-derived), and the worst performing pair was for the target mean corpuscular volume (AUROCs of 0.27 for manual selection and 0.14 for gaze-derived). Statistical differences between each pair of models were calculated from the 95% confidence interval of the average difference in AUROC values using bootstrapping. In 14 instances, the manual selection models performed statistically significantly better, and in 8 instances, the gaze-derived model performed statistically significantly better (see the red triangles in Figure 2 and footnote in Table 4). In the remaining 46 instances, there was no statistically significant difference in AUROCs between the two models (α=.05).

On the Wilcoxon signed-rank test (R software, wilcox.test function), the AUROC values of the manual selection models were similar to the values of the gaze-derived models (*P*=.40). The Wilcoxon signed-rank test is a nonparametric statistical test for comparing two groups of continuous measures.

Table 3. Summary of machine learning methods, imputation methods, number of models of each combination of machine learning and imputation methods, and number of features per model.

Machine learning and imputation method	Manual selection		Gaze-derived	
	Number of models	Features per model mean	Number of models	Features per model mean
Logistic regression				
Median	18	207.4	24	176.3
Regression	10	659.2	15	304.9
Support vector classifier				
Median	18	2366.3	9	2108.9
Regression	0	—	0	—
Random forests				
Median	27	336.0	37	341.9
Regression	14	806.1	30	613.8

Figure 2. Performance of models used to predict the relevance of targets. The x-axis value and y-axis value of each point indicate the area under the receiver operating characteristic curve (AUROC) values of a pair of manual selection and gaze-derived models, respectively. The vertical and horizontal lines indicate 95% confidence intervals of the AUROC values. The diagonal line indicates equal performance between manual selection and gaze-derived models. Red triangles indicate model pairs where the AUROC value of one model is significantly different than that of the other model (α=.05).

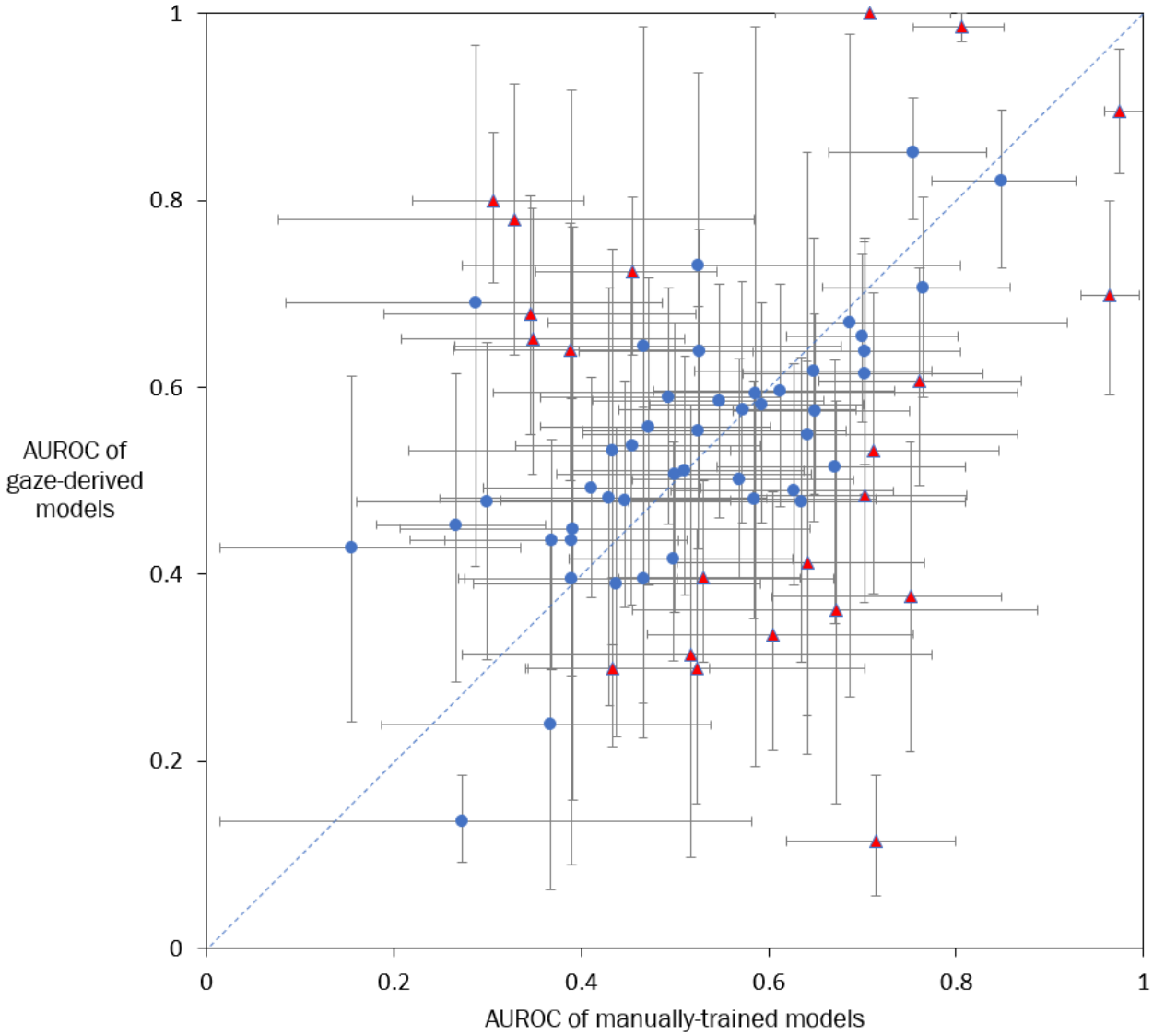


Table 4. Area under the receiver operating characteristic curve values (with 95% confidence intervals) on the evaluation data set of manual selection and gaze-derived models for predicting relevance of targets. Rows are sorted by manual selection performance.

Target	Number of positive samples in evaluation data set	Manual selection models	Gaze-derived models
Alanine aminotransferase	10	0.97[a] (1.00, 0.96)	0.90 (0.96, 0.83)
Aspartate aminotransferase	10	0.96[a] (1.00, 0.93)	0.70 (0.80, 0.59)
Norepinephrine	13	0.85 (0.93, 0.77)	0.82 (0.90, 0.73)
Levothyroxine	1	0.81 (0.85, 0.75)	0.99[a] (1.00, 0.97)
Fraction of inspired oxygen	27	0.77 (0.86, 0.66)	0.71 (0.80, 0.59)
Vancomycin	17	0.76[a] (0.87, 0.65)	0.61 (0.73, 0.50)
Total bilirubin	12	0.75[a] (0.85, 0.60)	0.38 (0.54, 0.21)
Bicarbonate, arterial	1	0.75 (0.83, 0.66)	0.85 (0.91, 0.78)
Aspirin	3	0.72[a] (0.80, 0.62)	0.12 (0.19, 0.06)
pH	6	0.71[a] (0.85, 0.56)	0.53 (0.70, 0.38)
Troponin	3	0.71 (0.79, 0.61)	1.00[a] (1.00, 1.00)
Lactate	11	0.70 (0.83, 0.57)	0.62 (0.76, 0.49)
Temperature	42	0.70[a] (0.81, 0.58)	0.48 (0.61, 0.37)
Lorazepam	3	0.70 (0.80, 0.62)	0.65 (0.74, 0.56)
Ventilator mode	20	0.70 (0.80, 0.58)	0.64 (0.76, 0.52)
Dextrose in water	3	0.69 (0.92, 0.36)	0.67 (0.98, 0.27)
Partial pressure of oxygen	6	0.67[a] (0.89, 0.45)	0.36 (0.59, 0.15)
Bilirubin direct	9	0.67 (0.81, 0.55)	0.52 (0.63, 0.35)
Heparin	16	0.65 (0.77, 0.52)	0.62 (0.76, 0.46)
Ionized calcium	1	0.65 (0.75, 0.56)	0.58 (0.68, 0.49)
Propofol	6	0.64 (0.87, 0.40)	0.55 (0.85, 0.25)
Piperacillin-tazobactam	15	0.64 (0.81, 0.45)	0.48 (0.63, 0.31)
Famotidine	6	0.64[a] (0.77, 0.50)	0.41 (0.63, 0.21)
Potassium	26	0.63 (0.73, 0.50)	0.49 (0.63, 0.39)
White blood cells	50	0.61 (0.74, 0.48)	0.60 (0.71, 0.47)
Bands	15	0.60[a] (0.75, 0.47)	0.34 (0.49, 0.21)
Vancomycin trough	4	0.59 (0.87, 0.31)	0.59 (0.99, 0.19)
Blood urea nitrogen	30	0.59 (0.70, 0.47)	0.58 (0.69, 0.46)
Intravenous base solution	19	0.58 (0.72, 0.45)	0.48 (0.61, 0.35)
Oxygen saturation	29	0.57 (0.69, 0.46)	0.50 (0.63, 0.40)
Intake and output	43	0.57 (0.69, 0.44)	0.58 (0.71, 0.46)
Heart rate	42	0.55 (0.66, 0.41)	0.59 (0.71, 0.46)
Prothrombin time	4	0.53 (0.81, 0.27)	0.73 (0.94, 0.48)
Pantoprazole	9	0.53 (0.71, 0.40)	0.64 (0.77, 0.48)
Bicarbonate, venous	15	0.53 (0.68, 0.40)	0.55 (0.69, 0.43)
Glomerular filtration rate	1	0.53[a] (0.63, 0.44)	0.40 (0.50, 0.31)
Hydrocortisone	5	0.52[a] (0.77, 0.27)	0.31 (0.58, 0.10)
Metoprolol	7	0.52[a] (0.70, 0.34)	0.30 (0.45, 0.16)

Target	Number of positive samples in evaluation data set	Manual selection models	Gaze-derived models
Glucose	21	0.51 (0.64, 0.39)	0.51 (0.63, 0.38)
Insulin	13	0.50 (0.65, 0.37)	0.51 (0.67, 0.36)
Platelets	30	0.50 (0.63, 0.39)	0.42 (0.54, 0.31)
Respiratory rate	17	0.49 (0.61, 0.36)	0.59 (0.71, 0.45)
Central venous pressure	2	0.47 (0.68, 0.27)	0.64 (0.99, 0.26)
Creatinine	59	0.47 (0.67, 0.28)	0.40 (0.58, 0.23)
Magnesium	9	0.47 (0.60, 0.36)	0.56 (0.72, 0.39)
Albumin	1	0.46 (0.55, 0.35)	0.72[a] (0.80, 0.63)
Phosphate	12	0.45 (0.59, 0.33)	0.54 (0.72, 0.37)
Sodium	41	0.45 (0.56, 0.31)	0.48 (0.61, 0.36)
Chloride	12	0.44 (0.59, 0.29)	0.39 (0.56, 0.23)
Blood pressure	61	0.43 (0.64, 0.22)	0.53 (0.75, 0.33)
Chlorhexidine topical	6	0.43 (0.60, 0.25)	0.48 (0.71, 0.26)
Partial thromboplastin time	1	0.43[a] (0.54, 0.34)	0.30 (0.40, 0.22)
Hemoglobin	31	0.41 (0.53, 0.30)	0.49 (0.61, 0.38)
Neutrophils	4	0.39 (0.64, 0.21)	0.45 (0.77, 0.16)
Fentanyl	14	0.39 (0.53, 0.26)	0.64[a] (0.78, 0.50)
Metronidazole	13	0.39 (0.51, 0.22)	0.44 (0.59, 0.29)
Furosemide	3	0.39 (0.50, 0.27)	0.40 (0.92, 0.09)
Calcium	2	0.37 (0.54, 0.19)	0.24 (0.43, 0.06)
Sodium chloride	28	0.37 (0.50, 0.25)	0.44 (0.54, 0.30)
International normalized ratio	11	0.35 (0.52, 0.19)	0.68[a] (0.81, 0.55)
Midazolam	2	0.35 (0.51, 0.21)	0.65[a] (0.79, 0.51)
Alkaline phosphatase	5	0.33 (0.58, 0.08)	0.78[a] (0.93, 0.64)
Acetaminophen	1	0.31 (0.40, 0.22)	0.80[a] (0.87, 0.71)
Ventilator tube status	9	0.30 (0.45, 0.16)	0.48 (0.65, 0.31)
Albuterol-ipratropium	5	0.29 (0.49, 0.09)	0.69 (0.97, 0.41)
Mean corpuscular volume	2	0.27 (0.58, 0.02)	0.14 (0.19, 0.09)
Partial pressure of carbon dioxide	4	0.27 (0.36, 0.18)	0.45 (0.61, 0.29)
Ventilator status	2	0.16 (0.34, 0.02)	0.43 (0.61, 0.24)

[a]Indicates statistically significant difference at α=.05.

Discussion

Principal Findings

Current EMR systems capture large amounts of patient data; however, they generally do not capitalize on the opportunity to prioritize display of data in a relevant and context-sensitive fashion. We proposed an intelligent method to identify, display, and focus user attention on relevant medical record data. The rate-limiting step while developing a prototype LEMR system has been the collection of training data. Instead of relying on manual annotations, we proposed using eye tracking as an automated, high-throughput method for capturing physician information-seeking behavior. LEMR models trained on gaze-derived target data performed similarly to models trained on manual selection target data.

Eye Tracking in Health Care: Pros and Cons

Eye tracking is an alluring method for collecting LEMR training data. First, the availability of inexpensive eye tracking devices makes broad implementation of the devices feasible. With devices installed on many of a hospital's computer monitors, vast amounts of training data could be recorded. Locally and continuously recorded data could be used to train a LEMR system that is specific to that location's standards of practice and adaptive to any changes in practice patterns that occur over

time. Second, eye tracking is less disruptive to workflow than requesting that physicians rate or annotate the EMR data they are viewing. Finally, eye tracking is more advantageous than manual annotation because it measures what physicians view rather than relying on what they perceive to be relevant, which, if collected retrospectively, could be more affected by recall and attention biases.

Despite its advantages, eye tracking has some drawbacks when used to capture information-seeking behavior. First, to record behavior, the user's eyes must be within the trackable range (ie, tracking box) of the eye-tracking device. Incorrect positioning led to some data loss during this study and could be exaggerated in a clinical setting where users are often standing or talking with colleagues. Second, to map physician gaze coordinates onto the EMR data shown on the monitor, a stream of the information layout is also needed. We instrumented the EMR interface used in this study to store a stream of the information layout. Adding the same instrumentation to a vendor-based EMR system may be technically feasible but administratively challenging. We are exploring this possibility and other alternatives, such as continuously recording what is on the monitor and parsing the recording using optical character recognition. Finally, we are assuming that a physician views the information they perceive to be relevant, which may not always be the case. More work is needed to determine in which situations gaze patterns are a reasonable approximation for the information physicians seek.

Models of Information-Seeking Behavior

Performance of the models was mixed and ranged from those with AUROCs above 0.90 to others with AUROCs below 0.50. The mediocre performance of some of the models was likely due to the relatively small sizes of the training data sets and the small number of positive targets. In previous work, we demonstrated that model performance improved as the number of training cases increased, and it is likely that larger training data sets will improve model performance [5]. Moreover, larger training data sets will increase target coverage (ie, the number of targets modeled).

The context of this study was limited to a single clinical task (ICU physicians preparing to present at morning rounds) and two clinical conditions (AKF and ARF). Even in this limited context and in a laboratory setting, there were variations in the information-seeking behaviors of the reviewers leading to differences in annotations for the same patient cases (eg, cases in the evaluation data set were each evaluated by four physicians). These differences partly stem from length of experience and depth of clinical knowledge, with some reviewers in fellowship training with less than two years of experience in critical care and others who are attendings with extensive experience. In active clinical environments, other factors such as urgency, workload, and interruptions will potentially influence the information-seeking behavior, leading to even more variation. In this study, the models were trained using annotations from all reviewers who were weighted equally. However, it may be desirable to either use data from only experienced physicians or weight their data more heavily for training models. Further research is needed to better characterize variability among the information-seeking behavior of physicians.

While prioritizing data can reduce cognitive load, with imperfect model performance, any reduction in cognitive load could be offset by the need to check unprioritized data to ensure that important information is not missed. With improved model performance, a LEMR system could reach high levels of accuracy, leading to a net reduction in cognitive load.

Limitations and Future Work

We used manual selection as the gold standard in the evaluation data set. This choice provides an advantage to models trained on manual selection training data. Despite the advantage, our results show eye tracking to be a promising, higher-throughput alternative.

The patients included in this study were chosen because they had either AKF or ARF. Future studies are needed to evaluate the effectiveness of machine learning methods when applied to patients with a wider range of clinical problems, comorbidities, and clinical tasks.

The eye-tracking device we used has several limitations, including reduced accuracy at the edges of the monitor and a limited tracking box (the three-dimensional space where a user's head can be positioned while still capturing coordinates of the user's eye gaze). To address the device's limitations, we designed the LEMR user interface to have adequate separation between data items. Further research is needed to determine the performance and limitations of inexpensive eye-tracking devices in an active clinical setting where users are likely to be more active and frequently distracted compared to a laboratory setting. Using LEMR models in conjunction with a commercial EMR poses additional challenges, such as obtaining patient data in real time. One promising approach to explore in the future is to deploy LEMR as a SMART (Substitutable Medical Applications, Reusable Technologies) on FHIR (Fast Healthcare Interoperability Resource) application [27].

Conclusions

We proposed and evaluated using eye tracking as a novel method to train a LEMR system that prioritizes the display of relevant medical record data. Our results support eye tracking as being a viable method for automatically capturing physician information-seeking behavior. Further studies are needed to evaluate the effectiveness of eye tracking in the clinical environment and to advance it as a practical approach for acquiring large amounts of training data for a LEMR system.

Acknowledgments

The research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under award numbers T15 LM007059 and R01 LM012095 and by the National Institute of General Medical Sciences of the

National Institutes of Health under award number R01 GM088224. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank Ms Yu Zhang for assistance with data preprocessing.

Conflicts of Interest

None declared.

References

1. Manor-Shulman O, Beyene J, Frndova H, Parshuram CS. Quantifying the volume of documented clinical information in critical illness. J Crit Care 2008 Jun;23(2):245-250. [doi: 10.1016/j.jcrc.2007.06.003] [Medline: 18538218]
2. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. JAMA Intern Med 2013 Apr 22;173(8):702-704 [FREE Full text] [doi: 10.1001/2013.jamainternmed.61] [Medline: 23460235]
3. Levick D, Osheroff J. A clinical decision support implementation guide: practical considerations. In: Greenes RA, editor. Clinical Decision Support: The Road to Broad Adoption, Second Edition. Cambridge: Academic Press; 2014:689-709.
4. King AJ, Cooper GF, Hochheiser H, Clermont G, Visweswaran S. Development and preliminary evaluation of a prototype of a learning electronic medical record system. AMIA Annu Symp Proc 2015;2015:1967-1975 [FREE Full text] [Medline: 26958296]
5. King AJ, Cooper GF, Hochheiser H, Clermont G, Hauskrecht M, Visweswaran S. Using machine learning to predict the information seeking behavior of clinicians using an electronic medical record system. AMIA Annu Symp Proc 2018;2018:673-682 [FREE Full text] [Medline: 30815109]
6. Institute of Medicine. In: Smith M, Saunders R, Stuckhardt L, McGinnis MJ, editors. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington: National Academies Press; 2013.
7. Blondon K, Wipfli R, Lovis C. Use of eye-tracking technology in clinical reasoning: a systematic review. Stud Health Technol Inform 2015;210:90-94. [Medline: 25991108]
8. Doberne JW, He Z, Mohan V, Gold JA, Marquard J, Chiang MF. Using high-fidelity simulation and eye tracking to characterize EHR workflow patterns among hospital physicians. AMIA Annu Symp Proc 2015:1881-1889 [FREE Full text] [Medline: 26958287]
9. Calvitti A, Weibel N, Hochheiser H, Liu L, Weir C, Ashfaq S, et al. Can eye tracking and electronic health record mouse activity tell us when clinicians are overloaded? Hum Factors Quart. Hum Factors Q 2014;8:1.
10. Asan O, Yang Y. Using eye trackers for usability evaluation of health information technology: a systematic literature review. JMIR Hum Factors 2015 Apr 14;2(1):e5 [FREE Full text] [doi: 10.2196/humanfactors.4062] [Medline: 27026079]
11. Rick S, Calvitti A, Agha Z, Weibel N. Eyes on the clinic: accelerating meaningful interface analysis through unobtrusive eye tracking. 2015 Presented at: 9th Int Conf Pervasive Comput Technol Healthcare; 2015 May 20; Istanbul p. 213-216. [doi: 10.4108/icst.pervasivehealth.2015.259276]
12. Moacdieh N, Sarter N. Clutter in electronic medical records: examining its performance and attentional costs using eye tracking. Hum Factors 2015 Jun;57(4):591-606. [doi: 10.1177/0018720814564594] [Medline: 25850110]
13. Gold JA, Stephenson LE, Gorsuch A, Parthasarathy K, Mohan V. Feasibility of utilizing a commercial eye tracker to assess electronic health record use during patient simulation. Health Informatics J 2016 Sep;22(3):744-757. [doi: 10.1177/1460458215590250] [Medline: 26142432]
14. King AJ, Hochheiser H, Visweswaran S, Clermont G, Cooper GF. Eye-tracking for clinical decision support: a method to capture automatically what physicians are viewing in the EMR. AMIA Jt Summits Transl Sci Proc 2017;2017:512-521 [FREE Full text] [Medline: 28815151]
15. King A, Hochheiser H, Cooper G, Visweswaran S. LEMR Interface. URL: https://github.com/ajk77/LEMRinterface [accessed 2020-01-26]
16. Specifications for EyeX. URL: https://help.tobii.com/hc/en-us/articles/212818309-Specifications-for-EyeX [accessed 2020-01-26]
17. Python bindings for Tobii Gaze SDK. 2015. URL: https://github.com/balancana/gazesdk [accessed 2020-01-26]
18. Gibaldi A, Vanegas M, Bex PJ, Maiello G. Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. Behav Res Methods 2017 Jun;49(3):923-946 [FREE Full text] [doi: 10.3758/s13428-016-0762-9] [Medline: 27401169]
19. King A, Visweswaran S, Hochheiser H, Cooper G. EyeBrowserPy. URL: https://github.com/ajk77/EyeBrowserPy [accessed 2020-01-26]
20. King AJ, Cooper GF, Clermont G, Hochheiser H, Hauskrecht M, Sittig DF, et al. Using machine learning to selectively highlight patient information. J Biomed Inform 2019 Dec;100(2):103327-103250. [doi: 10.1016/j.jbi.2019.103327] [Medline: 31676461]
21. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull 1998 Nov;124(3):372-422. [doi: 10.1037/0033-2909.124.3.372] [Medline: 9849112]

22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830. [Medline: 1000044560]
23. King A, Visweswaran S, Hochheiser H, Hauskrecht M, Cooper G. PatientPy. URL: https://github.com/ajk77/PatientPy [accessed 2020-01-26]
24. King A, Visweswaran S, Hochheiser H, Cooper G. RegressiveImputer. URL: https://github.com/ajk77/RegressiveImputer [accessed 2020-01-26]
25. King A, Hauskrecht M, Visweswaran S, Hochheiser H, Cooper G. PatientPyFeatureSelection. URL: https://github.com/ajk77/PatientPyFeatureSelection [accessed 2020-01-26]
26. King A, Cooper G, Clermont G, Hochheiser H, Hauskrecht M, Sittig D, et al. jmir-2020-king. URL: https://github.com/ajk77/jmir-2020-king [accessed 2020-01-26]
27. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Feb 17;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]

Abbreviations

AKF: acute kidney failure
ARF: acute respiratory failure
AUROC: area under the receiver operating characteristic curve
EMR: electronic medical record
FHIR: Fast Healthcare Interoperability Resource
ICU: intensive care unit
LEMR: Learning Electronic Medical Record
SMART: Substitutable Medical Applications, Reusable Technologies

Edited by G Eysenbach; submitted 14.08.19; peer-reviewed by W Hogan, A Davoudi, R Zhang; comments to author 24.09.19; revised version received 04.12.19; accepted 23.01.20; published 02.04.20.

Please cite as:
King AJ, Cooper GF, Clermont G, Hochheiser H, Hauskrecht M, Sittig DF, Visweswaran S
Leveraging Eye Tracking to Prioritize Relevant Medical Record Data: Comparative Machine Learning Study
J Med Internet Res 2020;22(4):e15876
URL: https://www.jmir.org/2020/4/e15876
doi:10.2196/15876
PMID:32238342

Original Paper

Adoption of an Electronic Patient Record Sharing Pilot Project: Cross-Sectional Survey

Jingxuan Wang[1], BSocSci, MPhil; Junjie Huang[1], MSc, MD; Clement Shek Kei Cheung[2], MB ChB, DCH; Wing Nam Wong[2], MBBS, MFSEM, DCH; Ngai Tseung Cheung[2], BSc (Med), MSc, MBBS; Martin CS Wong[1], MPH, MD

[1]The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, New Territories, Hong Kong

[2]Information Technology and Health Informatics Division, Hospital Authority, Hong Kong, China

Corresponding Author:
Martin CS Wong, MPH, MD
The Jockey Club School of Public Health and Primary Care
Faculty of Medicine
Chinese University of Hong Kong
New Territories
Hong Kong
Phone: 852 2252 8782
Email: wong_martin@cuhk.edu.hk

Abstract

Background: The Public Private Interface–Electronic Patient Record (PPI-ePR) system was implemented as a new electronic platform to facilitate collaboration between the public and private sectors in Hong Kong. However, its barriers to participate and benefits have not been comprehensively assessed.

Objective: This study aimed to evaluate the awareness, acceptance, perceived benefits, and obstacles to participation among private doctors and the general public.

Methods: From December 2012 to January 2013, 2435 telephone interviews were performed by trained interviewers to survey randomly selected patients who were enrolled or not enrolled in the PPI-ePR system. In addition, self-administered surveys were sent by postal mail to 4229 registered doctors in Hong Kong. The questionnaires for both patients and doctors contained questions on subjects' awareness, acceptance, and perceptions of the PPI-ePR, perceived benefits and obstacles of participating in the program, reasons for not using the system after enrolling, and perceived areas for service improvement of the system.

Results: More than 53.1% (266/501) of enrolled patients believed that the PPI-ePR system would improve health care quality by reducing duplicate tests and treatments, while more than 76.8% (314/409) of enrolled doctors emphasized timely access to patients' medical records as the biggest benefit of their enrollment. Among nonenrolled patients, unawareness of the project was the most popular obstacle to enrolling in the PPI-ePR system (483/1200, 40.3%). Regarding nonenrolled doctors, the complicated registration process hindered them from participating in the program the most (95/198, 48.0%). Television, newspaper, and magazine advertisements and medical profession newsletters or journals were suggested as the most effective means to encourage participation in the program among surveyed patients (1297/1701, 76.2%) and doctors (428/610, 70.2%), respectively. Lack of clinical indication requiring data extraction from other hospitals was the main reason for low level of PPI-ePR use.

Conclusions: This study comprehensively assessed the popularity, perceived benefits, and hindering factors of enrolling in the PPI-ePR system in Hong Kong. Low levels of awareness, few privacy concerns, and inactive use of the PPI-ePR system were among the key features for patients and physicians. Public promotions, simplified logistics, and a user-friendly online interface were suggested to improve the coverage and effectiveness of health information exchange between private and public health care sectors.

(J Med Internet Res 2020;22(4):e13761) doi:10.2196/13761

KEYWORDS

health information exchange; shared electronic health record; online platform; public-private partnership

Introduction

The electronic health record systems (EHRs) and electronic medical records (EMRs) have been widely discussed in Western societies [1-5]. Promotion of these systems is believed to facilitate the communication between doctors and patients, reduce health care costs, enhance medical efficacy, activate patients to join in their holistic care, and support patient self-management of health [6-10]. Assessment of factors affecting participation in and adoption of local electronic systems, however, has scarcely been done, especially in Asian societies [11-15].

The health care system in Hong Kong runs on a dual-track basis encompassing the public and private sectors. While the 44 public hospitals provide approximately 90% of hospital medical service and 29% of outpatient medical service, the private health care sector provides personalized choices and more accessible services to those who are willing and can afford to pay [16]. High mobility between hospitals was observed among both patients and doctors [17]. The Hospital Authority (HA) of Hong Kong, the statutory administrative body that manages all public hospitals and health institutes in Hong Kong, determined that the overburdened public hospitals had long waiting times and ambiguous rules and procedures on patient referral between private and public sectors [16]. Paper-based data exchange among hospitals and clinics required a cumbersome process and was subject to safety and quality issues in many cases. Australia, Malaysia, and many Western countries have developed hospital information management and exchange systems to facilitate cooperation between private and public health care sectors [18]. In Hong Kong, the Public-Private Interface-Electronic Patient Record (PPI-ePR) program was thus implemented in 2006 to promote a public-private partnership and enhance timely, large-scale, secure data exchange between health care providers in the two sectors [19]. Selected health records in the HA's electronic patient record (ePR) system were shared with other public and private health care organizations upon express consent of the patient. The information included patient identity card number, age, gender, diagnosis, procedures codes, discharge summaries, laboratory and radiology reports, medication orders, allergies, and future appointments. Details on the design and implementation of the PPI-ePR have been published elsewhere [17]. Mass media and social media promotions, campaigns, professional seminars, and financial incentives were adopted to implement the project.

As of March 12, 2016, the program covers all 44 public hospitals, 11 private hospitals, and 72 nongovernmental organizations providing health services in 403 residential centers or institutions. ePRs have been accessed more than 1,462,000 times [20]. However, 46% of enrolled patients and 23% of enrolled health care providers had not made use of the system to access patient records since enrollment [21]. Our study, therefore, aimed to investigate the acceptance, awareness, perception, and satisfaction toward the system, supporting further improvement of the eHR in Hong Kong.

Methods

Objectives

This project aimed to conduct surveys on the PPI-ePR with the following objectives: (1) evaluate the awareness, acceptance, and perceived benefits of the PPI-ePR among enrolled users; (2) study factors hindering the participation of private doctors and patients in the PPI-ePR; (3) assess the reasons for not using the system after having enrolled; and (4) collect residents' suggestions on facilitating public acceptance and use of the PPI-ePR. Our findings would provide direction for the design, development, and operation of the eHR system in Hong Kong.

Sampling

Regarding patients, we adopted simple random sampling to survey enrollees and nonenrollees of the PPI-ePR. Enrollees were randomly selected based on an enrollment list provided by the HA (n=246,000), while nonenrollees were selected from the Hong Kong telephone directory. Computer-generated numbers were used for subject recruitment. Sample sizes were calculated by the formula $n=4\times p(1-p)/[\text{precision}]^2$, where p=proportion of interests. We assumed 50% as the proportion of interests in all outcomes which would give the maximum sample size since the project is the first electronic system to integrate private and public health records in Hong Kong. The minimum sample size was 400 for enrolled patients, given an assumed precision level of .05. Among patients not previously enrolled in the PPI-ePR, a higher precision level was needed due to the heterogeneity of this patient group. Setting a precision level of .03, the minimum sample size was 1111 for nonenrolled patients.

Regarding doctor subjects, self-administered surveys were sent by postal mail to all 4229 registered private doctors (postal addresses were provided by the HA) in Hong Kong. Response rate of Hong Kong doctors in previous surveys was as low as 15% [17]. To maximize our sample, multiple postal mails were sent to doctors who did not respond in the first round. A total of 10,285 survey invitations were mailed.

Survey Instruments

Our survey was developed based on previously validated surveys and scales. The following constructs were involved in both patient and doctor surveys:

- Knowledge, attitude, and practice questions of enrolled and nonenrolled subjects. People's awareness and use level of the information-sharing platform were part of our focus
- Constructs of the health belief model were used to understand enroller and nonenroller behavior. Perceived barriers and benefits of participating or using the PPI-ePR program, self-efficacy among doctors, and cues to registration were the main sections of our questionnaires
- User satisfaction levels and doctor suggestions on improving the online platform were collected to inform the design and promotion of future territory-wide patient information sharing projects
- Sociodemographic information was collected for subgroup analysis

Drafted questionnaires were face-validated by a panel of epidemiologists, doctors, nursing professionals, public health practitioners, and academics and subsequently pilot-tested on 20 doctors and 20 patients. Revisions were made after the pilot testing to promote feasibility and item comprehensiveness. Patient and doctor surveys were available in both Chinese and English versions. Our study is nonanonymous. Participants were informed that all information presented would be at the aggregate level, which could not identify any individuals. Consent was sought verbally for phone interviews and was signed by the participants of postal surveys.

Statistical Analysis

All surveys were checked for their completeness and the presence of participant consent before data entry and analysis with SPSS Statistics version 18.0 (IBM Corp). As part of quality control, validity, quality, and accuracy were randomly checked for both doctor and patient data. Thereafter, items in both surveys were stratified according to the status of enrollment. Descriptive statistics including proportions, means, and standard deviations were presented for doctor users, doctor nonusers, patient users, and patient nonusers separately.

Results

Participant Characteristics

The response rates were 90.3% (501 completed surveys/555 telephone number dialed) and 73.4% (1191/1623) for enrolled and nonenrolled patients, respectively. Nonusers were on average younger than users of the PPI-ePR system. Compared with users, a higher percentage of nonusers had tertiary education (nonusers: 26.8% [322/1200] vs users: 18.4% [92/501]) or were a student (11.8% [59/501] vs 1.6% [19/1200]), housewife (17.4% [87/501] vs 15.0% [180/1200]) or full-time employee (43.3% [217/501] vs 33.2% [398/1200]). The median household incomes per month were HK $10,000 (US $1180) and HK $20,000 (US $2361) for enrolled and nonenrolled respondents, respectively (Multimedia Appendix 1).

A total of 10,285 postal invitations were sent to a list of all 4229 registered private doctors in Hong Kong. We received 610 completed postal surveys consisting of 409 enrolled doctors, 198 nonenrolled doctors, and three with unknown enrollment status. The response rate was 14.4% (610 completed surveys/4229 private doctors in Hong Kong as of August 2012). In general, the majority of enrolled doctors were aged 30 to 50 years, while the majority of nonenrolled doctors were aged 51 years or above, indicating a younger group of users compared with nonusers of the system. In addition, nonenrolled doctors were more likely to have longer practice experience after medical school graduation (51.0% [101/198] having more than 30 years) than enrolled doctors (29.1% [119/409] having more than 30 years). The percentage being in solo practice was higher among nonenrolled doctors (71/109, 65.4%) than enrolled doctors (227/409, 55.5%). More enrolled doctors were academy fellows than nonenrolled doctors (68.2% [135/198] versus 56.6% [56/99]; Multimedia Appendix 1).

Awareness, Acceptance, and Perceived Benefits and Obstacles of Using the Public-Private Interface–Electronic Patient Record

More than 99% (499/501, 99.6%) of enrolled patients were aware of the operation of the PPI-ePR system compared with only 26.2% (314/1200) among nonenrolled patients. Compared with nonenrolled patients, enrolled patients enjoyed more ways to discover the project, among which family doctor's recommendation (305/501, 60.9%) and seminars organized by the HA (85/501, 17.0%) were widely praised; yet for nonenrolled patients, TV and newspapers dominated their understanding toward the system. The sharp disparity of awareness between users and nonusers was not observed among doctors—74.7% (148/198) of the surveyed doctors were aware of the project although they had not enrolled. Peers in the health care sector and posters or leaflets from the HA were the most effective means to promote PPI-ePR among both enrolled and nonenrolled doctors.

To further understand subject enrollment and use of PPI-ePR, we asked enrolled patients and doctors about their perceived benefits of joining the system and asked nonenrolled patients and doctors about factors hindering their enrollment.

Improved health care quality, convenience, and economic incentives were among the benefits of the PPI-ePR system emphasized by enrolled patients (Table 1). Specifically, about half of project users perceived reduced repetition of health assessment and information provision (266/501, 53.1%), as well as easier access to doctors' recommendations (217/501, 43.3%). Improved health care quality was the most frequently reported benefit of the PPI-ePR system among enrolled doctors, followed by improved patient safety. Specifically, the system facilitated timely access to patient medical records (314/409, 76.8%), enhanced continuity of patient care (259/409, 63.3%), and smoothed delivery of health care services (249/409, 60.9%).

Table 1. Perceived benefits of the Public-Private Interface–Electronic Patient Record among users.

Perceived benefits among users	Persons, n (%)
Enrolled patients (n=501)	
Improved health care	
Reduce duplicated tests and treatments	266 (53.1)
Access doctor recommendation	217 (43.3)
Comprehensive medical records for better patient care	144 (28.7)
Fluent information flow between private and public health care sectors	100 (20.0)
Facilitate continuity of patient care	5 (1.0)
Convenience	
No need to bring medical reports	48 (9.6)
Economic incentive	
Souvenirs	7 (1.4)
Enrolled doctors (n=409)	
Improved health care quality	
Timely access to patient medical records in the Hospital Authority	314 (76.8)
Facilitate continuity of patient care	259 (63.3)
Delivery of health care service	249 (60.9)
Safety	
Improved patient safety	209 (51.1)

Lack of awareness was the most common obstacle preventing nonenrolled patients (483/1200, 40.3%) from joining the system (Table 2), followed by not being clear about project objectives (229/1200, 19.1%) and high levels of self-perceived health status (157/1200, 13.1%). Among nonenrolled doctors, feasibility and benefits of the system were widely challenged; 48.0% (95/198) of nonenrolled doctors complained about the complicated enrollment procedure and 40.4% (80/198) were hindered by the additional workload of migrating data from paper records to computers.

Table 2. Top five hindering factors for enrolling in the Public-Private Interface–Electronic Patient Record system.

Perceived barriers among nonusers	Persons, n (%)
Nonenrolled patients (n=1200)	
Low awareness	
Unaware of the project	483 (40.3)
Unclear about project objectives	229 (19.1)
Low necessity	
High levels of self-perceived health	157 (13.1)
Safety reasons	
Concern about safety of personal data and privacy	150 (12.5)
Time-consuming	
I don’t have enough time	93 (7.8)
Nonenrolled doctors (n=198)	
Low feasibility	
Complicated procedure to join the project	95 (48.0)
Concerns about additional workload for data migration from paper records to computer	80 (40.4)
Viewing electronic medical records is time-consuming	60 (30.3)
Unclear benefits	
Use of the system does not assist clinical operation and provide significant benefits	58 (29.3)

A Hong Kong Government report showed that a large proportion of subjects had not made use of the system to access patient records since enrollment [20]. Hence, our study explored the reasons people registered but did not use the system. A total of 44.5% (223/501) of surveyed patients had not used the system after enrollment as they did not have a clinical indication requiring PPI-ePR access (153/223, 68.6%), did not know how the system works (22/223, 9.9%), and/or their doctor had not enrolled (22/223, 9.9%; Table 3). Among the 409 enrolled doctors, 22 (5.4%) had never accessed their patients' medical records via the PPI-ePR. For 22 doctors who had not accessed the PPI-ePR after joining and gave a reason, the most commonly reported reason was the absence of clinical indication requiring PPI-ePR access (8/21, 38.1%) and forgotten log-in password (5/21, 23.8%; Table 3).

Table 3. Reasons for not using the Public-Private Interface–Electronic Patient Record system after enrollment.

Reasons for not using system after enrolling	Persons, n (%)
Enrolled patients (n=223)	
No clinical indication of need for accessing PPI-ePR[a] data	153 (68.6)
Don't know how to use	22 (9.9)
My family doctor does not participate	22 (9.9)
Password forgotten	16 (7.2)
Enrolled doctors (n=22)	
No clinical indication of need for accessing PPI-ePR data	8 (38.1)
Password forgotten	5 (23.8)
Patient has not joined PPI-ePR project	4 (19.0)
Patient failed or refused to provide log-in details to authorize the access	4 (19.0)
Cannot afford time to trace the records	1 (4.8)

[a]PPI-ePR: Public-Private Interface–Electronic Patient Record.

User Satisfaction

Enrolled patients, in general, had higher satisfaction levels than enrolled doctors ($P < 0.001$). Around 10% ([40+10]/409, 12.2%) of enrolled doctors were dissatisfied or very dissatisfied with the PPI-ePR online system, which is far higher than that among enrolled patients ([1+1]/501, 0.4%; Table 4).

Table 4. Users' satisfaction levels with the Public-Private Interface–Electronic Patient Record online system.

Level of satisfaction	Doctors (N=409), n (%)	Patients (n=501), n (%)
Very satisfied	40 (9.8%)	18 (16.5%)
Satisfied	238 (58.1%)	300 (59.9%)
Neutral	81 (19.9%)	116 (23.2%)
Dissatisfied	40 (9.8%)	1 (0.2%)
Very dissatisfied	10 (2.4%)	1 (0.2%)

Suggested Cues to Registration

Our survey also asked patients and doctors for suggestions on increasing the PPI-ePR registration rate. Television, newspaper, and magazine advertisements were regarded as the most effective means to enhance public knowledge and encourage participation in the program among the users and nonusers (1297/1701, 76.2%), followed by recommendation from health care professionals (248/1701, 14.6%; Table 5). Among doctors, medical profession newsletters or journals (428/610, 70.2%) were reported as the most popular promotional strategies. Onsite promotional activities (217/610, 35.6%); television, newspaper, and magazine advertisements; and websites (183/610, 30.0%) were also commonly suggested.

Table 5. Suggested cues to promote the registration of Public-Private Interface–Electronic Patient Record.

Suggested cues by user group	Persons, n (%)
Patients (n=1701)	
Television, newspaper, and magazine advertisements	1297 (76.2)
Recommendations from health care professionals	248 (14.6)
Posters/leaflets	245 (14.4)
Organized seminars	118 (6.9)
Governmental and Hospital Authority websites	78 (4.6)
Souvenirs	5 (0.3)
Doctors (n=610)	
Medical profession newsletters and journals	428 (70.2)
Onsite promotion activities	217 (35.6)
Television, newspaper, and magazine advertisements	183 (30.0)
Website	172 (28.2)
Posters/leaflets	104 (17.0)
Social media (eg, Facebook)	66 (10.8)
Incentives and souvenirs	61 (10.0)

Doctor Suggestions on Improving Public-Private Interface–Electronic Patient Record Online Platform

The extension of sharable data scope (224/409 enrolled doctors agreed, 54.8%) has been frequently reported as an area for future improvement among enrolled doctors, followed by a more user-friendly interface (61/198, 30.8%). Nonusers of the system believed simplification of the enrollment process (107/198, 54.0%) and technical support for operations (90/198, 45.5%) would be the most effective ways to improve the project (Table 6).

Table 6. Doctor suggestions for improving the Public-Private Interface–Electronic Patient Record Project system.

Suggestions for improving the online platform	Enrolled doctors (n=409), n (%)	Nonenrolled doctors (n=198), n (%)
Simplification of enrollment process	91 (22.2)	107 (54.0)
Technical support for system operation	54 (13.2)	90 (45.5)
Develop user-friendly interface	126 (30.8)	61 (30.8)
Extend sharable data scope	224 (54.8)	11 (5.6)

Discussion

Principal Findings

Territory-wide, large degree, timely, legal, and secure health information exchange (HIE) is necessary to improve health care quality, especially in Hong Kong where the private and public health care sectors are autonomous and independent units with little or no experience in information sharing. This study is the first to provide a comprehensive assessment on the awareness, perceptions, obstacles, and suggestions for registering and using the PPI-ePR online interface among both patients and doctors.

We revealed a much lower awareness level of HIE in Hong Kong than in other Asia-Pacific regions emphasizing computerized information exchange among hospitals [22]. Countries like Australia, Malaysia, the United Kingdom, and the United States had widely adopted, existing infrastructures of cloud computing and/or basic eHR platforms before the implementation of the territory-wide platform [18]. In Hong Kong, however, a large proportion of doctors were deeply involved in scattered, paper-based health records before joining the PPI-ePR project. The implementation of health information technology hence started from the idea germination stage, requiring an even longer time and more multidisciplinary promotions [23]. Mass media may be the most effective strategy to increase public awareness, according to our survey results. Western experience, however, emphasized the power of significant others, including family doctors and peers, in motivating patient enrollment [24]. Literature on health behaviors discussed both types of promotions. It has been summarized that public promotions were more effective in increasing individual awareness, especially in the earlier stages of innovation implementation [25]. Among populations with high levels of awareness and acceptance of the innovation, strong ties and small-group interactions were more essential for behavior modification [22]. Our survey results echoed this idea. Although both ways were praised by enrolled and nonenrolled patients, enrolled patients reported family and doctor recommendations as the most popular way of knowing about

PPI-ePR, while nonenrolled patients rated television and newspaper as the most popular ways.

Regarding perceived barriers, although concerns about privacy and data security have been widely discussed in global literature [26,27], few patients and doctors in Hong Kong were worried about data confidentiality issues. Scholars in Bangladesh, where the dual-track health care system is very similar to Hong Kong, also found no association between data privacy and the use of eHealth systems [28]. Patients' low bargaining power in the policy decision process and the absence of health privacy awareness were suggested as reasons for the low concerns in data confidentiality. Another qualitative study in the United States suggested that patient concerns about privacy and data security were mostly nonspecific, gut-level emotions that may be properly soothed by public education [29]. The low level of worries in Hong Kong may benefit from the repeated stress of encrypted and safeguard algorithms and constant emphasis on patients' legal rights in most promotional materials. All functions and powers of PPI-ePR are required under the Electronic Health Record Sharing System Ordinance (Chapter 625), with a well-developed complaint handling policy [30].

Our results echoed previous literature [31,32] reporting reducing duplicate tests and treatments and timely access to patients' medical records as the main benefits of HIE projects like PPI-ePR. Despite the perceived benefits, we noted low use levels of the information system, which has also been observed in studies in Australia, Europe, and some states in the United States [31,33]. Such low use levels hinder the effectiveness of building a health information system. A recent systematic review of 22 population-based studies summarized that regardless of the compatibility and efficacy of EHR systems, emergency department patients with severe symptoms, elderly patients with chronic disease, and those who had a recent admission history were more likely to have encounter-related use [34]. Future development and implementation of HIE projects in Hong Kong may consider more on these populations.

Our study also revealed that patient and doctor levels of use were dependent on each other. Nonenrollment of doctors is a popular reason for patients to not make use of the PPI-ePR platform. A lack of clinical indication required data extraction from other hospitals was among the main reasons for the low level of PPI-ePR use. The current PPI-ePR project only allows the exchange of limited types of data and does not allow patients access to their own records. Studies in the United Kingdom, United States, and Germany emphasized multiple functions of EHR which may encourage active use. These functions included recommendations for daily exercise, prescription refill, outpatient reservation, and appointment reminders [33,35,36]. Their experience may be adapted to Hong Kong to attract patient and doctor use. The HA in Hong Kong may consider enlisting enrolled doctors to clarify the simplicity of the enrollment procedure, as peers in the health care sectors were shown to be common means of learning about the PPI-ePR.

Among enrolled users of the PPI-ePR, over 76% of patients and 67% of doctors were satisfied with its overall performance. A user-friendly interface, simple enrollment procedures, good training, technical information support, and smooth transition with little migration efforts were regarded as important for improving doctors' satisfaction. Scholars from Malaysia also noted these needs especially among elder doctors who lacked computer skills, even 30 years after the EHR implementation [37], indicating the continuous needs of training and supports. Meanwhile, the dilemma between simplified data exchange procedure and data confidentiality was discussed in global studies. The current PPI-ePR system adopted a 2-factor authentication process to safeguard patient privacy. Each access to patient health records required passwords from both patients and authorized doctors, while both static password and a dynamic password generated from the security token were required on the doctors' side. Such a procedure has been found to be satisfactory following both external and internal audits [20] but required extra time and efforts from care givers. Nevertheless, ongoing modification of the online portal based on feedback from medical staff is the key to increasing use and satisfaction levels among doctors [38]. It is noted that HA and the Electronic Health Record Registration Office in Hong Kong have been considering doctors' suggestions and formulating technical details and logistic arrangements for a more effective migration plan.

Strengths and Limitations

This study is the first to comprehensively assess the acceptance and practice of the EHR system in Hong Kong. Although the benefits, facilitators, and barriers of EHRs have been widely discussed in countries including the United States [38], United Kingdom [33], Germany [39], and Australia [40] in the past decade, much effort is needed in the Eastern world, where EHRs have generally been established only in the past few years [40,41]. Previous assessments mainly focused on either the patient or doctor side [8,10,42,43]. Our research involved both sides, comparing the perceptions among enrolled patients, nonenrolled patients, enrolled doctors, and nonenrolled doctors. Results provided comprehensive evaluations of the first EHR project interface between the public and private health sectors in Hong Kong.

This study has some limitations. Our survey combined widely used questions on knowledge, attitudes, and practices studies and health belief model studies [44,45]. Some questions were rephrased to ensure their fitness in Hong Kong and have only undergone face validity by our expert panel. Results of this localized survey may only be adapted to Hong Kong society and may not be generalizable. For doctors' surveys, the sampling frame included all private medical doctors in Hong Kong from a comprehensive list of postal invitations. The response rate was, nevertheless, only 14.4%, with a large difference between enrolled and nonenrolled doctors. Until 2016, fewer than 1930 private doctors had enrolled in the PPI-ePR system. Our sample covered over 21.1% of the enrolled population. Yet for nonenrolled doctors, the coverage rate was less than 10%, introducing a form of response bias later studies may pay attention to.

Implication

The PPI-ePR project provided a backbone for the territory-wide Electronic Health Record Sharing System (eHRSS), which was launched in March 2016 [46]. Since the development of eHRSS

was heavily influenced by the PPI-ePR, further studies may compare the acceptance, barriers, facilitators, and satisfaction among various stakeholders between the PPI-ePR and eHRSS systems and assess trends of eHR adaptation in Hong Kong.

Conclusions

This study explored the attitudes, practices, barriers, and facilitators of joining PPI-ePR in Hong Kong. Findings were consistent with some global challenges of promoting territory-wide eHRs while specific features in Hong Kong were also noted. For one, comparing the high levels of satisfaction among users and limited awareness among nonusers, efforts should be paid at this stage on the behavioral change of nonusers. On the other hand, since the PPI-ePR is migrating toward the eHRSS, continuous engagement of existing users is equally important. The benefits of sharing health records between the private and public health care sectors, as well as that between patients and doctors, should be further promoted through medical journals, newspapers, and television advertisements. Meanwhile, the information technology system should be more user-friendly with more functions involved to encourage continuous use after enrollment. More promotion and simplified information technology operations can help to ensure the acceptance and effectiveness of eHR sharing in Hong Kong.

Acknowledgments

We are grateful that this study was funded by the Hospital Authority of Hong Kong Special Administrative Region. We acknowledge the support from the telephone service center led by Professor Joseph Lau, head of the Division of Health Improvement, School of Public Health and Primary Care, Chinese University of Hong Kong. We thank all the doctors, patients, and health care users for their participation in this survey.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient and doctor characteristics.
[DOCX File , 23 KB - jmir_v22i4e13761_app1.docx]

References

1. Bisbal J, Berry D. An analysis framework for electronic health record systems: interoperation and collaboration in shared healthcare. Methods Inf Med 2011;50(2):180-189. [doi: 10.3414/ME09-01-0002] [Medline: 19936438]
2. Duftschmid G, Dorda W, Endel G, Fröschl K, Gall W, Grossmann W, et al. Fragmentation of diabetes treatment in Austria: an indicator for the need for shared electronic health record systems. Stud Health Technol Inform 2012;180:667-671. [Medline: 22874275]
3. Duftschmid G, Rinner C, Kohler M, Huebner-Bloder G, Saboor S, Ammenwerth E. The EHR-ARCHE project: satisfying clinical information needs in a shared electronic health record system based on IHE XDS and archetypes. Int J Med Inform 2013 Dec;82(12):1195-1207 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.08.002] [Medline: 23999002]
4. Lenert L, Dunlea R, Del Fiol G, Hall LK. A model to support shared decision making in electronic health records systems. Med Decis Making 2014 Nov;34(8):987-995 [FREE Full text] [doi: 10.1177/0272989X14550102] [Medline: 25224366]
5. Rinner C, Grossmann W, Sauter SK, Wolzt M, Gall W. Effects of shared electronic health record systems on drug-drug interaction and duplication warning detection. Biomed Res Int 2015;2015:380497 [FREE Full text] [doi: 10.1155/2015/380497] [Medline: 26682218]
6. Kemper AR, Uren RL, Clark SJ. Adoption of electronic health records in primary care pediatric practices. Pediatrics 2006 Jul;118(1):e20-e24. [doi: 10.1542/peds.2005-3000] [Medline: 16818534]
7. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
8. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, Canadian Committee for Patient Accessible Health Records. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]
9. Castro V, Clements C, Murphy S, Gainer V, Fava M, Weilburg J, et al. QT interval and antidepressant use: a cross sectional study of electronic health records. BMJ 2013 Jan 29;346:f288 [FREE Full text] [doi: 10.1136/bmj.f288] [Medline: 23360890]
10. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
11. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 03;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]

12. Li P, Xie C, Pollard T, Johnson AEW, Cao D, Kang H, et al. Promoting secondary analysis of electronic medical records in China: summary of the PLAGH-MIT critical data conference and health datathon. JMIR Med Inform 2017 Nov 14;5(4):e43 [FREE Full text] [doi: 10.2196/medinform.7380] [Medline: 29138126]
13. Liu C, Cheng T. Exploring critical factors influencing physicians' acceptance of mobile electronic medical records based on the dual-factor model: a validation in Taiwan. BMC Med Inform Decis Mak 2015 Feb 07;15:4 [FREE Full text] [doi: 10.1186/s12911-014-0125-3] [Medline: 25889506]
14. Chae YM, Yoo KB, Kim ES, Chae H. The adoption of electronic medical records and decision support systems in Korea. Healthc Inform Res 2011 Sep;17(3):172-177 [FREE Full text] [doi: 10.4258/hir.2011.17.3.172] [Medline: 22084812]
15. Ochieng OG, Hosoi R. Factors influencing diffusion of electronic medical records: a case study in three healthcare institutions in Japan. Health Inf Manag 2006;34(4):120-129. [doi: 10.1177/183335830503400405] [Medline: 18216416]
16. Kong X, Yang Y, Gao J, Guan J, Liu Y, Wang R, et al. Overview of the health care system in Hong Kong and its referential significance to mainland China. J Chin Med Assoc 2015 Oct;78(10):569-573 [FREE Full text] [doi: 10.1016/j.jcma.2015.02.006] [Medline: 26144022]
17. Sek ACH, Cheung NT, Choy KM, Wong WN, Tong AYH, Fung VH, et al. A territory-wide electronic health record—from concept to practicality: the Hong Kong experience. Stud Health Technol Inform 2007;129(Pt 1):293-296. [Medline: 17911725]
18. Ratnam KA, Dominic P. Factors associating the IT capability in healthcare services and its perceived transformation impact. IJBIS 2016;22(1):82. [doi: 10.1504/ijbis.2016.075719]
19. Cheung N, Fung V, Wong WN, Tong A, Sek A, Greyling A, et al. Principles-based medical informatics for success: how Hong Kong built one of the world's largest integrated longitudinal electronic patient records. Stud Health Technol Inform 2007;129(Pt 1):307-310. [Medline: 17911728]
20. What is Public Private Interface–Electronic Patient Record sharing Pilot Project?. Hong Kong: Hospital Authority URL: https://www3.ha.org.hk/ppp/ppiepr.aspx?lang=eng [accessed 2020-02-25]
21. Report No. 58: Hospital Authority: Public-Private Partnership (PPP) programmes. Hong Kong: Audit Commission; 2012 Mar 28. URL: https://www.aud.gov.hk/pdf_e/e58ch03.pdf [accessed 2019-02-21]
22. Lei J, Sockolow P, Guan P, Meng Q, Zhang J. A comparison of electronic health records at two major Peking University Hospitals in China to United States meaningful use objectives. BMC Med Inform Decis Mak 2013 Aug 28;13:96 [FREE Full text] [doi: 10.1186/1472-6947-13-96] [Medline: 23984797]
23. Blumenthal D, Glaser JP. Information technology comes to medicine. N Engl J Med 2007 Jun 14;356(24):2527-2534. [doi: 10.1056/NEJMhpr066212] [Medline: 17568035]
24. Koskela T, Sandström S, Mäkinen J, Liira H. User perspectives on an electronic decision-support tool performing comprehensive medication reviews: a focus group study with physicians and nurses. BMC Med Inform Decis Mak 2016 Jan 22;16(1):6. [doi: 10.1186/s12911-016-0245-z]
25. Müller AM, Khoo S, Morris T. Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial. J Med Internet Res 2016 Jan 07;18(1):e5 [FREE Full text] [doi: 10.2196/jmir.5235] [Medline: 26742999]
26. Esmaeilzadeh P. The effects of public concern for information privacy on the adoption of health information exchanges (HIEs) by healthcare entities. Health Commun 2019 Sep;34(10):1202-1211. [doi: 10.1080/10410236.2018.1471336] [Medline: 29737872]
27. Ozair FF, Jamshed N, Sharma A, Aggarwal P. Ethical issues in electronic health records: a general overview. Perspect Clin Res 2015;6(2):73-76 [FREE Full text] [doi: 10.4103/2229-3485.153997] [Medline: 25878950]
28. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: a patient's perspective. Inform Health Soc Care 2017 Jan;42(1):1-17. [doi: 10.3109/17538157.2015.1075541] [Medline: 26865037]
29. Simon SR, Evans JS, Benjamin A, Delano D, Bates DW. Patients' attitudes toward electronic health information exchange: qualitative study. J Med Internet Res 2009 Aug 06;11(3):e30 [FREE Full text] [doi: 10.2196/jmir.1164] [Medline: 19674960]
30. Personal data (privacy) ordinance and electronic health record sharing system (points to note for healthcare providers and healthcare professionals). Hong Kong: Office of the Privacy Commissioner for Personal Data; 2016 Feb. URL: https://tinyurl.com/tyxctqt [accessed 2019-02-21]
31. Vest JR, Gamm LD, Ohsfeldt RL, Zhao H, Jasperson J. Factors associated with health information exchange system usage in a safety-net ambulatory care clinic setting. J Med Syst 2012 Aug;36(4):2455-2461 [FREE Full text] [doi: 10.1007/s10916-011-9712-3] [Medline: 21523428]
32. Hamp AD, Doshi RK, Lum GR, Allston A. Cross-jurisdictional data exchange impact on the estimation of the HIV population living in the District of Columbia: evaluation study. JMIR Public Health Surveill 2018 Aug 13;4(3):e62 [FREE Full text] [doi: 10.2196/publichealth.9800] [Medline: 30104182]
33. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Nov 16;341:c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
34. Bowden T, Coiera E. The role and benefits of accessing primary care patient records during unscheduled care: a systematic review. BMC Med Inform Decis Mak 2017 Sep 22;17(1):138. [doi: 10.1186/s12911-017-0523-4]

35. Silvestre A, Sue VM, Allen JY. If you build it, will they come? The Kaiser Permanente model of online health care. Health Aff (Millwood) 2009;28(2):334-344. [doi: 10.1377/hlthaff.28.2.334] [Medline: 19275988]
36. Jha AK, Doolan D, Grandt D, Scott T, Bates DW. The use of health information technology in seven nations. Int J Med Inform 2008 Dec;77(12):848-854. [doi: 10.1016/j.ijmedinf.2008.06.007] [Medline: 18657471]
37. Karim N, Ahmad M. An overview of electronic health record (EHR) implementation framework and impact on health care organizations in malaysia: a case study. 2010 Presented at: IEEE International Conference on Management of Innovation & Technology; 2010; Singapore.
38. Morton M, Wiedenbeck S. A framework for predicting EHR adoption attitudes: a physician survey. Perspect Health Inf Manag 2009 Sep 16;6:1a [FREE Full text] [Medline: 20169013]
39. Ueckert F, Goerz M, Ataian M, Tessmann S, Prokosch H. Empowerment of patients and communication with health care professionals through an electronic health record. Int J Med Inform 2003 Jul;70(2-3):99-108. [doi: 10.1016/s1386-5056(03)00052-2] [Medline: 12909161]
40. Sittig DF. Personal health records on the internet: a snapshot of the pioneers at the end of the 20th Century. Int J Med Inform 2002 Apr;65(1):1-6. [doi: 10.1016/s1386-5056(01)00215-5] [Medline: 11904243]
41. Abd Ghani M, Bali RK, Naguib RN, Marshall IM, Wickramasinghe NS. Electronic health records approaches and challenges: a comparison between Malaysia and four East Asian countries. Int J Electron Healthc 2008;4(1):78-104. [doi: 10.1504/IJEH.2008.018922] [Medline: 18583297]
42. Luchenski SA, Reed JE, Marston C, Papoutsi C, Majeed A, Bell D. Patient and public views on electronic health records and their uses in the United kingdom: cross-sectional survey. J Med Internet Res 2013 Aug 23;15(8):e160 [FREE Full text] [doi: 10.2196/jmir.2701] [Medline: 23975239]
43. Powell J, Buchan I. Electronic health records should support clinical research. J Med Internet Res 2005 Mar 14;7(1):e4 [FREE Full text] [doi: 10.2196/jmir.7.1.e4] [Medline: 15829476]
44. Lyu Y, Hu C, Sun L, Qin W, Xu P, Sun J, et al. Impact of an intervention programme on knowledge, attitudes and practices of population regarding severe fever with thrombocytopenia syndrome in endemic areas of Lu'an, China. Epidemiol Infect 2018 Jan;146(1):125-136. [doi: 10.1017/S0950268817002679] [Medline: 29173207]
45. Heid C, Knobloch MJ, Schulz LT, Safdar N. Use of the health belief model to study patient perceptions of antimicrobial stewardship in the acute care setting. Infect Control Hosp Epidemiol 2016 May;37(5):576-582 [FREE Full text] [doi: 10.1017/ice.2015.342] [Medline: 26809477]
46. Keung K, Lee C, Ng K, Leung S, Choy K. An empirical study on patients' acceptance and resistance towards electronic health record sharing system: a case study of Hong Kong. IJKSS 2018;9(2):1-27. [doi: 10.4018/ijkss.2018040101]

Abbreviations

EHR: electronic health record
eHRSS: Electronic Health Record Sharing System
EMR: electronic medical record
ePR: electronic patient record
HA: Hospital Authority
HIE: health information exchange
PPI-ePR: Public-Private Interface–Electronic Patient Record

Edited by G Eysenbach; submitted 18.03.19; peer-reviewed by E Brainin, J Reponen, K Nazi; comments to author 03.10.19; revised version received 26.11.19; accepted 09.02.20; published 06.04.20.

Please cite as:
Wang J, Huang J, Cheung CSK, Wong WN, Cheung NT, Wong MCS
Adoption of an Electronic Patient Record Sharing Pilot Project: Cross-Sectional Survey
J Med Internet Res 2020;22(4):e13761
URL: https://www.jmir.org/2020/4/e13761
doi:10.2196/13761
PMID:32250279

Original Paper

Epidemiological Characterization of a Directed and Weighted Disease Network Using Data From a Cohort of One Million Patients: Network Analysis

Kyungmin Ko[1,2*], MD; Chae Won Lee[1,3*], BSc; Sangmin Nam[4*], MD, PhD; Song Vogue Ahn[5], MD, PhD; Jung Ho Bae[6], MD; Chi Yong Ban[1,3], BA; Jongman Yoo[3,7], MD, PhD; Jungmin Park[8], PhD; Hyun Wook Han[1,3], MD, PhD

[1]Department of Biomedical Informatics, CHA University of Medicine, Seongnam, Republic of Korea

[2]Department of Pathology, Medstar Georgetown University Hospital, Washington, DC, WA, United States

[3]Institute of Basic Medical Sciences, School of Medicine, CHA University, Seongnam, Republic of Korea

[4]Department of Ophthalmology, CHA Bundang Medical Center, Seongnam, Republic of Korea

[5]Department of Health Convergence, Ewha Womans University, Seoul, Republic of Korea

[6]Department of Internal Medicine, Healthcare Research Institute, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Republic of Korea

[7]Department of Microbiology, CHA University School of Medicine, Seongnam, Republic of Korea

[8]Department of Nursing, School of Nursing, Hanyang University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyun Wook Han, MD, PhD
Department of Biomedical Informatics
CHA University of Medicine
Pangyo-ro, 335
Seongnam, KS009
Republic of Korea
Phone: 82 31 881 7109
Email: stepano7@gmail.com

Abstract

Background: In the past 20 years, various methods have been introduced to construct disease networks. However, established disease networks have not been clinically useful to date because of differences among demographic factors, as well as the temporal order and intensity among disease-disease associations.

Objective: This study sought to investigate the overall patterns of the associations among diseases; network properties, such as clustering, degree, and strength; and the relationship between the structure of disease networks and demographic factors.

Methods: We used National Health Insurance Service-National Sample Cohort (NHIS-NSC) data from the Republic of Korea, which included the time series insurance information of 1 million out of 50 million Korean (approximately 2%) patients obtained between 2002 and 2013. After setting the observation and outcome periods, we selected only 520 common Korean Classification of Disease, sixth revision codes that were the most prevalent diagnoses, making up approximately 80% of the cases, for statistical validity. Using these data, we constructed a directional and weighted temporal network that considered both demographic factors and network properties.

Results: Our disease network contained 294 nodes and 3085 edges, a relative risk value of more than 4, and a false discovery rate-adjusted *P* value of <.001. Interestingly, our network presented four large clusters. Analysis of the network topology revealed a stronger correlation between in-strength and out-strength than between in-degree and out-degree. Further, the mean age of each disease population was related to the position along the regression line of the out/in-strength plot. Conversely, clustering analysis suggested that our network boasted four large clusters with different sex, age, and disease categories.

Conclusions: We constructed a directional and weighted disease network visualizing demographic factors. Our proposed disease network model is expected to be a valuable tool for use by early clinical researchers seeking to explore the relationships among diseases in the future.

(J Med Internet Res 2020;22(4):e15196) doi:10.2196/15196

KEYWORDS

cohort studies; data science; longitudinal studies; statistical data interpretation; medical informatics

Introduction

Traditionally, clinical researchers have pushed forward to explore a number of risk factors that affect a single disease [1-3], and any diseases previously diagnosed are considered important clinical indicators to predict the disorder under investigation [4,5]. Among various methods for unearthing disease relationships, the concept of network medicine could be better suited to understand health and disease [6-8]. Likewise, a disease network was introduced a decade ago as a useful method to study the complex relationships among diseases [9-17].

Under the assumption that diseases are caused by genetic defects, many disease networks were constructed using genomic data [9,11,14]. For example, Li et al constructed a network to investigate disease relationships according to the genes of their shared pathways [14]. Nonetheless, according to the disease lists of the International Statistical Classification of Diseases, 10th revision (ICD-10), many diseases, such as traumatic bone fracture attributed to a traffic accident, are not related to genetic mutations. As such, genome-based disease networks alone are inevitably limited for accurately representing the complex pathogenesis of the relationships among diseases [18].

Thus, disease networks were later constructed using shared clinical information, such as symptoms and comorbidities [12,17]. Zhou et al generated a symptom-based network of human diseases that was based on the similarity of symptoms [17], whereas Hidalgo et al and Barabási et al constructed a comorbidity network using the Medicare database [7,12]. Because these efforts were focused on demonstrating the relationships among shared diseases or symptoms occurring or present at a single point in time, the networks did not take into account investigations of the temporal order of disease manifestations [19].

Recently, researchers have suggested that disease networks should consider temporal directionality when exploring the connections among diseases [13]. For instance, Jensen et al analyzed temporal disease progression patterns according to disease trajectory using the Danish National Patient Registry. In this study, we constructed a directional and weighted disease network visualizing the effects of demographic factors, such as sex, age, and disease outbreak size, according to the relative risk (RR) among diseases using the National Health Insurance Service-National Sample Cohort (NHIS-NSC) of South Korea, which includes epidemiological time series data of 12 years for approximately 1 million patients.

Finally, we investigated the overall patterns of the associations among diseases; network properties, such as clustering, degree, and strength; and the relationship between the structure of the disease network and demographic factors.

Methods

Construction and Visualization of the Disease Network

South Korea is a representative country implementing national health insurance services. The NHIS-NSC contains time insurance information of 1 million out of 50 million Korean (approximately 2%) patients, which was collected between 2002 and 2013. Thus, clinical information can be tracked for 12 years for every patient.

To examine the risk factors for diseases that a patient already had at the beginning of the cohort study, we needed to set an initial period before the main study period to serve as the medical history period. For most chronic diseases, the recommended follow-up interval rarely exceeds 2 years. Therefore, we set the observation period as 2002 through 2003 and the outcome period as 2004 through 2013.

From the sample of 1,016,580 patients who were eligible for National Health Insurance in 2004, we selected 885,125 patients who had at least one record of a medical visit during the aforementioned observation period. We defined this group of patients as the sample cohort. In South Korea, diagnoses are coded in the Korean Classification of Diseases sixth revision (KCD-6), an extension of the ICD-10. The only difference between the KCD-6 and ICD-10 is that the diagnosis codes for Korean medicine are included in the KCD-6 using U20-U99 codes.

To simplify the study, we truncated the KCD-6 codes beyond their third digit, in effect, grouping subcategories of conditions together. In total, the KCD-6, when used between 2002 and 2013, consisted of 2,097 unique diagnoses at the third digit level, and of these, 1,971 diagnoses were included in our data.

Ultimately, we chose only 520 common KCD-6 codes that were the most prevalent diagnoses, covering approximately 80% of the cases for statistical validity.

Support Offered by the Clinical Evidence From Relationships Among Diseases

All statistical analyses and visualizations were performed using the R package “igraph” (version 3.4.4) and Cytoscape. For calculation of the RR, we sought to obtain *P* values against the null hypothesis, which states that any two diseases present occur independently of one another in the sample cohort. False-discovery rate (FDR) corrections were performed using the Bonferroni method.

Clusters of associated diseases were identified using the random walktrap community detection algorithm [20,21]. This method detects clusters purely according to connectivity (unless specified to use weights) using random walks along edges. The demographic profiling of disease clusters was carried out by pooling the patients identified with at least one of the diagnoses in the cluster.

As a result, patient pools for each cluster are not exclusive but instead overlap somewhat with other clusters. The age distribution of the patient pools was calculated at the beginning of the observation period. An enrichment analysis of the clusters for the KCD categories was performed using the Fisher exact test for adjusted *P* values <.05.

Topological Characteristics of the Disease Network

In graph theory, the degree of a node is the total number of connections with other nodes. In a directed network, the out-degree of a node is the number of connections with that node as the source, whereas the in-degree of a node is the number of connections with that node as the target. Hence, the degree can be thought of as a measure of the level of disease risk in our network.

In contrast, the strength of a node is the sum of the RRs to achieve connections with other nodes. For example, the out-strength and in-strength of node *i* are defined, respectively, as follows:

$$s_{out}(i)=\sum_j RR_{ij} \quad (1)$$

$$s_{in}(i)=\sum_j RR_{ji} \quad (2)$$

where RR_{ij} is the weight of the edge from node *i* to node *j*, and RR_{ji} is the weight of the edge from node *j* to node *i*. The out-strength is a measure of the magnitude of disease morbidity, whereas the in-strength is a measure of the magnitude of a disease's tendency to follow from other diseases.

Characterization of Large Clusters Throughout Computational Clustering

To calculate the risk ratio from a risk disease D_1 to an outcome disease D_2 ($D_1 \rightarrow D_2$), we need to first identify the group of patients at risk of acquiring D_2. We regarded a patient as being at risk of disease D_2 if that patient had no record of being diagnosed with D_2 during the observation period. Patients were considered to be exposed if they had been diagnosed at least once with disease D_1 during the observation period. The RR of $D_1 \rightarrow D_2$ was defined using the following formula:

$$RR=(a / [a + b]) / (c / [c + d]) \quad (3)$$

where a is the number of patients exposed to D_1 in the initial period and D_2 in the outcome period; b is the number of patients exposed to D_1 in the initial period but not exposed to D_2 in the outcome period; c is the number of patients not exposed to D_1 in the initial period but exposed to D_2 in the outcome period; and d is the number of patients not exposed to either D_1 in the initial period or D_2 in the outcome period (Table 1).

Since a single misdiagnosis can cause a very large error in the RR value if the numbers in the contingency table are small, we established a minimum size of 947 patients for each group. For example, the diagnosis with the highest prevalence in the initial period was "J20: acute bronchitis," with 355,045 patients diagnosed at least once in the observation period.

The lowest diagnosis was "R80: isolated proteinuria," with 947 patients diagnosed during the observation period. Consequently, the at-risk group sizes ranged from 530,080 (885,125 − 355,045) for acute bronchitis to 884,178 (885,125 − 947) for isolated proteinuria.

To select the cuto value for the RR, we chose the closest integer to the top percentile (ie, the closest integer to x where *P* [RR > x] = .01], which was 4. Therefore, we selected disease relationships with an RR of more than 4 and an FDR-corrected *P* value of <.001 to construct our final network.

Accordingly, the prevalence and at-risk group sizes were large enough to accurately determine the RR. Since the self-interaction in this study was not the subject, the total number of theoretical interactions of a total of 520 nodes was found to be 269,880.

Table 1. Contingency table for disease-disease risk ratio calculation.

Risk disease in 2002-2003	Outcome disease in 2004-2013	
Exposed	Not exposed	
Exposed	Value[a]	Value[b]
Not exposed	Value[c]	Value[d]

[a]Number of patients exposed to the risk disease (D_1) and outcome disease (D_2).

[b]Number of patients exposed to D_1 but not exposed to D_2.

[c]Number of patients not exposed to D_1 but exposed to D_2.

[d]Number of patients not exposed to either D_1 or D_2.

Results

Construction and Visualization of the Disease Network

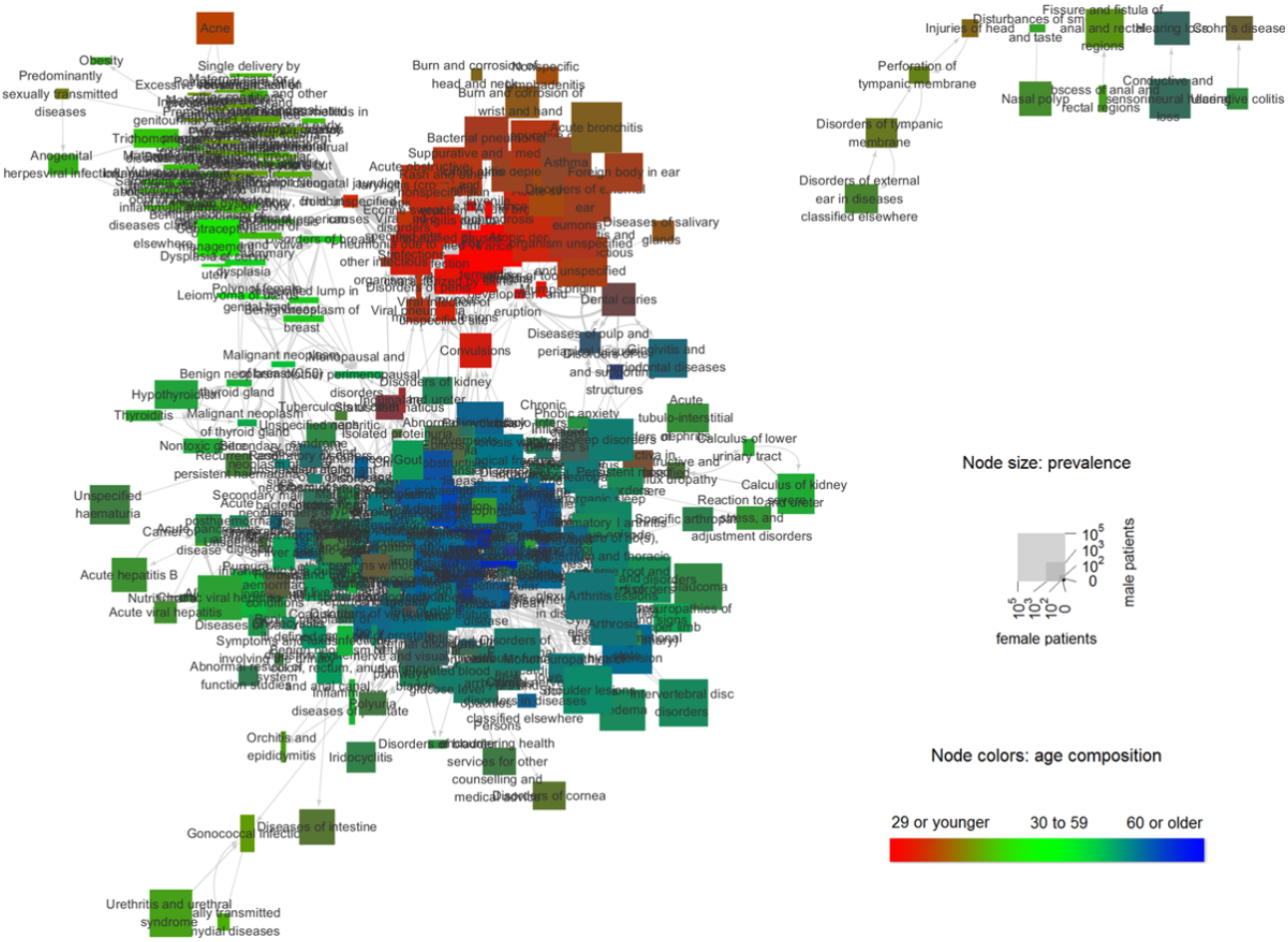
Initially, for the construction and visualization of our final disease network, we selected an RR of more than 4 and an FDR-adjusted *P* value of <.001. As a result, we were able to obtain a disease network with four clusters, 294 nodes, and 3085 edges (Figure 1).

For better clinically intuitive visualization, we designed a visualization scheme such that the color of the disease node would reflect the age of the patient affected with the disease and that the outbreak size would reflect the relative number of patients. The shape of the node was indicated by a rectangle. Node widths represented the number of female patients, whereas the heights represented the number of male patients. For node colors, the intensity of the red channel was proportional to the ratio of patients younger than 30 years, the intensity of the green

channel was proportional to the ratio of patients aged between 30 and 59 years, and the intensity of the blue channel was proportional to the ratio of patients aged 60 years or older.

Meanwhile, to indicate the directionality and weight of the node, the edges were represented by arrows and the relative thickness was represented in gray. The number of patients in each sex and age group for each disease was calculated at the beginning of the outcome period based on the included patients' histories during the observation period.

Figure 1. Visualization of the disease network. The network is constructed using the prefuse force-directed layout.



Support Offered by the Clinical Evidence From Relationships Among Diseases

To determine whether the correlations among inferred diseases in our disease network model were clinically meaningful, we investigated the available literature concerning the top six disease-disease associations with the highest RR values from the established disease network.

Most results appeared in agreement with previously known associations among diseases (Table 2). A substantial nosologic and biologic overlap exists between bipolar disorder and schizophrenia [22,23]. Further, long-standing hypertension is known to be an important cause and consequence of chronic kidney disease [24]. It is also well-known that anemia develops into chronic kidney disease and portends an unfavorable prognosis [25].

Interestingly, the association between diabetes mellitus in pregnancy and neonatal jaundice was also very high, despite the fact that neonates are never pregnant. This outcome is possible because diagnoses for infants who are not yet in the national registry are filled out under the mother's account for insurance [26-28]. Another interesting aspect of our results was the fact that neonatal jaundice and diaper dermatitis were strongly associated with one another, which was not observed in previous epidemiological studies.

Table 2. Top relative risk values.

Risk factor disease	Outcome disease	RR[a]	References
Bipolar affective disorder	Schizophrenia	34.4	[22,23]
Chronic kidney disease	Hypertensive renal disease	31.9	[24]
Diabetes mellitus in pregnancy	Neonatal jaundice	29.1	[26-28]
Neonatal jaundice	Diaper dermatitis	28.1	N/A[b]
Chronic kidney disease	Anemia in chronic disease	27.4	[25]
Hemorrhage in early pregnancy	Neonatal jaundice	26.1	[27]

[a]RR: relative risk.

[b]N/A: not applicable.

Topological Characteristics of the Disease Network

We investigated the in- and out-degree distributions of our constructed network. Like many other networks, the in- and out-degrees of our network followed a power-law distribution with a long tail [29] (Figure 2). However, in contrast with the degrees, neither in- nor out-strength followed the power-law distribution (Figure 3).

Figure 2. Distribution of the network for in- and out-degrees.

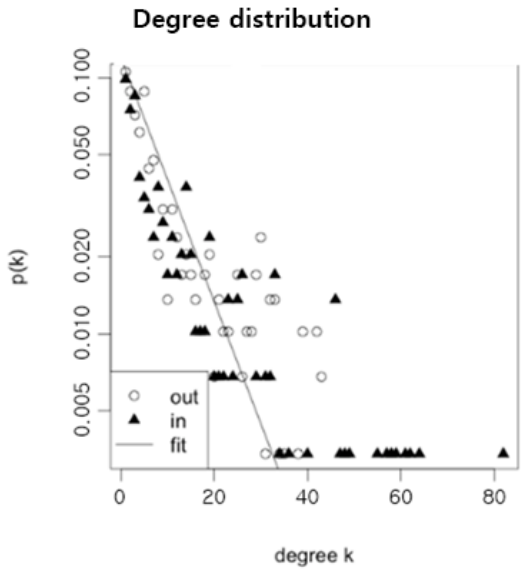


Figure 3. Distribution of the network for in- and out-strength.

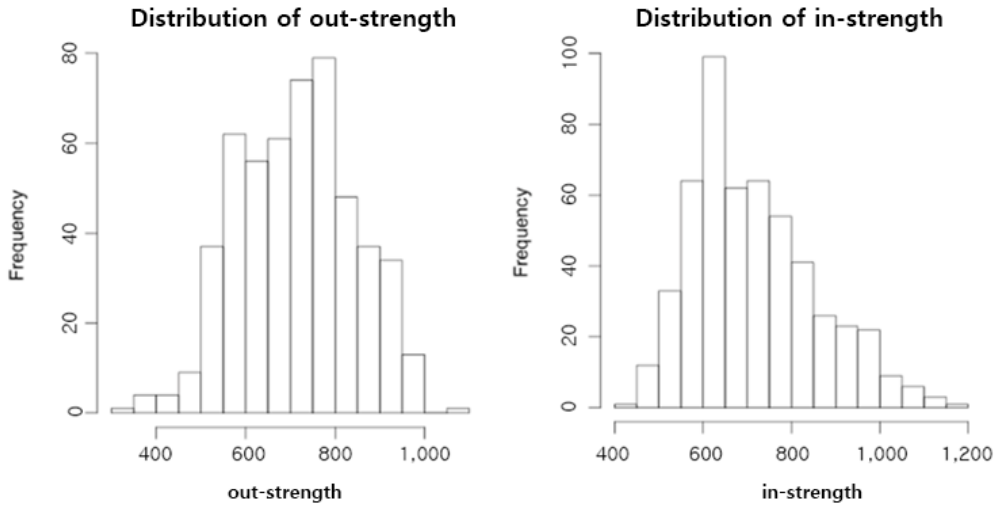


Table 3 shows the top six diseases with the highest out-degree, in-degree, out-strength, and in-strength results. The top out-degree diseases included diseases that are known to affect many other conditions such as chronic kidney disease and essential hypertension.

The top in-degree diseases are known to be associated with long-term hospitalization or immunocompromise, which are statuses that can arise from various diseases. The top out-degree and top out-strength diseases had considerable overlap, with polyneuropathy, senile cataract, and retinal disorders all being both high out-degree and high out-strength diseases. Patients with these diseases may be at greater risk for developing multiple comorbidities.

The top in-degree and top in-strength diseases included Parkinson disease, chronic kidney disease, anemia in chronic disease, and osteoporosis with pathological fracture. This suggests that many different diseases can have a strong tendency for coverage onto these diseases. Subsequently, we explored the relationships between out-degree and in-degree and between out-strength and in-strength results.

The correlation between the out-strength and in-strength findings (Pearson correlation coefficient: 0.72) was stronger than that between the out-degree and in-degree findings (Pearson correlation coefficient: 0.57) (Figure 4). This means that diseases show strong tendencies to develop from other diseases. For better characterization, we color-coded the diseases in the out-/in-strength plot according to the age composition of the patients (Figure 5). This revealed that mean age was related with positioning along the regression line of the out-/in-strength plot.

Table 3. Top out-/in-degree diseases and top out-/in-strength diagnoses.

KCD[a] code and disease	Degree
Top out-degree diseases	
G63: polyneuropathy	43
C61: malignant neoplasm of the prostate	43
H25: senile cataract	43
H36: retinal disorders	42
N18: chronic kidney disease	42
I10: essential hypertension	39
Top in-degree diseases	
G20: Parkinson disease	82
M80: osteoporosis with pathological fracture	64
N18: chronic kidney disease	62
D63: anemia in chronic diseases	61
A41: sepsis	59
Top out-strength diseases	
G63: polyneuropathy	1057
H36: retinal disorders	998
M48: spondylopathies	992
H25: senile cataract	992
M81: osteoporosis without pathological fracture	981
M17: arthrosis of the knee	979
Top in-strength diseases	
G20: Parkinson disease	1197
N18: chronic kidney disease	1135
D63: anemia in chronic diseases	1123
M80: osteoporosis with pathological fracture	1120
I12: hypertensive renal disease	1100
H27: disorder of the lens	1089

[a]KCD: Korean Classification of Diseases.

Figure 4. Correlations between in- and out-degrees and in- and out-strengths.

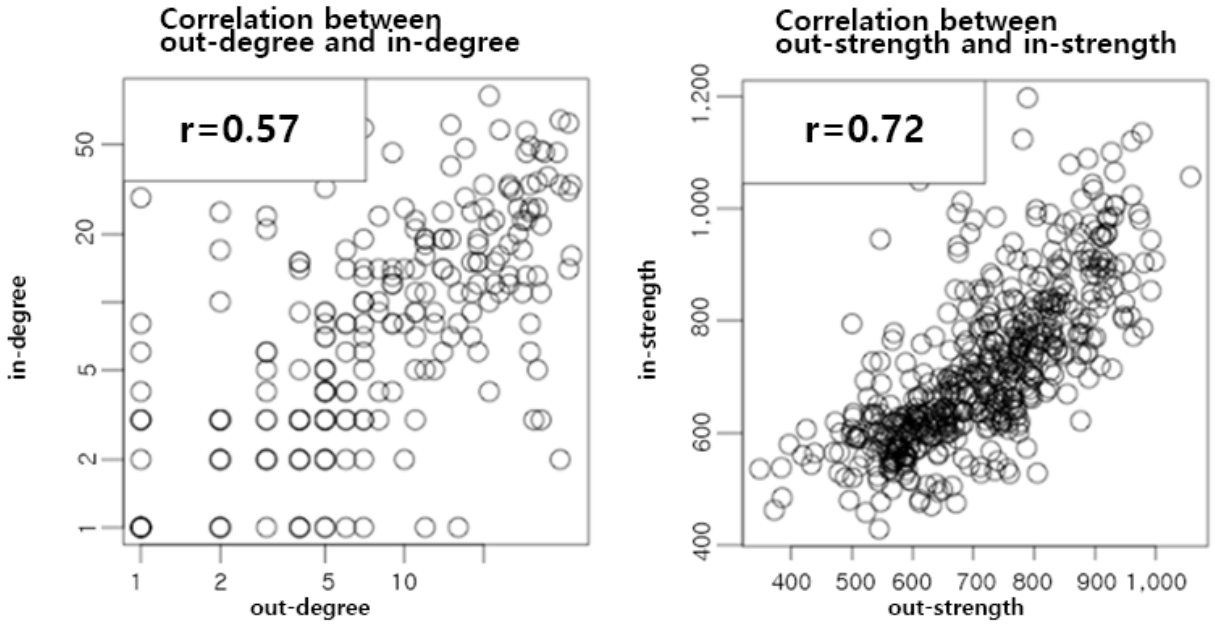
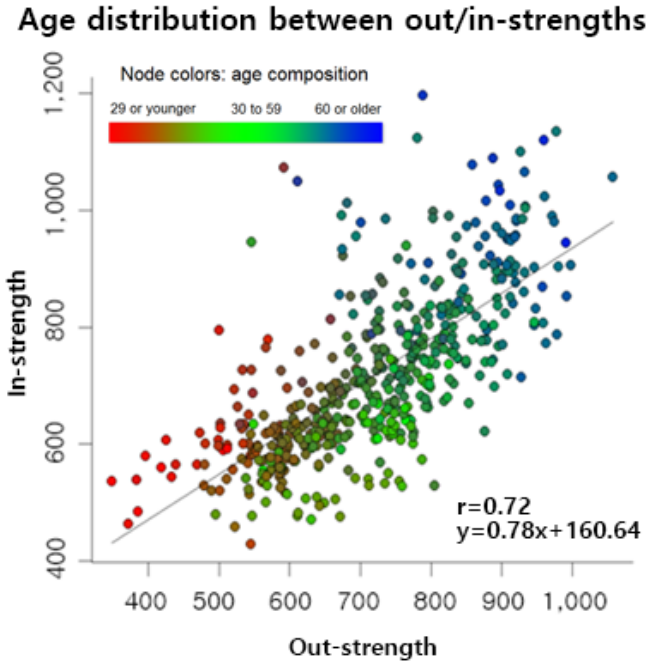


Figure 5. Out- and in-strengths plotted according to the age composition of the patients.



Characterization of Large Clusters Throughout Computational Clustering

To confirm whether the visually observed clustering in Figure 1 was an artifact of the layout algorithm, we employed a random walktrap algorithm for network clustering [20,21].

A total of 19 clusters were detected, including four large clusters of a size greater than 38 and 15 small clusters of a size less than 13. When we color-coded the network using these four major clusters, we could see that the top right and top left clusters were almost exactly as visualized, but the largest cluster was detected as two large subclusters (Figure 6).

This confirmed that disease associations grouped diseases into a few distinct clusters and that this occurred independently of the prefuse force-directed layout. Interestingly, the modularity score for the random walktrap algorithm (0.53) was more than twice the score for the KCD categories (0.24). To see whether the four major clusters actually had the characteristics that we noticed in the visualization, we profiled the clusters with respect to the age distribution and sex ratio of the affected patients (Figure 7).

Patients diagnosed with diseases in clusters 1 and 3 were relatively older (mean age of 47.4 [SD 18.22] years and 48.19 [SD 18.66] years, respectively). The diseases in cluster 2 were

dominated by women of reproductive age (the ratio of males to females was 1:18.67; mean age: 39.38 [SD 13.08] years). Cluster 4 included patients who were relatively young, with slightly more females (the ratio of males to females was 1:1.22; mean age: 31.7 [SD 21.56] years).

We profiled the KCD classes of each cluster and performed an enrichment analysis to investigate the types of diseases that were enriched in each cluster (P=.05). Although every cluster contained its own disease groups (Multimedia Appendix 1), the enrichment analysis revealed that each of the four major clusters was enriched with nonoverlapping sets of KCD categories (Multimedia Appendix 2). Since each cluster had distinct characteristics, we labeled the major clusters from 1 to 4, according to their most prominent features, as "chronic debilitation," "women's disease," "hemato-oncology," and "infectious disease" clusters, respectively.

Figure 6. Four major clusters of the network.

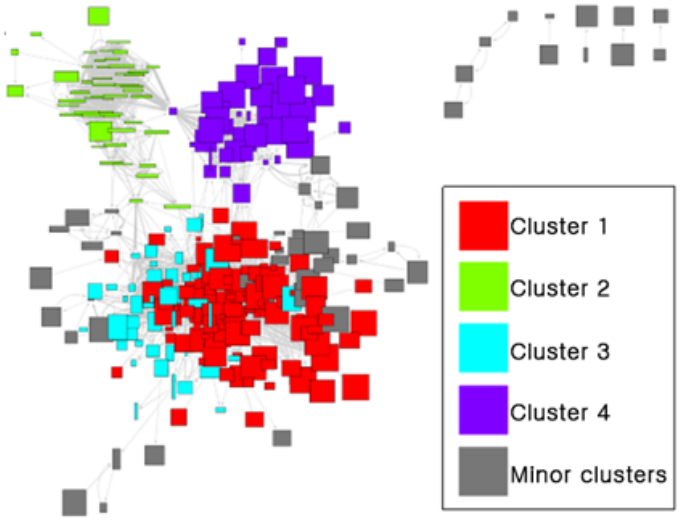
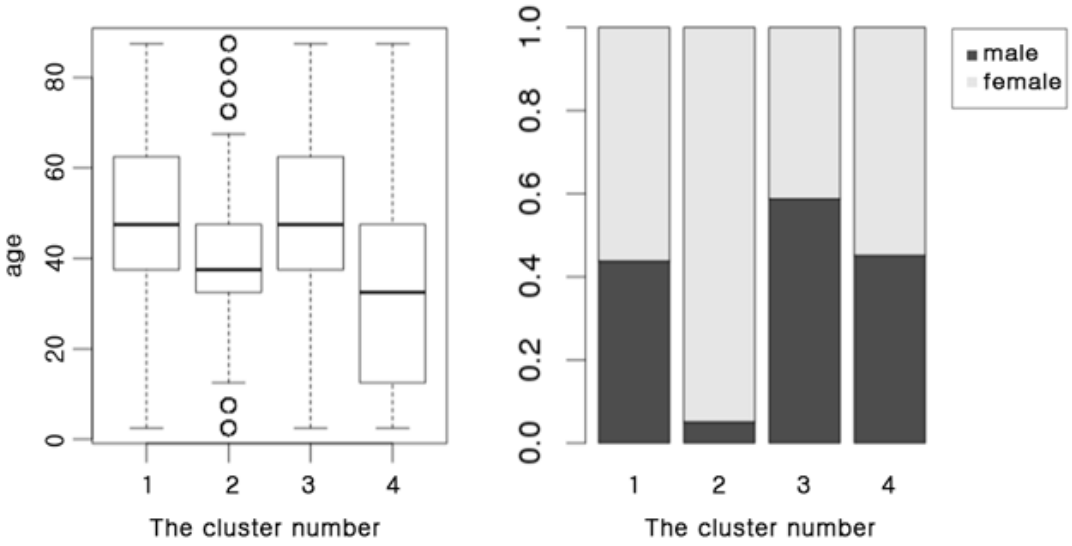


Figure 7. Age distribution and sex ratio for the four major clusters.



Discussion

Principal Findings

In this study, we proposed a comprehensive method for modeling a disease network with directionality and weight of edges using medical claims data. We selected only the most common diagnoses to avoid an overestimation of RR among rare diseases. The φ correlation is also useful to avoid overestimation of associations among rare diseases [12], but it is clinically less intuitive and unnecessary for the purpose of studying the overall pattern of common disease associations. Epidemiological factors, such as age and sex, are important inducers of disease development [30-33]; they are, in effect, the most critical clinical factors affecting the prevalence and classification of diseases.

Another purpose of this study was to dissolve these various factors in the disease network and to see how various factors

affect the structure and dynamics of the disease network. In addition, these factors were reflected in the visualization of the disease network. In our disease network model, we proposed an intuitive visualization method that maximizes clinical usability.

Nodes indicate the patient outbreak size, and at the same time, represent the relative proportion of width (women) and height (men) in a rectangle. In addition, each node is divided into red for young patients, green for middle-aged patients, and blue for old patients. Conversely, the RR and direction among diseases can be intuitively grasped through the arrow and the thickness of the edge.

As a result, our visualization method of a disease network can help to intuitively identify the direction and RR among diseases and can help to effectively understand the age distribution, sex ratio, and disease outbreak size. The directionality of the disease relationship is a consequence of the study design being longitudinal with a chronological order. Strong RR values support disease association in a chronological order, which is a prerequisite for causality among diseases in clinical research [34]. Because of this, our network can be a starting point to investigate causality among diseases. Here, we examined the literature on disease relationships with high RRs.

Limitations

The NHIS-NSC includes the proportional stratified sampling data of 1,025,340 patients from among 47,851,928 patients. These patients were randomly extracted by age group, sex, eligibility status, and income level using a proportional stratified sampling method [35]. In general, NHIS-NSC data are representative, but some rare diseases may lose their representativeness owing to the difficulty in obtaining statistical significance. Owing to these limitations, this study excluded disease groups with a small sample size, and rare diseases that have not been assessed in this study are expected to be evaluated in future studies.

As a result, we have shown that our network can provide clues to reveal the causal relationships among diseases. In our network, neonatal jaundice and diaper dermatitis presented a statistically meaningful association (RR=28.1, $P<.001$), but we did not find other supporting evidence of such an association during the literature search. Nonetheless, this does not mean that our network had an incorrect result. Rather, it suggests the possibility of associations that researchers have not yet discovered. In a strict sense, it can be difficult to say that this is a causal relationship because directionality can only be thought of as the natural progression of a disease, the outcome of a treatment, or the process of making a diagnosis.

Conclusions

During our research, we investigated network topologies, such as degree and strength. Both in- and out-degrees followed power-law distribution like other biological networks; however, strength distributions did not. Since the RR values did not indicate causality, we cannot say that a certain disease is the cause of many other diseases by only looking at out-degrees and out-strengths. Despite this, patients with high out-degree and high out-strength diseases (eg, polyneuropathy, retinal disorders, and senile cataract) are worthy of special attention for secondary prevention purposes. Similarly, diseases with high in-degree and high in-strength findings, such as Parkinson disease, osteoporosis with pathological fracture, and chronic kidney disease, can be seen as common comorbidities of many different diseases.

We found stronger correlations between in- and out-strengths than between in- and out-degrees. Moreover, a stronger risk associated with a disease tended to be related to older affected patients. The association between age and strength suggested that the previously discovered correlation between disease connectivity and mortality could be explained by the phenomenon of increased risk strength.

Through clustering of the network, we found four major disease clusters with distinct demographic characteristics. Interestingly, each cluster was exclusively enriched in KCD categories and had a different mean age and sex ratio. The clustering patterns analyzed using our network suggest that KCD categories, age, and sex have strong influences on disease associations and highlight the importance of demographic factors. Since patients with diseases within a cluster tend to acquire other diseases within the same cluster, we may be able to minimize the onset of comorbidities through patient care by configuring specialty clinics to cater to clusters or subclusters of associated diseases, as is the case with obstetrics and gynecology.

In this regard, our proposed disease network model will likely serve as a valuable tool for early clinical researchers seeking to further explore the relationships of diseases in the future.

For future study attempts, we will take into account the dynamicity of network-considered time order and assess the network collapse point that can affect the overall network structure.

Acknowledgments

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2017R1E1A1A03070934) and the Ministry of Science & ICT (NRF-2019M3C7A1032262).

Authors' Contributions

HWH conceived the project; HWH, SVA, JHB, and NSM designed the research; KMK and JMP performed the research; KMK and JMY analyzed the data; and CWL, CYB, and HWH wrote the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Korean Classification of Diseases composition of disease nodes in each cluster. The composition of each disease cluster in terms of disease categories is shown in this bar chart.
[DOCX File , 238 KB - jmir_v22i4e15196_app1.docx]

Multimedia Appendix 2
Enrichment analyses revealed that each of the four major communities was enriched with nonoverlapping sets of Korean Classification of Diseases categories.
[DOCX File , 14 KB - jmir_v22i4e15196_app2.docx]

References

1. Mooe T, Björklund F, Graipe A, Huber D, Jakobsson S, Kajermo U, et al. The Nurse-Based Age Independent Intervention to Limit Evolution of Disease After Acute Coronary Syndrome (NAILED ACS) Risk Factor Trial: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2014 Aug 15;3(3):e42 [FREE Full text] [doi: 10.2196/resprot.3466] [Medline: 25131960]
2. Peng S, Shen F, Wen A, Wang L, Fan Y, Liu X, et al. Detecting Lifestyle Risk Factors for Chronic Kidney Disease With Comorbidities: Association Rule Mining Analysis of Web-Based Survey Data. J Med Internet Res 2019 Dec 10;21(12):e14204 [FREE Full text] [doi: 10.2196/14204] [Medline: 31821152]
3. Tluway F, Urio F, Mmbando B, Sangeda RZ, Makubi A, Makani J. Possible Risk Factors for Severe Anemia in Hospitalized Sickle Cell Patients at Muhimbili National Hospital, Tanzania: Protocol for a Cross-Sectional Study. JMIR Res Protoc 2018 Feb 28;7(2):e46 [FREE Full text] [doi: 10.2196/resprot.7349] [Medline: 29490896]
4. Linet M, Vajdic C, Morton L, de Roos AJ, Skibola C, Boffetta P, et al. Medical history, lifestyle, family history, and occupational risk factors for follicular lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr 2014 Aug;2014(48):26-40 [FREE Full text] [doi: 10.1093/jncimonographs/lgu006] [Medline: 25174024]
5. Silverman DT, Schiffman M, Everhart J, Goldstein A, Lillemoe KD, Swanson GM, et al. Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. Br J Cancer 1999 Aug;80(11):1830-1837 [FREE Full text] [doi: 10.1038/sj.bjc.6690607] [Medline: 10468306]
6. Tillmann T, Gibson AR, Scott G, Harrison O, Dominiczak A, Hanlon P. Systems Medicine 2.0: potential benefits of combining electronic health care records with systems science models. J Med Internet Res 2015 Mar 23;17(3):e64 [FREE Full text] [doi: 10.2196/jmir.3082] [Medline: 25831125]
7. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011 Jan;12(1):56-68 [FREE Full text] [doi: 10.1038/nrg2918] [Medline: 21164525]
8. Loscalzo J, Barabasi A. Systems biology and the future of medicine. Wiley Interdiscip Rev Syst Biol Med 2011;3(6):619-627 [FREE Full text] [doi: 10.1002/wsbm.144] [Medline: 21928407]
9. Al-Harazi O, Al Insaif S, Al-Ajlan MA, Kaya N, Dzimiri N, Colak D. Integrated Genomic and Network-Based Analyses of Complex Diseases and Human Disease Network. J Genet Genomics 2016 Jun 20;43(6):349-367. [doi: 10.1016/j.jgg.2015.11.002] [Medline: 27318646]
10. Davis DA, Chawla NV. Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. PLoS One 2011;6(7):e22670 [FREE Full text] [doi: 10.1371/journal.pone.0022670] [Medline: 21829475]
11. Goh K, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proc Natl Acad Sci U S A 2007 May 22;104(21):8685-8690 [FREE Full text] [doi: 10.1073/pnas.0701361104] [Medline: 17502601]
12. Hidalgo CA, Blumm N, Barabási AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol 2009 Apr;5(4):e1000353 [FREE Full text] [doi: 10.1371/journal.pcbi.1000353] [Medline: 19360091]
13. Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nat Commun 2014 Jun 24;5:4022 [FREE Full text] [doi: 10.1038/ncomms5022] [Medline: 24959948]
14. Li Y, Agarwal P. A pathway-based view of human diseases and disease relationships. PLoS One 2009;4(2):e4346 [FREE Full text] [doi: 10.1371/journal.pone.0004346] [Medline: 19194489]
15. Park J, Lee D, Christakis NA, Barabási AL. The impact of cellular networks on disease comorbidity. Mol Syst Biol 2009;5:262 [FREE Full text] [doi: 10.1038/msb.2009.16] [Medline: 19357641]
16. Zhang X, Zhang R, Jiang Y, Sun P, Tang G, Wang X, et al. The expanded human disease network combining protein-protein interaction information. Eur J Hum Genet 2011 Jul;19(7):783-788 [FREE Full text] [doi: 10.1038/ejhg.2011.30] [Medline: 21386875]
17. Zhou X, Menche J, Barabási AL, Sharma A. Human symptoms-disease network. Nat Commun 2014 Jun 26;5:4212. [doi: 10.1038/ncomms5212] [Medline: 24967666]

18. Roberts NJ, Vogelstein JT, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. The predictive capacity of personal genome sequencing. Sci Transl Med 2012 May 09;4(133):133ra58 [FREE Full text] [doi: 10.1126/scitranslmed.3003380] [Medline: 22472521]
19. Gross F. What systems biology can tell us about disease. Hist Philos Life Sci 2011;33(4):477-496. [Medline: 22662506]
20. Pons P, Latapy M. Computing communities in large networks using random walks. In: Computer and Information Sciences. Berlin, Heidelberg: Springer; 2005 Presented at: 20th International Conference on Computer and Information Sciences; October 26-28; Istanbul, Turkey p. 284-293. [doi: 10.1007/11569596_31]
21. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A 2008 Jan 29;105(4):1118-1123 [FREE Full text] [doi: 10.1073/pnas.0706851105] [Medline: 18216267]
22. Laursen TM, Agerbo E, Pedersen CB. Bipolar disorder, schizoaffective disorder, and schizophrenia overlap: a new comorbidity index. J Clin Psychiatry 2009 Oct;70(10):1432-1438. [doi: 10.4088/JCP.08m04807] [Medline: 19538905]
23. Pearlson GD. Etiologic, phenomenologic, and endophenotypic overlap of schizophrenia and bipolar disorder. Annu Rev Clin Psychol 2015;11:251-281. [doi: 10.1146/annurev-clinpsy-032814-112915] [Medline: 25581236]
24. Gargiulo R, Suhail F, Lerma EV. Hypertension and chronic kidney disease. Dis Mon 2015 Sep;61(9):387-395. [doi: 10.1016/j.disamonth.2015.07.003] [Medline: 26328515]
25. Iseki K, Kohagura K. Anemia as a risk factor for chronic kidney disease. Kidney Int Suppl 2007 Nov(107):S4-S9 [FREE Full text] [doi: 10.1038/sj.ki.5002481] [Medline: 17943141]
26. Alam M, Raza SJ, Sherali AR, Akhtar SM. Neonatal complications in infants born to diabetic mothers. J Coll Physicians Surg Pak 2006 Mar;16(3):212-215. [Medline: 16542603]
27. Wang Z, Kanguru L, Hussein J, Fitzmaurice A, Ritchie K. Incidence of adverse outcomes associated with gestational diabetes mellitus in low- and middle-income countries. Int J Gynaecol Obstet 2013 Apr;121(1):14-19. [doi: 10.1016/j.ijgo.2012.10.032] [Medline: 23321368]
28. Yang J, Cummings EA, O'connell C, Jangaard K. Fetal and neonatal outcomes of diabetic pregnancies. Obstet Gynecol 2006 Sep;108(3 Pt 1):644-650. [doi: 10.1097/01.AOG.0000231688.08263.47] [Medline: 16946226]
29. Albert R. Scale-free networks in cell biology. J Cell Sci 2005 Nov 01;118(Pt 21):4947-4957 [FREE Full text] [doi: 10.1242/jcs.02714] [Medline: 16254242]
30. Cohen P, Cohen J, Kasen S, Velez CN, Hartmark C, Johnson J, et al. An epidemiological study of disorders in late childhood and adolescence--I. Age- and gender-specific prevalence. J Child Psychol Psychiatry 1993 Sep;34(6):851-867. [doi: 10.1111/j.1469-7610.1993.tb01094.x] [Medline: 8408371]
31. Farrer L, Cupples L, Haines J, Hyman B, Kukull W, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA 1997;278(16):1349-1356. [Medline: 9343467]
32. Meisinger C, Thorand B, Schneider A, Stieber J, Döring A, Löwel H. Sex differences in risk factors for incident type 2 diabetes mellitus: the MONICA Augsburg cohort study. Arch Intern Med 2002 Jan 14;162(1):82-89. [doi: 10.1001/archinte.162.1.82] [Medline: 11784224]
33. Piccinelli M, Wilkinson G. Gender differences in depression. Critical review. Br J Psychiatry 2000 Dec;177:486-492. [doi: 10.1192/bjp.177.6.486] [Medline: 11102321]
34. Bang S, Kim J, Shin H. Causality modeling for directed disease network. Bioinformatics 2016 Sep 01;32(17):i437-i444. [doi: 10.1093/bioinformatics/btw439] [Medline: 27587660]
35. Lee J, Lee JS, Park S, Shin SA, Kim K. Cohort Profile: The National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. Int J Epidemiol 2017 Apr 01;46(2):e15. [doi: 10.1093/ije/dyv319] [Medline: 26822938]

Abbreviations

FDR: false-discovery rate
ICD-10: International Statistical Classification of Diseases, 10th version
KCD-6: Korean Classification of Diseases, sixth revision
NHIS-NSC: National Health Insurance Service-National Sample Cohort
RR: relative risk

Edited by G Eysenbach; submitted 28.06.19; peer-reviewed by G Kolostoumpis, A Mavragani; comments to author 14.08.19; revised version received 08.10.19; accepted 24.01.20; published 09.04.20.

Please cite as:
Ko K, Lee CW, Nam S, Ahn SV, Bae JH, Ban CY, Yoo J, Park J, Han HW
Epidemiological Characterization of a Directed and Weighted Disease Network Using Data From a Cohort of One Million Patients: Network Analysis
J Med Internet Res 2020;22(4):e15196
URL: https://www.jmir.org/2020/4/e15196
doi:10.2196/15196
PMID:32271154

Original Paper

Influence of Health Literacy on Effects of Patient Rating Websites: Survey Study Using a Hypothetical Situation and Fictitious Doctors

Peter Johannes Schulz[1], PhD; Fabia Rothenfluh[2], PhD

[1]Università della Svizzera Italiana, Lugano, Switzerland

[2]Comparis Company, Zurich, Switzerland

Corresponding Author:
Peter Johannes Schulz, PhD
Università della Svizzera Italiana
Via G Buffi, 13
Lugano, 6900
Switzerland
Phone: 41 58 666 4724
Email: schulzp@usi.ch

Abstract

Background: Physician rating websites (PRWs) are a device people use actively and passively, although their objective capabilities are insufficient when it comes to judging the medical performance and qualification of physicians. PRWs are an innovation born of the potential of the Internet and boosted very much by the longstanding policy of improving and encouraging patient participation in medical decision-making. A mismatch is feared between patient motivations to participate and their capabilities of doing so well. Awareness of such a mismatch might contribute to some skepticism of patient-written physician reviews on PRWs.

Objective: We intend to test whether health literacy is able to dampen the effects that a patient-written review of a physician's performance might have on physician choice.

Methods: An experiment was conducted within a survey interview. Participants were put into a fictitious decision situation in which they had to choose between two physicians on the basis of their profiles on a PRW. One of the physician profiles contained the experimental stimulus in the form of a friendly and a critical written review. The dependent variable was physician choice. An attitude differential, trust differential, and two measures of health literacy, the newest vital sign as an example of a performance-based measure and eHealth Literacy Scale as an example of a perception-based measure, were tested for roles as intermediary variables. Analysis traced the influence of the review tendency on the dependent variables and a possible moderating effect of health literacy on these influences.

Results: Reviews of a physician's competence and medical skill affected participant choice of a physician. High health literacy dampened these effects only in the case of the perception-based measure and only for the negative review. Correspondingly, the effect of the review tendency appeared to be stronger for the positive review. Attitudes and trust only affected physician choice when included as covariants, considerably increasing the variance explained by regression models.

Conclusions: Findings sustain physician worries that even one negative PRW review can affect patient choice and damage doctors' reputations. Hopes that health literacy might raise awareness of the poor basis of physician reviews and ratings given by patients have some foundation.

(J Med Internet Res 2020;22(4):e14134) doi:10.2196/14134

KEYWORDS

physician rating websites; warning messages; experiment; physician competence assessment; patient feedback

Introduction

Physician Rating Websites

Physician rating websites on the internet (PRWs) came about as a synergy between technological innovation and social reform. The innovation, of course, was the advent and rapid development of the internet. The social reform linked to it was the longstanding public health policy of encouraging patients to participate more actively and autonomously in their health care and preservation.

A key element of the Web 2.0 is electronic world of mouth (eWOM), through which consumers have the chance to evaluate any conceivable product or brand [1] and read evaluations posted by others. This increasingly affects consumer choices [2] and shifts the consumer into a new, more powerful position [3,4]. Some have seen this as generating more equality and democracy in the relationship between seller and customer [5].

eWOM devices soon reached into the field of health care, with PRWs as a prominent example. PRWs are promoted as a medium to increase transparency by communicating health care consumer experiences physicians to a large audience. These websites are becoming increasingly popular; as many as 59% of participants in a representative American study indicated that PRWs were important when choosing a doctor [6], while 25% of Germans [7] have searched for a doctor on the internet. Physician reviews are becoming more and more commonplace, and awareness of PRWs seems to be high in the general population [6,8]. The simple, easily comprehensible, narrative nature of physician rating websites seems to catch users' attention and appeal to them more than formal quality information such as academic qualifications, degrees, or areas of specialization [9]. Yet, it is largely one-sided communication. Less than 2% of the reviews are responded to by physicians. The share has been growing recently, but is still very low [10].

The benefits of PRWs for health care consumers' physician choice have been debated [11-13]. Physicians often feel discomfited with open and nonregulated patient feedback platforms [14,15], mainly because they assume health care consumers do not have enough knowledge to pass judgment on a trained physician's diagnosis and therapeutic recommendations [16].

Assessment of the quality of health care by medical laity differs from expert assessments and is likely to be erroneous, ill-founded, or irrelevant. Decisions based on these assessments, or so health professionals fear, run a high risk of not being in the patient's best interest as health professionals would see it. In helping the development of PRWs, the ideal of health literacy may have created a form of health communication that is likely to have many drawbacks.

On a more general level, patient satisfaction is not necessarily related to objective outcome-of-care assessments [17-20]. PRW reviews or ratings were found to be only selectively [21,22], weakly [23], or not at all associated to objective measures of quality such as mortality rates [24,25] or surgeon volume. Furthermore, patient judgment of health care quality may be clouded by circumstantial factors. For example, a study by Swiss researchers found in a pre-post test assessment that the renovation of a medical practice led patients to give better ratings not only to the practice infrastructure but also the quality of the staff and the care received, even though personnel and care remained the same [26].

Physicians and patients alike doubt the qualification of patients to pass judgment of the medical qualification of physicians or the technical and outcome aspects of care [27-29]. Patients preferred recommendations from experts when choosing a physician because they perceived them to be more trustworthy and of higher expertise than reviews from patients [30].

As a patient review of a physician's medical capabilities often is communication from a source whose insight into medical knowledge cannot level with a university-trained physician, the question is raised, does it matter? It will matter if the reviews have effects, for instance, on physician choice. This study investigated the effect of favorable and unfavorable patient reviews of a physician's medical capabilities on patient attitudes toward a physician and their choice of physician. A second concern was possible ways to dampen these effects, and here health literacy comes into view.

Patient Participation and Health Literacy

A more active patient participation in health care is generally wished for, for medical reasons—it accelerates, some claim, the process of healing and improves health outcomes [31]—as well as for social values that grant individuals mastery over their own lives, as a well-known definition of empowerment has it [32]. Limited health literacy is found in substantial minorities of populations [33,34] and associated with bad self-care [35-37], bad general health [38,39], and premature death [40]. High health literacy is associated with a number of positive consequences, such as improved disease management [41].

As a young concept, health literacy was restricted to basic understanding of health matters. It was defined as the ability "to obtain, process, and understand basic health information and services needed to make appropriate health decisions" [42]. Later conceptualizations included more demanding abilities in the notion, among them the ability to critically assess the trustworthiness and credibility of the sources of health information [43]. Such ability would make, one can argue, people with high health literacy recognize the questionable nature of the source of physician ratings and consequently activate their skepticism of the credibility of the raters. This should assuage the effects of the reviews.

Hypotheses

The most fundamental relationship to be considered is the effect of positive and negative PRW text reviews of a physician's performance. The effect considered foremost is on health care consumers' choices of a physician. This research interest is captured in the first hypothesis:

- Hypothesis 1: Health care consumers who read a positive text review on the competence of a fictional physician will choose this physician more often than consumers who read a negative text review.

Studies from adjacent fields suggest that negative information is especially powerful in impacting internet users' opinion on a given subject. Specifically, negative information catches people's attention more than positive content as it refers to hazards, the warning against which was crucial for human survival from an evolutionary perspective [44]. A stronger effect of negative reviews was also found in eWOM for product choices [45] as well as credibility [46,47]. Some medical research, however, finds an equal effect on PRW users' choices of positive and negative reviews. An American study, for instance, found that 37% of PRW users avoided a physician when they were confronted with bad reviews, while 35%

decided to consult a physician when facing positive evaluations [6,7,48-50]. These inconclusive findings suggest a research question on the influence of review valence on the strength of the effect of negative and positive reviews.

- Research Question 1: Do negative or positive text reviews have a stronger effect on patient choice of a physician?

The analysis must be mindful of the fact that PRW reviews are predominantly positive (ranging from 63% to 88%) as studies from websites in the United States [51-54], Germany [55], the United Kingdom [56], Poland [57], and China [58] report.

People with high health literacy acknowledge that a lay person's ability to assess a physician's medical competence is limited and do not consider reviews when they make their physician choice. Technically speaking, this means high health literacy would diminish the relationship between review tendency and physician choice and attitudes. In other words: health literacy would negatively moderate the effect of the review. This is the major relationship we want to demonstrate in this research (Hypothesis 2).

- Hypothesis 2: High health literacy will weaken the effect of review tendency on attitudes and physician choice (negative moderation effect).

The association between review tendency and behavioral intentions, physician choice in our case, might be made more complex by attitudes or trust. Several models are conceivable. Attitudes and trust may mediate the effect of review tendency on physician choice, and attitudes and trust might interfere in the assuaging role of health literacy if there is such.

- Research Question 2: How do attitudes and trust affect the influences on patient choice of a physician?

Methods

Data Collection and Sample Characteristics

An online survey was conducted in March 2017 via a contracting firm (Qualtrics) specialized in survey administration and market research data collection. The ethics committee at the Università della Svizzera Italiana confirmed the study was outside the committee's jurisdiction (CE 2017-1). For inclusion in the study, participants had to be (1) aged 18 years or older, (2) residing in the German part of Switzerland, and (3) fluent in German. Data collection was anonymous, and participants could only participate once. Based on the ultimate of three pretests (n=24 took part in total), the median survey completion time was 14 minutes. To ensure validity, participants who used less than two-thirds of the median time (9 minutes) were excluded. Participants who did not pass the manipulation check, in which they were asked about the number of text reviews present on each of the two physician profiles, were screened out as well. A total of 258 participants passed the time and manipulation checks. When the data were screened for automatic response behavior, 4 more cases were excluded, yielding a final sample of 254 cases. The sample was equally split in terms of gender (129/254, 50.8% female), had education levels that are comparable to the Swiss general population [59], and participants were on average 47.8 (SD 16.05) years old (range 18-85 years; see Table 1).

Table 1. Sociodemographic characteristics of participants by experimental group (n=254).

Characteristic	Positive review n=126	Negative review n=128	*P* value
Gender, n (%)			**.71**[a]
Male	59 (46.8)	66 (51.6)	—
Female	67 (53.2)	62 (48.4)	—
Education level, n (%)			**.32**[a]
Low (primary or secondary school or apprenticeship)	65 (51.6)	61 (47.7)	—
Medium (high school or similar)	31 (24.6)	24 (18.8)	—
High (applied science or university degree)	30 (23.8)	43 (33.6)	—
Age in years, mean (SD)	45.75 (14.46)	51.66 (15.85)	.09[b]

[a]Chi-square test.

[b]One-way analysis of variance.

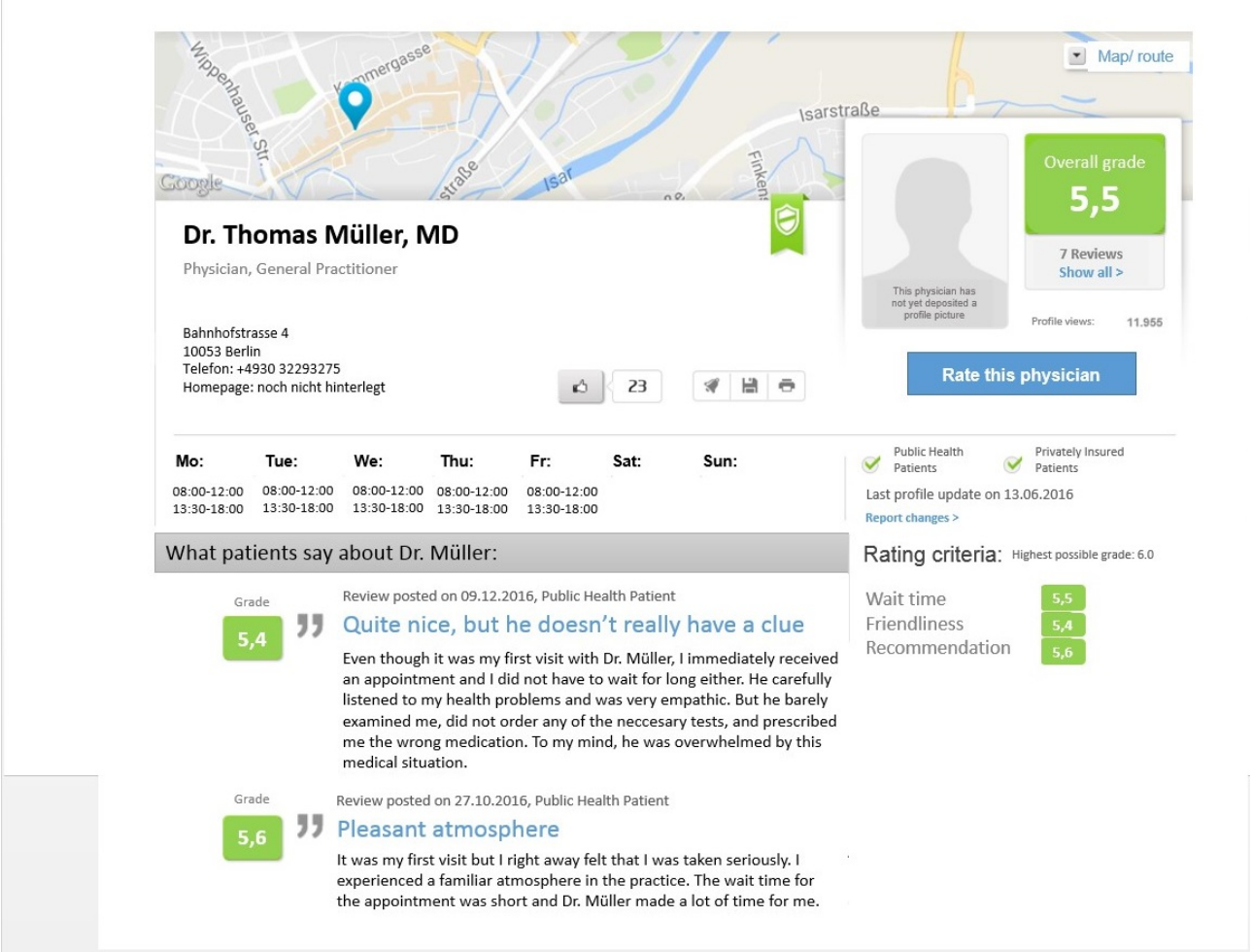
Procedures

In the first part of the survey, participants were asked about their search behaviors on the internet and basic sociodemographic questions (Table 1). Next, they were exposed to a hypothetical scenario to introduce them to a physician selection task. The scenario was developed in cooperation with a medical doctor to ensure its reasonability. Specifically, participants were asked to imagine to have been hurt by a nail during their move to a new town. Subsequently, the wound got infected. Because they had not met anyone yet to ask for a recommendation of a physician, they would look for one online and find the PRWs of two doctors, among whom they would need to choose one. The complete instructions can be found in Multimedia Appendix 1. The PRWs were designed as to resemble real PRWs as much as possible.

The fictional doctors were named Dr med Thomas Müller and Dr med Michael Schmidt, both common German names to provide a realistic scenario. Based on the finding that users pay more attention to written than numerical reviews [9], one of two written reviews in Dr Müller's profile was manipulated

(assessing his competence positively or negatively), and participants were randomly assigned to either one of two conditions. Recruitment and randomization to experimental and control conditions followed Qualtrics procedures based on random numbers. As research finds the number of written reviews on PRWs rather small [60,61], we offered two of them on Dr Müller's rating site, one of which was manipulated, and none on Dr Schmidt's profile. The numerical scores on both PRWs varied slightly but produced the same average assessment. Dr Schmidt's PRW was not manipulated and, borrowing an idea from Li and colleagues [62], added solely for the purpose of creating a decision situation (see Figure 1 for an example of the profiles). All relevant information on the website (eg, overall numerical rating score, number of numeric reviews) except the manipulated text review were similar across conditions. We chose the reviews rather than the ratings for manipulation because they are more variable, which will be helpful in future studies. The two profiles differed only minimally in terms of information content to make the choice situation more realistic (eg, house number of the practice address, phone number). The display order of the two doctors' profiles was randomized to account for primacy effects. There were actually more differences as the experiment was in a 2×2 factorial design with the presence or absence of a warning message as second independent variable, but this is of no relevance to the analyses reported in this article and is therefore not mentioned further.

Figure 1. Stimulus material for unfavorable review condition.



Dependent and Intervening Measures

For physician choice, the major dependent variable in all analyses, participants were asked which one of the two physicians they would choose (7=definitely Dr Müller, 1=definitely Dr Schmidt). As only Dr Müller had received reviews, the scale values are referred to as preference for Dr Müller. The attitudinal variables pertaining to RQ2 included an attitude differential and a trust differential. Attitude toward each of the profiles was surveyed based on semantic differential 5-point scales consisting of 8 items. Adjective pairs included, for example, helpful versus useless, realistic versus forged, reliable versus unreliable. One item was reversed to check users' attention to account for automatic response behavior. The average score of the 8 items was calculated and the difference between the two physicians calculated and used in the analysis. This was done as Dr Schmidt's evaluation was of no interest, while the comparison between the two doctors was.

Trust in the respective physician, specifically the impression participants formed of him, was measured by three items: Dr X makes a good impression on me, Dr X convinces me, I would trust Dr X (1=not at all true, 7=entirely true). A cumulative score was calculated and subsequently averaged after reliability checks were found to be satisfactory. All scales achieved satisfactory internal reliability with Cronbach alpha above .85.

As was done for attitudes, the variable used was the difference between the two physicians' cumulative scores. An overview of the variables and scale details can be found in Table 2, and the complete questionnaire in Multimedia Appendix 2.

For the measurement of health literacy, two types of indicators are available. The classic measures are based on the performance of patients, as in their ability to correctly pronounce medical terms (Rapid Estimate of Adult Literacy in Medicine) [63] or complete omissions in sentences describing medical matters (Short Test of Functional Health Literacy in Adults) [64]. As the concept was extended by more demanding cognitive abilities, new measures were introduced, based on the (self)perception of health care consumers. While performance-based measures can claim a high degree of objectivity, perception-based measures are undoubtedly subjective and therefore prone to be affected by a person's biases, ideals, and needs. We chose two measures, one of each type: the eHealth Literacy Scale (eHEALS) measure [65], in a validated German translation [66], as an example of the perception-based type and the newest vital sign (NVS [67]), also in the German version, for the performance-based measure.

The original eHEALS consists of 8 items; we used 7 of them, probably not wisely, because the item "I know what health resources are available on the internet" did not work well in some of our studies, but did in others [68]. Users rated the items on 5-point scales (1=strongly agree, 5=strongly disagree); the items were averaged and had a satisfactory reliability (Cronbach alpha=.86).

The NVS consisted of items assessing reading comprehension and calculation questions referring to the information given on a fictitious ice cream nutrition label. For every correct answer, one point was assigned, adding up to a final score between 0 and 6, with higher scores indicating higher literacy levels. Applying the standard procedure for translating such materials, the German version was translated from English independently by two native speakers of German and then backtranslated independently by two native speakers of English, discussing and resolving discrepancies at both stages. A separate formal validation was not conducted. For an overview of the scales, see Table 2.

Table 2. Independent and moderator variables, their function and descriptive statistics, scale, and item properties (n=254).

Variable	Function in analysis	Items n	Scale range	Mean (SD)	Reliability Cronbach alpha
Performance-based health literacy (NVS[a])	Moderator in analysis pertaining to H2[b]	6	0 to 6	4.29 (1.72)	n/a
Perception-based health literacy (eHEALS[c])	Moderator in analysis pertaining to H2	7	1 to 5	3.65 (0.68)	.86
Physician choice (7=Dr Müller, 1=Dr Schmidt)	Dependent variable	1	1 to 7	3.48 (1.71)	n/a
Attitude differential Dr Müller over Dr Schmidt	Intermediary	8	–4 to 4	0.29 (0.79)	.89
Trust differential Dr Müller over Dr Schmidt	Intermediary	3	–6 to 6	0.29 (1.68)	.88
Skepticism of eWOM[d,e]	Not in hypotheses but helps explain findings	10	1 to 5	2.96 (0.69)	.88
Usefulness of PRW[f] information[g]	Not in hypotheses but helps explain findings	3	1 to 7	4.10 (1.59)	.95

[a]NVS: Newest Vital Sign.

[b]H2: hypothesis 2.

[c]eHEALS: eHealth Literacy Scale.

[d]eWOM: electronic word of mouth.

[e]Adapted from Grabner-Kräuter and Waiguny [69].

[f]PRW: physician rating website.

[g]Adapted from Diviani et al [68].

Table 2 lists two variables not entered in the hypotheses and research questions but later used for interpretation of results. The perceived usefulness of reviews on websites was assessed (1=not at all, 7=very useful) based on 3 items adapted from Grabner-Kräuter and Waiguny [69] (see Multimedia Appendix 2 for questionnaire). A cumulative average score was calculated after the scale was checked for internal reliability.

Skepticism toward eWOM adapted to the PRW context [70] was assessed on a 5-point scale (1=completely disagree, 5=completely agree) applied to 8 items inquiring about participants' agreement with statements. Individual item scores were also added up and averaged after internal reliability checks were found to be satisfactory.

All variables were coded in a direction so that correlations between hypothesized or intuitively associated variables were positive. In particular, high scale values indicate a friendly review tendency, a stronger preference for Dr Müller over Dr Schmidt, high level of health literacy (on both measures), more favorable attitudes toward Dr Müller compared with Dr Schmidt (attitude differential), and higher trust in Dr Müller compared

with Dr Schmidt (trust differential). The hypothesized negative moderator role of health literacy on the association between review tendency and preference for Dr Müller would show in negative effect coefficients.

Analysis

The collected data were analyzed quantitatively using SPSS Statistics 23.0 software package (IBM Corp). First, data were analyzed for uni- and multivariate outliers, nonnormality, and missing data [71]. H1 was tested and RQ1 answered by comparing means, and *t* tests, chi-square difference tests, and one-way analysis of variance (ANOVA) were applied to assess potential differences between the experimental groups. To test the moderating effect of health literacy on choice behavior (H2), moderation analyses (model 1) based on the PROCESS macro for SPSS version 2.16.3 (AF Hayes) were applied [72]. For the analysis of the final model, which includes attitudes and trust as covariates (RQ2), we added these covariates to the PROCESS command line in model 1, the cov option).

Results

Effect of Review Tendency on Physician Choice and Attitudes

H1 held that health care consumers who encountered a positive text review of Dr Müller's competence would choose him more often than consumers who read a negative review. Individuals who read a positive assessment of Dr Müller's competence were more willing to choose him than participants who saw the negative assessment. ANOVA for testing significance found an impact of the review tendency on physician choice, with a significant main effect. The average scores and details of significance tests are shown in Table 3. These findings indicate that the tendency of reviews of a physician's competence impact users' choices. H1 is supported: reviews matter.

The group confronted with a negative review of Dr Müller rated their attitudes toward the two physicians similarly. When there was a positive review of Dr Müller, his mean attitudinal rating was clearly higher than Dr Schmidt's. Trust in Dr Müller exceeded trust in Dr Schmidt by 1.02 (SD 1.49) scale points when Dr Müller was reviewed positively and fell short by –0.43 scale points when the rating was negative. Both differences were highly significant (attitudinal differential $F_{1,252}$=41.151, P<.001; trust differential $F_{1,252}$=134.656, P<.001). These results indicate that, aside from physician choice, health care consumers' attitudes and trust can also be affected by review tendency. For a simple overview of the difference review tendency makes, see Table 3.

RQ1, on the magnitude of effect of negative and positive reviews, can be tentatively answered in favor of positive reviews. The physician choice measure ranged from 1 to 7 and has 4 as the meaningful middle point, which indicates similar likelihood to choose either one of the doctors. In the condition of a negative review, the score fell 0.31 scale points down from the middle on the preference for Dr Schmidt's side, while with a positive review it was 1.37 points on the preference for Dr Müller's side. As just shown, attitude and trust differentials as indicators of the evaluation of the two physicians reacted significantly more strongly to positive than negative reviews (0.58 vs 0.01 in case of the attitude differential, and 1.02 vs –0.43 in case of the trust differential). A similar test for the more obvious outcome variable, physician preference, is not possible, as the difference was asked in just one question and the deviation from the scale mean is good for illustration but is no basis for significance testing. Without a strict testing for significance, these results suggest rather than validly support that in our setting the positive review seemed to have had the stronger effect on preference, too.

Table 3. Preference and assessment of fictitious doctors in different conditions.

Experimental condition	Line #	Prefer Dr Muller over Dr Schmidt	Objects and attitudes of trust	Attitudes	Trust
Negative review on Müller profile	1	3.69 (1.63)	Schmidt	3.18 (0.53)	3.96 (1.08)
Negative review on Müller profile	2	—	Müller	3.17 (0.64)	3.53 (1.22)
Negative review on Müller profile	3	—	Δ (2–1)[a]	–0.01 (0.75)	–0.43 (1.56)
Positive review on Müller profile	4	5.37 (1.34)	Schmidt	3.15 (0.62)	3.91 (1.07)
Positive review on Müller profile	5	—	Müller	3.73 (0.64)	4.94 (1.17)
Positive review on Müller profile	6	—	Δ (5–4)[b]	0.58 (0.72)	1.02 (1.47)
Grand difference as indicator of difference between the conditions Δ (4–1)[c,d]	7	1.68	Δ (6–3)[e]	0.59	1.45
Significance	n/a[f]	—	—	—	—
F/t[g]	—	80.55	—	6.415	6415
df[h]	—	1250	—	252	252
P value	—	<.001	—	<.001	<.001
η_p[i]	—	0.244	—	—	—
R^2[i]	—	.235	—	—	—

[a]Difference between line 2 and line 1 values for attitudes and trust.

[b]Difference between line 5 and line 4 values for attitudes and trust.

[c]Difference between line 4 and line 1 values for prefer Dr Muller over Dr Schmidt.

[d]Positive values indicate higher preference for Dr Müller, more confidence in own decision, and higher trust and better attitudes when the Müller review was friendly.

[e]Difference between line 6 and line 3 values for attitudes and trust.

[f]Not applicable.

[f]F/t: indicator of model fit.

[g]*df*: degree of freedom.

[h]η_p: indicator of association between dependent and independent variables.

[i]R^2: indicator of contribution of independent on dependent variables.

Adding Health Literacy to the Picture

Our question was whether the effect of review tendency on doctors' choice was dependent on the level of health literacy. The logic of the moderation analysis is to calculate and assess the impact of the interaction of the independent and moderator variables on top of the main effect of both. We can only speak of moderation if the interaction can be shown to exert a significant influence beyond the main effect of review tendency as the hypothesized independent variable and health literacy as the hypothesized moderator. Effects were only found for the eHEALS. Therefore, the analysis proceeds with this measure as indicator of health literacy. The hypothesis was that the more literate the consumer was, the weaker the effect would be of review tendency on the preference for the reviewed doctor. The reason for that is that highly literate people (particularly those who consider themselves as being more familiar with the internet) are less willing to believe in the face value of reviews of doctors as they are posted online. There were two ways to approach this question: (1) use regression models to identify a moderating role of health literacy on the effect of review tendency on preference for the reviewed doctor (Table 4) or (2) analyze the two experimental groups separately.

Regression model 1 shows the effect of the experimental condition on the preference for Dr Müller. When the presence of a negative review was coded 0 and a positive review 1, the constant corresponds to the average of the preference variable under the negative review condition and X is the difference of preference under the positive review condition. This is another way to demonstrate the support that H1 received in our data.

Model 2 adds an influence of health literacy on preference for Dr Müller. This was included not for testing a substantial hypothesis but merely as a built-up to model 3. Unexpectedly, however, model 2 revealed a significance of the association between health literacy and preference. Irrespective of the review tendency, people with a higher health literacy showed a higher preference for Dr Müller over Dr Schmidt. The correlation was r=.104; it failed to reach significance in bivariate analysis (P=.10) other than in model 2.

Model 3 adds an interaction term to test for the dampening influence of health literacy as formulated in H2. The effect was there, as expected with a negative sign. Model 4 added attitudes, trust, and gender as covariates, which almost doubled the R^2 value from 26.6% to 52.4%. Other models were tested that

conceptualized attitude and trust in various mediator and moderator roles. None of these showed any significant contribution of attitudes, trust, and sociodemographic variables.

Technically speaking, as eHEALS health literacy increases by one unit, the difference in preference for Dr Müller according to model 3 between those exposed to a positive review and those exposed to a negative review decreases by –.546. So to say, the moderator quantifies a difference between differences. Other moderation analyses based on the PROCESS macro (model 1) and addressing the influence of attitudes and trust were applied, as well as some according to model 2 with more than one moderator. These analyses did not show any significant moderating effects (RQ2) [72].

The separate analysis of the two experimental groups (positive vs negative review) can be summarized as follows. For those who were exposed to a positive review, preference for the reviewed doctor did not depend on their health literacy levels; whether they considered their health literacy high or low, the share among them who preferred the applauded doctor remained the same. This changed with respect to the participants who were exposed to negative reviews: the lower their literacy, the lower the preference for the criticized doctor would be. In different perspective: the higher the health literacy levels participants showed, the less they were influenced by the negative review to disregard the doctor. As a matter of fact, people with high literacy levels exposed to negative reviews came closer in their choice to those who had seen a positive review of the doctor.

A simple way to illustrate the difference is correlation analysis. For the group that saw a positive review, the correlation between health literacy and preference for Dr Müller was nonexistent (r=.01, P=.91), while the two variables were associated in the group that saw the negative review (r=.228, P<.01).

Table 4. Various regression models estimating preference for Dr Müller over Dr Schmidt.

Independent variables	Coefficient	Standard error	*t*	*P* value
Model 1 (R^2=.243)				
Constant	3.687	0.132	27.914	<.001
Review tendency (X)	1.686	0.188	8.986	<.001
Model 2 (R^2=.255)				
Constant	2.675	0.520	5.143	<.001
Review tendency (X_1)	1.690	0.186	9.061	<.001
eHEALS[a] (X_2)	0.277	0.138	2.012	.045
Model 3 (R^2=.266)				
Constant	1.621	0.740	2.189	<.05
Review tendency (X)	3.679	1.017	3.616	<.001
eHEALS (M)	0.566	0.199	2.836	<.001
X*M	–0.546	0.274	–1.989	.048
Model 4 (final model; R^2=.524)				
Constant	2.151	0.650	3.309	<.001
Review tendency (X)	2.384	0.838	2.846	<.01
eHEALS (M)	0.322	0.165	1.952	.05
X*M	–0.428	0.225	–1.906	.06
Covariate trust	0.346	0.063	5.493	<.001
Covariate attitudes	0.515	0.134	3.846	<.001
Covariate sex	0.358	0.154	2.318	<.05

[a]eHEALS: eHealth Literacy Scale.

Discussion

Principal Findings

Our study showed that when a physician has a patient-written review on their online profile that describes their technical skill and competence negatively, internet users are less willing to choose that doctor and hold more critical attitudes about and trust the doctor less. The opposite is true when reviews are positive. These findings underscore the results from previous studies, which reported that PRW users rely heavily on text reviews when selecting a doctor [1], and confirm that review tendency impacts their perception of the review content [62].

The manipulated review came along with a neutral second one that was not manipulated. The experimental stimulus was inconspicuously placed on the manipulated profile, and the

quantitative information on the profile was untouched. Hence, the effects found should be judged considerable. This confirms physicians' worries about the impact of even one negative, potentially unjustified review [16,73]. As a consequence, the need for better physician protection mechanisms against fake or unjustified reviews remains a priority. For example, if reviews go online only when a quorum of them is available, say 10, the weight of a single negative one will diminish and a more realistic picture of a physician's true competence [8,73] will emerge. Despite awareness of such balancing efforts, a recent content analysis on PRWs found that only a small minority of PRWs demand a quorum before reviews go online [74], and many physicians in the country have far fewer than 10 reviews on their websites [60].

On the other hand, our data and analyses do not show the slightest hint of halo effects. That is to say, attitudes toward Dr Schmidt were not affected by the reviews Dr Müller received, neither critical nor friendly. Such halo effects are not inconceivable, and they make health care consumer reactions even more unpredictable. Their absence, moreover, does not mean that a physician whose performance was not reviewed remains unaffected. The effect is on patient behavior that leads to increased or deteriorating demand for the services a physician offers.

The stronger effect of positive compared with negative reviews rests on the assumption that Dr Müller without review would receive judgments similar to Dr Schmidt without review. Given that the profiles are virtually identical, this assumption is reasonable, but we cannot be 100% sure as there were no profiles of Dr Müller without reviews. Data even show a slightly better liking of Dr Müller (MMüller 3.45, MSchmidt 3.16, t_{253}=5.755, P<.001 in paired samples t test) and somewhat more trust put in him across the total sample (MMüller 4.23, MSchmidt 3.94, t_{253}=5.755, P<.001). One difference that comes to mind could contribute to these findings. The profiles of the two physicians differed in two aspects, tendency and presence of reviews, and presence exerts an influence distinct from that of tendency, which might explain part of what appears as tendency effects in our analysis. Testing this must be left to further research. At least the stronger effect of the positive versus the negative review on physician choice corresponds well to the limitation of the dampening effect by health literacy to the group that read the negative review. If health literacy assuages the effect of negative reviews, it is no wonder that the effect of the positive review is stronger.

The major motive for this study was to see whether health literacy might dampen the effects of reviews written and uploaded by health care consumers, who might just not know enough to assess the medical side of a physician's performance. The answer is, "Yes, but…" The dampening effect was found for subjective health literacy only. Our bet beforehand was that the expected effect was more likely to occur with performance-based health literacy measures.

The background for the distinction between performance-based and perception-based health literacy is the increasing scholarly insight that health literacy is a two-faced concept that contains not only willingness to participate in health decisions but also the capability to do it. Increasing patient participation in health care will be beneficial only if patients are not only willing but also capable of playing a more active role. If you give patients more of a say in their own health care while they lack the necessary capabilities, health decisions might worsen. The relationship between the two is often blurred, but it is imperative to distinguish them properly [75].

Perception-based measures of health literacy are associated with motivation and willingness, while performance-based measures aim at the objective side, at ability. Therefore we had, as said, the bet on the performance-based NVS measure as moderating the effects of patient-written physician reviews negatively. For the perception-based measures, an explanation close at hand might be that some transfer occurred. Higher values on these measures would indicate an overestimation not only of one's own but also of other people's communication competence. As a consequence, there would be a positive moderation effect of subjective health literacy rather than the negative effect we find.

Diverging results from applying these measures raise the question of whether the objective, performance-based indicators do indeed measure the same concept as the subjective, perception-based indicators do [76]. The lack of any correlation between the two measures supports the assumption of two concepts.

The finding of different effects of health literacy on preference for Dr Müller can be explained in various ways. More literate people, particularly those who consider themselves as being more familiar with health information on the internet, are on a general level more experienced with negative reviews (not just in the context of health) and less willing to take negative information at face value. Overall, negative reviews are quite common. A presentation of the doctor as "Quite nice, but he doesn't really have a clue," combined with the critique that the doctor did not properly examine the patient and prescribed the wrong medication impressed literate people only to a minor degree. This reasoning could be challenged by the fact that most reviews on PRWs are positive.

Another explanation is that people with higher literacy skills consider themselves more capable of constructively dealing with doctors who treat them in a deficient way. Showing a higher level of self-efficacy may make literate people believe they can handle critical situations as presented in the negative review. In one word, they are less afraid of the impact a negative review might have on them, while for less literate people, the negative effect looms larger, to use the description of the value curve from prospect theory [77]. This second point implies there is no moderation effect of the NVS, the other indicator of health literacy used in this study. That is to say, this performance-based measure, other than the perception-based eHEALS measure, is free of any self-efficacy component.

If health literacy, in spite of our results, should have a potential to increase skepticism of physician reviews written by health care consumers and published anonymously, then some of this potential should show in our interview data beyond those used in the study so far. Usefulness and skepticism toward PRWs correlate with objective and subjective health literacy differently. Perceived skepticism of eWOM in the field of health, for

instance, is negatively correlated with the perception of the usefulness of PRWs (r=–.485, P<.001). That means people who are skeptical of PRWs consequently find them less useful. Performance-based health literacy is not at all correlated with skepticism (r=–.02, P=.80) or perceived usefulness (r=.04, P=.54). That means it is not true, as one could assume, that people with an objectively higher level of health literacy get more skeptic of PRWs and find them less useful. To the contrary, people found to score high on perception-based health literacy find PRWs more useful (r=.22, P<.001) and are less skeptical of them (r=–.21, P<.001), which stands in contrast to our finding that they are less willing to follow the latent advice contained in reviews. All in all, there is not much reason to assume health literacy can contribute to solve a problem it has helped create: a more autonomous patient who at times is not equipped with the abilities and knowledge necessary to get involved in decision making.

The subject of this research, PRWs and their role in physician choice, is different from most other health decisions that are studied. Health decisions are usually demanded by a patient who is ill or has some prevention concern. The doctor provides the good, but under the rule of participation policies, asks and/or encourages the patient to get involved. If the wrong decision is made, that will primarily be injurious to the patient. Physician choice is a decision the patient makes on their own, at least there is usually no physician involved. The primary damage is the physician's—the loss of business—and the decision need not even be wrong. Wrong physician choice decisions can also damage the patient if there is a wide qualification differential between the physicians the patient does not notice or heed. One would think such cases are the exception rather than the rule. If physicians are concerned about choice decisions not being based on expertise, their concern might be fueled by the fact that they can be the victim of unreasonable patient decisions themselves. If patient input in usual health care decisions is suboptimal, the damage is the patient's and the doctor is there to correct. It is no big surprise that physicians are less concerned about the quality of patient input in these cases.

Patients are skeptical of their ability to make judgments on the professional expertise of physicians and yet still pay attention to and draw conclusions from them. With PRWs, however, health care consumers talk to people who are just like themselves. That is a rare thing or was until the internet and especially Web 2.0 apps. "I might have some qualms about my and my fellow patients' ability to rate doctors, but I share with them all the goodwill, all my experience with doctors, and a lot of suffering when doctors could not help." Feelings like these might indirectly give physician ratings some legitimacy and value in the eyes of patients.

As to the results on the moderation effect of health literacies, they foremost suggest that performance-based and perception-based health literacies are to be separated more strictly than has been done in the past. One of the drawbacks of patient participation policy might well be that some patients actually lack a basis for adequate interference or involvement but still insist on having a say in medical decisions in the widest sense. It is this group that might be most at risk of suffering detrimental consequences.

Limitations

This study faces limitations concerning its design and data collection. Specifically, the data were collected via an online platform where participants take part in studies to gain rewards (eg, vouchers, airline miles). Hence, it cannot be ruled out that such data collection attracts a specific audience. Even though our sample corresponded to the Swiss population in terms of education, age, and gender [59], results should not be generalized. Additionally, in order to check the attention and comprehension of the participants, we automatically screened out individuals who did not pass the manipulation check, in which they were asked about the number of text reviews present on each of the two physician profiles. By these means we ensured that the review on the assessment of the doctor's competence was noticed. As the contracting firm Qualtrics removed those participants automatically, we are not able to report on the number of individuals who did not pass the manipulation check.

Another limitation of the study was based on randomization. Even though we randomized the order in which the two physician profiles were presented, we did not randomize the order of the text reviews on Dr Müller's profile. This may have impacted participant choice due to primacy effects [62]. In order to account for primacy effects, future studies should control for this or randomize the order.

In addition, the subjective health literacy measure explicitly aims at online capabilities and is thus substantially close to online physician rating, while the objective measure of health literacy by understanding a food label has no such proximity. The subjective measure better applies to the context of online information and credibility judgments than does the NVS. So if people are not completely amiss in assessing their own skills, the relationship between high eHEALS scores and reasonable reaction to one bad review of Dr Müller might be the consequence of the closeness of the situation and measure rather than the subjective nature of the measure. In hindsight, it was probably not a good idea to employ only 7 of the 8 items on the eHEALS, although we doubt that the omission of one item in an 8-item measure would have changed the substance of the analysis much.

Conclusion

Text reviews that assess a physician's competence weigh heavily on the choice of physician on PRWs, even though health care consumers have voiced skepticism toward the truthfulness of such reviews and doubt their own and fellow patients' capabilities to assess health care and physicians accurately [28,29]. Our study evidenced that patient-written reviews impact health care consumer choices and attitudes toward a physician and affect their perceptions of the doctor's skill and trustworthiness. These findings sustain physician worries that even one negative PRW review can be highly damaging to their reputation [71]. Furthermore, these results put forward that internet users select health care providers based on patient-written reviews that contain information only weakly, selectively, or not at all related to objective measures of care quality according to various studies [24,78]. Hopes that health literacy might raise awareness of the insufficient basis of the

rating of physicians' medical performance by patients have mostly not been sustained. Rather, the divergence of subjective and objective health literacy might compound the problem.

Acknowledgments

The authors are grateful to all participants for their precious time and inputs. We would also like to thank the Mütter- und Väterberatung Luzern, as well as the nurseries Eichhörnli, Maluum, Müsliburg, and Rösslispiel for their support in the recruitment of participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Instructions presented to participants before viewing the physician profiles (translated from German).
[DOCX File , 12 KB - jmir_v22i4e14134_app1.docx]

Multimedia Appendix 2
Questionnaire examining users' response to physicians' reviews and rating on websites.
[DOCX File , 1339 KB - jmir_v22i4e14134_app2.docx]

References

1. Dellarocas C. The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag Sci 2003 Oct;49(10):1407-1424. [doi: 10.1287/mnsc.49.10.1407.17308]
2. Steffes EM, Burgee LE. Social ties and online word of mouth. Internet Res 2009 Jan 30;19(1):42-59. [doi: 10.1108/10662240910927812]
3. Christodoulides G. Branding in the post-internet era. Marketing Theory 2009 Mar;9(1):141-144. [doi: 10.1177/1470593108100071]
4. Gamble J, Gilmore A. A new era of consumer marketing? Eur J Marketing 2013 Nov 11;47(11/12):1859-1888. [doi: 10.1108/ejm-10-2011-0584]
5. Kuehn K. "There's got to be a review democracy": communicative capitalism, neoliberal citizenship and the politics of participation on the consumer evaluation website Yelp.com. Int J Comm 2013;7(1):607-625.
6. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
7. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
8. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
9. Schlesinger M, Kanouse DE, Martino SC, Shaller D, Rybowski L. Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. Med Care Res Rev 2014 Oct;71(5 Suppl):38S-64S [FREE Full text] [doi: 10.1177/1077558713496321] [Medline: 23999489]
10. Emmert M, Sauter L, Jablonski L, Sander U, Taheri-Zadeh F. Do physicians respond to web-based patient ratings? An analysis of physicians' responses to more than one million web-based ratings over a six-year period. J Med Internet Res 2017 Jul 26;19(7):e275 [FREE Full text] [doi: 10.2196/jmir.7538] [Medline: 28747292]
11. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ 2009;338:b1030. [Medline: 19293223]
12. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ 2009;338:b1033. [Medline: 19293224]
13. Samora JB, Lifchez SD, Blazar PE. Physician-rating web sites: ethical implications. J Hand Surg Am 2016 Jan;41(1):104-110. [doi: 10.1016/j.jhsa.2015.05.034] [Medline: 26304734]
14. Strech D. [Physician rating sites from an ethical viewpoint: a roadmap]. Z Evid Fortbild Qual Gesundhwes 2010;104(8-9):674-681. [doi: 10.1016/j.zefq.2010.09.019] [Medline: 21129706]
15. Holliday AM, Kachalia A, Meyer GS, Sequist TD. Physician and patient views on public physician rating websites: a cross-sectional study. J Gen Intern Med 2017 Jun;32(6):626-631. [doi: 10.1007/s11606-017-3982-5] [Medline: 28150098]
16. Patel S, Cain R, Neailey K, Hooberman L. General practitioners' concerns about online patient feedback: findings from a descriptive exploratory qualitative study in England. J Med Internet Res 2015 Dec 08;17(12):e276 [FREE Full text] [doi: 10.2196/jmir.4989] [Medline: 26681299]

17. Glickman SW, Boulding W, Manary M, Staelin R, Roe MT, Wolosin RJ, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):188-195. [doi: 10.1161/CIRCOUTCOMES.109.900597] [Medline: 20179265]
18. Kane RL, Maciejewski M, Finch M. The relationship of patient satisfaction with care and clinical outcomes. Med Care 1997 Jul;35(7):714-730. [doi: 10.1097/00005650-199707000-00005] [Medline: 9219498]
19. Sacks GD, Lawson EH, Dawes AJ, Russell MM, Maggard-Gibbons M, Zingmond DS, et al. Relationship between hospital performance on a patient satisfaction survey and surgical quality. JAMA Surg 2015 Sep;150(9):858-864. [doi: 10.1001/jamasurg.2015.1108] [Medline: 26108091]
20. Chang JT, Hays RD, Shekelle PG, MacLean CH, Solomon DH, Reuben DB, et al. Patients' global ratings of their health care are not associated with the technical quality of their care. Ann Intern Med 2006 May 2;144(9):665-672. [Medline: 16670136]
21. Emmert M, Adelhardt T, Sander U, Wambach V, Lindenthal J. A cross-sectional study assessing the association between online ratings and structural and quality of care measures: results from two German physician rating websites. BMC Health Serv Res 2015;15(1):414 [FREE Full text] [doi: 10.1186/s12913-015-1051-5] [Medline: 26404452]
22. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
23. Segal J. Physician News. 2009 May 05. Online doctor ratings or physician bashing: there has got to be a better way URL: https://physiciansnews.com/2009/05/05/online-doctor-ratings-or-physician-bashing-there-has-got-to-be-a-better-way/ [accessed 2020-01-22]
24. Chen J, Presson AP, Zhang C, Ray DM, Finlayson SR, Glasgow RE. Online physician scorecards measure patient satisfaction and perception of experience rather than quality of care. J Am Coll Surg 2016 Oct;223(4):S113. [doi: 10.1016/j.jamcollsurg.2016.06.234]
25. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
26. Tièche R, da Costa BR, Streit S. Patient satisfaction is biased by renovations to the interior of a primary care office: a pretest-posttest assessment. BMC Health Serv Res 2016 Aug 11;16(1):373. [doi: 10.1186/s12913-016-1647-4]
27. Ludolph R, Allam A, Schulz PJ. Manipulating Google's knowledge graph box to counter biased information processing during an online search on vaccination: application of a technological debiasing strategy. J Med Internet Res 2016 Jun 02;18(6):e137 [FREE Full text] [doi: 10.2196/jmir.5430] [Medline: 27255736]
28. Rothenfluh F, Germeni E, Schulz PJ. Consumer decision-making based on review websites: are there differences between choosing a hotel and choosing a physician? J Med Internet Res 2016 Jun 16;18(6):e129 [FREE Full text] [doi: 10.2196/jmir.5580] [Medline: 27311623]
29. Rothenfluh F, Schulz PJ. Physician rating websites: what aspects are important to identify a good doctor, and are patients capable of assessing them? a mixed-methods approach including physicians' and health care consumers' perspectives. J Med Internet Res 2017 May 01;19(5):e127 [FREE Full text] [doi: 10.2196/jmir.6875] [Medline: 28461285]
30. Jans LCH, Kranzbühler A. The influence of rating volume in the effects of expert versus patient online ratings. Acta Orthop Belg 2015 Dec;81(4):662-667. [Medline: 26790788]
31. Cho YI, Lee SD, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. Soc Sci Med 2008 Apr;66(8):1809-1816. [doi: 10.1016/j.socscimed.2008.01.003] [Medline: 18295949]
32. Bensing J. Bridging the gap: the separate worlds of evidence based medicine and patient-centered medicine. Patient Educ Couns 2000 Jan;39(1):17-25. [doi: 10.1016/s0738-3991(99)00087-7]
33. Kutner M, Greenburg E, Jin Y, Paulsen C. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. Washington: National Center for Education Statistics; 2006.
34. Rudd RE. Health literacy skills of U.S. adults. Am J Health Behav 2007;31 Suppl 1:S8-S18. [doi: 10.5555/ajhb.2007.31.supp.S8] [Medline: 17931141]
35. Federman AD, Wolf MS, Sofianou A, Martynenko M, O'Connor R, Halm EA, et al. Self-management behaviors in older adults with asthma: associations with health literacy. J Am Geriatr Soc 2014 May;62(5):872-879 [FREE Full text] [doi: 10.1111/jgs.12797] [Medline: 24779482]
36. van der Heide I, Uiters E, Rademakers J, Struijs JN, Schuit AJ, Baan CA. Associations among health literacy, diabetes knowledge, and self-management behavior in adults with diabetes: results of a dutch cross-sectional study. J Health Commun 2014;19 Suppl 2:115-131. [doi: 10.1080/10810730.2014.936989] [Medline: 25315588]
37. White S. Relationship of preventive health practices and health literacy: a national study. Am J Health Behav 2008;32(3):227-242. [doi: 10.5993/ajhb.32.3.1]
38. Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. Arch Intern Med 2005 Sep 26;165(17):1946-1952. [doi: 10.1001/archinte.165.17.1946] [Medline: 16186463]
39. van der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J Health Commun 2013;18 Suppl 1:172-184 [FREE Full text] [doi: 10.1080/10810730.2013.825668] [Medline: 24093354]

40. Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med 2007 Jul 23;167(14):1503-1509. [doi: 10.1001/archinte.167.14.1503] [Medline: 17646604]
41. Thai AL, George M. The effects of health literacy on asthma self-management. J Asthma Allerg Educ 2010 Apr 30;1(2):50-55. [doi: 10.1177/2150129710368850]
42. Nielsen-Bohlman L, Panzer A, Kindig D. Health Literacy: A Prescription to End Confusion. Washington: National Academies Press; 2004.
43. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
44. Vaish A, Grossmann T, Woodward A. Not all emotions are created equal: the negativity bias in social-emotional development. Psychol Bull 2008 May;134(3):383-403 [FREE Full text] [doi: 10.1037/0033-2909.134.3.383] [Medline: 18444702]
45. Lee M, Youn S. Electronic word of mouth (eWOM). Int J Advert 2015 Jan 07;28(3):473-499. [doi: 10.2501/s0265048709200709]
46. Herr PM, Kardes FR, Kim J. Effects of word-of-mouth and product-attribute information on persuasion: an accessibility-diagnosticity perspective. J Consum Res 1991 Mar;17(4):454-462. [doi: 10.1086/208570]
47. Lee K, Koo D. Effects of attribute and valence of e-WOM on message adoption: moderating roles of subjective knowledge and regulatory focus. Comput Hum Behav 2012;28:1974-1984 [FREE Full text] [doi: 10.1016/j.chb.2012.05.018]
48. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015;15:416 [FREE Full text] [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]
49. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Parental awareness and use of online physician rating sites. Pediatrics 2014 Oct;134(4):e966-e975 [FREE Full text] [doi: 10.1542/peds.2014-0681] [Medline: 25246629]
50. Rischer R, Emmert M, Palmowski S, Meszmer N. Spotlight Gesundheit—Thema: Arztbewertungen. 2017. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV_SpotGes_Arztbewertung_dt_web.pdf [accessed 2020-01-22]
51. Black EW, Thompson LA, Saliba H, Dawson K, Black NMP. An analysis of healthcare providers' online ratings. Inform Prim Care 2009;17(4):249-253 [FREE Full text] [Medline: 20359403]
52. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 May 13;25(9):942-946. [doi: 10.1007/s11606-010-1383-0]
53. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
54. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a sports medicine surgeon's. HSS J 2016 Oct;12(3):272-277. [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
55. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? An analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
56. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
57. Walczewski M, Wieczorek M. Identifying patients' criteria for assessment of doctors on Polish physician rating websites. J Med Sci 2015;84(3):189-196.
58. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
59. Bundesamt für Statistik. 2015. Bildungsstand der Bevölkerung: Bildungsstand nach Wohnbevölkerung nach Alter und Geschlecht URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungsstand-kompetenzen/bevoelkerung.html [accessed 2017-04-12]
60. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. JAMA 2017 Dec 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]
61. McLennan S, Strech D, Reimann S. Developments in the frequency of ratings and evaluation tendencies: a review of German physician rating websites. J Med Internet Res 2017 Aug 25;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]
62. Li S, Feng B, Chen M, Bell RA. Physician review websites: effects of the proportion and position of negative reviews on readers' willingness to choose the doctor. J Health Commun 2015 Apr;20(4):453-461. [doi: 10.1080/10810730.2014.977467] [Medline: 25749406]
63. 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]
64. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [Medline: 14528569]

65. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
66. Soellner R, Huber S, Reder M. The concept of ehealth literacy and its measurement: German translation of the eHEALS. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
67. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
68. Diviani N, Dima AL, Schulz PJ. A psychometric analysis of the Italian version of the eHealth Literacy Scale using item response and classical test theory methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
69. Grabner-Kräuter S, Waiguny MKJ. Insights into the impact of online physician reviews on patients' decision making: randomized experiment. J Med Internet Res 2015 Apr 09;17(4):e93 [FREE Full text] [doi: 10.2196/jmir.3991] [Medline: 25862516]
70. Zhang X, Ko M, Carpenter D. Development of a scale to measure skepticism toward electronic word-of-mouth. Comput Hum Behav 2016 Mar;56:198-208. [doi: 10.1016/j.chb.2015.11.042]
71. Tabachnick B, Fidell L. Cleaning up your act: Screening data prior to analysis. In: Using Multivariate Statistics, 5th Edition. Boston: Pearson Education, Inc; 2008:60-114.
72. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, 2nd Edition. New York: Guilford Press; 2018.
73. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
74. Rothenfluh F, Schulz PJ. Content, quality, and assessment tools of physician-rating websites in 12 countries: quantitative analysis. J Med Internet Res 2018 Jun 14;20(6):e212 [FREE Full text] [doi: 10.2196/jmir.9105] [Medline: 29903704]
75. 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]
76. Kiechle ES, Bailey SC, Hedlund LA, Viera AJ, Sheridan SL. Different measures, different outcomes? A systematic review of performance-based versus self-reported measures of health literacy and numeracy. J Gen Intern Med 2015 Oct;30(10):1538-1546 [FREE Full text] [doi: 10.1007/s11606-015-3288-4] [Medline: 25917656]
77. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979 Mar;47(2):263-291. [doi: 10.2307/1914185]
78. Adelhardt T, Emmert M, Sander U, Wambach V, Lindenthal J. Can patients rely on results of physician rating websites when selecting a physician? A cross-sectional study assessing the association between online ratings and structural and quality of care measures from two German physician rating websites. Value in health? J Int Soc Pharmacoeconom Outcomes Res 2015 Nov;18(7):A545. [doi: 10.1016/j.jval.2015.09.1732]

Abbreviations

ANOVA: analysis of variance
eHEALS: eHealth Literacy Scale
eWOM: electronic word of mouth
NVS: newest vital sign
PRW: physician rating website

Edited by G Eysenbach; submitted 25.03.19; peer-reviewed by C Meppelink, S McLennan, S Bidmon; comments to author 05.06.19; revised version received 10.07.19; accepted 19.12.19; published 06.04.20.

Please cite as:
Schulz PJ, Rothenfluh F
Influence of Health Literacy on Effects of Patient Rating Websites: Survey Study Using a Hypothetical Situation and Fictitious Doctors
J Med Internet Res 2020;22(4):e14134
URL: https://www.jmir.org/2020/4/e14134
doi:10.2196/14134
PMID:32250275

Original Paper

Exploring the Factors Influencing Consumers to Voluntarily Reward Free Health Service Contributors in Online Health Communities: Empirical Study

Junjie Zhou[1], PhD; Fang Liu[2], PhD; Tingting Zhou[3], MPhil

[1]Shantou University Business School, Shantou, China

[2]China Life Property & Casualty Insurance Company Limited, Beijing, China

[3]Henan Foreign Trade School, Zhengzhou, China

Corresponding Author:
Fang Liu, PhD
China Life Property & Casualty Insurance Company Limited
No 16 Jinrong Street
Beijing
China
Phone: 86 1066190179
Email: liu-fang@chinalife-p.com.cn

Abstract

Background: Rewarding health knowledge and health service contributors with money is one possible approach for the sustainable provision of health knowledge and health services in online health communities (OHCs); however, the reasons why consumers voluntarily reward free health knowledge and health service contributors are still underinvestigated.

Objective: This study aimed to address the abovementioned gap by exploring the factors influencing consumers' voluntary rewarding behaviors (VRBs) toward contributors of free health services in OHCs.

Methods: On the basis of prior studies and the cognitive-experiential self-theory (CEST), we incorporated two health service content–related variables (ie, informational support and emotional support) and two interpersonal factors (ie, social norm compliance and social interaction) and built a proposed model. We crawled a dataset from a Chinese OHC for mental health, coded it, extracted nine variables, and tested the model with a negative binomial model.

Results: The data sample included 2148 health-related questions and 12,133 answers. The empirical results indicated that the effects of informational support (β=.168; *P*<.001), emotional support (β=.463; *P*<.001), social norm compliance (β=.510; *P*<.001), and social interaction (β=.281; *P*<.001) were significant. The moderating effects of social interaction on informational support (β=.032; *P*=.02) and emotional support (β=−.086; *P*<.001) were significant. The moderating effect of social interaction on social norm compliance (β=.014; *P*=.38) was insignificant.

Conclusions: Informational support, emotional support, social norm compliance, and social interaction positively influence consumers to voluntarily reward free online health service contributors. Social interaction enhances the effect of informational support but weakens the effect of emotional support. This study contributes to the literature on knowledge sharing in OHCs by exploring the factors influencing consumers' VRBs toward free online health service contributors and contributes to the CEST literature by verifying that the effects of experiential and rational systems on individual behaviors can vary while external factors change.

(J Med Internet Res 2020;22(4):e16526) doi:10.2196/16526

KEYWORDS

telemedicine; health services; social media; reward; social interaction; social support; pay-what-you-want

Introduction

Background

With the development of information and communication technologies (ICTs), the sharing economy (SE) has emerged as a market for collaborative consumption in which peer communities gain access to a pool of shared knowledge and resources [1-3]. Health services, a typical kind of knowledge-intensive service [4,5], has recently become increasingly popular worldwide on many noncommercial web-based SE platforms. Such services emerge in online health communities (OHCs)—a special kind of online forums that links health care professionals and normal users [6-10]. In OHCs, health care professionals and consumers collaborate with each other to generate new health knowledge, such as disease symptoms and routine daily care discussions, health self-management experiences, or suggestions on treatments [5,11-19]. The generated knowledge will become available to the public and can be freely accessed by every consumer on online SE platforms [20,21].

Similar to many other noncommercial web-based SE platforms, OHCs are facing the sustainability issue (ie, the provision of free health knowledge and health services) [6,22-25]. In OHCs, health care professionals or enthusiastic consumers generally provide free health knowledge and health services. They voluntarily contribute their time and knowledge to the community [11,22,26]. However, both health care professionals and other free health service contributors have their own professional burnouts, duties, and responsibilities [22,27,28]. They are likely to stop providing health knowledge and health services if they lose their passion to contribute or they become busy with other duties.

Objectives

To keep the sustainable provision and sharing of free health knowledge and health services, some OHCs have designed a new feature that allows consumers to voluntarily reward free health service contributors. Such rewarding behavior is particularly important for OHCs to thrive, because the rewards act as monetary incentives that can stimulate health service providers to continuously contribute high-quality health knowledge and free health services [28-35]. However, given that the voluntary reward feature is new and consumers' rewarding behaviors are emerging, we still have little knowledge on the following questions:

1. What are the factors that motivate consumers to voluntarily reward free health service contributors in OHCs?
2. How do those factors motivate consumers to voluntarily reward free health service contributors in OHCs?

This study aimed to address the abovementioned questions. We adopted the cognitive-experiential self-theory (CEST) as the theoretical foundation and proposed seven hypotheses. We crawled an objective dataset from an OHC for mental health and verified most of the hypotheses. The empirical results indicate that informational support, emotional support, social norm compliance, and social interaction positively influence consumers to voluntarily reward free online health service contributors. Social interaction enhances the effect of informational support but weakens the effect of emotional support. These findings provided several important theoretical contributions and practical implications.

Methods

Literature Review

We reviewed two streams of related studies to address the research questions. Specifically, we reviewed the literature on free health services in OHCs to describe the characteristics of free online health services. We reviewed the literature on pay-what-you-want to understand the theories, variables, and models that were used to explain consumers' voluntary rewarding behaviors (VRBs). In this section, we have summarized the implications of prior studies.

Free Health Services in Online Health Communities

There are different types of OHCs (eg, peer communication for health care professionals, physician-patient interaction communities, and patient-patient interaction communities), and activities in different OHCs are organized differently [7,11,30,36]. In this study, we have specifically focused on freemium problem-solving communities (eg, health-related question and answer forums), in which both health care professionals and patients can participate [17,22,36,37]. In those communities, health services—eg, users' collaborative behaviors to generate new health knowledge, help consumers meet their health needs, and help consumers to reach a better state of health [8,38-41]—are usually free, and both health service providers and consumers create new values in a collaborative way [9,15,21,38]. As a voluntary behavior, providing health services is mainly motivated by prosocial factors. Prosocial factors are those factors relating to a broad range of actions intended to benefit one or more people other than oneself, such as trust, enjoyment, altruism, empathy, and reciprocity [9,11], and such factors are usually salient in noncommercial web-based communities [2,42]. Such factors are important because they enable the sustainable provision of free health services in OHCs [11,22,24].

Free web-based health services provide consumers many benefits. Consumers can conduct health-related activities in OHCs, such as health knowledge sharing and seeking (eg, recommending treatment plans and seeking health care suggestions) and health self-management [5-7,12,36,43]. They can manage their embarrassing conditions or stigmatized illnesses in OHCs and access health services without physically appearing in hospitals [22,36]. Free online health services meet consumers' needs and help them to achieve better health outcomes, such as higher electronic health literacy, increased patient empowerment, and a better quality of life [6,8,39,44-46].

The nature of free health services in OHCs can be treated as social support [6-8,19,36,47,48]. Social support refers to the individual's perceptions and experiences that they feel they are being cared for [49]. Social support could be divided into five subtypes: informational support, emotional support, network support, esteem support, and tangible support [49]. In this paper, we have particularly focused on freemium problem-solving

communities, and in such communities, consumers mainly exchange emotional support (eg, show or receive sympathy and make new friends or companionships) and informational support (eg, health knowledge seeking or contributing), whereas network support, esteem support, and tangible support are less salient. For example, many studies have shown that members in such communities do not form new subnetworks [16,36,47]. As a result, informational support and emotional support become the two most crucial aspects in the literature relating to freemium problem-solving communities [8,9,36,47,50-52]. In this paper, we have followed prior studies and adopted informational support and emotional support to describe the contents of free online health services.

Pay-What-You-Want and Voluntary Rewarding

Voluntarily rewarding free health services belongs to an emerging business model that gives consumers full control in monetizing free web-based knowledge/goods/services [33,34]. Consumers can choose to pay any amount of money or pay nothing but still enjoy free knowledge/services. Such behavior is similar to the pay-what-you-want behavior, which is "a new participative pricing mechanism in which consumers have maximum control over the price they pay" [3]. According to existing literature, firms can use pay-what-you-want pricing for two different goals: (1) commercial profit or (2) free promotion to increase knowledge and service provision on the internet [53]. In this paper, we believe that voluntary behavior is noncommercial behavior that is similar to the pay-what-you-want behaviors for the promotion of increasing knowledge/service on the internet. In such a situation, exchange partners build their relationships according to prosocial exchange norms (eg, norms of reciprocity, norms of cooperation, or norms of distribution) [54,55]. Thus, we have referred to the studies on pay-what-you-want behaviors and investigated consumers' voluntary reward behaviors from a prosocial motivation perspective [2,3,54,56,57]. We reviewed some related studies and summarized them in Table 1.

Table 1. Key constructs related to the pay-what-you-want behaviors in prior studies.

Sources	Contexts	Theory	Independent variables	Dependent variables
Kim et al [3]	Restaurant, cinema, and delicatessen	Equity theory	• Fairness, altruism, satisfaction, and loyalty • CVs[a]: price consciousness and income	Final price paid
Jang and Chu [58]	Experiments for consumers	Equity theory	• Fairness motives of individuals, self-signaling, and norm conformity	Willing to pay
León et al [59]	Travel company	Game theory	• Customer characteristics, the influence of subjective factors, and product characteristics	Payments in El trato
Hilbert and Suessmair [60]	A laboratory experiment about a travel mug	N/A[b]	• Social interaction and social norm compliance	Willing to pay
Regner [57]	An online survey about the online music label/store, Magnatune	N/A	• Social preferences, reciprocity, guilt, social norms, altruism, fairness, and social image concerns	Willing to pay
Barone et al [61]	A leadership questionnaire	N/A	• Consumer power, perceived value, and perceived self-reliance	Purchase intentions
Dorn and Suessmair [62]	Survey in several countries under three hypothetical situations where a McDonald's Big Mac was offered	N/A	• Satisfaction, income, price consciousness, reference price, high level of reputation, loyalty, altruism, fairness, social acceptance, and social norm compliance	Willing to pay
Narwal and Nayak [63]	Scenario-based online experimental approach on purchase intention	N/A	• Quality of product/services, satisfaction, types of products/services, self-image, and fairness perception • Moderators: communication message, interaction, and reference prices	Pay-what-you-want
Viglia et al [64]	Service	Fairness theory	• Timing and uncertainty reduction	Consumers' chosen payments

[a]CV: control variable.

[b]Not applicable.

Implications of Prior Literature for This Study

We concluded three useful findings according to the literature review. First, pay-what-you-want is a result of consumers' positive experiences with the services via direct interactions with service providers [59,65]. The experiences are related to factors in three domains: (1) consumer characteristics, eg, fairness motivation, income, or self-image [3,57,62], (2) product or service content–related factors, eg, price, quality, or value of services [59,61,66], and (3) interpersonal factors, eg, social interaction or social norm compliance [60,62,66].

Specific to this study, we proposed that consumers' VRBs toward free health service contributors are a result of consumers' positive experiences with the services via direct interactions with service providers in OHCs [59,65]. In addition, we incorporated informational support and emotional support as service content–related factors, social norm compliance as interpersonal factors, and social interaction to describe the communication between service providers and consumers [60,66].

Second, research focuses are shifting with time. As discussed above (please see the timeline of prior studies in Table 1 and the first paragraph of section: Implications of Prior Literature for This Study), early studies have adopted the experimental approach and mainly focus on consumer characteristics, whereas recent studies have paid more attention to service content–related factors and interpersonal factors (see also Table 1). For example, scholars have verified that the ways in which online health services are delivered are crucial in the era of ICTs [39,67], and consumers can easily be influenced by peers or friends their age [62,63]. In addition, new methodologies, such as online surveys [57,62] and econometric modeling based on objective data, are emerging [22,30]. We sought to adopt new methodologies in this paper.

Finally, there is a lack of conceptual frameworks in analyzing consumers' pay-what-you-want behaviors. Scholars tend to analyze this issue from a prosocial motivation perspective. They have adopted theories such as the equity theory and fairness theory to select influencing factors (see Table 1) rather than using them to build proposed research models. Scholars should build a conceptual framework to better explain consumers' pay-what-you-want behaviors [68].

Theoretical Foundations and Logic for Model Development

Cognitive-Experiential Self-Theory

As there is a lack of conceptual frameworks to explain consumers' pay-what-you-want behaviors [68], we incorporated a new theory, the CEST, to explore how the four selected factors influence consumers' VRBs in OHCs.

CEST is a psychological theory that argues that human beings operate with two systems: an experiential/intuitive system (hereafter referred to as the experiential system) and a rational/analytical system (hereafter referred to as the rational system) [69-71]. We noted in persuasion literature that scholars also refer to the dual-process models (ie, the elaboration-likelihood model and the heuristic-systematic model) [72,73]. These models also mentioned controlled vs automatic processes. However, these models are limited to validity-seeking persuasion contexts [73], which are not suitable in our research contexts. Specifically, they assume that the primary goal of recipients is to assess the validity of persuasive messages [73], but in our research contexts, the rewarding behaviors are voluntary, and people post the answers and discussions in OHCs to help rather than to persuade users to reward. Compared with assessing the validity, assessing the helpfulness is more important for recipients. The experiential system operates in an automatic, nonverbal, imagistic, rapid, and effortless manner, which is associated with affect or emotions. This system has also been called an automatic system [74], a natural system [75], and system 1 [76]. Compared with the experiential system, the rational system is a reasoning system that operates in a conscious, verbal, abstract, slow, and effortful manner, which is affect-free and demanding of cognitive resources [70,71]. This system has also been called an intentional system [74], an extensional system [75], and system 2 [76].

CEST is being widely used to explain consumers' web-based behaviors, including their web purchase–related decisions. For example, consumers' reactions to experiential information demonstrates a contagion effect: experiential information at the early stage can cause more similar information in the following stage, and normal consumers like to follow opinion leaders who post experiential information [77]. To avoid consumers being influenced by negative experiential information, operation teams should enhance the information or topic management in their communities [78]. In their study, Kim and Lennon [79] applied CEST to explain the effects of different product presentation formats (visual vs verbal) on consumers' attitudes toward products and their purchase intentions in an electronic-commerce context. Previous research has verified that consumers' involvement and their consequential behaviors (eg, attitude and purchase attention) are conditional upon the amount of experiential information provided by web-based sellers [80]. The abovementioned studies indicate that consumers' money-related decisions could be explained with CEST. Therefore, it is appropriate to use CEST to explain consumers' VRBs in OHCs.

Key Logic for Model Development

We built our research model based on the following logic.

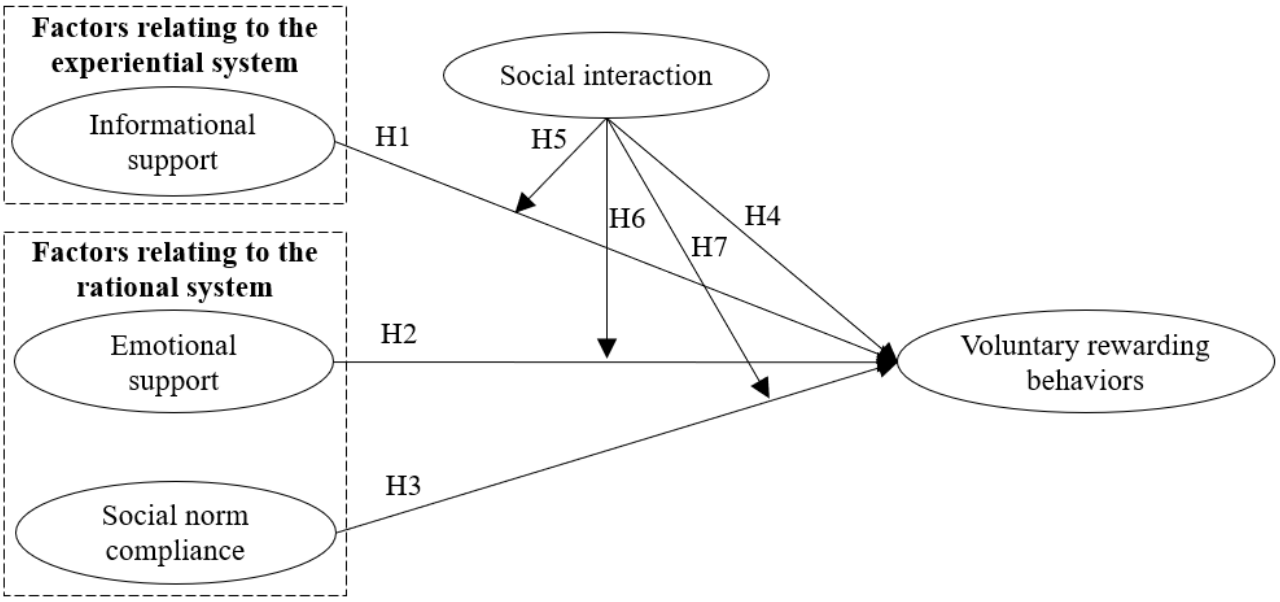
According to CEST, the rational is a verbal reasoning system—it suggests that human behaviors are driven by logic inferences from the information or evidence received [70]. As discussed earlier, informational support is one of the most important aspects of free health services in OHCs. Consumers evaluate the quality of health services they receive (eg, whether the services include useful health knowledge) and then decide how to react to these services (eg, whether to reward or not). This is a reasonable and logic-directed process. We thus used the impact of informational support to reflect the rational processing [70].

According to CEST, the experiential system is an affect-driven system—it suggests human behaviors are directed by pursuing positive feelings and avoiding negative feelings [70]. On one hand, emotional support is closely related to affect, because emotional support is a typical feeling of experience and intuition [36]. As a result, we used the impact of emotional support to reflect the experiential processing. On the other hand, consumers can observe what others do and comply with others to avoid negative results [57,62]. We thus used the impact of social norm compliance to reflect the experiential processing.

CEST also argues that the relative influence of both systems varies along a dimension of complete dominance by one system to complete dominance by the other [70]. Previous studies have verified that external factors could change the effects of experiential and rational systems [81,82], which is followed in

this study. Considering that consumers' experience of health services is influenced by the interaction between service providers and consumers (ie, social interaction) [5,39,83], we treated social interaction as an external factor and proposed that social interaction can change the effects of emotional support, social norm compliance, and informational support (see Figure 1).

Figure 1. Hypotheses and research model.



Hypotheses

Direct Effects Relating to the Rational System

Informational support refers to the overall quality and usefulness of the information received in OHCs. According to CEST, the rational system is verbal and based on the information received, so users tend to rely on rational processing when receiving informational support. Service providers and consumers usually collaboratively generate health services in the form of question and answers in OHCs. Consumers post their questions and respondents address these questions. They discuss health-related issues and generate new health knowledge in OHCs. CEST also suggests that by rational processing, consumers behave based on the logical inference from information/evidence received [70]. As a result, to better help consumers achieve logical inference, the information or knowledge quality provided in OHCs becomes important. High quality usually causes positive results. For example, high-quality information can satisfy consumers' informational needs [84,85] and motivates users to purchase [86] or to continue using web-based services [87]. Specific to the context of health services, consumers will evaluate the quality of health information they receive from free health services. As CEST suggests, if the consumers' rational processing of information suggests that it is logical and can meet their health-related needs, they will be more likely to reward these services [70]. Thus, we hypothesized the following:

> *H1: Informational support expressed in free health service threads positively influences consumers' voluntary rewarding behaviors in OHCs.*

Direct Effects Relating to the Experiential System

Emotional support refers to sympathy, ie, perceiving, understanding, and reacting to others' distress or needs [88]. According to CEST, the experiential system is emotional [70], so users tend to rely on experiential processing when receiving emotional support. As the experiential system suggests that users are motivated to pursue positive emotions and avoid negative emotions [70] when receiving emotional support, consumers' consequential behaviors (eg, rewarding decisions) are directed by their experiential processing [70]. Specifically, consumers participate in OHCs to look for patients similar to them. They can share personal experiences and exchange emotional support. Expressing care and concern could make others feel that they are being taken care of and are valued [36]. Emotional support is especially important for consumers with diseases who rely less on physical treatments. For example, consumers with mental health conditions can be alleviated with emotional direction and confession and can move to a better state of health [22,36]. Consumers in turn are likely to reward these services that provide them useful emotional support. Thus, we hypothesized the following:

> *H2: Emotional support expressed in free health service threads positively influences consumers' voluntary rewarding behaviors in OHCs.*

Social norm compliance refers to conformity to a set of norms that are accepted by a significant number of people in a social surrounding, community, or society [60,62]. The detailed norms in prior studies include altruism, reciprocity, and fairness [3,57,60]. According to CEST, the experiential system learns from prior experience, belief, or norms [70], so consumers tend to rely on experiential processing when they feel they need to comply with some social norms. Normative messages can influence people and promote prosocial behaviors [89,90]. In a free service and voluntary reward context, service providers help others without expecting economic rewards. This is a prosocial behavior and can activate the service consumers' sense

of reciprocity and fairness. CEST also suggests that the experiential system influences consumers to pursue positive emotions and avoid negative emotions [70]; therefore, we believed that social norm compliance can positively influence consumers to conduct voluntary reward behaviors to pursue positive feelings and avoid negative feelings [57,60,62,70]. Thus, we hypothesized the following:

H3: Social norm compliance positively influences consumers to voluntarily reward free health service contributors in OHCs.

The Direct Effect of Social Interaction

Social interaction refers to the observed strength of relationships, the amount of time spent, or the communication frequency among health service providers and consumers in a health service thread [39,91,92]. The application of ICTs in health care has significantly changed the context in which health service is delivered and experienced [5,83]. Consumers need to interact with service providers to better understand professional health knowledge and know how to apply it [39]. More frequent social interactions between service providers and consumers can better deliver health services and make consumers have better health outcomes [93,94]. Consumers could be grateful to service providers and thus choose to reward those free health services to feel less guilty [60,68]. Namely, social interaction drives consumers to pursue positive feelings and avoid negative feelings [70]. Thus, we hypothesized the following:

H4: Social interaction between service providers and consumers motivates consumers to voluntarily reward online free health service contributors in OHCs.

Moderating Effects of Social Interaction

CEST suggests that the extent to which people think or behave primarily according to the experiential system or rational system depends on the situation [70]. The relative influence of both systems varies along a dimension of complete dominance by one system to complete dominance by the other [70,95]. Previous studies have verified that external factors could change the effects of experiential and rational systems [81,82]. For example, in a conflict - handling context, constructive thinking together with the experiential system and rational system influences consumers' conflict - handling style [81]. In an online shopping context, consumers' involvement changes the effects of experiential information on their product attitude and purchase intention [80]. We followed the above findings and proposed that social interaction as an external factor changes the effects of the experiential system and rational system on consumers' VRBs.

OHCs are web-based social networks in which health-related stakeholders with common interests, goals, or practices interact to share health information and knowledge, communicate health services, and engage in social interaction [7,91]. It is the nature of social interaction and the resources embedded within social interaction networks that sustain the communities [91]. In OHCs, social interaction links different community members and provides them opportunities to discuss health information and knowledge [93,96]. We proposed that higher levels of social interaction can facilitate consumers to think or behave in a manner that is based more on the rational system. This is because by interacting with others, consumers can clearly express their health condition and needs, which also helps knowledge providers to better understand their needs and thus offer more useful suggestions. Consumers can then also carefully compare different information they receive. During the above mentioned process, they take time to think and logically evaluate the quality of informational support, which also slows down their decision-making process. Given that the rational system is slow and more logical, consumers' VRBs rely more on their rational system [70,95], meaning they rely more on informational support. Thus, we proposed the following:

H5: Social interaction positively moderates the effect of informational support on consumers' voluntary rewarding behaviors in OHCs.

As discussed earlier, both emotional support and social norm compliance are factors relating to the experiential system. According to CEST, because the relative influence of the experiential system and rational system varies from complete dominance by one to complete dominance by the other [70], when consumers rely more on their rational system to decide whether or not to reward, they tend to rely less on their experiential system, ie, although higher levels of social interaction make consumers rely more on informational support, it also makes users rely less heavily on emotional support and social norm compliance. In addition, when consumers are highly involved in social interaction, they pay more attention to the informational support they receive; therefore, they tend to be less influenced by their emotions and social norms [91]. Thus, we proposed the following:

H6: Social interaction negatively moderates the effect of emotional support on consumers' voluntary rewarding behaviors in OHCs.

H7: Social interaction negatively moderates the effect of social norm compliance on consumers' voluntary rewarding behaviors in OHCs.

Data Collection

To test the hypothesized model, we crawled an objective dataset from the question and answer forum on a Chinese OHC for mental health (the question and answer forum on YiXinLi). YiXinLi is a leading web-based health community for mental health in China. We focused on mental health because without mental health there can be no true physical health [97]; in addition, mental health services in China are relatively limited [98,99], and consumers usually read books or use the internet for health-related knowledge or services [100].

YiXinLi was set up in 2011 and aims to promote mental health services in China. The question and answer forum on YiXinLi, which was launched in 2014, provides free mental health services for consumers. Consumers can post their health-related questions in the question and answer system and wait for free answers. However, with the emerging trend of knowledge monetizing [33,34], the YiXinLi website launched a new feature, "Voluntary Reward," that supports the consumers' decision to reward the answers as they desire. As rewarding is voluntary, we were curious about the factors motivating consumers to

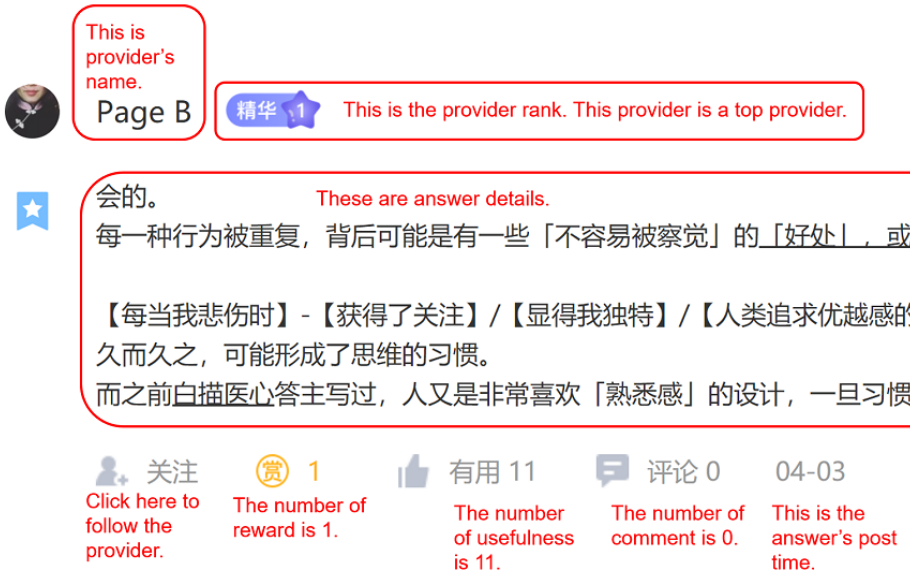
voluntarily reward free health services and the impact of those factors.

We used a spider program (named Locoy Spide) and crawled all the threads on the YiXinLi question and answer forum on January 12, 2019. We treated a question and answer thread (ie, a question and its answers) as the basic analysis unit. We cleaned the data by deleting 12 inconsistent threads—the threads in which the actual number of answers was less than the number shown on the web page because one or more answers were deleted by the providers (the number of answers displayed on the web page includes all the answers that have been provided. However, if a provider deletes his or her answer, the number shown on the web page does not change, but the actual number of answers we crawled would be less than the number shown on the web page). After cleaning the data, we had 2148 data samples, including 2148 questions and 12,133 answers. Figure 2 shows detailed information on a question posted in a question and answer thread.

As shown in Figure 2, the question and answer thread web page displays question-related information (eg, question title, question content, post time, number of page views, number of answers received, number of hugs received, and number of times favorited) at the top of the page. Figure 3 shows detailed information on answers in a question and answer thread.

Figure 2. A sample of a question.

Figure 3. A sample of an answer.



As shown in Figure 3, the question and answer thread web page displays answer-related information (eg, provider ID, provider rank, answer details, post time, number of rewards, usefulness rank, and number of comments) following the question.

Data Coding

We coded nine variables that were used for data analysis. We treated consumers' *voluntary rewarding behaviors* as the *dependent variable*. *Voluntary rewarding behaviors* was measured by the number of times a thread is rewarded. There were four key *independent variables*: *informational support*, *emotional support*, *social norm compliance*, and *socialinteraction*. Other factors such as *answer length* [101], *date of exposure*, *page view* [102], and *provider reputation* [103] also might influence consumers' rewarding behaviors and were treated as control variables in this study. Table 2 shows the details of all variables.

The descriptive statistical results of different variables are shown in Table 3.

Table 2. Variables and measurement.

Variable	Value, mean (SD)	Measurement
VRB[a]	2.141 (3.334)	• The VRB is measured by the rewarding times of a thread received. For example, the answers of the sample thread in Figure 3 received four rewards (3+1+0=4). Therefore, the value of VRB is 4 • We did not use the sum of real money that all answers received. In fact, we cannot capture the actual sum of rewarded money in a thread
IS[b]	4.375 (3.991)	• On YiXinLi, consumers can evaluate the answer quality with the feature, *usefulness*. We measured IS with the average answer quality in a thread • For example, the answer in Figure 3 have received 25 times of usefulness. And if there is another answer for the same question received 14 times of usefulness; in total, the question and answer thread received 25 times of usefulness. The value of IS is assigned as 8.333 (ie, 25/3=8.333)
ES[c]	3.274 (1.467)	• On YiXinLi, providers and other consumers can use the feature, *hugs*, to show their sympathy for help-seekers • We thus use the volume of *hugs* in a thread to measure the emotional support that help-seekers received. For example, there are six hugs in Figure 2. Thus, the value of emotional support is 6 • Although *hugs* in a thread are expressed to the help-seeker (ie, the thread poster), the empathy mechanism (ie, feeling there are patients like me) makes other consumers who have similar conditions feel that they are being taken care of and loved by others
SNC[d]	0.536 (0.61)	• SNC is measured by the percentage of people interested in the question who finally reward the question. Such a measurement reflects the peer pressure the consumers feel when they find that others have rewarded the thread they viewed. We designed this measurement according to industrial practice and prior studies. Previous literature suggests that other consumers' purchase behavior (number of goods purchased) acting as social norms influences a focal consumer's intention [104]. For example, in the electronic commerce context, Amazon designed a notification stating "15% of consumers who viewed this item have bought this item" to incent other consumers' purchase intention/behaviors; in the tax auditing context, some scholars used the rate of taxpaying (tax paid/tax owed) to measure the compliance rate (ie, other people's paying behaviors) and verified that individuals' taxpaying intention will increase when they can see a higher compliance rate [105]. We followed the above studies and measured SNC with the following equation: *SNC=VRB/(favorite+1)* • Specifically, VRB refers to the number of rewarding. The volume of *favorite* (see Figure 2) represents the number of consumers who are interested in a question. "1" represents the help-seeker himself/herself, and *(favorite+1)* represents all the people who are interested in a question. The result of *VRB/(favorite+1)* therefore represents the compliance rate (ie, the percentage of people interested in the question who finally reward the question) • For example, there are five favorites in Figure 2. Thus, the value of SNC is 0.83 (ie, 5/(5+1)=0.83)
SI[e]	8.75 (8.757)	• SI is measured by the interaction frequency between service providers and consumers in a thread. On YiXinLi, providers can respond to a question by posting their answers. Providers and consumers can also discuss a particular answer via the feature *comment* (see Figure 3). Social interaction is evaluated by the sum of answer volume and comment volume • For example, there are three answers and 0 comments in Figure 3. Thus, the value of SI is 3 (ie, 3+0=3)
AL[f]	188.4 (120.866)	• AL refers to the average text length of all answers in a thread. We calculated the character numbers of all answers and then divided the volume of answers in a thread • For example, there are three answers in a thread. The first one has 200 characters, the second one has 300 characters, and the last one has 400 characters. Thus, the value of AL is 300 (ie, (200+300+400)/3=300)
DoE[g]	73.17 (135.115)	• DoE is measured by comparing the time a question is posted with the time we crawled the dataset
PV[h]	647.985 (1918.211)	• PV refers to how many times a thread is read. • For example, the thread in Figure 2 was read 171 times. Thus, the value of *PV* is 171.
PR[i]	0.835 (0.193)	• On YiXinLi, there are 3 rank levels for a service provider, ie, normal provider, higher-rank provider, and top provider. The rank level is related to how many times their answers were set as best answers. We used the rate of higher rank/top providers of all providers in a thread to measure the PR • For example, the three providers in a thread include one normal provider, one higher-rank provider, and one top provider. Thus, the value of PR is 0.667 (ie, 2/3=0.667).

[a]VRB: voluntary rewarding behavior.

[b]IS: informational support.

[c]ES: emotional support.
[d]SNC: social norm compliance.
[e]SI: social interaction.
[f]AL: answer length.
[g]DoE: date of exposure.
[h]PV: page view.
[i]PR: provider reputation.

Table 3. Results of descriptive statistics and the covariance matrix.

Variables	Value, mean (SD)	Min	Max	VRB[a]	AL[b]	PV[c]	DoE[d]	PR[e]	IS[f]	ES[g]	SNC[h]	SI[i]
VRB	2.141 (3.334)	0	37	1	0.025	0.216[j]	−0.019	0.013	0.562[j]	0.376[j]	0.464[j]	0.490[j]
AL	188.46 (120.866)	9	894	0.025	1	−0.025	0.015	−0.071[k]	0.081[j]	0.001	0.029	0.104[j]
PV	647.985 (1918.211)	17	46,173	0.216[j]	−0.025	1	0.350[j]	−0.136[j]	0.374[j]	0.110[j]	0.026	0.302[j]
DoE	73.170 (135.115)	0	2457	−0.019	0.015	0.350[j]	1	−0.234[j]	0.113[j]	−0.052[l]	−0.031	0.164[j]
PR	.835 (.193)	0	1	0.013	−0.071[k]	−0.136[j]	−0.234[j]	1	−0.189[j]	−0.002	0.094[j]	−0.230[j]
IS	3.274 (1.467)	1	54	0.562[j]	0.081[j]	0.374[j]	0.113[j]	−0.189[j]	1	0.357[j]	0.052[l]	0.135[j]
ES	4.375 (3.991)	0	14.5	0.376[j]	0.001	0.110[j]	−0.052[l]	−0.002	0.357[j]	1	0.055[l]	0.116[j]
SNC	.536 (.610)	0	5.333	0.464[j]	0.029	0.026	−0.031	0.094[j]	0.052[l]	0.055[l]	1	0.589[j]
SI	8.75 (8.757)	1	88	0.490[j]	0.104[j]	0.302[j]	0.164[j]	−0.230[j]	0.135[j]	0.116[j]	0.589[j]	1

[a]VRB: voluntary rewarding behavior.
[b]AL: answer length.
[c]PV: page view.
[d]DoE: date of exposure.
[e]PR: provider reputation.
[f]IS: informational support.
[g]ES: emotional support.
[h]SNC: social norm compliance.
[i]SI: social interaction.
[j]$P<.001$.
[g]$P<.01$.
[j]$P<.05$.

Data Analysis

As our dependent variable (ie, *voluntary rewarding behavior*) is count data, we used count data models for our analysis [106]. As the variance value of VRB (11.114) is greater than its mean value (2.141), the distribution of the dependent variable was overdispersed, and a negative binomial (NB) model is preferred over a Poisson model [107]. NB regression relies on a log-transformation of the conditional expectation of the dependent variable and requires an exponential transformation of the estimated coefficients for assessing and interpreting the effect sizes [108]. Following econometric modeling guidelines and based on Stata 15 [106], we tested our hypotheses by using the *nbreg* model with the following equation:

$$Log(\lambda(VRB_i)) = \beta_0 + \beta_1 ArticleLength_i + \beta_2 PageView_i + \beta_3 DoE_i + \beta_4 ProviderReputation_i + \beta_5 InformationSupport_i + \beta_6 EmotionalSupport_i + \beta_7 SocialNormComplaince_i + \beta_8 SocialInteraction_i + \varepsilon_i$$

Where $\lambda_i = \exp(x_i + offset_i)$, represents a vector of parameters for the model predictors, x_i represents the i^{th} predictor, and ε_i represents the i^{th} error term.

Results

Hypothesis Test

We ran the NB model with the volume of *voluntary rewarding behaviors* as the dependent variable. The overall results indicated a good fit with a highly significant log likelihood ratio ($P<.001$ for Wald[2]; see Table 4).

Table 4. Results of the negative binomial model (N=2148).

Indices[a,b,c]	Results			
	Coefficient	SE	Z test	*P* value>Z test value
Constant	0.367[d]	0.021	17.180	<.001
Response length	−0.033[e]	0.019	−1.780	.07
Page view	0.072[d]	0.017	4.220	<.001
Date of exposure	−0.050[f]	0.022	−2.250	.02
Provider reputation	0.135[d]	0.023	5.960	<.001
Informational support	0.168[d]	0.020	8.540	<.001
Emotional support	0.463[d]	0.023	20.490	<.001
Social norm compliance	0.510[d]	0.018	28.150	<.001
Social interaction	0.281[d]	0.021	13.230	<.001
Social interaction×informational support	0.032[f]	0.013	2.410	.02
Social interaction×emotional support	−0.086[d]	0.006	−13.600	<.001
Social interaction×social norm compliance	0.014[g]	0.016	0.880	.38

[a]Log likelihood=−3130.778.

[b]Likelihood ratio$^2_{11}$=2178.5 (*P* value<.001).

[c]Pseudo R^2=0.258.

[d]*P*<.001.

[e]*P*<.1.

[f]*P*<0.05

[g]Nonsignificant.

Findings

As shown in Table 4, most hypotheses were supported (our tests are 2-tailed tests and the degree of freedom is 11). The four direct effects were significant. Informational support (β=.168; t_{11}=8.540), emotional support (β=.463; t_{11}=20.490), social norm compliance (β=.510; t_{11}=28.150), and social interaction (β=.281; t_{11}=13.230) positively influenced consumers' VRBs in OHCs. H1, H2, H3, and H4 were supported. The moderating effects of social interaction on informational support (β=.032; t_{11}=2.410) and emotional support (β=−.086; t_{11}=13.600) were significant. H5 and H6 were supported. The moderating effect of social interaction on social norm compliance (β=.014; t_{11}=0.880) was insignificant. H7 was unsupported.

Although we proposed that social interaction negatively moderates the effect of social norm compliance on consumers' VRBs, our results did not support this hypothesis. This may be because although CEST indicates such a negative moderating effect [70], other literature suggest that social interaction can provide consumers an opportunity to observe what others do [39,91], ie, the more frequently health service providers and consumers interact, the more consumers feel social pressure from others and the expectation to fit within social norms. This may be likely to enhance the effects of social norm compliance to some extent and that is why we did not observe a significant relationship empirically.

Discussion

On the basis of prior related studies and grounding our research in CEST, this study has identified two health service content–related factors and two interpersonal factors and explored how these factors influence consumers' VRBs toward free health service contributors in OHCs. Our empirical findings have demonstrated that informational support, emotional support, social norm compliance, and social interaction positively influence consumers to voluntarily reward free health service contributors. In addition, social interaction enhances the effect of informational support but weakens the effect of emotional support on consumers' VRBs toward free health service contributors in OHCs.

Theoretical Contribution

This paper makes two theoretical contributions. First, we contribute to the literature on knowledge sharing in OHCs. As noncommercial web-based SE platforms are becoming increasingly popular, scholars have begun to examine health care professionals' or consumers' health knowledge–sharing behaviors [6,9,11,22,32]. However, few studies have explored the factors influencing consumers' VRBs, which is an effective way of promoting the sustainable provision of health services in OHCs. This study has addressed this gap. On the basis of prior studies, we identified two health service content–related factors (ie, informational support and emotional support) and

two interpersonal factors (ie, social norm compliance and social interaction). On the basis of CEST, we verified that informational support, emotional support, and social norm compliance positively influence consumers' VRBs, and social interaction, as an external factor, also positively influences consumers' VRBs. Social interaction enhances the effect of informational support but weakens the effect of emotional support. Given that the VRBs toward free web-based health service contributors is so new that it has not been studied well, the abovementioned findings contribute to the research on knowledge sharing by identifying and explaining how different factors motivate consumers to voluntarily reward free health services in OHCs.

Second, our research is based on CEST and also contributes to CEST. Specifically, CEST mentioned that the extent to which individuals behave primarily according to one of the systems varies based on situations or the person himself or herself [70,95], but it did not specifically study which factor can affect such changes. Some later studies have verified the abovementioned proposition in different situations and found that external factors (eg, attraction effect and constructive thinking) do change the effects of experiential and rational systems [81,82]. This study has verified the abovementioned proposition in an OHC context. We found that social interaction together with emotional support negatively influences consumers' VRBs, but together with informational support, it positively influences consumers' VRBs. This finding extends the literature on CEST by verifying the moderating roles of a new external factor (social interaction) in a new context (OHCs).

Practical Implication

This paper has identified and verified the effects of four main variables on consumers' VRBs on free health services in OHCs. We contributed to noncommercial web-based SE platforms by providing these platform operators strategies on how to motivate consumers to voluntarily reward free service contributors.

First, platform operators could optimize their platform feature design. They can optimize the platform communication features and encourage service providers and consumers to interact with each other. In addition, they can design and implement new rewarding systems. For example, they can display the rewarding messages such as "consumer XX just rewarded provider YY some money." These rewarding messages might cause more consumers to comply with others and choose to reward free service contributors.

Second, platform operators should encourage service providers to contribute professional knowledge and generate high-quality services. They can invite more professionals or experts to use their platforms. They can help enthusiastic consumers to improve professional capabilities. The engagement of professionals and enthusiastic consumers can guarantee the quality of services on noncommercial SE platforms and can in turn attract more consumers to use their platforms and reward free service contributors.

Limitations for Future Studies

We address two potential limitations. First, we did not test the effects of consumers' sociodemographic variables and consumer characteristics. As the dataset was crawled in a public community, we could not obtain consumers' sociodemographic information and their characteristics. In addition, we measured all variables with the objective data, namely an indirect measurement approach. Second, different from prior studies that use the actual volume of money as dependent variables, we used the number of times a thread is being rewarded as the dependent variables. We are not sure whether these points undermine our conclusions or not. We appeal that more studies be conducted through the econometric modeling approach and also suggest a mixed method approach of combining objective data and subjective data in future studies.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (71501062), Key Projects of Philosophy and Social Sciences Research of Chinese Ministry of Education (grant number 19JZD021), Guangdong Provincial Science and Technology Research Project (grant number 2019A101002110), and Shantou University Scientific Research Initiation Grant (STF18011). Full control of all primary data is with the authors, and the data will be available if requested.

Conflicts of Interest

None declared.

References

1. Botsman R, Rogers R. What's Mine Is Yours: How Collaborative Consumption is Changing the Way We Live. New York City, NY: Collins; 2011.
2. Schneider D. Rewarding Prosociality on Non-Commercial Online Sharing Platforms. In: Proceedings of the 25th European Conference on Information Systems. 2017 Presented at: ECIS'17; June 5-10, 2017; Guimarães, Portugal p. 2269-2284 URL: https://pdfs.semanticscholar.org/4a63/506e761fd12e35aa2914777e5ee1fadba009.pdf
3. Kim JY, Natter M, Spann M. Pay what you want: a new participative pricing mechanism. J Mark 2009;73(1):44-58. [doi: 10.1509/jmkg.73.1.44]
4. Abidi SSR. Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. Int J Med Inform 2001 Sep;63(1-2):5-18. [doi: 10.1016/s1386-5056(01)00167-8] [Medline: 11518661]

5. Paul DL. Collaborative activities in virtual settings: a knowledge management perspective of telemedicine. J Manag Inf Syst 2006;22(4):143-176. [doi: 10.2753/mis0742-1222220406]
6. Chen L, Baird A, Straub D. Fostering participant health knowledge and attitudes: an econometric study of a chronic disease-focused online health community. J Manag Inf Syst 2019;36(1):194-229. [doi: 10.1080/07421222.2018.1550547]
7. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
8. Yao T, Zheng Q, Fan X. The impact of online social support on patients' quality of life and the moderating role of social exclusion. J Serv Res 2015;18(3):369-383. [doi: 10.1177/1094670515583271]
9. Zhao J, Wang T, Fan X. Patient value co-creation in online health communities: social identity effects on customer knowledge contributions and membership continuance intentions in online health communities. J Serv Manag 2015;26(1):72-96. [doi: 10.1108/josm-12-2013-0344]
10. 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 Mar 13;19(3):e74 [FREE Full text] [doi: 10.2196/jmir.7002] [Medline: 28288953]
11. Zhang X, Liu S, Deng Z, Chen X. Knowledge sharing motivations in online health communities: a comparative study of health professionals and normal users. Comput Hum Behav 2017;75:797-810. [doi: 10.1016/j.chb.2017.06.028]
12. Zhou J. Factors influencing people's personal information disclosure behaviors in online health communities: a pilot study. Asia Pac J Public Health 2018 Apr;30(3):286-295. [doi: 10.1177/1010539518754390] [Medline: 29405740]
13. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
14. 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]
15. 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]
16. Zhou J, Wang G, Zhou T, Fan T. The role of off-topic discussions in online health support groups: insights from a content analysis of an online rectal cancer group. Support Care Cancer 2019 Nov 13:-. [doi: 10.1007/s00520-019-05159-4] [Medline: 31720804]
17. Zhou J, Wang C. Improving cancer survivors' e-health literacy via online health communities (OHCs): a social support perspective. J Cancer Surviv 2019 Dec 09:-. [doi: 10.1007/s11764-019-00833-2] [Medline: 31820215]
18. Huang KY, Chengalur-Smith I, Pinsonneault A. Sharing is caring: social support provision and companionship activities in healthcare virtual support communities. Manag Inf Syst Q 2019;43(2):395-423. [doi: 10.25300/misq/2019/13225]
19. Plinsinga ML, Besomi M, Maclachlan L, Melo L, Robbins S, Lawford BJ, et al. Exploring the characteristics and preferences for online support groups: mixed method study. J Med Internet Res 2019 Dec 3;21(12):e15987 [FREE Full text] [doi: 10.2196/15987] [Medline: 31793893]
20. Zhang X, Wang C. Network positions and contributions to online public goods: the case of Chinese Wikipedia. J Manag Inf Syst 2012;29(2):11-40. [doi: 10.2753/mis0742-1222290202]
21. Kanthawala S, Vermeesch A, Given B, Huh J. Answers to health questions: internet search results versus online health community responses. J Med Internet Res 2016 Apr 28;18(4):e95 [FREE Full text] [doi: 10.2196/jmir.5369] [Medline: 27125622]
22. Zhou J, Zuo M, Ye C. Understanding the factors influencing health professionals' online voluntary behaviors: evidence from YiXinLi, a Chinese online health community for mental health. Int J Med Inform 2019 Oct;130:103939. [doi: 10.1016/j.ijmedinf.2019.07.018] [Medline: 31434043]
23. Ho SY, Rai A. Continued voluntary participation intention in firm-participating open source software projects. Inform Syst Res 2017;28(3):603-625. [doi: 10.1287/isre.2016.0687]
24. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013 Jun 11;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
25. Kokkodis M, Lappas T, Ransbotham S. From lurkers to workers: predicting voluntary contribution and community welfare. SSRN Journal 2020:Online First. [doi: 10.2139/ssrn.3467256]
26. Yada N, Head M. Attitudes toward health care virtual communities of practice: survey among health care workers. J Med Internet Res 2019 Dec 4;21(12):e15176 [FREE Full text] [doi: 10.2196/15176] [Medline: 31799934]
27. 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]
28. Wu H, Liu L, Wang Y, Gao F, Zhao X, Wang L. Factors associated with burnout among Chinese hospital doctors: a cross-sectional study. BMC Public Health 2013 Aug 29;13:786 [FREE Full text] [doi: 10.1186/1471-2458-13-786] [Medline: 23985038]
29. Burtch G, Hong Y, Bapna R, Griskevicius V. Stimulating online reviews by combining financial incentives and social norms. Manage Sci 2017;64(5):2065-2082. [doi: 10.1287/mnsc.2016.2715]

30. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manag Inf Syst 2017;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
31. Maynard A. Is doctors' self interest undermining the National Health Service. Br Med J 2007 Feb 3;334(7587):234 [FREE Full text] [doi: 10.1136/bmj.39066.452847.68] [Medline: 17272562]
32. Wang JN, Chiu YL, Yu H, Hsu YT. Understanding a nonlinear causal relationship between rewards and physicians' contributions in online health care communities: longitudinal study. J Med Internet Res 2017 Dec 21;19(12):e427 [FREE Full text] [doi: 10.2196/jmir.9082] [Medline: 29269344]
33. Zhao YC, Peng X, Liu Z, Song S, Hansen P. Factors that affect asker's pay intention in trilateral payment‐based social Q&A platforms: from a benefit and cost perspective. J Assoc Inf Sci Technol 2019 Jun 20:Online First. [doi: 10.1002/asi.24262]
34. Jing D, Jin Y, Liu J. The impact of monetary incentives on physician prosocial behavior in online medical consulting platforms: Evidence from china. J Med Internet Res 2019 Jul 26;21(7):e14685 [FREE Full text] [doi: 10.2196/14685] [Medline: 31350834]
35. Yang H, Zhang X. Investigating the effect of paid and free feedback about physicians' telemedicine services on patients' and physicians' behaviors: panel data analysis. J Med Internet Res 2019 Mar 22;21(3):e12156 [FREE Full text] [doi: 10.2196/12156] [Medline: 30900997]
36. Yan L, Tan Y. Feeling blue? Go online: an empirical study of social support among patients. Inf Syst Res 2014 Dec;25(4):667-891. [doi: 10.1287/isre.2014.0538]
37. Khurana S, Qiu L, Kumar S. When a doctor knows, it shows: an empirical analysis of doctors' responses in a Q&A forum of an online healthcare portal. Inf Syst Res 2019;30(3):872-891. [doi: 10.1287/isre.2019.0836]
38. Barrett M, Oborn E, Orlikowski W. Creating value in online communities: the sociomaterial configuring of strategy, platform, and stakeholder engagement. Inf Syst Res 2016;27(4):704-723. [doi: 10.1287/isre.2016.0648]
39. Zhou J, Fan T. Understanding the factors influencing patient e-health literacy in online health communities (OHCS): a social cognitive theory perspective. Int J Environ Res Public Health 2019 Jul 10;16(14):pii: E2455 [FREE Full text] [doi: 10.3390/ijerph16142455] [Medline: 31295924]
40. Costello RE, Anand A, Evans MJ, Dixon WG. Associations between engagement with an online health community and changes in patient activation and health care utilization: longitudinal web-based survey. J Med Internet Res 2019 Aug 29;21(8):e13477 [FREE Full text] [doi: 10.2196/13477] [Medline: 31469082]
41. Carron-Arthur B, Ali K, Cunningham JA, Griffiths KM. From help-seekers to influential users: a systematic review of participation styles in online health communities. J Med Internet Res 2015 Dec 1;17(12):e271 [FREE Full text] [doi: 10.2196/jmir.4705] [Medline: 26627369]
42. Batson CD, Powell AA. Altruism and prosocial behavior. In: Millon T, Lerner MJ, Weiner IB, editors. Handbook of Psychology: Personality and Social Psychology, Volume 5. Hoboken, NJ: John Wiley & Sons Inc; 2003:463-484.
43. Bordoloi P, Islam N. Knowledge management practices and healthcare delivery: a contingency framework. Electron J Knowl Manag 2012;10(2):110-120 [FREE Full text]
44. Anderson RM, Funnell MM. Patient empowerment: myths and misconceptions. Patient Educ Couns 2010 Jun;79(3):277-282 [FREE Full text] [doi: 10.1016/j.pec.2009.07.025] [Medline: 19682830]
45. Barbour L, Armstrong R, Condron P, Palermo C. Communities of practice to improve public health outcomes: a systematic review. J Knowl Manag 2018;22(2):326-343. [doi: 10.1108/jkm-03-2017-0111]
46. Bianchini C, Truccolo I, Bidoli E, Mazzocut M. Avoiding misleading information: a study of complementary medicine online information for cancer patients. Librar Inf Sci Res 2019;41(1):67-77. [doi: 10.1016/j.lisr.2019.02.006]
47. Huang KY, Chengalur-Smith I, Ran W. Not just for support: companionship activities in healthcare virtual support communities. Commun Assoc Inf Syst 2014;34:561-594. [doi: 10.17705/1cais.03429]
48. Choi MJ, Kim SH, Lee S, Kwon BC, Yi JS, Choo J, et al. Toward predicting social support needs in online health social networks. J Med Internet Res 2017 Aug 2;19(8):e272 [FREE Full text] [doi: 10.2196/jmir.7660] [Medline: 28768609]
49. Cutrona CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Commun Res 1992;19(2):154-174. [doi: 10.1177/009365092019002002]
50. Nambisan P, Gustafson DH, Hawkins R, Pingree S. Social support and responsiveness in online patient communities: impact on service quality perceptions. Health Expect 2016 Feb;19(1):87-97 [FREE Full text] [doi: 10.1111/hex.12332] [Medline: 25556856]
51. Oh HJ, Ozkaya E, LaRose R. How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction. Comput Hum Behav 2014;30:69-78. [doi: 10.1016/j.chb.2013.07.053]
52. Yan LL, Tan Y. The consensus effect in online health-care communities. J Manag Inf Syst 2017;34(1):11-39. [doi: 10.1080/07421222.2017.1296742]
53. Groening C, Mills P. A guide to pay-what-you-wish pricing from the consumer's viewpoint. Bus Horiz 2017;60(4):441-445. [doi: 10.1016/j.bushor.2017.03.004]
54. Heyman J, Ariely D. Effort for payment. A tale of two markets. Psychol Sci 2004 Nov;15(11):787-793. [doi: 10.1111/j.0956-7976.2004.00757.x] [Medline: 15482452]

55. Shampanier K, Mazar N, Ariely D. Zero as a special price: The true value of free products. Mark Sci 2007;26(6):742-757. [doi: 10.1287/mksc.1060.0254]
56. Krzyżanowska M, Tkaczyk J. Giving customers full control over prices: state of the research in marketing. Handel Wewnętrzny 2018;4(375):177-185 [FREE Full text]
57. Regner T. Why consumers pay voluntarily: evidence from online music. J Behav Exp Econ 2015;57:205-214. [doi: 10.1016/j.socec.2014.10.006]
58. Jang H, Chu W. Are consumers acting fairly toward companies?: an examination of pay-what-you-want pricing. J Macromarketing 2012;32(4):348-360. [doi: 10.1177/0276146712448193]
59. León FJ, Noguera JA, Tena-Sánchez J. How much would you like to pay? Trust, reciprocity and prosocial motivations in El trato. Soc Sci Inform 2012;51(3):389-417. [doi: 10.1177/0539018412441756]
60. Hilbert LP, Suessmair A. The effects of social interaction and social norm compliance in pay-what-you-want situations. Am J Ind Bus Manag 2015;5(8):548-556. [doi: 10.4236/ajibm.2015.58054]
61. Barone MJ, Bae TJ, Qian S, d'Mello J. Power and the appeal of the deal: how consumers value the control provided by Pay What You Want (PWYW) pricing. Mark Lett 2017;28(3):437-447. [doi: 10.1007/s11002-017-9425-6]
62. Dorn T, Suessmair A. Determinants in pay-what-you-want pricing decisions—a cross-country study. Am J Ind Bus Manag 2017 Feb;7(2):115-142. [doi: 10.4236/ajibm.2017.72010]
63. Narwal P, Nayak JK. How consumers respond to social norms: an evidence from pay-what-you-want (PWYW) pricing. J Consum Mark 2019;36(4):494-505. [doi: 10.1108/jcm-05-2018-2677]
64. Viglia G, Maras M, Schumann J, Navarro-Martinez D. Paying before or paying after? Timing and uncertainty in pay-what-you-want pricing. J Serv Res 2019;22(3):272-284. [doi: 10.1177/1094670519835308]
65. Weisstein FL, Kukar-Kinney M, Monroe KB. Determinants of consumers' response to pay-what-you-want pricing strategy on the internet. J Bus Res 2016;69(10):4313-4320. [doi: 10.1016/j.jbusres.2016.04.005]
66. Chung JY. Price fairness and PWYW (pay what you want): a behavioral economics perspective. J Revenue Pricing Manag 2017;16(1):40-55. [doi: 10.1057/s41272-017-0078-0]
67. Lee D, Hosanagar K, Nair HS. Advertising content and consumer engagement on social media: evidence from Facebook. Manag Sci 2018;64(11):5105-5131. [doi: 10.1287/mnsc.2017.2902]
68. Kunter M. Exploring the pay-what-you-want payment motivation. J Bus Res 2015;68(11):2347-2357. [doi: 10.1016/j.jbusres.2015.03.044]
69. Dane E, Pratt MG. Exploring intuition and its role in managerial decision making. Acad Manag Rev 2007;32(1):33-54. [doi: 10.5465/amr.2007.23463682]
70. Epstein S. Demystifying intuition: what it is, what it does, and how it does it. Psychol Inq 2010;21(4):295-312. [doi: 10.1080/1047840x.2010.523875]
71. Epstein S. Cognitive-experiential self-theory: an integrative theory of personality. In: Curtis RC, editor. The Relational Self: Theoretical Convergences in Psychoanalysis and Social Psychology. New York, New York, United States: The Guilford Press; 1991:111-137.
72. Cacioppo J, Petty R. The elaboration likelihood model of persuasion. Advances in Consumer Research 1984;11(1):673-675. [doi: 10.1016/s0065-2601(08)60214-2]
73. Chaiken S, Eagly AH, Liberman A. Heuristic and systematic information processing within and beyond the persuasion context. In: Uleman JS, Baugh JA, editors. Unintended Thought. New York, NY: Guilford Press; 1989:212-252.
74. Bargh JA, Chartrand TL. The unbearable automaticity of being. Am Psychol 1999;54(7):462-479. [doi: 10.1037/0003-066x.54.7.462]
75. Tversky A, Kahneman D. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol Rev 1983;90(4):293-315. [doi: 10.1037/0033-295x.90.4.293]
76. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. Am Psychol 2003 Sep;58(9):697-720. [doi: 10.1037/0003-066X.58.9.697] [Medline: 14584987]
77. Huang L. Social contagion effects in experiential information exchange on bulletin board systems. J Mark Manag 2010;26(3-4):197-212. [doi: 10.1080/02672571003594770]
78. Thompson K, Haigh L. Perceptions of equitation science revealed in an online forum: improving equine health and welfare by communicating science to equestrians and equestrian to scientists. J Veterinary Behav 2018;25:1-8. [doi: 10.1016/j.jveb.2018.02.002]
79. Kim M, Lennon S. The effects of visual and verbal information on attitudes and purchase intentions in internet shopping. Psychol Mark 2008;25(2):146-178. [doi: 10.1002/mar.20204]
80. Chang ML, Wu WY. Revisiting perceived risk in the context of online shopping: an alternative perspective of decision-making styles. Psychol Mark 2012;29(5):378-400. [doi: 10.1002/mar.20528]
81. Cerni T, Curtis GJ, Colmar SH. Cognitive - experiential self theory and conflict - handling styles: rational and constructive experiential systems are related to the integrating and compromising conflict - handling styles. Int J Confl Manag 2012;23(4):362-381. [doi: 10.1108/10444061211267263]
82. Mao W, Oppewal H. The attraction effect is more pronounced for consumers who rely on intuitive reasoning. Mark Lett 2012;23(1):339-351. [doi: 10.1007/s11002-011-9157-y]

83. Ostrom AL, Parasuraman A, Bowen DE, Patrício L, Voss CA. Service research priorities in a rapidly changing context. J Serv Res 2015;18(2):127-159. [doi: 10.1177/1094670515576315]
84. Nicolaou AI, McKnight DH. Perceived information quality in data exchanges: effects on risk, trust, and intention to use. Inf Syst Res 2006;17(4):332-351. [doi: 10.1287/isre.1060.0103]
85. Salaün Y, Flores K. Information quality: Meeting the needs of the consumer. Int J Inf Manag 2001;21(1):21-37. [doi: 10.1016/s0268-4012(00)00048-7]
86. Lee GG, Lin HF. Customer perceptions of e‐service quality in online shopping. Int J Retail Distrib Manag 2005;33(2):161-176. [doi: 10.1108/09590550510581485]
87. Zheng Y, Zhao K, Stylianou A. The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation. Decis Sup Syst 2013;56:513-524. [doi: 10.1016/j.dss.2012.11.008]
88. Chung JE. Social networking in online support groups for health: how online social networking benefits patients. J Health Commun 2014;19(6):639-659. [doi: 10.1080/10810730.2012.757396] [Medline: 23557148]
89. Gächter S, Nosenzo D, Sefton M. Peer effects in pro-social behavior: social norms or social preferences? J Eur Econ Assoc 2013 Jun 3;11(3):548-573 [FREE Full text] [doi: 10.1111/jeea.12015] [Medline: 28553193]
90. Gerber AS, Rogers T. Descriptive social norms and motivation to vote: everybody's voting and so should you. J Poilt 2009;71(1):178-191. [doi: 10.1017/s0022381608090117]
91. Chiu CM, Hsu MH, Wang ET. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. Decis Sup Syst 2006;42(3):1872-1888. [doi: 10.1016/j.dss.2006.04.001]
92. Park H, Park SJ. Communication behavior and online knowledge collaboration: evidence from Wikipedia. J Knowl Manag 2016;20(4):769-792. [doi: 10.1108/jkm-08-2015-0312]
93. Chandwani R, De' R. Doctor-patient interaction in telemedicine: logic of choice and logic of care perspectives. Inf Syst Front 2017;19(4):955-968. [doi: 10.1007/s10796-016-9643-0]
94. Jiang Z, Chan J, Tan BC, Chua WS. Effects of interactivity on website involvement and purchase intention. J Assoc Inform Syst 2010;11(1):1-46. [doi: 10.17705/1jais.00218]
95. Hammond KR. Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice. New York, NY: Oxford University Press; 2000.
96. Wang J, Zhang R, Hao JX, Chen X. Motivation factors of knowledge collaboration in virtual communities of practice: a perspective from system dynamics. J Knowl Manag 2019;23(3):466-488. [doi: 10.1108/jkm-02-2018-0061]
97. Kolappa K, Henderson DC, Kishore SP. No physical health without mental health: lessons unlearned? Bull World Health Organ 2013 Jan 1;91(1):3-3A [FREE Full text] [doi: 10.2471/BLT.12.115063] [Medline: 23397342]
98. Saxena S. Disparity between burden and budget for mental health. Lancet Public Health 2019 Feb;4(2):e75-e76 [FREE Full text] [doi: 10.1016/S2468-2667(18)30238-X] [Medline: 30446417]
99. Patel V, Xiao S, Chen H, Hanna F, Jotheeswaran AT, Luo D, et al. The magnitude of and health system responses to the mental health treatment gap in adults in India and China. Lancet 2016 Dec 17;388(10063):3074-3084. [doi: 10.1016/S0140-6736(16)00160-4] [Medline: 27209149]
100. CBNData.: CBNData; 2018. 2018 Psychological Counseling Industry Insight Report URL: https://www.cbndata.com/report/1275/detail?isReading=report&page=1 [accessed 2020-02-07]
101. Yang Z, Fang X. Online service quality dimensions and their relationships with satisfaction. Int J Serv Ind Manag 2004;15(3):302-326. [doi: 10.1108/09564230410540953]
102. Chu W, Choi B, Song MR. The role of on-line retailer brand and infomediary reputation in increasing consumer purchase intention. Int J Electron Commer 2005;9(3):115-127. [doi: 10.1080/10864415.2005.11044336]
103. Kim JY, Kaufmann K, Stegemann M. The impact of buyer–seller relationships and reference prices on the effectiveness of the pay what you want pricing mechanism. Mark Lett 2014;25(4):409-423. [doi: 10.1007/s11002-013-9261-2]
104. Cheung CM, Xiao BS, Liu IL. Do actions speak louder than voices? The signaling role of social information cues in influencing consumer purchase decisions. Decis Sup Syst 2014;65:50-58. [doi: 10.1016/j.dss.2014.05.002]
105. Alm J, Jackson BR, McKee M. Getting the word out: enforcement information dissemination and compliance behavior. J Public Econ 2009;93(3-4):392-402. [doi: 10.1016/j.jpubeco.2008.10.007]
106. Stata Corporation. Stata Base Reference Manual Release 16. College Station, Texas: Stata Press; 2019.
107. Winkelmann R, Zimmermann KF. Recent developments in count data modelling: theory and application. J Econ Surv 1995;9(1):1-24. [doi: 10.1111/j.1467-6419.1995.tb00108.x]
108. Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. J Manag Inform Syst 2013;29(4):217-248. [doi: 10.2753/mis0742-1222290408]

Abbreviations

CEST: cognitive-experiential self-theory
ICT: information and communication technology
NB: negative binomial

OHC: online health community
SE: sharing economy
VRB: voluntary rewarding behavior

Edited by G Eysenbach; submitted 07.10.19; peer-reviewed by Z Deng, J Mou; comments to author 13.11.19; revised version received 07.01.20; accepted 24.01.20; published 14.04.20.

Please cite as:
Zhou J, Liu F, Zhou T
Exploring the Factors Influencing Consumers to Voluntarily Reward Free Health Service Contributors in Online Health Communities: Empirical Study
J Med Internet Res 2020;22(4):e16526
URL: http://www.jmir.org/2020/4/e16526/
doi:10.2196/16526
PMID:32286231

Original Paper

The Buffering Effect of Health Care Provider Video Biographies When Viewed in Combination With Negative Reviews: "You Can't Fake Nice"

Evan K Perrault[1], PhD; Grace M Hildenbrand[1], MA

Brian Lamb School of Communication, Purdue University, West Lafayette, IN, United States

Corresponding Author:
Evan K Perrault, PhD
Brian Lamb School of Communication
Purdue University
100 North University Street
West Lafayette, IN, 47907
United States
Phone: 1 7654966429
Email: perrault@purdue.edu

Abstract

Background: Patients seek information from numerous sources before choosing a primary care provider; two of the most popular sources are providers' own online biographies and patient rating websites. However, prior research has generally only examined how these sources influence patients' decisions in isolation.

Objective: This study aimed to determine how primary care providers' online biographies and online patient ratings interact to affect patients' decision making, especially in the face of negative reviews.

Methods: An 8-condition online experiment (n=866) was conducted, manipulating patient ratings and the timing of viewing a provider's online biographical video (pre- or postrating viewing).

Results: When participants were shown a short video introduction of a provider after reading predominantly negative reviews a positive expectancy violation occurred, which was also related to more positive perceptions of the provider. When exposed to all negative reviews, 43% of participants indicated they would still choose to make an appointment with the provider, with many indicating that the video provided the evidence needed to help make up their own minds.

Conclusions: These findings are especially relevant to health care organizations seeking to combat a recent rise in fake patient reviews. Providing patients with realistic expectations of the care that clinicians can offer via their own online biographical videos can help counteract negative patient comments online.

(J Med Internet Res 2020;22(4):e16635) doi:10.2196/16635

KEYWORDS

patient ratings; health care providers; video; biographies; expectancy violations; thin slice

Introduction

Background

Selecting a new primary care provider, "one of the most important health-related decisions a patient makes" [1], can be a daunting task with numerous qualities to consider. For example, can I easily get an appointment; will the provider treat me with respect; will I feel comfortable communicating with this provider? As a result of governments' and health care organizations' directives to provide greater levels of patient-centered care (PCC), patients arguably have more information at their fingertips than ever before to help make this decision.

Outside of recommendations from others, one source that prospective patients use to gain information about providers is through biographies on health care systems' websites [2]. Other sources some patients consult are physician online ratings websites, in which patients give reviews in the form of numerical ratings and include narrative comments about their experiences with providers [3].

Given the importance of selecting a provider, patients likely consult information from multiple sources. However, previous

research has generally examined patients' decision-making processes and their perceptions of providers based on visiting only one type of information source. For example, Perrault and Silk [4,5] examined prospective patients' attitudes when experimentally manipulating content within providers' online biographies. Others have solely investigated the effect of providers' online ratings on patients' beliefs and decision making [6-8]. What happens when prospective patients examine multiple sources of information to make their decisions? If researchers continue to only study individual sources of information in isolation, a complete understanding of the impact that the totality of this information can offer will never be fully realized.

Therefore, guided by expectancy violations theory [9,10] and the concept of thin slicing [11], this research sought to understand how both the information provided by providers' own online biographies and that offered on rating websites might interact in influencing prospective patients' perceptions of a primary care provider and patients' decision making. The results could also provide important information to health care organizations on the strong influence their providers' own online biographies may have, especially in the face of negative reviews.

The Growing Information About Providers

Providing PCC is becoming a growing necessity in the health care industry. In the near future, some degree of provider reimbursements is going to be tied to the Agency for Healthcare Research and Quality's (AHRQ's) Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey measures [12], an assessment that among other things measures patient-centered experiences (eg, ease of access to health care services and provider communication). A part of providing greater levels of PCC is offering patients information to enable them to be more informed decision makers, especially in helping them choose a provider or practice that is most likely to meet their individualized needs [13]. The industry is therefore seeing increased innovation and research into the development and improvement of information available to patients to help them make decisions on choosing providers. Two of these information sources are providers' own online biographies and third-party rating websites.

Online Biographies

In a recent survey of almost 4000 people, Perrault and Hildenbrand [2] found that the most popular source from which prospective patients sought information about providers was through online biographies provided by health care organizations. In an industry that is increasingly becoming more consumer-centric [14], with systems competing for patients, there has been greater attention paid toward finding ways to help health care organizations improve outward-facing communication about their providers to prospective patients. After all, the information provided by health care systems on their own websites about providers is under their complete and direct control [15]. Although patients desire the technical expertise of providers, which is often displayed within providers' online biographies through the articulation of degrees and fellowships [16], patients strongly care about the communication qualities of providers. Specific qualities valued by patients include a provider engaging in active listening, being friendly, explaining information in an understandable way, and having a good bedside manner [17,18]. These types of qualities can usually be conveyed within online biographies through philosophy of care statements or even through video introductions of providers [5,19], and it is therefore likely why these types of communicative qualities are also usually displayed on rating websites.

Rating Websites

There are a growing number of online provider rating websites for patients to choose from when seeking information about prospective providers [3,20]. Most often these rating sites consist of quantitative ratings and narrative comments in which patients rate or describe personal experiences with particular providers [3]. These websites request the patient's feedback in categories such as physician's knowledge, timeliness, and interpersonal skills [21].

However, although large numbers of patients do seek information from these rating websites [3,22], most do so with caution. "Americans do not seem to put much stock in overall rating systems of doctors or other care providers" [17]. Only about 10% of Americans "completely", or "trust very much", the provider quality information provided by ratings websites, and only about 30% would trust quality ratings from patients who are surveyed anonymously about the quality of their care [17]. In addition, most ratings on rating websites tend to be positive [6], indicating that patients who consult these sites may not be receiving a fully representative picture of other patients' experiences [23]. In other words, patients who solely make decisions based on rating websites may choose providers who violate the patients' expectations.

Expectancy Violations Theory

Expectancy violations theory is rooted in the belief that everyone has or develops expectations about what a future interaction with someone should and will be like [10]. In the case of seeking a new health care provider, prospective patients could develop these expectations by reading other patients' comments on rating websites discussing their experiences with a specific provider. Even though prospective patients do not place much confidence in patient rating websites to select providers, about two-thirds of those surveyed in a nationally representative sample indicated that patients' ratings of providers' communication are an important factor [17]. Patients are less likely to visit doctors when they are rated negatively, and this is especially the case when negative reviews are shown before positive reviews [7]. In other words, once a negative expectation of the provider is set, that negative expectation may persist and influence a patient's ultimate decision to not visit the provider.

However, expectancy violations theory also posits that a person may modify his or her perceptions of a target when the target's actual communication runs counter to what is expected [9,10]. As Burgoon and LePoire [24] found, negatively induced preinteractional expectancies about a target could be overcome after having a pleasant conversation with that target: "To the extent that uncertainty is introduced by mixed expectancies...perceivers should be motivated instead to attend

more to the actual behavioral evidence" [24]. One way that health care organizations are beginning to provide this behavioral evidence to prospective patients is through the development of short video introductions of providers to place within online biographies [19]. Therefore, if given a short video introduction of a provider showcasing positive communication skills, that video may be able to override the initial negative expectancies induced by negative reviews.

Hypothesis 1: Participants who first view predominantly negative reviews (all negative or two-third negative) will have more significant expectancy violations of the provider after subsequently viewing a short video of the provider than those exposed to predominantly positive reviews.

Video Biographies as Thin Slices

People's ability to accurately predict attributes of others after only viewing short video clips has been termed *thin slicing* [11]. Thin-slice research has found that participants are able to make accurate judgments of targets from as little as 6-second silent videos [25]. Others have found that attributes such as sexual orientation can be predicted from as little as 10-second clips [26], the level of altruism from 20-second clips [27], and personality traits from 30-second clips [28].

In addition, in watching short video biographies of providers, prospective patients are able to actually see the providers' personality traits, thereby helping them better predict how the provider might interact in a consultation [19]. For example, one participant in Perrault's [19] study of provider videos indicated the videos to which she was exposed "helped me see if I would feel comfortable with that person."

Therefore, viewing a video after reading predominantly negative reviews (ie, a positive violation) might actually repair the initial negative perceptions prospective patients had about the provider such as provider liking, trustworthiness, expertise, anticipated patient satisfaction, and anticipated medical care quality—qualities that prior research finds are important to patients [29-31]. Conversely, viewing a video before reading negative comments also might provide a protective effect, lessening any negative impact those comments could have had if simply viewed in isolation. Therefore, we hypothesized the following:

Hypothesis 2: A significant interaction between the viewing order of provider content (video and patient reviews) and the valence of the reviews viewed (all positive, two-third positive, all negative, and two-third negative) on the dependent variables of provider liking, trustworthiness, expertise, anticipated patient satisfaction, and anticipated quality of medical care will be observed.

Choosing a Provider

After considering all the information available, patients ultimately have to make a choice [32]. To the best of the authors' knowledge, this is the first study to have participants view both health care system–controlled biographical information of the provider and third-party patient ratings before making decisions. Therefore, a series of research questions (RQs) were posed:

- RQ1: How will the information viewing condition be related to provider selection?
- RQ2: What information influences people's decision the most regarding whether or not they would want to select the provider?

Finally, given that we predicted that an expectancy violation will occur when people who are exposed to negative reviews subsequently see the provider's video, we believed that there will be some people who will choose to visit the provider even in the face of all negative reviews. Therefore, we are curious how people will explain their decisions when provided all negative reviews with the following RQ:

- RQ3: When exposed to a condition containing all negative reviews, what reasons do people provide for wanting, or not wanting, to choose to visit the provider?

Methods

This study took the form of a 4 (provider ratings: all positive, two-third positive, all negative, two-third negative) x 2 (viewing order of provider content) mixed design experiment, where the provider ratings (all positive, two-third positive, all negative, and two-third negative) was the between-subjects factor and the viewing order (video first–reviews second vs reviews first–video second) was the within-subjects factor.

Procedures and Scenario

Upon consenting and indicating that they were using a device in which they could view the video and listen to the audio, participants were recruited into the study. Participants were asked to imagine themselves as patients who had recently moved across the country for a new job and had fallen ill. After a few days of rest and not feeling any better, they decided it was time to go to a health care provider. They went online to look for a nearby clinic and provider who fit with their health insurance. One half of the participants were told that their first stop online was the health care provider's own website where a video of the provider they were considering could be found. The other half were told that their first stop online was a popular website where patients' ratings of health care providers existed, and they looked up the ratings of a provider they were considering. At the end of the study, participants were asked to rate on a one-item measure (1=strongly disagree and 7=strongly agree) on how realistic they thought the scenario regarding selecting a new provider seemed. A one-sample *t* test revealed a mean score significantly above the scale's midpoint (mean 5.99, SD 0.94; t_{852}=61.73; P<.001), indicating that participants thought this scenario was realistic.

Provider Content—Experimental Manipulations

The Video

Participants were exposed to a 68-second video of a nurse practitioner who was interviewed discussing her philosophy of care, what a normal consultation with her is like, and what she likes to do when she is not at the clinic. The practitioner was shot in an interview style, with her head and shoulders in the frame. The majority of the interview footage was covered up with B-roll of the provider actually interacting with a patient.

For example, a participant could see the provider actually asking questions to the patient, performing a brief examination, and then discussing treatment options with the patient. This B-roll of the provider was included within the video as prior research indicated that prospective patients would like B-roll included in video introductions [19]. The video was produced and edited by the first author who is also a former television reporter. To try and ensure that the participants actually viewed the video, participants were not able to continue with the survey until 68 seconds (the length of the video) had elapsed. Underneath this video was a brief biography of the provider that only provided her name, photo, specialty, and educational credentials (see Multimedia Appendix 1). This basic content is usually provided within most online provider biographies [16].

Likeability Induction of the Provider

All participants who viewed the video first (n=413) rated the provider as likable (see the Measures section). A one-sample two-tailed *t* test found responses significantly above the midpoint of the scale (mean 6.38, SD 0.75; t_{412}=64.52; *P*<.001). Therefore, this video succeeded at showing participants a provider who is initially perceived as likable by individuals without any information to the contrary.

The Reviews

Reviews of the provider were developed by simulating the page of a provider rating website (see Multimedia Appendix 2). Patients' comments were developed by the second author who viewed hundreds of real patients' comments on numerous rating websites to create the most realistic comments possible. All comments were solely focused on the communication between that patient and the provider. This is because patients' ratings of providers' communication play an important role in prospective patients' decision making [17,33]. In addition, a majority of clinicians and patients agree that providers should not be evaluated by patients on the clinicians' technical skills but do agree that patients have the knowledge to evaluate the clinicians' communication skills [34].

Each review had 3 patient comments, followed by a 5-point star rating. Four sets of patient comments were developed to which a participant could have been randomly assigned: all 3 positive comments, all 3 negative comments, 2 positive comments out of 3 comments (middle comment negative), and 2 negative comments out of 3 comments (middle comment positive). Positive comments were all given five stars, and negative comments were all given one star. Negative comments were an exact opposite translation of the positive comments. For example, if a positive comment said that the provider "always pays attention to me," the negative version of that same comment said that the provider "never pays attention to me."

Viewing Order

To test this study's hypotheses and RQs, the viewing order of the video and reviews was also randomly assigned. Half of the participants were randomly assigned to view the video first and the other half randomly assigned to view the patients' reviews first. After viewing each portion of the provider content, participants completed a series of survey measures.

Measures

Unless otherwise noted, all variables were measured at two timepoints, once after each exposure to the video and patient reviews.

Liking

Provider liking was measured with four items adapted from a study by Jayanti and Whipple [35]. Participants rated their level of agreement on a 7-point Likert scale (1=strongly disagree and 7=strongly agree) with the following statements: This provider seems likable, pleasant, friendly, like a nice person (alpha=.992 and alpha=.984 for the first and second times, respectively).

Trust

Provider trust was measured with six 7-point, semantic differential items adapted from source credibility scales of McCroskey and Teven [36] and Ohanian [37]. Participants were asked to rate how dishonest-honest, undependable-dependable, unreliable-reliable, insincere-sincere, untrustworthy-trustworthy, and phony-genuine they perceived the provider to be (alpha=.980 and alpha=.977 for the first and second times, respectively). Higher scores indicated greater levels of trust.

Expertise

Expertise was measured with six 7-point, semantic differential items also adapted from the source credibility scales of McCroskey and Teven [36] and Ohanian [37]. Participants rated how they believed the provider to be an expert/not an expert, inexperienced/experienced, incompetent/competent, unqualified/qualified, unskilled/skilled, and stupid/smart (alpha=.971 and alpha=.969 for the first and second times, respectively). Higher scores indicated greater levels of expertise.

Anticipated Patient Satisfaction

Anticipated patient satisfaction was measured with three 7-point, semantic differential items from Richmond et al's [38] satisfaction with the physician scale. Participants were asked to indicate how displeased-pleased, dissatisfied-satisfied, uncomfortable-comfortable they would be with their visit (alpha=.987 and alpha=.974 for the first and second times, respectively). Higher scores indicated greater satisfaction.

Anticipated Quality of Medical Care

The anticipated quality of medical care was measured using four items adapted from Richmond et al's [38] perceived quality of medical care measure. Participants were asked to indicate where they would fall along the 7-point continuum for the following word pairs regarding the kind of medical care they would obtain from the provider: impersonal-personal, uncaring-caring, unconcerned-concerned, and unsatisfactory-satisfactory (alpha=.984 and alpha=.977 for the first and second times, respectively). Higher scores indicated greater perceptions of care quality.

Expectancy Violation

Expectancy violation was measured only once, after participants were exposed to the video. Expectancy violation was measured using three items adapted from a study by Klingle and Burgoon [39]. Participants rated on a 7-point Likert scale (1=strongly disagree and 7=strongly agree) their level of agreement to the

following prompt and statements: Based on the video you just saw of the provider, please rate your level of agreement with each statement: Kris communicated in a way that I expected; Kris' communication style is what I anticipated it would be; I expected that Kris would interact with the patient in the way she did (alpha=.960). Lower scores indicated a greater expectancy violation.

Decision Making

After viewing both pieces of provider content (video and reviews), participants were asked to indicate (yes/no) their decision regarding the following question: "Based on all the information that you saw, would you decide to make a medical appointment with Kris?"

Influence of Content

Participants were then asked to indicate via a closed-ended response about what information influenced them the most regarding whether or not they would make an appointment with the provider. Participants could select from the following three options: provider's video and biography, patients' ratings of the provider, or both the video/biography and patients' ratings.

Rationale for Decision

Participants were then asked to respond to an open-ended question inviting them to indicate why they would or would not choose to make an appointment with the provider.

Participants and Data Cleaning

Participants were recruited utilizing Amazon Mechanical Turk (Seattle, WA) in October 2018 and paid US $1 for participating. Participants were only recruited from the United States. To ensure integrity of the data, multiple procedures were utilized to clean the dataset. An initial captcha item was used to ensure humans, and not machines, were the actual participants. Initially, 1716 surveys were completed. Utilizing procedures outlined by Dennis et al [40] regarding how to identify and remove participants who circumvent initial screening methods (eg, using server farms to circumvent country of residence), the following methods were used. Responses that originated from duplicate GPS coordinates were initially removed (n=301); next, those originating from duplicate internet protocol addresses were also removed (n=105). Three people were removed who did not watch the video, and 138 were removed for not completing more than half of the study's questions. An additional 87 people were removed for indicating that they had previously seen the provider, 96 people who worked in the health care industry were removed, and 109 people who took less than 8 min to complete the study (an approximate time to reasonably view all stimuli and answer questions) were removed. Finally, 11 participants were removed who did not logically answer the open-ended question asking why they would (or would not) choose to make an appointment, eg, "for meet her" and "I make." This left a final participant pool of 866 valid responses.

Demographics

The average age of participants was 39.2 years (SD 12.6; range 18-82). A little more than half of the participants identified as female (n=493). Most participants (626/866, 72.3%) identified as Caucasian, followed by African American (92/866, 10.6%), Hispanic (63/866, 7.3%), Asian (56/866, 6.5%), Native American (8/866, 0.9%), Pacific Islander (2/866, 0.2%), and other (16/866, 1.8%). Participants came from all states except North Dakota. Six participants reported their highest level of education as never completing high school; other participants reported completing a high school diploma/general education diploma (225/866, 26.0%), 2-year college degree (182/866, 21.0%), 4-year degree (336/866, 38.8%), and an advanced college degree (114/866, 13.2%).

Results

Hypothesis 1

Hypothesis 1 predicted that participants who first viewed predominantly negative reviews of the provider (ie, all negative or two-third negative) would have a more significant expectancy violation of the provider after subsequently viewing the provider's video than those who were initially exposed to predominantly positive reviews. To test this hypothesis, a one-way analysis of variance (ANOVA) was conducted where the provider's review condition was the independent variable and expectancy violation was the dependent variable. The analyses revealed a significant finding: $F_{7,858}$=93.11 and P<.001. Post hoc comparisons using the Tukey honestly significant difference test at P<.05 showed that those who initially saw all negative reviews had the most significant expectancy violation after viewing the video (mean 3.34, SD 1.76), followed by those who saw two negative comments and then the video (mean 4.14, SD 1.51). All the other six conditions were not statistically different from one another (see Table 1). Therefore, hypothesis 1 was supported. Those who viewed predominantly negative reviews of the provider and then viewed her video had the most significant expectancy violations.

Table 1. Expectancy violations by condition.

Dependent variable	Condition, mean (SD)								*F* test (*df*)	*P* value
	Video first,...then reviews				...Reviews first, then video					
	All positive (n=105)	Two-third positive (n=101)	All negative (n=106)	Two-third negative (n=101)	All positive (n=126)	Two-third positive (n=100)	All negative (n=114)	Two-third negative (n=113)		
Expectancy violation	5.72[a] (0.91)	5.85[a] (0.97)	5.87[a] (0.95)	5.88[a] (0.93)	6.26[a] (0.78)	5.84[a] (0.94)	3.34[b] (1.76)	4.14[c] (1.51)	93.11 (7, 858)	<.001

[a-c]Means with different superscripts differ at $P<.05$ using the Tukey honestly significant difference test. Expectancy violation was only measured after viewing the video of the provider.

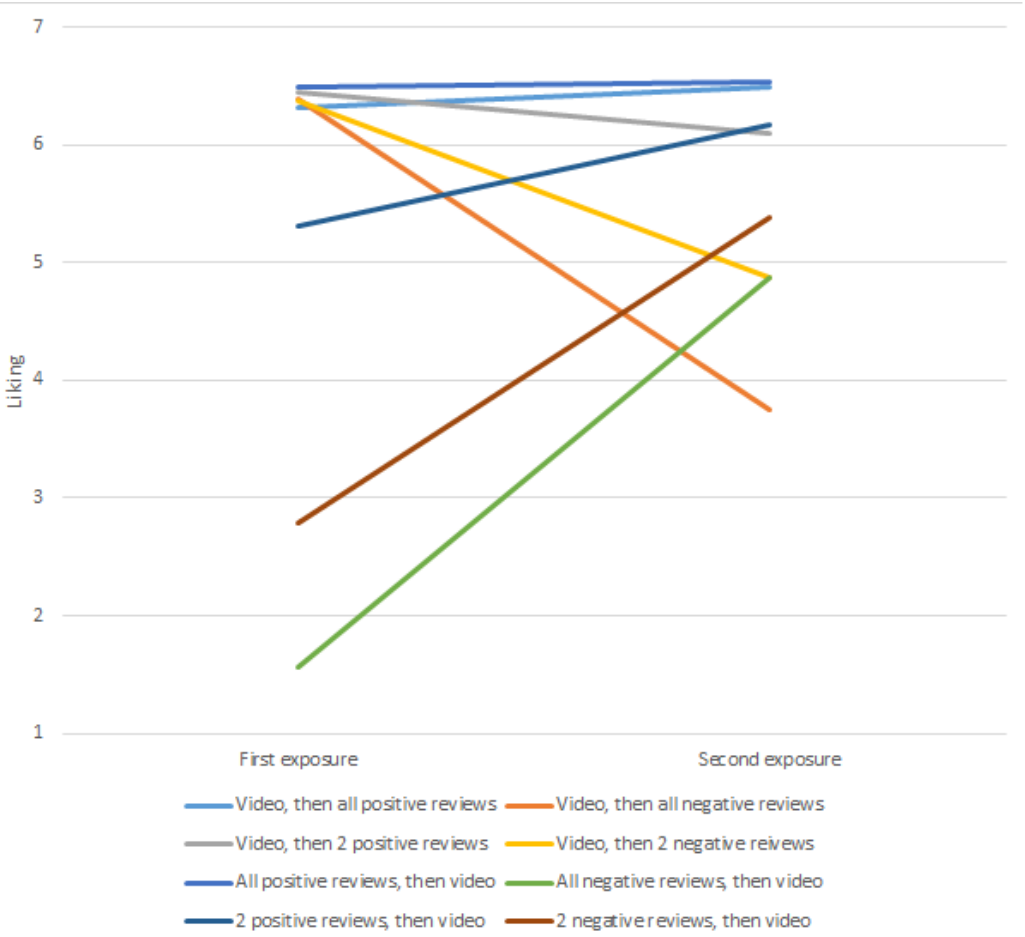
Hypothesis 2

Hypothesis 2 predicted that a significant interaction would arise between the viewing order of the provider content and the valence of the reviews viewed on the dependent variables of provider liking, trustworthiness, expertise, anticipated patient satisfaction, and quality of medical care. A series of mixed ANOVA for each dependent variable was conducted, where the valence of reviews was the between-subjects factor and the viewing order of the provider content (video/reviews either first/second) was the within-subjects factor. The analyses revealed significant interactions for all five dependent variables in the same pattern (see Table 2 for descriptive data and analyses and Figure 1 for a visual depiction of one of the interactions). In general, those who viewed predominantly negative comments first saw their attitudes toward the provider increase significantly after viewing the video of the provider. In fact, in every instance these participants' attitudes increased to a point significantly above the midpoint of the 7-point scale according to one-sample *t* tests. Conversely, participants who viewed the video first and then were exposed to predominantly negative reviews saw their attitudes significantly decrease. However, in none of these instances did subsequent attitudes decrease significantly below the midpoint of the 7-point scale according to one-sample *t* tests. Therefore, although the videos were not able to hold participants' initial positive attitudes stable in the face of subsequently viewing predominantly negative reviews, these participants did not ultimately hold negative attitudes toward the provider, or the care that could be provided, after viewing the negative reviews.

Table 2. Descriptive statistics and mixed analysis of variance results.

Dependent variables and exposure	Condition, mean (SD)								Condition×exposure		Partial η^2
	Video first, then...reviews				...Reviews first, then video				*F* test (*df*)	*P* value	
	All positive (n=105)	Two-third positive (n=101)	All negative (n=106)	Two-third negative (n=101)	All positive (n=126)	Two-third positive (n=100)	All negative (n=114)	Two-third negative (n=113)			
Liking									296.64 (7, 858)	<.001	0.708
First	6.32 (0.78)	6.44 (0.78)	6.39 (0.75)	6.37 (0.69)	6.49 (0.59)	5.31 (0.78)	1.56 (0.87)	2.79 (0.99)			
Second	6.49 (0.72)	6.10 (0.75)	3.76 (1.81)	4.88 (1.34)	6.54 (0.75)	6.17 (0.69)	4.87 (1.27)	5.39 (1.04)			
Trust									179.79 (7, 858)	<.001	0.595
First	6.31 (0.84)	6.52 (0.70)	6.44 (0.92)	6.47 (0.78)	6.36 (0.79)	5.64 (0.87)	3.05 (1.43)	3.81 (1.19)			
Second	6.47 (0.82)	6.26 (0.80)	4.17 (1.83)	5.14 (1.39)	6.59 (0.69)	6.32 (0.79)	5.16 (1.23)	5.57 (1.09)			
Expertise									101.61 (7, 858)	<.001	0.453
First	6.21 (0.99)	6.30 (0.85)	6.24 (0.97)	6.33 (0.76)	6.25 (0.81)	5.79 (0.95)	3.69 (1.47)	4.52 (1.14)			
Second	6.35 (0.88)	6.22 (0.83)	4.89 (1.61)	5.45 (1.23)	6.46 (0.73)	6.24 (0.96)	5.49 (1.16)	5.79 (1.02)			
Patient satisfaction									249.18 (7, 856)	<.001	0.671
First	6.28 (0.91)	6.47 (0.68)	6.34 (1.03)	6.32 (0.94)	6.45 (0.66)	5.40 (0.91)	1.66 (0.87)	2.81 (1.10)			
Second	6.47 (0.81)	6.20 (0.85)	3.85 (1.86)	4.81 (1.55)	6.60 (0.68)	6.18 (0.98)	4.84 (1.48)	5.40 (1.26)			
Medical care quality									268.70 (7, 858)	<.001	0.687
First	6.32 (0.78)	6.46 (0.69)	6.43 (0.81)	6.35 (0.87)	6.50 (0.65)	5.54 (0.85)	1.71 (1.05)	2.71 (1.14)			
Second	6.49 (0.76)	6.23 (0.82)	3.79 (1.89)	4.76 (1.62)	6.59 (0.70)	6.26 (0.87)	4.90 (1.52)	5.41 (1.19)			

Figure 1. Interaction effect between content viewing order and valence of reviews.



Research Question 1

RQ1 was interested in determining how the information viewing condition was related to whether the participants would decide to make an appointment with the provider. A chi-square analysis resulted in a significant finding: χ^2_7=216.1 (n=866) and *P*<.001. In only two instances (ie, both conditions where participants saw all negative reviews) did the number of people who indicated not wanting to make an appointment with the provider outnumber those who would. However, even in these two conditions, a large number of participants (95/220, 43.2%) indicated that they would make an appointment.

Research Question 2

This RQ sought to understand what information influenced people the most regarding their decision to select the provider. A chi-square analysis resulted in a significant finding: χ^2_{14}=288.5 (n=866) and *P*<.001. In only one condition (ie, video first, then all negative comments), the participants reported that the patients' ratings had the most significant influence. When the participants were exposed to two negative comments or all negative comments first, the video that was viewed influenced the participants more (see Table 3). In conditions where there was agreement in the content that was viewed (ie, reviews predominantly matched the pleasantness of the provider viewed in the video), participants indicated that a combination of the ratings and video were equally influential.

Table 3. Participants' decision making by condition.

Participant decision	Condition, n								Row total
	Video first, then...reviews				...Reviews first, then video				
	All positive (n=105)	Two-third positive (n=101)	All negative (n=106)	Two-third negative (n=101)	All positive (n=126)	Two-third positive (n=100)	All negative (n=114)	Two-third negative (n=113)	
Participant wants to make an appointment									
Yes	98	94	39	59	119	94	56	86	645
No	7	7	67	42	7	6	58	27	221
Participant was most influenced by									
Video/biography	28	38	32	44	30	54	52	71	349
Patients' ratings	5	2	58	34	8	2	39	16	164
Video/biography and patients' ratings	72	61	16	23	88	44	23	26	353

Research Question 3

To more thoroughly investigate why participants exposed to all negative reviews would, or would not, choose to visit the provider, a content analysis of the participants' open-ended responses was conducted. The two researchers utilized a thematic analysis approach [41] where both researchers independently read all 220 open-ended responses for why participants decided to either want to make, or not make, an appointment with the provider. General themes were developed into a formal coding scheme. Both researchers independently coded the responses with a high level of initial overall agreement (kappa for each category >0.7) and then met to resolve disagreements until 100% agreement was reached. The following themes emerged.

Participants Who Would Make an Appointment

A total of 95 people who were exposed to all negative comments mentioned that they would make an appointment with the provider. The following four relevant themes emerged from these participants' rationales. See Table 4 for all frequencies.

Table 4. Rationales for selection of participants exposed to all negative reviews.

Reasoning behind the decision	Participants, n (%)	Example comments
Would visit the provider (n=95)		
Personality of the provider	50 (52.6)	• "she seemed genuine" • "seems very nice and caring" • "she seemed very warm"
Do not trust reviews	50 (52.6)	• "many people online can be dishonest about their visit" • "I don't pay attention to reviews about people, products yes. Some people just grate on each other."
Video made the difference	47 (49.5)	• "the video tells all" • "I liked the way she was in the video"
Expertise	44 (46.3)	• "she seems very competent" • "she has experience"
Other	10 (10.5)	• "she sounded like the kind of doctor I would want"
Would not visit the provider (n=125)		
Reviews were bad	113 (90.4)	• "the reviews she has were all bad" • "I think past patient reviews say a lot and all of hers were negative"
Did not like the providers' communication style in the video	12 (9.6)	• "I thought her communication was poor in the video" • "her tone in the video seemed very cold"
She is not a doctor	8 (6.4)	• "I prefer to see a doctor rather than a nurse practitioner"
Other	7 (5.6)	• "I am male and prefer to speak with a male"

Do Not Trust Reviews

A little more than half of the participants who mentioned that they would visit the provider (50/95, 53%) indicated not putting much trust in the patients' reviews when making decisions. Examples included statements like "I do not always believe what is written by patients/clients/consumers on review sites. I just do not trust the general public when it comes to impartial opinions," and "I would want to make up my mind on my own instead of relying on the opinions of strangers...There could be a lot of reasons why someone would leave a negative review, sometimes out of spite or because they did not get their way."

Personality of the Provider

A little more than half also indicated that their choice was because of the positive personality characteristics perceived (50/95, 53%). Examples included statements such as "she seemed sincere," "she seems genuine," and "she seemed to be very nice, compassionate."

Video Made a Difference

Nearly half of these participants (47/95, 50%) explicitly mentioned the video as a deciding factor. For example, "I would make an appointment with her after seeing the video," "I am relying on my own judgement from her behavior in the video," and "I feel like I know much more about what to expect after seeing the video."

Expertise

Just less than half (44/95, 46%) referenced the provider's expertise as a deciding factor. Perceived expertise was seen in comments such as "she has a long career with good experience," "she seemed smart, capable," and "I liked her credentials, her experience."

Multiple themes could have been present in each statement. For example, 14 responses contained all themes. An example of one of these was "I would make an appointment with her after seeing the video because she seems like a very nice and experienced person. Originally, I thought that she was going to be very rude and unprofessional based on the reviews I had seen. However, I now feel that the reviews were wrong."

Participants Who Would Not Make an Appointment

A total of 125 participants who were exposed to all negative reviews indicated that they did not want to make an appointment with the provider. The following three relevant themes emerged.

Reviews Were Bad

Overall, 90.4% (113/125) of these participants indicated that the negative reviews played a deciding factor. Examples include statements such as "I take reviews from people with experience very seriously, and they were all negative," "I trust patient reviews more," "based on reviews, I believe those that have seen her, especially when the reviews are so consistent," and "based on the people that saw her and talked about her I would not care to be involved with her at all."

Did Not Like the Communication Style in the Video

Furthermore, 9.6% (12/125) of these participants also indicated not liking the communication style the provider displayed in the video. For example, "she seemed insincere when interacting with the patient," "watching the video confirmed [for] me that she is not that friendly," and "I feel like I can read people well by body language and facial expressions. Based on that alone, the vibe she gives me is still impersonal and not very warm."

She Is Not a Doctor

Overall, 6.4% (8/125) of the participants indicated not wanting to make an appointment because the provider was not a doctor. Examples included statements such as "I would prefer an M.D.," "nurse practitioners are not doctors," and "I prefer to see a doctor rather than a nurse practitioner."

Discussion

Principal Findings

This study strove to determine how provider-controlled content (ie, providers' online biographies/videos) and uncontrolled content (ie, online patient reviews) interact to influence patients' attitudes and decision-making processes. The findings revealed that the initial deleterious effects of viewing negative patient comments can be significantly reversed when provided with a realistic preview of the provider through a short video introduction. In other words, any initially negative attitudes toward the provider after viewing negative reviews did not persist after viewing the provider's video. Although the participants' attitudes did not reach the same heights as when people only viewed positive comments, this research does show health care organizations that hosting videos of providers on their websites can provide a significant buffering effect to negative comments that might exist online via third-party rating websites.

The only conditions in which the provider's video had the least amount of impact on choosing to visit the provider was when participants viewed comments that comprised all negative reviews. However, even in these conditions, just under half of the participants (43%) chose to go against the reviews and indicated wanting to make an appointment. In these instances, about half of the participants indicated that the video played an important role with one participant stating that "the reviews must have been fake because she seems genuine, compassionate, and capable." Prior research supports the claim that patients seek providers whom patients perceive as having good interpersonal skills [42]. Therefore, providing prospective patients with videos can offer a realistic preview into how an interaction with the provider might unfold, allowing patients to make up their own minds even in the face of contradictory reviews.

Extending Expectancy Violations Theory

Although expectancy violations theory was originally applied to nonverbal behaviors [10], it was later extended to verbal behaviors in the context of face-to-face interactions [9]. Since then, the theory has been applied in computer-mediated [43] and mass-mediated settings [44] as well as health settings such as health campaigns [45] and patients' expectations for communication with a physician [46]. This study breaks new ground in the application of expectancy violations theory by incorporating the comparison between multiple sources—a

provider's online video introduction and online patient ratings—to demonstrate how providers' videos can be used to generate positive expectancy violations in the case of negative provider reviews.

Limitations and Future Directions

The first limitation of this study was that only a positively perceived provider was utilized in this study, allowing for only positive expectancy violations to take place. However, given that the majority of patient reviews found online are positive [6], providers may exist who could induce negative expectancy violations. In other words, future studies may want to provide conditions where a provider receives positive reviews but appears grumpy and gruff in his or her video. Future studies may also want to vary video length to determine how short of a video (ie, how thin of a slice) could work to still be effective at providing a significant buffering effect to negative comments.

This study also only tested reviews appearing on the extremes with comments being bipolar opposites of one another (eg, five stars or one star). Future studies might want to test the effects of reviews that are more middle of the road, ie, combinations of two, three, and four stars, and how these ratings interact with providers' online biographical content provided by health care organizations. In addition, in this study, the content of the reviews focused on the provider's communication. Future research may also want to include comments discussing other qualities of the provider (eg, credentials and technical competence).

Conclusions

The rise of health care consumerism today means that patients are shopping around for providers more than ever before. A recent report of top health industry issues reveals that more than three-quarters of consumers desire a "menu of care options offered by multiple providers, allowing them to choose care from local providers or virtual care from specialists across the country." [47] As Perrault and Hildenbrand [2] found, two of the most popular sources patients are using to seek this information are providers' own online biographies and patient reviews online. Therefore, continuing to only research the impact of each of these channels in isolation on patients' perceptions will only provide limited conclusions.

More importantly, recent media reports indicate that fake reviews of medical providers are on the rise, possibly attributing them to competing offices, disgruntled former employees, or even image repair companies seeking to make a profit [48]. Doctors can even pay large sums of money to hide negative reviews or hire reputation management firms [49,50]. However, what this research found is that there is a much less expensive solution to combat potentially false negative reviews online—offering a realistic preview of the provider through short video introductions on providers' own profiles. As this study revealed, providing a video to participants initially exposed to primarily negative reviews can produce a positive expectancy violation and turn initially negative perceptions into positive ones. Most importantly, more than 40% of the participants exposed to all negative reviews indicated wanting to choose to visit the provider anyway, with many of those indicating that the video helped in making this decision.

In the increasingly competitive world of health care, if providers continue to only offer prospective patients limited information about themselves (eg, text biographies that only provide credentials)—information that is currently the norm [16], patients' perceptions may become overly clouded by reviews that they read online, whether they are genuine or not. However, if health care organizations decide to offer patients videos that can actually showcase how providers communicate, these organizations may just find that patients are willing to trust their own intuitions. As one participant stated, "you can't fake nice."

Acknowledgments

This research was funded by startup funds provided by Purdue University, College of Liberal Arts.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video and biography presentation.
[PNG File , 293 KB - jmir_v22i4e16635_app1.png]

Multimedia Appendix 2
Provider ratings (all positive reviews).
[PNG File , 55 KB - jmir_v22i4e16635_app2.png]

References

1. Mercado F, Mercado M, Myers N, Hewit M, Haller NA. Patient preferences in choosing a primary care physician. J Prim Care Community Health 2012 Apr 1;3(2):125-131. [doi: 10.1177/2150131911421802] [Medline: 23803456]
2. Perrault EK, Hildenbrand GM. Primary care confusion-public knowledge of NP and PA duties and their information gathering behaviors. J Gen Intern Med 2018 Nov;33(11):1857-1858 [FREE Full text] [doi: 10.1007/s11606-018-4580-x] [Medline: 30022415]

3. Holliday AM, Kachalia A, Meyer GS, Sequist TD. Physician and patient views on public physician rating websites: a cross-sectional study. J Gen Intern Med 2017 Jun;32(6):626-631 [FREE Full text] [doi: 10.1007/s11606-017-3982-5] [Medline: 28150098]
4. Perrault EK, Silk KJ. Reducing communication apprehension for new patients through information found within physicians' biographies. J Health Commun 2015;20(7):743-750. [doi: 10.1080/10810730.2015.1018569] [Medline: 25942070]
5. Perrault EK, Silk KJ. The uncertainty reducing capabilities of primary care physicians' video biographies for choosing a new doctor: is a video worth more than two hundred words? Health Commun 2016 Dec;31(12):1472-1481. [doi: 10.1080/10410236.2015.1082457] [Medline: 27054561]
6. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015 Sep 26;15:416 [FREE Full text] [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]
7. Li S, Feng B, Chen M, Bell RA. Physician review websites: effects of the proportion and position of negative reviews on readers' willingness to choose the doctor. J Health Commun 2015 Apr;20(4):453-461. [doi: 10.1080/10810730.2014.977467] [Medline: 25749406]
8. Li S, Lee-Won RJ, McKnight J. Effects of online physician reviews and physician gender on perceptions of physician skills and primary care physician (PCP) selection. Health Commun 2019 Oct;34(11):1250-1258. [doi: 10.1080/10410236.2018.1475192] [Medline: 29792519]
9. Burgoon JK. Interpersonal expectations, expectancy violations, and emotional communication. J Lang Soc Psychol 1993;12(1-2):30-48. [doi: 10.1177/0261927X93121003]
10. Burgoon JK, Hale JL. Nonverbal expectancy violations: model elaboration and application to immediacy behaviors. Commun Monogr 1988;55(1):58-79. [doi: 10.1080/03637758809376158]
11. Ambady N, Rosenthal R. Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. Psychol Bull 1992;111(2):256-274. [doi: 10.1037//0033-2909.111.2.256]
12. Agency for Healthcare Research and Quality. Section 2: Why Improve Patient Experience? URL: https://www.ahrq.gov/cahps/quality-improvement/improvement-guide/2-why-improve/index.html [accessed 2019-10-09]
13. Davis K, Schoenbaum SC, Audet AM. A 2020 vision of patient-centered primary care. J Gen Intern Med 2005 Oct;20(10):953-957 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0178.x] [Medline: 16191145]
14. Cohen SB, Grote KD, Pietraszek WE, Laflamme F. Increasing consumerism in healthcare through intelligent information technology. Am J Manag Care 2010 Dec;16(12 Suppl HIT):SP37-SP43 [FREE Full text] [Medline: 21314219]
15. Mostaghimi A, Crotty BH, Landon BE. The availability and nature of physician information on the internet. J Gen Intern Med 2010 Nov;25(11):1152-1156 [FREE Full text] [doi: 10.1007/s11606-010-1425-7] [Medline: 20544300]
16. Perrault EK, Smreker KC. What can we learn from physicians' online biographies to help in choosing a doctor? Not much. A content analysis of primary care physician biographies. J Commun Healthc 2013 Nov 15;6(2):122-127. [doi: 10.1179/1753807612Y.0000000027]
17. Associated Press: NORC Center for Public Affairs Research. Finding Quality Doctors: How Americans Evaluate Provider Quality in the United States URL: http://www.apnorc.org/projects/Pages/finding-quality-doctors-how-americans-evaluate-provider-quality-in-the-united-states.aspx [accessed 2019-10-09]
18. McGlone TA, Butler ES, McGlone VL. Factors influencing consumers' selection of a primary care physician. Health Mark Q 2002;19(3):21-37. [doi: 10.1300/J026v19n03_03] [Medline: 12077807]
19. Perrault EK. Patients' preferences for family medicine physicians' video biographies. Health Mark Q 2016;33(4):342-352. [doi: 10.1080/07359683.2016.1238654] [Medline: 27782785]
20. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. J Am Med Assoc 2017 Feb 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]
21. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
22. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. J Am Med Assoc 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
23. Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Finucane ML, et al. Taking patients' narratives about clinicians from anecdote to science. N Engl J Med 2015 Aug 13;373(7):675-679. [doi: 10.1056/NEJMsb1502361] [Medline: 26267629]
24. Burgoon JK, Le Poire BA. Effects of communication expectancies, actual communication, and expectancy disconfirmation on evaluations of communicators and their communication behavior. Hum Commun Res 1993;20(1):67-96. [doi: 10.1111/j.1468-2958.1993.tb00316.x]
25. Ambady N, Rosenthal R. Half a minute: predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. J Pers Soc Psychol 1993;64(3):431-441. [doi: 10.1037/0022-3514.64.3.431]
26. Ambady N, Hallahan M, Conner B. Accuracy of judgments of sexual orientation from thin slices of behavior. J Pers Soc Psychol 1999 Sep;77(3):538-547. [doi: 10.1037//0022-3514.77.3.538] [Medline: 10510507]

27. Fetchenhauer D, Groothuis T, Pradel J. Not only states but traits — humans can identify permanent altruistic dispositions in 20 s. Evol Hum Behav 2010;31(2):80-86. [doi: 10.1016/j.evolhumbehav.2009.06.009]
28. Oltmanns TF, Friedman JN, Fiedler ER, Turkheimer E. Perceptions of people with personality disorders based on thin slices of behavior. J Res Pers 2004;38(3):216-229. [doi: 10.1016/s0092-6566(03)00066-7]
29. Choi K, Cho W, Lee S, Lee H, Kim C. The relationships among quality, value, satisfaction and behavioral intention in health care provider choice: a South Korean study. J Bus Res 2004 Aug;57(8):913-921. [doi: 10.1016/s0148-2963(02)00293-x]
30. Leach B, Gradison M, Morgan P, Everett C, Dill MJ, de Oliveira JS. Patient preference in primary care provider type. Healthc (Amst) 2018 Mar;6(1):13-16. [doi: 10.1016/j.hjdsi.2017.01.001] [Medline: 28602803]
31. Lings P, Evans P, Seamark D, Seamark C, Sweeney K, Dixon M, et al. The doctor-patient relationship in US primary care. J R Soc Med 2003 Apr;96(4):180-184 [FREE Full text] [doi: 10.1258/jrsm.96.4.180] [Medline: 12668705]
32. Schlesinger M, Kanouse DE, Martino SC, Shaller D, Rybowski L. Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. Med Care Res Rev 2014 Oct;71(5 Suppl):38S-64S [FREE Full text] [doi: 10.1177/1077558713496321] [Medline: 23999489]
33. Fung CH, Elliott MN, Hays RD, Kahn KL, Kanouse DE, McGlynn EA, et al. Patients' preferences for technical versus interpersonal quality when selecting a primary care physician. Health Serv Res 2005 Aug;40(4):957-977 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00395.x] [Medline: 16033487]
34. Rothenfluh F, Schulz PJ. Physician rating websites: What aspects are important to identify a good doctor, and are patients capable of assessing them? A mixed-methods approach including physicians' and health care consumers' perspectives. J Med Internet Res 2017 May 1;19(5):e127 [FREE Full text] [doi: 10.2196/jmir.6875] [Medline: 28461285]
35. Jayanti RK, Whipple TW. Like Me … Like Me Not: the role of physician likability on service evaluations. J Mark Theory Pract 2008;16(1):79-86. [doi: 10.2753/MTP1069-6679160106]
36. McCroskey JC, Teven JJ. Goodwill: a reexamination of the construct and its measurement. Commun Monogr 1999;66(1):90-103. [doi: 10.1080/03637759909376464]
37. Ohanian R. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 1990;19(3):39-52. [doi: 10.1080/00913367.1990.10673191]
38. Richmond VP, Heisel AM, Smith Jr RS, McCroskey JC. The impact of communication apprehension and fear of talking with a physician on perceived medical outcomes. Commun Res Rep 1998;15(4):344-353. [doi: 10.1080/08824099809362133]
39. Klingle RS, Burgoon M. Patient compliance and satisfaction with physician influence attempts: a reinforcement expectancy approach to compliance-gaining over time. Commun Res 1995;22(2):148-187. [doi: 10.1177/009365095022002002]
40. Dennis SA, Goodson BM, Pearson C. Mturk workers' use of low-cost 'Virtual Private Servers' to circumvent screening methods: a research note. SSRN Journal 2018:-. [doi: 10.2139/ssrn.3233954]
41. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
42. Fanjiang G, von Glahn T, Chang H, Rogers WH, Safran DG. Providing patients web-based data to inform physician choice: if you build it, will they come? J Gen Intern Med 2007 Oct;22(10):1463-1466 [FREE Full text] [doi: 10.1007/s11606-007-0278-1] [Medline: 17653810]
43. Kalman YM, Rafaeli S. Online pauses and silence: chronemic expectancy violations in written computer-mediated communication. Commun Res 2011;38(1):54-69. [doi: 10.1177/0093650210378229]
44. Walther-Martin W. Media-generated expectancy violations: a study of political humor, race, and source perceptions. Western J Commun 2015;79(4):492-507. [doi: 10.1080/10570314.2015.1072233]
45. Campo S, Cameron KA, Brossard D, Frazer MS. Social norms and expectancy violation theories: assessing the effectiveness of health communication campaigns. Commun Monogr 2004;71(4):448-470. [doi: 10.1080/0363452042000307498]
46. Burgoon M, Birk TS, Hall JR. Compliance and satisfaction with physician-patient communication: an expectancy theory interpretation of gender differences. Hum Commun Res 1991 Dec;18(2):177-208. [doi: 10.1111/j.1468-2958.1991.tb00543.x]
47. PricewaterhouseCoopers. Top Health Industry Issues of 2019: The New Health Economy Comes of Age URL: https://www.pwc.com/us/en/industries/health-services/pdf/pwc-us-healthcare-top-health-industry-issues-2019.pdf [accessed 2019-10-09]
48. Fleischer J, Yarborough R, Piper J. NBC New York. 2018 Oct 29. Doctors: Get a Second Opinion Before Trusting Online Medical Reviews URL: https://www.nbcnewyork.com/news/national-international/doctors-get-a-second-opinion-before-trusting-online-medical-reviews/2064660/ [accessed 2020-02-07]
49. Taylor P. The Globe and Mail. 2018 May 31. Doctors Can Pay to Hide Negative Reviews on Websites Like RateMDs.com. Should We Use Them? URL: https://www.theglobeandmail.com/life/health-and-fitness/article-doctors-can-pay-to-hide-negative-reviews-on-websites-like-ratemdscom/ [accessed 2019-10-09]
50. Wang SS. The Wall Street Journal. 2017 Jun 25. What Doctors Are Doing About Bad Reviews Online: Reputation-Management Firms Help Physicians Undo the Damage Done by Critical Feedback URL: https://www.wsj.com/articles/what-doctors-are-doing-about-bad-reviews-online-1498442580 [accessed 2019-10-09]

Abbreviations

ANOVA: analysis of variance

PCC: patient-centered care
RQ: research question

Edited by G Eysenbach; submitted 09.10.19; peer-reviewed by K Willes, P Schulz; comments to author 18.11.19; revised version received 05.12.19; accepted 27.01.20; published 14.04.20.

Please cite as:
Perrault EK, Hildenbrand GM
The Buffering Effect of Health Care Provider Video Biographies When Viewed in Combination With Negative Reviews: "You Can't Fake Nice"
J Med Internet Res 2020;22(4):e16635
URL: https://www.jmir.org/2020/4/e16635
doi:10.2196/16635
PMID:32286234

Review

Effectiveness of Serious Games to Increase Physical Activity in Children With a Chronic Disease: Systematic Review With Meta-Analysis

Daniël Bossen[1*], PhD; Aline Broekema[1*], MSc; Bart Visser[1], PhD; Annette Brons[2], MSc; Annieck Timmerman[3], PhD; Faridi van Etten-Jamaludin[4], BSc; Katja Braam[1], PhD; Raoul Engelbert[1,5,6], PhD

[1]ACHIEVE Center of Applied Research, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[2]Digital Life Centre, Faculty of Digital Media and Creative Industries, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[3]Topsport Amsterdam, Amsterdam, Netherlands

[4]Amsterdam University Medical Center, Medical Library, University of Amsterdam, Amsterdam, Netherlands

[5]Amsterdam University Medical Center, Rehabilitation, University of Amsterdam, Amsterdam, Netherlands

[6]Amsterdam University Medical Center, Pediatrics, University of Amsterdam, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Daniël Bossen, PhD
ACHIEVE Center of Applied Research
Faculty of Health
Amsterdam University of Applied Sciences
Tafelbergweg 51
Amsterdam, 1105 BD
Netherlands
Phone: 31 621156637
Email: d.bossen@hva.nl

Abstract

Background: Physical activity (PA) is important for children with a chronic disease. Serious games may be useful to promote PA levels among these children.

Objective: The primary purpose of this systematic review was to evaluate the effectiveness of serious games on PA levels in children with a chronic disease.

Methods: PubMed, EMBASE, PsycINFO, ERIC, Cochrane Library, and CINAHL were systematically searched for articles published from January 1990 to May 2018. Both randomized controlled trials and controlled clinical trials were included to examine the effects of serious games on PA levels in children with a chronic disease. Two investigators independently assessed the intervention, methods, and methodological quality in all articles using the Cochrane risk of bias tool. Both qualitative and quantitative analyses were performed.

Results: This systematic review included 9 randomized controlled trials (886 participants). In 2 of the studies, significant between-group differences in PA levels in favor of the intervention group were reported. The meta-analysis on PA levels showed a nonsignificant effect on moderate to vigorous PA (measured in minutes per day) between the intervention and control groups (standardized mean difference 0.30, 95% CI –0.15 to 0.75, *P*=.19). The analysis of body composition resulted in significantly greater reductions in BMI in the intervention group (standardized mean difference –0.24, 95% CI –0.45 to 0.04, *P*=.02).

Conclusions: This review does not support the hypothesis that serious games improve PA levels in children with a chronic disease. The meta-analysis on body composition showed positive intervention effects with significantly greater reductions in BMI in favor of the intervention group. A high percentage of nonuse was identified in the study of serious games, and little attention was paid to behavior change theories and specific theoretical approaches to enhance PA in serious games. Small sample sizes, large variability between intervention designs, and limited details about the interventions were the main limitations. Future research should determine which strategies enhance the effectiveness of serious games, possibly by incorporating behavior change techniques.

(J Med Internet Res 2020;22(4):e14549) doi:10.2196/14549

KEYWORDS

video games; computer games; pediatrics; chronic disease; exercise therapy; health education

Introduction

Worldwide, there are many children who have been diagnosed with a chronic disease [1,2]. A disease or condition is considered chronic in childhood if all 4 of the following conditions are met: (1) it occurs in children aged 0-18 years; (2) the diagnosis is based on medical scientific knowledge and can be established using reproducible and valid methods or instruments according to professional standards; (3) it is not (yet) curable or, for mental health conditions, it is highly resistant to treatment; and (4) it has been present for >3 months; it will, very probably, last >3 months; or it has occurred ≥3 times during the past year and will probably recur again [3]. Having a chronic disease during childhood can impact all elements of growth and development, including physical, psychosocial, and emotional functioning [4]. Physical activity (PA) is important for general health and to minimize the impact of chronic diseases on children. Unlike healthy children, children with chronic diseases are more often restricted in PA, such as playing, running and leisure activities. As a consequence, children with a chronic disease are less physically active than their healthy peers [5,6]. Participation in PA is of particular importance for children with a chronic disease [7-12]. In children with a chronic disease with a physical cause such as juvenile idiopathic arthritis, PA can improve muscle strength and physical fitness without exacerbating joint pain [8]. And, for children with type 1 diabetes mellitus, a physically active lifestyle is beneficial for glycemic control [7] and insulin sensitivity [11]. In addition, participation in sports and PA improves social functioning and mental health for all children [10].

To increase PA levels in children with a chronic disease, various physical exercise interventions have been developed, but not all have had substantial and significant effects [13-15]. The growing popularity of videogames has led to the development of serious games to promote PA in children. Serious games are especially suitable for children as these interventions use interactive and visual strategies that match the learning style of these 'digital natives' [16]. Well-designed games are adjustable in content, provide a feeling of satisfaction, and are challenging; they match the personal interest, motor skills, and cognitive levels of children [16]. Serious games are accessed through a personal computer, game console, tablet, or a smartphone, all of which are commonly used in a general family's daily life. According to Bergeron [17], serious games are defined as interactive computer applications, with or without a hardware component, that have challenging goals, are fun to play, are engaging, incorporate concepts of scoring, and impart skills, knowledge, or attitudes to the user that can be applied in the real world. Whereas the hardware is comparable between games, the program itself and working mechanisms may differ. For example, Dance Dance Revolution (Konami of America, Inc., Redwood City, CA) is an exergame available on different platforms that uses cameras, motion sensors, and force sensors to encourage dancing. Other games, such as Reumaatjes@work, motivate and educate children through an interactive website with films, puzzles, and brain twisters to better cope with having childhood juvenile idiopathic arthritis and promote PA [18].

According to previous reviews, serious games have the potential to promote a physically active lifestyle in children. Especially in studies assessing the effectiveness of serious games in overweight children, first results indicate that serious games can help increase energy expenditure, heart rate, and program compliance [19]. Other studies have shown improvements in specific health outcomes, such as lung function and glycated hemoglobin A1C [20,21]. The evidence from reviews in healthy children and adolescents also supports the argument that serious games effectively promote quantitative [22-25] and qualitative PA. [26,27]. However, the long-term effectiveness on PA maintenance is not well known [22-24].

Despite the potential of serious games for healthy and overweight children, less is known about the effectiveness of serious games that promote PA in children with a chronic disease. Compared with healthy and overweight children, children with a chronic disease have unique perceived barriers and other perspectives and desires with respect to PA and serious games. In children with juvenile idiopathic arthritis, for example, pain has a major impact on the performance of PA in daily life [28], while children with diabetes have a fear of hypoglycemia [29]. It would be worthwhile to further investigate the therapeutic options for serious games in children with a chronic disease, especially when targeting PA behavior. Therefore, the aim of the present review was to study the effectiveness of serious games that promote PA in children with a chronic disease on the outcome of PA, compared to any control group condition. It was hypothesized that children in the serious game groups displayed higher levels of PA than children in the control groups.

Methods

This review was conducted according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [30] (Multimedia Appendix 1).

Search Strategy

A comprehensive systematic search was performed using PubMed, EMBASE (Ovid), PsycINFO (Ovid), ERIC (Ovid), CINAHL (EBSCO), and Cochrane Central Register of Controlled Trials (CENTRAL) through the Cochrane Library. The search was conducted in May 2018. Since the use of health games started after 1990, the review was limited to studies published after 1990. We wanted to select chronic diseases with a physical cause. Therefore, a combination of the following constructs was used: Chronic Disease AND Child AND Serious games AND Motor activity AND Intervention study. Only publications in English were included. Multimedia Appendix 2 provides the search string used for each database.

Screening Process and Eligibility Criteria

The full review screening process was completed by two reviewers (AB and DB), who independently selected the titles and abstracts meeting the inclusion criteria. When an article met the inclusion criteria, full text articles were obtained for closer inspection. To complete the search, the reference lists were checked, and the titles and abstracts of conference proceedings were scanned. When a conference proceeding met the inclusion criteria, the author was contacted and asked for further information on the study process and possible publications. Any unsolved disagreement between the two reviewers was resolved through discussion with a third reviewer (BV).

Inclusion and Exclusion Criteria

Randomized controlled trials (RCTs) or controlled clinical trials (CCTs) examining the effects of a serious game on PA levels in children with a chronic disease were included. Participants needed to be 6-18 years old. The definition of a chronic disease was based on the criteria by Mokkink et al [3]. We included all serious games that focused on PA behavior and were designed to entertain children with a chronic disease. Serious game interventions fulfilling the criteria of Bergeron [17] were included. These included interventions that (1) had challenging goals, (2) were fun to play and engaging, (3) incorporated some concept of scoring, and (4) provided the player with skills, knowledge, or an attitude that could be applied in the real world. Studies were excluded when PA measurements were not used and when the children, apart from the chronic disease, had a serious intellectual disability. These exclusion criteria were used to increase the comparability of the intervention outcome and population by standardizing the knowledge and understanding to general age-matched education levels.

Data Extraction

Collected data comprised study, study population, and intervention characteristics. The study characteristics were first author, year of publication, country, study design, type of control group, and follow-up, drop-out, and adherence rates. The primary outcome measure for this review was the level of PA. Objective (eg, accelerometer, pedometer) and subjective (eg, questionnaires, PA diary) PA outcomes were systematically extracted. Other outcomes, including cardiorespiratory endurance, muscular strength, body composition, and quality of life, were also retrieved. Data about the study population (eg, disease, age, gender) and intervention characteristics (eg, type and structure of serious games) were extracted. Missing data were requested from study authors.

Quality Assessment

Two reviewers (AB and DB) independently assessed the methodological quality and risk of bias criteria (at the study level) of the articles using the Cochrane Collaboration's tool for assessing the risk of bias [31]. Items were rated as low risk, high risk, or unclear when there were no data to assess the criteria. Items scored as low risk received 1 point. Items scored as high risk or unclear received 0 points. In line with the PRISMA guidelines, an interrater process was adopted, and the degree of agreement was assessed to reduce the risk of bias. Disagreements between the authors on the risk of bias were resolved by discussion, with involvement of a third review author (BV) when necessary.

Data Synthesis and Statistical Analyses

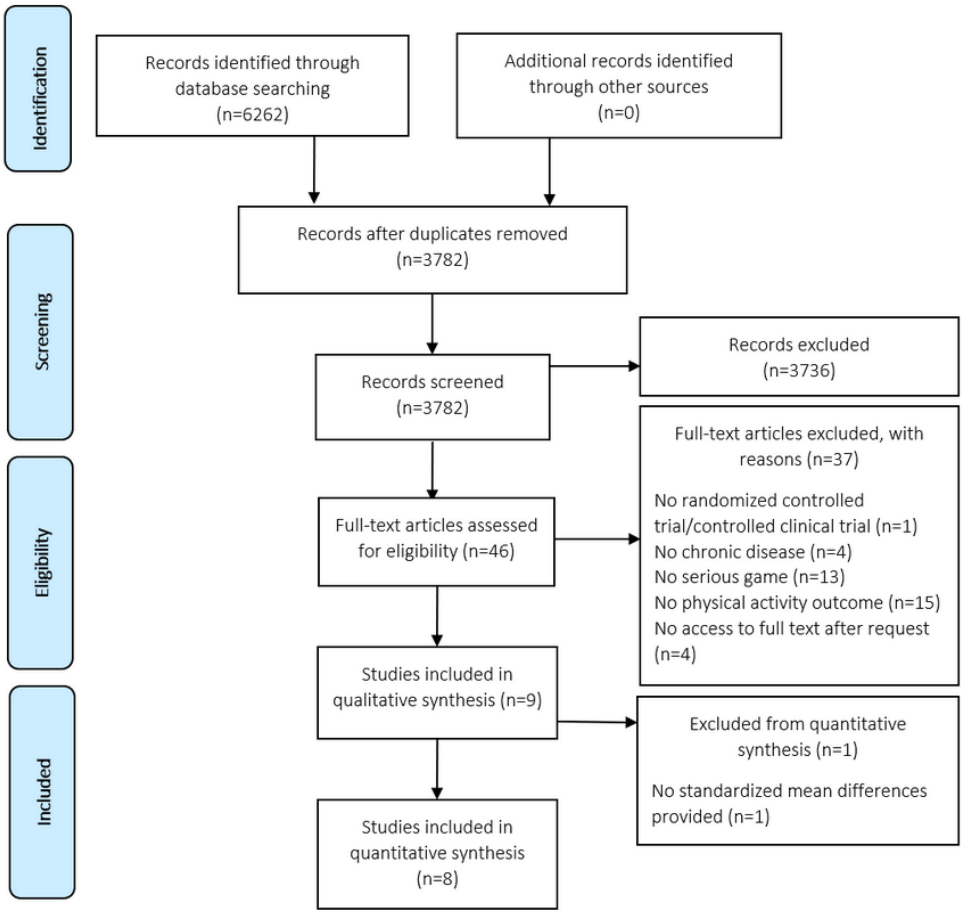
When possible, data were pooled to assess the combined effects of the studies on PA levels. The mean outcome difference and SD in PA were retrieved from the data of each study. Continuous outcomes are presented as standardized mean difference (SMD) scores with 95% CIs. When indicated, data were converted from measurement in weeks to days, and where moderate and vigorous PA were separate groups, they were combined into a moderate-to-vigorous PA group. Heterogeneity in the effect measures between the studies was assessed using both the Chi-square test and I^2 statistic [32] and interpreted following the thresholds in the Cochrane Handbook for Systematic Reviews of Interventions, where a P value <.1 and/or an I^2 value >50% may represent substantial heterogeneity [31]. A fixed-effects model was used when heterogeneity was low; otherwise, a random-effects model was applied. The meta-analysis was performed on the latest follow-up point for each study. In addition, a sensitivity analysis based on study quality was conducted. Since Web-based interventions contain game elements other than exergames, we performed a subgroup analysis with respect to the type of intervention. Moreover, we executed a subgroup analysis on diagnosis. The meta-analysis was performed using Review Manager (RevMan) version 5.3 (The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen) software. The protocol of this review is available in the International Prospective Register of Systematic Reviews: CRD42018070662.

Results

Search Results

The initial search yielded 6262 publications: 1577 from PubMed, 1939 from EMBASE, 355 from PsycINFO, 362 from ERIC, 1397 from Cochrane Central, and 632 from CINAHL. After duplicates were removed, 3782 articles remained for which the titles and abstract were screened. Finally, 46 full-text articles were assessed for eligibility. Consensus about the inclusion of one study was not reached [18]; after discussion with the third reviewer, this study was included. Eventually, the review included 9 studies, of which 8 could be used in the meta-analysis. The study by Armbrust et al [18] could not be included in the meta-analysis because it used median outcome scores. An overview of the search strategy is provided in Figure 1.

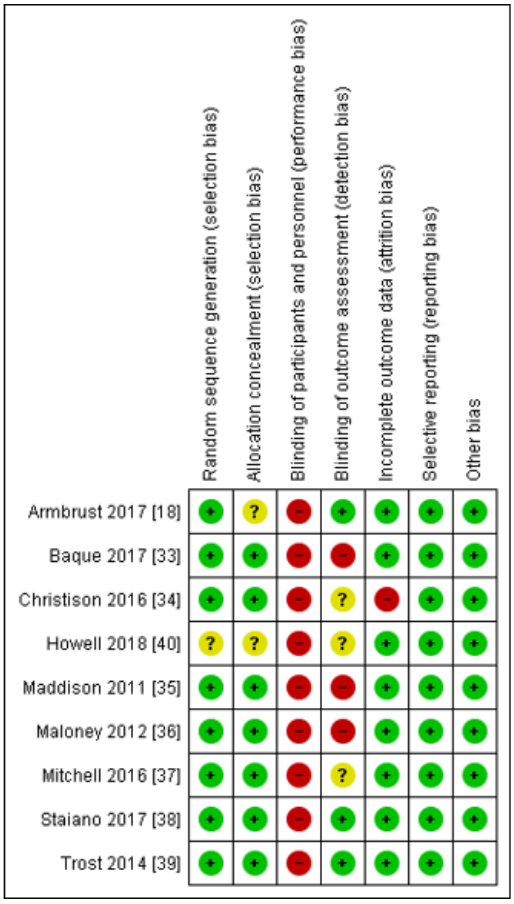
Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart showing the selection procedure for the studies in this systematic review and meta-analysis.



Methodological Quality Assessment of the Manuscripts

Methodological quality assessment resulted in 98% interrater agreement between the two assessors with a kappa value of 0.98 (95% CI 0.95 to 1.00). The majority of the studies showed no bias in the random sequence generation [18,33-39], allocation concealment [33-39], and complete outcome data [18,33,35-40]. None of the studies fulfilled the criteria of blinding the participants and personnel. In 6 studies, the outcome assessment was not blinded, or the blinding was unclear [33,40]. None of the studies had selective reporting. Of the 9 studies, 8 [18,33,35-40] had complete outcome data and therefore had a low risk of attrition bias. Regarding other forms of bias, commercial sponsorship was investigated. We did not detect any financial conflicts of interest or other forms of bias. The risk of bias is presented in Figure 2.

Figure 2. Risk of bias summary of the included studies.



Study Characteristics

Table 1 shows the characteristics of the included studies. All 9 studies used a 2-arm RCT design. All studies used an accelerometer to assess PA levels. In addition to the accelerometers, 4 studies measured objective PA using step counts with a pedometer [33,36]. Three studies used a combination of self-reported and objective PA data [18,36,38]. Other collected outcome measures were cardiorespiratory endurance [18,33-37], body composition [34-36,38,39], muscle strength [33,37,40], and health-related quality of life [18,40]. In 6 of 9 studies, the control group received usual care without any additional care [18,33,35-38]. In one study, the control group underwent a family-based intervention [39], and in 2 studies, the control group received education information [34,40]. Studies used different follow-up periods, and the total study duration varied from 12 weeks [37] to 12 months [18].

Table 1. Characteristics of the studies included in the systematic review and meta-analysis.

Study	Participants				Intervention		Outcomes		Drop-outs	
Author, year, country, study design	Sample size (n for I[a], n for C[b])	Gender (girls), n (%)	Age (years), mean (SD)	Chronic disorder	Treatment	Control	Physical activity measurements	Other outcomes	n for I, n for C	Timing (weeks)
Armbrust, 2017, Netherlands, 2-arm multicenter RCT[c] [18]	49 (28, 21)	33 (67%)	9.9 (8.7 to 11.3[d])	Juvenile idiopathic arthritis	Web-based intervention	Supervised group sessions	AM[e], 7-day activity diary	Cardiorespiratory endurance, quality of life	2, 1	14
Baque, 2017, Australia, 2-arm RCT [33]	60 (30, 30)	26 (45%)	12 (2.5)	Brain injury	Exergame	No intervention (waiting list)	AM, pedometer	Cardiorespiratory endurance, functional muscle strength	4, 3	20
Christison, 2016, USA, 2-arm RCT [34]	84 (60, 24)	46 (58%)	11.1 (1.3)	Obesity	Exergame	Didactic program with family	AM, pedometer	Cardiorespiratory endurance, body composition	24, 8	6[f]
Howell, 2018, USA, 2-arm RCT [40]	94 (63, 31)	43 (55%)	12.7 (1.1)	Cancer survivors	Web-based intervention	Activity monitor and educational materials	AM	Muscle strength, quality of life	10, 6	24
Maddison, 2011, New Zealand, 2-arm RCT [35]	322 (160, 162)	87 (27%)	11.6 (1.1)	Overweight & obesity	Exergame	No intervention (waiting list)	AM	Cardiorespiratory endurance, body composition	20, 12	24
Maloney, 2012, USA, 2-arm RCT [36]	64 (33, 31)	34 (53%)	12.3 (2.4)	Overweight & obesity	Exergame	Pedometers only	AM, pedometer, questionnaire	Body composition, cardiorespiratory endurance	0, 0	12
Mitchell, 2016, Australia, 2-arm RCT [37]	101 (51, 50)	48 (48%)	11.3 (2.5)	Cerebral palsy	Web-based intervention	No intervention (waiting list)	AM, pedometer, questionnaire	Cardiorespiratory endurance, muscle strength	4, 6	20
Staiano, 2017, USA, 2-arm RCT [38]	37 (19, 18)	37 (100%)	15.7 (1.3)	Overweight & obesity	Exergame	No intervention	AM, questionnaire	Body composition	5[g]*	13
Trost, 2014, USA, 2-arm multicenter RCT [39]	75 (34, 41)	41 (55%)	10.0 (1.7)	Overweight &obesity	Exergame	Family-based weight management program	AM	Body composition	3, 3	16

[a]I: intervention group.

[b]C: control group.

[c]RCT: randomized controlled trial.

[d]IQT: interquartile range.

[e]AM: accelerometer.

[f]months.

[g]I + C, not stratified by group.

Sample Characteristics

A total of 886 participants were included in this review. The sample consisted of 582 children diagnosed with obesity [34-36,38,39], 101 children diagnosed with cerebral palsy [37], 94 survivors of childhood cancer [40], 60 children with a previous brain injury [33], and 49 children diagnosed with juvenile idiopathic arthritis [18]. Per study, the sample size

ranged from 37 [38] to 322 patients [35]. Overall, 111 of the 886 participants did not complete the respective study. There were no large gender differences; an almost equal percentage of girls (44.6%; range 27%-100%) and boys (55.4%; range 0%-73%) participated. The age of the participants ranged from 8 to 18 years, with an estimated mean age of 11.8 years (SD 1.7 years).

Intervention Characteristics

Table 2 presents the PA and technology used during the serious game interventions. The first type of serious game was exergames. Exergames involve PA in gameplay, such as dancing, boxing, and balancing exercises using a Nintendo Wii, PlayStation, or Xbox [33-35,38,39]. The second type of serious game was Web-based educational interventions [18,37,40]. These interventions use puzzles, avatars, and brain twisters to encourage PA.

The duration of the interventions ranged from 10 [34] to 24 weeks [35] with a mean duration of 17.3 weeks (Table 3). In 2 studies, children were instructed to 'play' the serious game daily [33,37]; in another 2 studies, children were requested to play the game at least once a week [18,38]. The remaining 5 studies did not specify the frequency of playing [34-36,39,40].

Table 2. Description of the physical activity and technology used in the interventions.

Study	Setting	Type of technology	PA[a] elements	Serious game description
Armbrust et al [18]	Home environment	Web-based application	Arthritis and physical activity education, including barriers, PA benefits, and information about self-efficacy towards becoming more physically active	Films, animations, spoken text, puzzles, brain twisters, and assignments to promote PA; goal setting; email reminders to complete assignments; and a feedback loop to verify whether the child had read the information and finished the assignment. Cognitive behavioral theory was used.
Baque et al [33]	Home environment	Exergame with internet-connected computer and Microsoft Kinect	Gross motor activities combined with cognitive and visual perception and upper limb exercises	Gross motor and daily PA assignments represented on a computer. An example is to use cognitive and visual perception and move the upper limb to solve a mathematical equation. Persuasive elements consisted of feedback and positive reinforcement by parents/guardians.
Christison et al [34]	Research laboratory	Exergame through a PlayStation and Nintendo Wii	Aerobic and muscle strength exercises	A group activity with several games, including aerobic dance, interactive stationary biking, hitting/kicking targets, and boxing. The games used goal setting and documentation of PA in diaries rewarded with small incentives (not specified).
Howell et al [40]	Home environment	Web-based application	Promotion of moderate to vigorous physical activities	The goal was to progress the avatar through various levels on a website. Educational materials, an activity monitor, and access to an interactive website were used to encourage PA via rewards. Points could be redeemed for small prizes (eg, t-shirts, stickers) and/or gift cards.
Maddison et al [35]	Home environment	Exergame through a PlayStation	Promotion of light- to moderate-intensity physical activity	The games were Play3, Kinetic, Sport, and Dance Factory. This was combined with information and education about PA.
Maloney et al [36]	Home environment	Exergame through a PlayStation and Wii	Promotion of physical dancing	Games to encourage dancing.
Mitchell et al [37]	Home environment	Web-based application	Functional gross motor exercises such as sit-to-stand, squatting, and balancing	The Web-based exercises involved upper limb and visual-perceptual games. Examples of active video games are flying a spaceship while squatting and balancing on foam or lunging to shoot a pirate ship with a cannon ball.
Staiano et al [38]	Research laboratory	Exergame through an Xbox 360 console	Encouragement of whole-body movement and moderate-intensity energy expenditure.	Different dance games. Games, songs, dance mode, intensity level, and dance partner were self-selected by the participant.
Trost et al [39]	Schools and young men's Christian associations	Exergame through an Xbox 360 console	Not specified	During the second session, the JOIN for ME program was supplemented with an active sports game. A second active game was provided in week 9 of the JOIN for ME program. No explicit advice or goal was given regarding the use of the active gaming tool.

[a]PA: physical activity.

Table 3. Characteristics of the serious game interventions.

Study	Duration (weeks)	Frequency/ intensity	Guidance and supervision	Measurement points	Game adherence
Armbrust et al [18]	14	Weekly/not specified	4 supervised group sessions with parents and children and contact via email with research personnel. Parents were requested to participate.	Baseline, 3 months, and 12 months	Not specified
Baque et al [33]	20	30 minutes per day, 60 hours game play in total	Supervision by a caregiver and assessment of online adherence. Parents were requested to participate.	Baseline and 20 weeks	Mean 17.57 hours (SD 14.9 hours, range 0-46.14 hours) of 'Move it to improve it' (Mitii) training, average of 52.68 logins (SD 39.98 logins)
Christison et al [34]	10	Not specified	10 supervised group sessions, 4 monthly maintenance sessions by a dietitian or counselor, and medical students as facilitators. Parents were requested to participate.	Baseline, 10 months, and 6 months	Not specified
Howell et al [40]	24	Not specified	None	Baseline, 12 weeks, and 24 weeks	Not specified
Maddison et al [35]	24	Not specified	None	Baseline, 12 weeks, and 24 weeks	At 12 weeks, 15.5 minutes a day (SD 26.3 minutes a day); at 24 weeks, 10.2 minutes a day (SD 23.9 minutes a day)
Maloney et al [36]	12	Not specified	2-6 contacts with research personnel over a 20-week period. Contact by e-mail or fax was made if participants did not use the program for 2 weeks.	Baseline and 12 weeks	89 minutes per week over the 12-week period
Mitchell et al [37]	20	30 minutes a day on 6 days a week	Contact with therapists via email, telephone, or video conferencing for encouragement and technical support.	Baseline and 20 weeks	32.4 hours (SD 17.2 hours) of training over the 20-week period, logging in for 24.2 minutes (SD 5.5 minutes) on an average of 77.7 days
Staiano et al [38]	12	60 minutes a day, 3 times a week	Three “Gaming Coaches” were present to supervise the game sessions.	Baseline and 14 weeks	Not specified
Trost et al [39]	16	Not specified	The JOIN for Me program consisted of 16 weekly supervised sessions with groups of children and parents.	Baseline, 8 weeks, and 16 weeks	8 children discontinued the program

In 7 of 9 studies, the intervention was combined with personal guidance from a healthcare provider or researcher [18,33,34,36-39]. In 3 interventions, the guidance consisted of supervised group sessions [18,34,39]. In another 3 studies, one-on-one guidance was provided by a therapist or game coach [33,36,38], and 1 study used digital communication strategies, including email, telephone, or video conferencing [37]. Additionally, only 1 study (Armbrust et al [18]) reported the use of a health behavior theory (cognitive behavioral theory) to support the design of the intervention.

Main Outcomes

Narrative Analysis

Significant between-group differences in PA levels in favor of the intervention group were reported by 2 studies [38,39]. The first study by Trost et al [39] reported statistically significant increases in moderate-to-vigorous (7.4 minutes a day) and vigorous (2.8 minutes a day) PA. In the second study by Staiano et al [38], significant positive effects were observed for self-reported PA levels but not for accelerometer-measured PA. Although most other studies reported significant within-group results in favor of the intervention group [18,33,36,40], these differences were not significant when compared between the study groups.

Body composition was measured in 5 studies [34-36,38,39]. Of these studies, 3 [35,38,39] reported significant end-of-study measurement differences between the intervention and control groups. With respect to cardiorespiratory endurance, there were 6 studies reporting on this outcome [18,33-37]; only the trial by Mitchell et al [37] found significant improvements in the intervention group. Muscle strength was measured in 3 studies [33,37,40], of which 2 studies [33,37] reported a significant increase in muscle strength in favor of the intervention group. Two studies assessed health-related quality of life [18,40] and reported no between-group differences on this outcome measure.

Meta-analysis

For the primary outcome of PA, accelerometer and pedometer data were extracted to perform a meta-analysis to investigate the effectiveness of serious games on objective PA levels (Figures 3 and 4). The outcome was described as minutes per day of moderate to vigorous PA. The study by Christison et al [34] could not be included in the accelerometer meta-analysis, since this study used only step counts to measure PA. The meta-analysis results showed a positive but nonstatistically significant effect in favor of the intervention group (SMD 0.30, 95% CI –0.15 to 0.75, *P*=.19). Assessment of statistical heterogeneity indicated high study heterogeneity (χ^2_6=35.9, *P*<.001, I^2=83%), which supported the performance of a random effects meta-analysis. In addition to minutes per day of moderate to vigorous PA, we also performed a meta-analysis on step counts. The meta-analysis showed a nonsignificant negative effect on step counts [33,34,36,37], which means that the intervention group had lower step counts than the control group (SMD –0.22, 95% CI –0.69 to 0.26, *P*=.37) with statistical heterogeneity (χ^2_3=6.3, *P*=.10, I^2=52%).

Figure 3. Forest plot for the meta-analysis of moderate to vigorous physical activity (minutes per day).

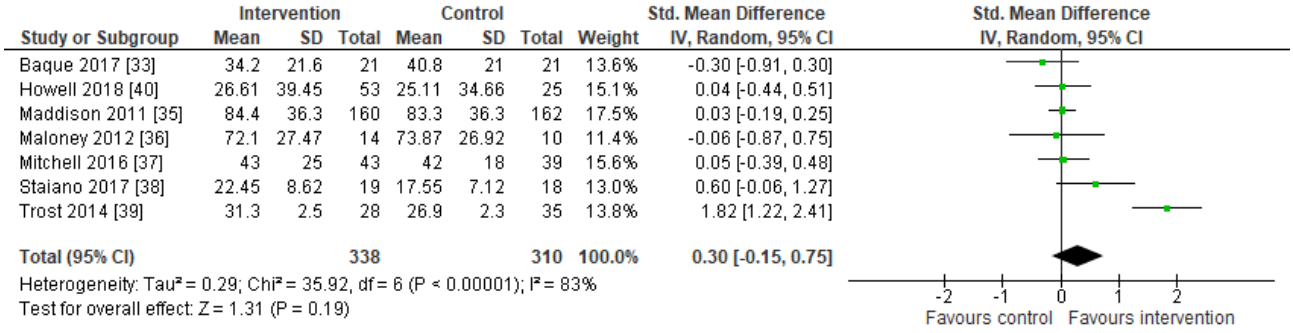
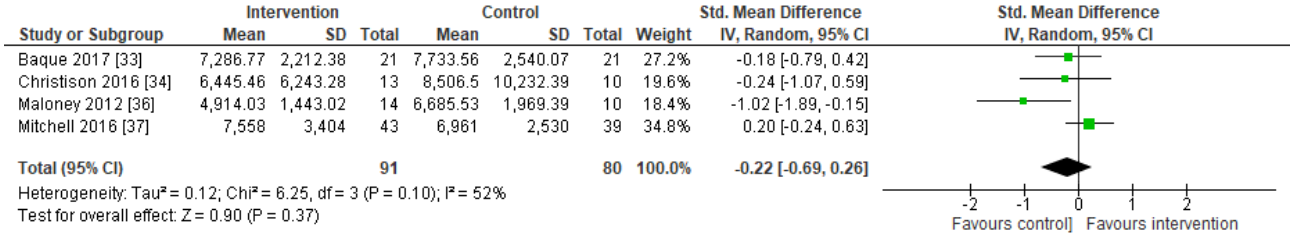


Figure 4. Forest plot for the meta-analysis of physical activity (step counts).



Regarding other outcomes, 3 other meta-analyses were performed on cardiorespiratory endurance, functional muscle strength, and body composition. The forests plots are provided in Figures 5-7. Compared to the control group, serious games had no significant influence on submaximally tested cardiorespiratory endurance (SDM 0.32, 95% CI –0.01 to 0.66, *P*=.06) or functional muscle strength (SMD 0.46, 95% CI –0.12 to 1.04, *P*=.12). However, there was a significant intervention effect for BMI; the studies reported a reduction in BMI in favor of the intervention group (SMD –0.24, 95% CI –0.45 to –0.04, *P*=.02). A meta-analysis on health-related quality of life could not be performed due to different testing instruments and outcome presentation [18,40].

Figure 5. Forest plot for the meta-analysis of cardiorespiratory endurance (6-minute walk test).

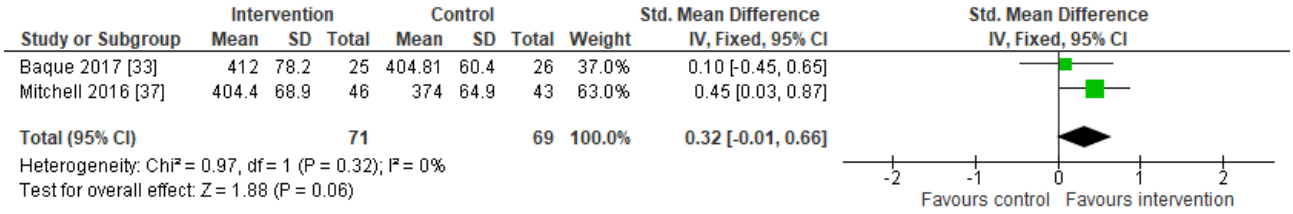


Figure 6. Forest plot for the meta-analysis of functional muscle strength.

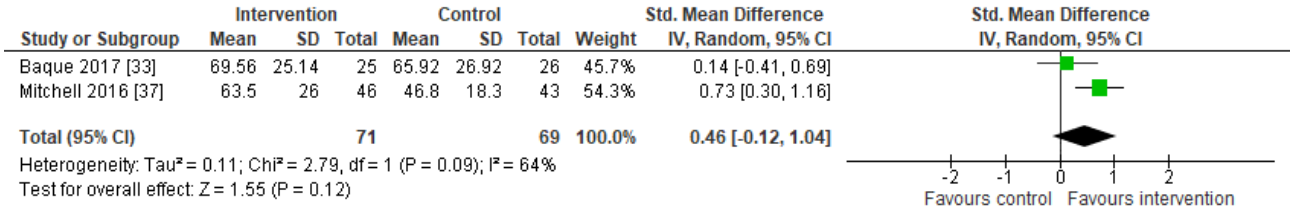
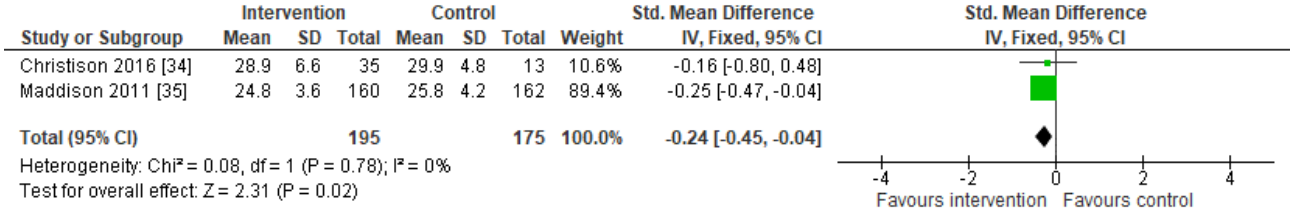


Figure 7. Forest plot for the meta-analysis of BMI.



Subgroup and Sensitivity Analysis

Sensitivity analysis was performed for the primary outcome measure PA by excluding the studies with the highest risk of bias. After removal, the direction and magnitude of the SMD did not markedly differ. We also performed two subgroup analyses based on disease stratification (obesity and overweight versus other chronic diseases) and type of intervention (Web-based interventions versus exergame interventions). Additionally, we performed a post-hoc analysis for age by excluding the study by Staiano et al (>15 years) [38]. These analyses showed no significant difference in intervention effectiveness (Multimedia Appendix 3).

Discussion

This review and meta-analysis showed no significant effects of serious games on PA levels in children with a chronic disease. Only the meta-analysis on BMI showed positive intervention effects, with a significantly greater reduction in BMI in the intervention group. This review is the first to assess the effectiveness of serious games on PA in children with a chronic disease, while former reviews studied the effectiveness of serious games in the promotion of PA levels in healthy children [21-24,26,27,41,42]. Results of these reviews were inconsistent. Some showed significant positive effects on PA [21-23,26,27], another was inconclusive [24], and others did not find positive effects [41,42].

We wondered why the serious games in this review failed to increase PA in children with a chronic disease. A possible explanation can be related to poor adherence rates and low intensity during use. In serious games, it is uncommon to have specific time frames or necessary duration of playtime. Serious games provide (a perception of) autonomy by allowing a child to choose their own exercise levels and own training intensity [43,44]. However, availability and simple promotion of serious games does not automatically lead to improved PA levels [45]. The success of serious games requires explicit instructions and active participation, and children must complete a certain amount of required game content to reach sufficient levels of PA in daily life. In training programs for children, the so-called F.I.T.T. factors (frequency, intensity, time, and type) are recommended [46]. According to these factors, the frequency of training should be at least twice per week for at least 12 weeks, the intensity should be higher than 66% of peak heart rate, and the duration should be between 20 and 60 minutes per session [47]. Of the 9 included studies, 4 studies reported substantial rates of nonuse [34,36-38], which might be the reason for a lack of change in PA behavior. These poor adherence rates are not limited to video games and are recognized as a universal problem in all types of PA exercise interventions [48].

Another factor that may have negatively influenced serious game effects in our review is the absence of information on behavior change techniques in the design of the interventions to enhance PA behavior. In general, behavior change techniques are recommended in the design of complex health service interventions [49] and are important in the promotion of PA [50]. There were only a few studies that mentioned the use of goal setting, rewards, and positive reinforcement to stimulate PA levels [18,33,40]. Other studies paid no or little attention to behavior change techniques [35-39].

Offering a serious game alone is not enough. Serious games should contain support from others in order to increase PA behavior in children. Parents are the most important role model for children and therefore can stimulate their PA participation [51,52]. In the study by Trost et al [39], parents had an active role during the intervention period. This was the study that reported positive effects of the intervention on PA during the assessment of a 16-week family-based pediatric weight management program. Due to the mixed intervention design in this study, it is difficult to tell which part of the intervention was the most important to increase PA. In 4 other studies, the parents had only an administrative role during the intervention period; this role included signing informed consent forms and providing input for data collection such as baseline characteristics and study administration purposes [35,36,38,40].

There are several limitations that need to be addressed. First, in this review we could include only a limited number of studies, almost all of which had small sample sizes. Therefore, it should be noted that the findings with respect to the other outcomes (ie, cardiorespiratory endurance, functional muscle strength, and body composition) must be interpreted with caution since these results were based on only 2 studies. Second, due to limited details about the game structure and exercises included in the serious games, the repeatability of the included studies is low. Third, the interventions were not homogeneous. This might be the consequence of the variation in the primary aim between interventions; some studies focused on a multicomponent intervention to achieve weight loss and increase PA, while others only focused on increasing PA. Last, 5 of the 9 studies [34-36,38,39] addressed overweight and obese participants. It could be that those studies were predominantly focused on body composition, rather than PA, making it difficult to compare those studies.

In line with the findings and limitations of this systematic review, there is a need for more high-quality studies with larger homogeneous sample sizes to better understand the long-term

impact of serious games. Furthermore, future research should determine which components, including game duration, intensity, genre, and behavior change techniques, enhance the effectiveness of serious games. Research targeting those components is essential for the development of effective serious games. In addition, studies should aim to determine which strategies are effective to improve game adherence. Studying dose-response relationships between game exposure and PA as the outcome would be useful. Since this review included heterogeneous interventions and samples with a diversity of diseases, future reviews should use a more homogeneous approach.

Acknowledgments

The authors acknowledge the SIMBA consortium partners for their participation in this study. This review is performed as part of the SIMBA study, which is funded by the directing body SIA, part of the Dutch Organization for Scientific Research (NWO).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 19 KB - jmir_v22i4e14549_app1.docx]

Multimedia Appendix 2
Keywords (PubMed version).
[DOCX File , 13 KB - jmir_v22i4e14549_app2.docx]

Multimedia Appendix 3
Subgroup and sensitivity analyses.
[DOCX File , 91 KB - jmir_v22i4e14549_app3.docx]

References

1. Van Cleave J, Gortmaker SL, Perrin JM. Dynamics of obesity and chronic health conditions among children and youth. JAMA 2010 Feb 17;303(7):623-630 [FREE Full text] [doi: 10.1001/jama.2010.104] [Medline: 20159870]
2. Wijlaars LPMM, Gilbert R, Hardelid P. Chronic conditions in children and young people: learning from administrative data. Arch Dis Child 2016 Oct 31;101(10):881-885 [FREE Full text] [doi: 10.1136/archdischild-2016-310716] [Medline: 27246068]
3. Mokkink LB, van der Lee JH, Grootenhuis MA, Offringa M, Heymans HSA, Dutch National Consensus Committee Chronic DiseasesHealth Conditions in Childhood. Defining chronic diseases and health conditions in childhood (0-18 years of age): national consensus in the Netherlands. Eur J Pediatr 2008 Dec;167(12):1441-1447. [doi: 10.1007/s00431-008-0697-y] [Medline: 18340463]
4. Blackman J, Gurka M, Gurka K, Oliver M. Emotional, developmental and behavioural co-morbidities of children with chronic health conditions. J Paediatr Child Health 2011 Oct;47(10):742-747. [doi: 10.1111/j.1440-1754.2011.02044.x] [Medline: 21449905]
5. Burghard M, de Jong NB, Vlieger S, Takken T. 2017 Dutch Report Card: Results From the First Physical Activity Report Card Plus for Dutch Youth With a Chronic Disease or Disability. Front Pediatr 2018;6:122 [FREE Full text] [doi: 10.3389/fped.2018.00122] [Medline: 29761094]
6. West SL, Banks L, Schneiderman JE, Caterini JE, Stephens S, White G, et al. Physical activity for children with chronic disease; a narrative review and practical applications. BMC Pediatr 2019 Jan 08;19(1):12 [FREE Full text] [doi: 10.1186/s12887-018-1377-3] [Medline: 30621667]
7. Aouadi R, Khalifa R, Aouidet A, Ben Mansour A, Ben Rayana M, Mdini F, et al. Aerobic training programs and glycemic control in diabetic children in relation to exercise frequency. J Sports Med Phys Fitness 2011 Sep;51(3):393-400. [Medline: 21904277]
8. Klepper S, Mano Khong TT, Klotz R, Gregorek AO, Chan YC, Sawade S. Effects of Structured Exercise Training in Children and Adolescents With Juvenile Idiopathic Arthritis. Pediatr Phys Ther 2019 Dec;31(1):3-21. [doi: 10.1097/PEP.0000000000000555] [Medline: 30557274]
9. Klijn PHC, Oudshoorn A, van der Ent CK, van der Net J, Kimpen JL, Helders PJM. Effects of anaerobic training in children with cystic fibrosis: a randomized controlled study. Chest 2004 Apr;125(4):1299-1305. [doi: 10.1378/chest.125.4.1299] [Medline: 15078738]

10. Murphy NA, Carbone PS, American Academy of Pediatrics Council on Children With Disabilities. Promoting the participation of children with disabilities in sports, recreation, and physical activities. Pediatrics 2008 May;121(5):1057-1061. [doi: 10.1542/peds.2008-0566] [Medline: 18450913]
11. Salem MA, AboElAsrar MA, Elbarbary NS, ElHilaly RA, Refaat YM. Is exercise a therapeutic tool for improvement of cardiovascular risk factors in adolescents with type 1 diabetes mellitus? A randomised controlled trial. Diabetol Metab Syndr 2010 Jul 11;2(1):47 [FREE Full text] [doi: 10.1186/1758-5996-2-47] [Medline: 20618996]
12. Selvadurai HC, Blimkie CJ, Meyers N, Mellis CM, Cooper PJ, Van Asperen PP. Randomized controlled study of in-hospital exercise training programs in children with cystic fibrosis. Pediatr Pulmonol 2002 Mar;33(3):194-200. [doi: 10.1002/ppul.10015] [Medline: 11836799]
13. Joschtel B, Gomersall SR, Tweedy S, Petsky H, Chang AB, Trost SG. Effects of exercise training on physical and psychosocial health in children with chronic respiratory disease: a systematic review and meta-analysis. BMJ Open Sport Exerc Med 2018;4(1):e000409 [FREE Full text] [doi: 10.1136/bmjsem-2018-000409] [Medline: 30305925]
14. Braam KI, van der Torre P, Takken T, Veening MA, van Dulmen-den Broeder E, Kaspers GJL. Physical exercise training interventions for children and young adults during and after treatment for childhood cancer. Cochrane Database Syst Rev 2016 Mar 31;3:CD008796 [FREE Full text] [doi: 10.1002/14651858.CD008796.pub3] [Medline: 27030386]
15. Clutterbuck G, Auld M, Johnston L. Active exercise interventions improve gross motor function of ambulant/semi-ambulant children with cerebral palsy: a systematic review. Disabil Rehabil 2019 May;41(10):1131-1151. [doi: 10.1080/09638288.2017.1422035] [Medline: 29303007]
16. Benzing V, Schmidt M. Exergaming for Children and Adolescents: Strengths, Weaknesses, Opportunities and Threats. J Clin Med 2018 Nov 08;7(11):E422 [FREE Full text] [doi: 10.3390/jcm7110422] [Medline: 30413016]
17. Bergeron B. Developing Serious Games, 1st edn. In: Developing Serious Games (game Development Series). Massachusetts: Charles River Media; 2020.
18. Armbrust W, Bos GJFJ, Wulffraat NM, van Brussel M, Cappon J, Dijkstra PU, et al. Internet Program for Physical Activity and Exercise Capacity in Children With Juvenile Idiopathic Arthritis: A Multicenter Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2017 Dec;69(7):1040-1049 [FREE Full text] [doi: 10.1002/acr.23100] [Medline: 27696793]
19. Gao Z, Chen S, Pasco D, Pope Z. A meta-analysis of active video games on health outcomes among children and adolescents. Obes Rev 2015 Sep;16(9):783-794. [doi: 10.1111/obr.12287] [Medline: 25943852]
20. Charlier N, Zupancic N, Fieuws S, Denhaerynck K, Zaman B, Moons P. Serious games for improving knowledge and self-management in young people with chronic conditions: a systematic review and meta-analysis. J Am Med Inform Assoc 2016 Jan;23(1):230-239. [doi: 10.1093/jamia/ocv100] [Medline: 26186934]
21. Dias JD, Domingues AN, Tibes CM, Zem-Mascarenhas SH, Fonseca LMM. Serious games as an educational strategy to control childhood obesity: a systematic literature review1. Rev Lat Am Enfermagem 2018 Sep 03;26:e3036 [FREE Full text] [doi: 10.1590/1518-8345.2509.3036] [Medline: 30183872]
22. Parisod H, Pakarinen A, Kauhanen L, Aromaa M, Leppänen V, Liukkonen TN, et al. Promoting Children's Health with Digital Games: A Review of Reviews. Games Health J 2014 Jun;3(3):145-156. [doi: 10.1089/g4h.2013.0086] [Medline: 26196173]
23. Biddiss E, Irwin J. Active video games to promote physical activity in children and youth: a systematic review. Arch Pediatr Adolesc Med 2010 Jul;164(7):664-672. [doi: 10.1001/archpediatrics.2010.104] [Medline: 20603468]
24. Pakarinen A, Parisod H, Smed J, Salanterä S. Health game interventions to enhance physical activity self-efficacy of children: a quantitative systematic review. J Adv Nurs 2017 Apr;73(4):794-811. [doi: 10.1111/jan.13160] [Medline: 27688056]
25. do Carmo J, Goncalves R, Batalau R, Palmeira AL. Active video games and physical activity in overweight children and adolescents. 2013 Presented at: IEEE 2nd International Conference on Serious Games and Applications for Health (SeGAH); May 2-3, 2013; Vilamoura, Portugal p. 1-5. [doi: 10.1109/segah.2013.6665323]
26. Medeiros PD, Capistrano R, Zequinão MA, Silva SAD, Beltrame TS, Cardoso FL. EXERGAMES AS A TOOL FOR THE ACQUISITION AND DEVELOPMENT OF MOTOR SKILLS AND ABILITIES: A SYSTEMATIC REVIEW. Rev Paul Pediatr 2017;35(4):464-471 [FREE Full text] [doi: 10.1590/1984-0462/;2017;35;4;00013] [Medline: 28977134]
27. Merino-Campos C, Del Castillo Fernández H. The Benefits of Active Video Games for Educational and Physical Activity Approaches: A Systematic Review. J New Appr Ed Res 2016 Jul 15;5(2):115-122. [doi: 10.7821/naer.2016.7.164]
28. Nijhof LN, Nap-van der Vlist MM, van de Putte EM, van Royen-Kerkhof A, Nijhof SL. Non-pharmacological options for managing chronic musculoskeletal pain in children with pediatric rheumatic disease: a systematic review. Rheumatol Int 2018 Nov 23;38(11):2015-2025 [FREE Full text] [doi: 10.1007/s00296-018-4136-8] [Medline: 30155667]
29. Jabbour G, Henderson M, Mathieu M. Barriers to Active Lifestyles in Children with Type 1 Diabetes. Can J Diabetes 2016 Apr;40(2):170-172. [doi: 10.1016/j.jcjd.2015.12.001] [Medline: 27038139]
30. 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]
31. Higgins P, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane handbook for systematic reviews of interventions The Cochrane Collaboration; 2018 URL: https://www.cochrane-handbook.org [accessed 2019-04-30] [WebCite Cache ID 781PrgwLe]

32. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
33. Baque E, Barber L, Sakzewski L, Boyd RN. Randomized controlled trial of web-based multimodal therapy for children with acquired brain injury to improve gross motor capacity and performance. Clin Rehabil 2017 Jun;31(6):722-732. [doi: 10.1177/0269215516651980] [Medline: 27271374]
34. Christison AL, Evans TA, Bleess BB, Wang H, Aldag JC, Binns HJ. Exergaming for Health: A Randomized Study of Community-Based Exergaming Curriculum in Pediatric Weight Management. Games Health J 2016 Dec;5(6):413-421. [doi: 10.1089/g4h.2015.0097] [Medline: 27911621]
35. Maddison R, Foley L, Ni MC, Jiang Y, Jull A, Prapavessis H, et al. Effects of active video games on body composition: a randomized controlled trial. Am J Clin Nutr 2011 Jul;94(1):156-163 [FREE Full text] [doi: 10.3945/ajcn.110.009142] [Medline: 21562081]
36. Maloney AE, Threlkeld KA, Cook WL. Comparative Effectiveness of a 12-Week Physical Activity Intervention for Overweight and Obese Youth: Exergaming with. Games Health J 2012 Apr;1(2):96-103. [doi: 10.1089/g4h.2011.0009] [Medline: 26193183]
37. Mitchell LE, Ziviani J, Boyd RN. A randomized controlled trial of web-based training to increase activity in children with cerebral palsy. Dev Med Child Neurol 2016 Jul;58(7):767-773 [FREE Full text] [doi: 10.1111/dmcn.13065] [Medline: 26877078]
38. Staiano AE, Beyl RA, Hsia DS, Katzmarzyk PT, Newton RL. Twelve weeks of dance exergaming in overweight and obese adolescent girls: Transfer effects on physical activity, screen time, and self-efficacy. J Sport Health Sci 2017 Mar;6(1):4-10 [FREE Full text] [doi: 10.1016/j.jshs.2016.11.005] [Medline: 28491483]
39. Trost SG, Sundal D, Foster GD, Lent MR, Vojta D. Effects of a pediatric weight management program with and without active video games a randomized trial. JAMA Pediatr 2014 May;168(5):407-413. [doi: 10.1001/jamapediatrics.2013.3436] [Medline: 24589566]
40. Howell CR, Krull KR, Partin RE, Kadan-Lottick NS, Robison LL, Hudson MM, et al. Randomized web-based physical activity intervention in adolescent survivors of childhood cancer. Pediatr Blood Cancer 2018 Dec;65(8):e27216. [doi: 10.1002/pbc.27216] [Medline: 29722481]
41. LeBlanc AG, Chaput J, McFarlane A, Colley RC, Thivel D, Biddle SJH, et al. Active video games and health indicators in children and youth: a systematic review. PLoS One 2013;8(6):e65351 [FREE Full text] [doi: 10.1371/journal.pone.0065351] [Medline: 23799008]
42. Drummond D, Monnier D, Tesnière A, Hadchouel A. A systematic review of serious games in asthma education. Pediatr Allergy Immunol 2017 Dec;28(3):257-265. [doi: 10.1111/pai.12690] [Medline: 27992659]
43. Hamlyn-Williams CC, Freeman P, Parfitt G. Acute affective responses to prescribed and self-selected exercise sessions in adolescent girls: an observational study. BMC Sports Sci Med Rehabil 2014;6:35 [FREE Full text] [doi: 10.1186/2052-1847-6-35] [Medline: 25285215]
44. Rose EA, Parfitt G. Exercise experience influences affective and motivational outcomes of prescribed and self-selected intensity exercise. Scand J Med Sci Sports 2012 Apr;22(2):265-277. [doi: 10.1111/j.1600-0838.2010.01161.x] [Medline: 20626702]
45. Baranowski T, Abdelsamad D, Baranowski J, O'Connor TM, Thompson D, Barnett A, et al. Impact of an active video game on healthy children's physical activity. Pediatrics 2012 Mar;129(3):e636-e642 [FREE Full text] [doi: 10.1542/peds.2011-2050] [Medline: 22371457]
46. Takken T, Hulzebos EH. Exercise testing and training in chronic childhood conditions. Hong Kong Physiotherapy Journal 2013 Dec;31(2):58-63. [doi: 10.1016/j.hkpj.2013.05.002]
47. Baquet G, van Praagh E, Berthoin S. Endurance training and aerobic fitness in young people. Sports Med 2003;33(15):1127-1143. [doi: 10.2165/00007256-200333150-00004] [Medline: 14719981]
48. Modi AC, Lim CS, Yu N, Geller D, Wagner MH, Quittner AL. A multi-method assessment of treatment adherence for children with cystic fibrosis. J Cyst Fibros 2006 Aug;5(3):177-185 [FREE Full text] [doi: 10.1016/j.jcf.2006.03.002] [Medline: 16679071]
49. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015 Sep;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
50. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: A systematic review and meta-analysis. Soc Sci Med 2017 Nov;192:125-133. [doi: 10.1016/j.socscimed.2017.09.021] [Medline: 28965003]
51. Ritchie LD, Welk G, Styne D, Gerstein DE, Crawford PB. Family environment and pediatric overweight: what is a parent to do? J Am Diet Assoc 2005 May;105(5 Suppl 1):S70-S79. [doi: 10.1016/j.jada.2005.02.017] [Medline: 15867900]
52. Stein D, Weinberger-Litman SL, Latzer Y. Psychosocial perspectives and the issue of prevention in childhood obesity. Front Public Health 2014;2:104 [FREE Full text] [doi: 10.3389/fpubh.2014.00104] [Medline: 25133140]

Abbreviations

AM: accelerometer.
CCT: controlled clinical trial.
PA: physical activity.
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
RCT: randomized controlled trial.
SMD: standardized mean difference.

Edited by G Eysenbach; submitted 01.05.19; peer-reviewed by P Arriaga, B Bock, R Oliveira; comments to author 03.10.19; revised version received 21.11.19; accepted 27.01.20; published 01.04.20.

Please cite as:
Bossen D, Broekema A, Visser B, Brons A, Timmerman A, van Etten-Jamaludin F, Braam K, Engelbert R
Effectiveness of Serious Games to Increase Physical Activity in Children With a Chronic Disease: Systematic Review With Meta-Analysis
J Med Internet Res 2020;22(4):e14549
URL: https://www.jmir.org/2020/4/e14549
doi:10.2196/14549
PMID:32234697

Original Paper

Spoken Animated Self-Management Video Messages Aimed at Improving Physical Activity in People With Type 2 Diabetes: Development and Interview Study

Colette van het Schip[1], MSc; Kei Long Cheung[2], PhD; Stan Vluggen[3], PhD; Ciska Hoving[3], PhD; Nicolaas C Schaper[4], Prof Dr; Hein de Vries[3], Prof Dr

[1]Department of Physiotherapy, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[2]Department of Clinical Sciences, College of Health and Life Sciences, Brunel University London, London, United Kingdom
[3]Caphri School of Public Health and Primary Care, Health Promotion, Maastricht University, Maastricht, Netherlands
[4]Department of Endocrinology and Internal Medicine, Maastricht University Medical Centre, Maastricht, Netherlands

Corresponding Author:
Hein de Vries, Prof Dr
Caphri School of Public Health and Primary Care
Health Promotion
Maastricht University
PO Box 616
Maastricht, 6200 MD
Netherlands
Phone: 31 433882410
Email: hein.devries@maastrichtuniversity.nl

Abstract

Background: Web-based tailored interventions are a promising approach to help people with type 2 diabetes successfully adopt regular physical activity. Spoken animation seems to be effective regardless of the characteristics of the user and may be a relevant strategy to communicate complex health information

Objective: The objectives of our study were to evaluate (1) pretesting communication elements and user appreciation, and (2) the applied behavior change techniques of the previously designed spoken animated video messages in a tailored self-management program for people with type 2 diabetes.

Methods: We conducted semistructured interviews with patients with type 2 diabetes recruited from general practices located in different socioeconomic status urban neighborhoods. Based on the pretesting key communication elements of Salazar's model, we asked participants about the spoken animated video messages' attractiveness, comprehensibility, acceptance, believability, involvement, and relevance and to what extent the video messages motivated them to become more physically active. We also assessed participants' intention to use the spoken animated video messages and to recommend them to others. To evaluate participants' appreciation of the different applied behavior change techniques, we conducted a post hoc analysis of the qualitative data using the MAXQDA program. Transcripts were coded by 2 coders using iterative qualitative content analysis methods to uncover key health communication issues.

Results: Of 23 patients who expressed an interest in participating, 17 met the inclusion criteria and 15 took part in the interviews. The positive appreciation of the comprehensibility, believability, and personalization was supported by participants' statements on behavior change techniques and other communication elements. Reinforcement of and feedback on participants' answers were positively evaluated as was the simplicity and concreteness of the spoken animated video messages. Most participants indicated reasons for not feeling motivated to increase their physical activity level, including being already sufficiently physically active and the presence of other impeding health factors.

Conclusions: Spoken animated video messages should be simple, short, concrete, and without the use of medical terminology. Providing positive reinforcement, feedback on participants' answers, examples that match user characteristics, and the possibility to identify with the animation figures will enhance involvement in the health message. To connect more with patients' needs and thereby increase the perceived relevance of and motivation to use an animated video program, we suggest offering the program soon after diabetes mellitus is diagnosed. We recommend piloting behavior change techniques to identify potential resistance.

(J Med Internet Res 2020;22(4):e15397) doi:10.2196/15397

KEYWORDS

diabetes mellitus, type 2; internet-based intervention; telemedicine; computer tailoring; self-management; exercise; animation; health literacy

Introduction

Background

Type 2 diabetes mellitus is a chronic metabolic disease characterized by poor regulation of blood glucose caused by insulin resistance and beta-cell impairment [1,2]. The prevalence of people with type 2 diabetes in the Netherlands was almost 1.2 million in 2018 and is expected to increase by 34% by 2030 [3]. Socioeconomically disadvantaged and minority populations are especially at risk [4]. As a result, the cost of diabetes care in the Netherlands amounted to €1.7 billion (about US $1.9 billion) in 2011 [3]. Without proper treatment, type 2 diabetes can lead to long-term complications, such as neuropathy, nephropathy, retinopathy, cardiovascular disease, and a lowered quality of life [5]. The treatment of patients with type 2 diabetes is largely dependent on patients' daily self-care, of which regular physical activity (PA), healthy diet, and medication adherence have been shown to be core elements [6,7]. Hence, diabetes self-management is recognized as the key factor of overall diabetes management [8]. Self-management enables patients to take control of their chronic disease by making their own decisions and performing self-chosen actions aimed at improving their health [5].

Scientific evidence shows the importance of regular PA as one way to reduce the risk of complications associated with type 2 diabetes [9,10]. Practicing regular PA may help to improve glycemic control, reduce visceral adiposity, lower plasma triglycerides, and reduce all-cause mortality risk [10]. According to the Dutch recommendations on PA, adults with type 2 diabetes should achieve at least 150 minutes of moderate- to high-intensity PA (≥3 metabolic equivalent tasks [METs]) per week spread over several days [11]. However, most people with type 2 diabetes are physically inactive and do not meet these recommendations [9]. For the Netherlands, in 2016, 40% of people with a chronic condition such as diabetes did not meet the national recommendations [12].

To promote PA self-management for people with type 2 diabetes, web-based computer-tailored support programs have been developed and can help people to improve their health behaviors and thereby their health outcomes [5,13,14]. Computer-tailored interventions are characterized by the provision of feedback (in the form of health messages) adapted to the users' characteristics to target health-related behaviors [15]. Evidence showed that these programs can have positive effects on glycemic control, food intake, and exercise [5,13,14] and increase the effectiveness and reach of clinical-based consultations [5,16]. The relative advantage concerns the ease with which the determinants can be assessed and behavior change techniques (BCTs) can be applied in formulating web-based health messages [17]. However, many of these programs used text-driven health messages that in particular experienced problems reaching people with a low educational level, who are at risk of developing type 2 diabetes [18,19].

To improve the effectiveness of web-based self-management support programs for people with type 2 diabetes, it is a prerequisite to develop health messages that match the user experience, which thus enhances appreciation and usage of the intervention. Tailored health messages can be presented in various delivery modes, such as short message service (SMS) text messaging, video messages, or animations [20]. Exploring different delivery modes is needed, as they have the potential to enhance the effects of computer-tailored interventions. This is illustrated by several previous studies investigating the effects of video tailoring as compared with text tailoring [18,21,22]. To further explore ways to improve the user experience of these tailored interventions, more research on delivery modes is of paramount importance.

Studies showed that the delivery mode of spoken animation, consisting of simple line drawings, is a promising way to communicate complex health information to people with low health literacy [23,24]. Health literacy implies the achievement of a level of knowledge, personal skills, and confidence to take action to improve personal and community health by changing personal lifestyles and living conditions [25].

Animations may be better able than video illustrations to support the creation of an adequate mental representation to enhance learning and understanding [24,26,27]. People with low health literacy are better able to recall and have a more positive attitude toward spoken messages than to written texts [24]. Additionally, since animations do not negatively influence high–health literate target groups, health messages can be adapted to low–health literacy groups, which can lead to a better comprehension [23,24].

Several BCTs may be effective to help people with type 2 diabetes change their unhealthy lifestyles and should thus be applied in formulating health messages. A review by van Vugt et al [5] of the use of BCTs in web-based self-management programs for patients with type 2 diabetes mellitus showed that providing feedback on performance, information on consequences of behavior, barrier identification, and problem solving, and self-monitoring of behavior were linked to positive outcomes for health behavior change, psychological well-being, or clinical parameters.

According to Salazar's model, animated health messages that are appreciated and positively evaluated on pretesting communication elements are a prerequisite to enhance the user experience of the intervention [28]. Health messages should catch the attention and be comprehensible, attractive, and relevant for the target group. BCTs target determinants of behavior and create the content of the messages. It is therefore relevant to evaluate health messages with respect to BCTs [28].

The My Diabetes Profile Intervention

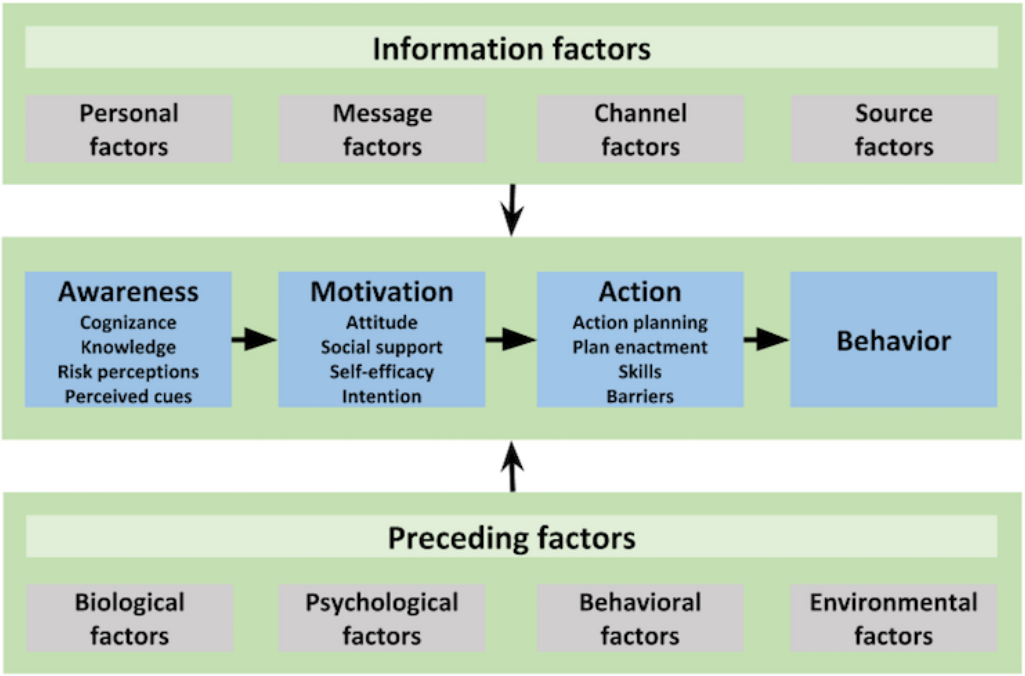
Development

Intervention development took place at Maastricht University. As an alternative to SMS text messaging, we previously developed tailored spoken animated health messages aimed at increasing PA behavior in people with type 2 diabetes [29]. These videos will be part of aweb-based computer-tailored program, My Diabetes Profile, supporting several parts of self-management (Dutch Trial Register NTR6840) [29].

Framework

We developed tailored spoken animated video messages on the basis of the integrated change (I-Change) model [1,30,31] and tailored to address important determinants [1,32,33] of PA in patients with type 2 diabetes. We used the I-Change model [30] as the theoretical framework for developing the video messages, since it integrates several social cognitive theories, including the attitude-social influence-efficacy model [31], social cognitive theory [32], the transtheoretical model [33], the health belief model [34], and implementation and goal-setting theories (Figure 1 [35]) [32,36,37]. In recent years, the use of the I-Change model in tailored interventions has resulted in long-term positive impacts on PA behavior among adults of various age groups [1,34,36]. The model identifies relevant determinants of PA in patients with type 2 diabetes [1,22,32,33].

Figure 1. The integrated change (I-Change) model for increasing self-management behavior. Adapted from De Vries [35].



The I-Change model encompasses 3 phases of the behavior change process: awareness, motivation, and action. It suggests that information factors such as the quality of messages, channels, and sources that are used will determine awareness, motivation, and behavior change, which adds up to the research question. We developed the spoken animated video messages based on the 3 indicated phases in the behavior change process: (1) to increase a person's awareness of the importance of changing behavior, (2) to motivate the person in favor of adopting regular PA, and (3) to transfer this motivation to the adoption of regular PA.

In the web-based computer-tailored program My Diabetes Profile [29], for which we developed the spoken animated video messages, users first receive a tailored health risk appraisal based on their answers to a baseline assessment (video messages on cognizance, knowledge, and risk perception). The health risk appraisal provides information on whether their perceived adherence to the PA recommendations matches with the objective guideline target. Then, if a participant does not meet the recommendations on PA, their intention to change is assessed. A participant who has a low intention to change their PA behavior is then directed to a second session that aims to increase the motivation to change (video messages on attitude, social influence, and self-efficacy) and raise awareness of the need to make improvements. If a participant has a high intention to change their PA behavior, they are directed to session 3, which shows the animated video messages on goal setting, action planning, and coping planning.

Construction of Spoken Animated Feedback Messages

We tailored spoken animated feedback messages according to participants' input considering their actual level of PA and the assessment of the key determinants of PA behavior. Depending on the construct of each concept, we developed 2 (intention, preparatory plans, coping plans), 3 (knowledge, attitude, self-efficacy), 6 (cognizance, risk perception), or 9 (social influence) types of animated video messages for people with a low, a neutral, and a high score on the different determinants. For example, people could indicate on a 2-point Likert scale

whether they did or did not have a coping plan (2 video messages). Attitude questions were answered on a 5-point Likert scale and recoded to a low, neutral, or high score (3 video messages). We combined cognizance, the degree to which people are aware of (the level of) their own health behavior (eg, "I am currently sufficiently physically active (yes/no)"), with a person's actual PA behavior (meeting the norm, or almost or not meeting the norm). We selected effective BCTs based on the I-Change model and earlier tailored studies using this model [30,31] and applied them in formulating animated video messages intending to positively modify the different determinants [1,32,33]. These techniques included repeating the given answer in the questionnaire, providing information on the consequences of a behavior, identifying barriers, and solving problems [5]. Table 1 provides an overview of applied BCTs for each concept. We used the computer program Go Animate (Vyond) [38] to construct the spoken video animations. First, we wrote a script for each determinant, then developed the animations consisting of 1 or more slides per applied BCT. We asked 1 man and 2 women to voice-over the script. We took into account important requirements [19,39] for the appreciation and effectiveness of animated video-tailored health messages. These requirements implied that animated health messages should be shorter than 5 minutes, should be clear and to the point, and should be convincing and motivating [19].

Table 1. Overview of applied behavior change techniques for each determinant of physical activity behavior.

Concept	Behavior change techniques
All health messages	• Repetition of answer • Positive reinforcement (high, medium[a]) • Showing empathy or respect (low)
Determinant	
Cognizance and behavior	• Feedback (meets the physical activity norm?) • Provision of information) • Persuasive communication
Knowledge	• Feedback on performance • Provision of information (chunking)
Risk perception (susceptibility and severity)	• Personalized risk • Provision of information on the consequences of diabetes mellitus or inactivity • Problem solving, self-efficacy–enhancing conclusion
Attitude pros	• Feedback on participants' perception of pros • Counterpersuasion (low, medium) • Gain framing • Affective arguments • Social modeling, peer reference
Attitude cons	• Counterpersuasion (low, medium) • Affective and cognitive arguments • Problem solving • Creation of positive image
Social influence (support, modeling, norm)	• Self-efficacy–enhancing questions to enable people to mobilize social support, find modeling examples, and think about their perceived social norm • Creation of positive image
Self-efficacy	• Provision of information • Barrier identification • Problem solving (create a plan)
Intention	• Respect of autonomy (low) • Counterpersuasion (low) • Provision of information on making an action plan or setting goals
Preparatory plans	• Provision of information on the pros of making a preparatory plan • Provision of information on how to make a plan (tips): • If-then format (goal setting) • SMART[b] principles (goal setting) • Graded activity • Summarizing the most important information
Coping plans	• Provision of information on how to make a coping plan (tips): • If-then format (goal setting) • Mobilization of social supports • Summarizing the most important information

[a]Low, medium, and high refer to the specific health message in which the behavior change technique is applied for people with a low, medium, or high score on the different determinants.

[b]SMART: specific, measurable, achievable, relevant, time-bound.

Objectives

The aim of this study was 2-fold: (1) to evaluate pretesting communication elements and user appreciation, and (2) to identify participants' perceptions of the applied BCTs of the previously designed spoken animated video messages in a tailored self-management program for people with type 2 diabetes.

Methods

Study Design

According to the UK Medical Research Council framework, pilot testing is a prerequisite for the success of an intervention [40]. Hence, to optimize the intervention, we conducted a descriptive qualitative study with a cross-sectional design [41] focused on participants' perceptions of and experiences with the previously developed spoken animated tailored video messages. To study the target group's appreciation of the messages and to evaluate the key elements of effective communication, we conducted a semistructured interview.

Recruitment and Study Sample

To develop effective tailored health messages for a diverse patient population, we aimed to recruit a sample that was heterogeneous in age, sex, socioeconomic status, and educational background. Data collection took place in the Dutch province of Utrecht. We selected 3 general practices located in different socioeconomic status neighborhoods in a city in the center of the Netherlands for participation. One general practitioner and 2 practice diabetes nurses agreed to recruit patients with type 2 diabetes mellitus who met the inclusion criteria. To be included, patients had to be between 40 and 70 years of age, as type 2 diabetes is relatively rare before the age of 40 years; have received a diagnosis of type 2 diabetes more than 12 months previously; and were using at least one diabetes tablet or insulin, or both. Participants also had to be able to understand and read the Dutch language and have access to a computer with internet connectivity.

Recruitment took place either during patient consultations with the practice nurse or general practitioner or by telephone. The recruiting health professionals were instructed to give a short explanation of the study and hand out the invitation letter and informed consent form to potential study participants.

Procedure

We collected data from January to April 2017. Semistructured face-to-face interviews were conducted by 1 researcher (CvhS). To minimize bias due to the interview setting, we collected data in the participant's natural home setting.

First, we gathered input to provide tailored advice, in line with the tailoring logic of the final intervention. One day before the interview, we sent participants a quantitative questionnaire assessing their current level of PA, as well as premotivational and motivational determinants of PA behavior derived from the I-Change model [18,37]. Questionnaires were sent by email 1 day prior to the appointment with the researcher, since we wanted to increase the chance that participants would remember the questions and their answers. We asked participants to fill in the questionnaire and send it back to the researcher that same day. Depending on the sum score (high, [medium], or low) of each concept, the corresponding spoken animated video messages were selected.

Second, a session was scheduled to first assess the demographic characteristics age, sex, marital status, nationality, educational level, and home internet use (hours per week). Subsequently, participants watched the spoken animated video messages one by one so that the researcher could ask and verify whether participants could summarize the essence of each message and thus comprehended the content. Keywords and sentences of these messages were written down by the researcher in advance and compared with participants' answers.

Third, in the same session a semistructured interview was conducted to be able to evaluate key elements of effective communication and end users' appreciation of the spoken animated video messages. Except for the comprehension element, we asked participants' opinions on the pretesting elements regarding the content and design of all video messages. At the start of each interview, participants were reminded of the topic, purpose, content, added value, and procedure of the research and why it was important for them to participate. The interviews were recorded and transcribed by 1 of the authors (CvhS) to be able to analyze the content. Interviews were conducted until no new answers were being given and no new codes could be created, indicating data saturation was achieved.

Measurement

Prior to conducting the interviews, we measured participants' current level of PA in minutes of weekly light and moderate to vigorous PA using the Short Questionnaire to Assess Health-Enhancing Physical Activity [42]. We calculated MET scores of participants' PA behavior based on age, PA expressed in MET values, and time spent on this PA to evaluate whether the participant met the norm on PA at the time of this study of at least half an hour of moderate- to vigorous- (≥3 or 4 METs depending on age) intensity PA on a minimum of 5 days of the week. We assessed the determinants cognizance, knowledge, risk perception (susceptibility, severity), attitude pros and cons, social influence (support, norm, modeling), self-efficacy, intention, preparatory plans, and coping plans using 3 or 4 statements per construct (eg, "My friends stimulate me to become more physically active"). We used a 5-point Likert scale (score range 1-5) to indicate the strength of agreement or disagreement, where a score of 1 corresponded to "strongly disagree" and a score of 5 corresponded to "strongly agree." We used 10 correct or incorrect answering options to measure participants' knowledge of type 2 diabetes mellitus. We recorded outcomes for each concept. For example, concerning attitude pro, the answer options "agree" and "strongly agree" were recoded as 1, and "strongly disagree," "disagree," and "neutral" scored 0. We calculated a sum score so as to assign each concept to a low, (medium), or high group. We labeled a sum score of 3 to 6 attitude pros as attitude pro high, 1 to 2 attitude pros to attitude pro medium, and 0 attitude pros to attitude pro low.

We used Salazar's pretesting elements [28] as a frame in developing the interview guide (Table 2). Salazar argued that, while pretesting communication material, one should gather information concerning various elements; that is, a health message should catch attention, be comprehensible, be credible, be attractive, and be relevant to and accepted by the target group. Also, users should be able to identify with the content of the health communication material in order to feel personally addressed and involved. The effect of the messages on behavior (motivation) should be explored and points for improvement

should be solicited to be able to enhance the content. For each pretesting element, we formulated open-ended questions on the basis of Salazar's model [28] (eg, "What do you think of the believability of the spoken animated video messages?"). In addition, for each element we asked participants to indicate to what extent he or she perceived the animated video messages to be, for example, believable, on a score from 1 and 10, where 1 indicated "not believable at all" and 10 indicated "very believable."

Table 2. Salazar's pretesting elements [28].

Pretesting element	Recommendation	Example questions
Attractiveness	Allow participants to compare alternative versions of materials.	What do you think about the animated video messages? What was the first thing that caught your attention?
Comprehension	Try to focus participants on the main idea of the message.	Can you indicate what you think is the most important message of this animated video message(s)? What words or sentences are difficult to read or understand?
Believability	Question whether the material is credible and realistic to the audience.	What do you think about the believability of the animated video messages?
Involvement	Question whether the audience can identify with the material.	To what extent were the animated video messages tailored to your personal situation?
Acceptance	Explore issues that could potentially be overlooked.	Is there anything about the animated video messages that you find offensive or annoying?
Relevance	Have participants confirm whether the material is appropriate for them.	What type of people should read or watch this? In what way are the people in the animated video messages like or different from you?
Motivation and persuasion	Explore the effect on behavior and desires.	What does these animated video messages make you want to do? How likely are you to do that?
Improvement	Find out other ways to enhance the material.	What new information did you learn? What do you think is missing?

Data Analysis

We computed descriptive statistics (median, mean, and range) for program rating, rating for the different key communication elements, and participants' demographic characteristics.

We used thematic analysis to interpret the data derived from the semistructured interviews. Deductive and inductive approaches were combined and used to produce the themes and analyze the data [43]. Deductive approaches are theory driven. Analysis is shaped by preexisting theory or concepts, whereas in an inductive approach the themes identified are strongly linked to the data. This means that data are coded without trying to fit the data into a preexisting theory or framework [44].

Analysis of the data started after the first interview was conducted. After subsequent interviews, the transcripts were read and reread to fill in gaps or find new themes. We analyzed the content using the constant comparative method. Data derived from the semistructured interviews were managed and analyzed using MAXQDA Standard (VERBI GmbH), a software package for qualitative data analysis. We followed the stages described in the guidelines for thematic analysis developed by Clarke and Braun [44]. First, the audiotaped material was transcribed, followed by organizing, indexing, and anonymizing the data. Next, gathered data were structured using Salazar's key elements of effective communication [28] as initial codes supplemented by other interesting features of the data. Subsequently, 1 researcher (CvhS) collated codes into potential themes, reviewed the codes, and generated a thematic map of the analysis (tree structure). Next, a second researcher (KLC) coded the data independently, whereafter both researchers discussed the coding scheme and analysis to aid data interpretation and formulate the findings. We calculated the initial interrater reliability, indicating the level of agreement or consensus between the 2 researchers, for each interview transcription, which ranged from 87.10% to 96.77%. A level of agreement of 80% is recommended as the minimum interrater reliability, whereby an interrater reliability between 82% and 100% is interpreted as almost perfect [45]. All discrepancies were solved by discussion until consensus was reached. The interview outcomes created an overview of key aspects of the spoken animated health messages that were appreciated by the target group as well as points for improvement.

Results

Participant Characteristics

We received contact information of 23 interested patients from the practice nurses. We telephoned these patients to explain the study, answer questions, and, for those who agreed to participate, make an appointment to conduct the interview. A total of 17 persons met the inclusion criteria, agreed to participate in the study, and were asked to complete a written informed consent form on the day of the appointment. We excluded 1 person who did not take antidiabetic medication, and 5 others decided not to participate. Reasons mentioned were not having enough time and personal circumstances.

We interviewed a total of 15 patients with type 2 diabetes mellitus, with interviews lasting between 35 and 50 minutes

each. Table 3 provides an overview of the participants' baseline characteristics. Most participants were senior adults; the median age was 63 years, with about half of them being female. About half of the participants were married or were living with a partner. The majority were Dutch; only 1 German woman participated. Their education levels were equally distributed. The median time spent using the internet at home was 7 hours per week.

Table 3. Participants' baseline characteristics (N=15).

Characteristic	Value
Age (years), median (range)	63 (49-70)
Sex, n (%)	
Male	7 (47)
Female	8 (53)
Marital status, n (%)	
Single	7 (47)
Married or living with partner	8 (53)
Nationality, n (%)	
Dutch	14 (93)
German	1 (7)
Education level, n (%)	
Primary school or basic vocational school	4 (27)
Secondary vocational school or high school degree	6 (40)
Higher professional degree or university degree	5 (33)
Home internet use (h/wk), median (range)	7 (1-24)

Interview Outcomes

The semistructured interviews we conducted to evaluate pretesting communication elements and user appreciation (aim 1) showed a need for various improvements concerning didactics, message content, animation figures, and language use.

Comprehensibility

All participants mentioned that the use of animations to enhance the spoken words was highly supportive for understanding the messages, and most (8/15) indicated that the explanation of the content was simple and concrete without any difficult medical terminology, making it comprehensible to a wide audience (eg, "The explanation often is very good I think, concrete, no complex medical stories." [male, age 61 years, patient ID D11]). Some (3/15) commented on the given examples as being recognizable and thereby helpful understanding the message. Except for the video message concerning cognizance, the majority of participants were able to identify the main message of each video construct, since the spoken words and sentences matched the keywords and sentences written down by the researcher in advance. Participants' rating of the comprehensibility of the videos ranged from 7 to 10 on a 10-point Likert scale (median 8).

Attractiveness

Participants were asked what first attracted their attention when watching the spoken animated video messages. Some (4/15) praised the positive tone that was maintained even for people who clearly had a less healthy level of PA, while a few (3/15) mentioned the hidden jokes that attracted their attention, or the mirth of the animated video messages. Others indicated that the immediate mention of the core of the spoken animated video message was important to attract attention; for example, some participants (3/15) found it very confrontational to instantly hear all the possible consequences of diabetes, realizing they were at risk. Some participants (2/15) considered the tips on making an action or coping plan the most attractive, while 1 explicitly mentioned that the repetition of his answer (questionnaire) at the start of each spoken animated video message was important to capture his attention ("The personalized feedback at the start of the video ensures that you are drawn into the video within 10 seconds." [male, age 61 years, D11]).

Animation Figures

Opinions on the animation figures varied greatly, with some participants (5/15) appreciating their diversity, as they reflected society, enabling identification, while others (2/15) noticed an absence of cultural diversity. Several participants (6/15) liked the mimics and movements of the figures because they clarify the meaning of the message. Others (2/15) had no affinity with the animation figures, as they found them old-fashioned, not modern enough. Some (4/15) thought that the animation figures fit the posture of people with type 2 diabetes mellitus, whereas others (2/15) considered them stigmatizing.

Text (Written and Spoken) and Voice-Overs

The font size, use of keywords, and time the text remained in the picture were mentioned as factors enhancing readability. The amount of information provided was considered "not too

much and not too little." One participant commented that, in general, each sentence contained 1 message, which for him made the content of the messages understandable and easy to follow. All participants found both male and female voices friendly and pleasant to listen to, the pace of speaking easy to follow, and words well articulated.

Background and Music

In general, participants (11/15) were unaware of the backgrounds and music used in the spoken animated video messages and indicated that they were not distracting but neutral, tranquil, or matching the animation, allowing participants to focus on the content of the message. Opinions differed on the colors used. Some preferred neutral, serene colors, while others thought that bright colors emphasized the importance of the message.

Duration and Amount of Video Messages

Participants (13/15) regarded the spoken animated video messages as short but powerful, to the point, and of "Facebook length." The opinions concerning the amount of video messages varied. The majority (10/15) considered them all necessary to create an action plan, clear, educational, and crossing the t's. However, some participants (4/15) commented that there were too many, perceiving many similarities between the different messages. One participant stated that the amount of messages in 1 animated video is more important than the total amount of animated video messages and concluded that there should be 1 clear message per video.

Believability

The spoken animated video message contents' similarity to the information provided by the health professional and correspondence to the participants' own knowledge contributed to the believability (7/14). Several participants (8/14) indicated that the "clarity," "concreteness," and "logical explanation" of the animated video messages made it credible to them. One person noted that the factual information provided without exaggeration or deterrent effects made it believable. Participants' rating of the believability of the animated health messages ranged from 7 to 10 (median 9) and was graded higher than the comprehensibility.

Acceptance

Some participants (2/15) felt that the spoken animated video message on social influence was compelling and an interference to their own life. Others (2/15) thought that the question "If you would get these complaints how disagreeable would you find that?," which queries the perceived severity as part of risk perception, was strange and rhetorical (eg, "Of course I'll find it disagreeable!" [female, age 59 years, D14]). A few (2/15) indicated that the spoken animated video messages contained many fat animation figures, which they considered stigmatizing. One participant felt annoyed by the repetition of the same message "over and over again."

Involvement (Personalization)

Most participants (10/15) indicated that they recognized themselves in the feedback on their questionnaire responses, the examples provided, or the animation figures, whereby they felt personally addressed (eg, "The confirmation, since the videos you showed were really like 'you filled in the questions this way'...that was confirmed in the movie...so you feel...I'm watching something that really addresses me." [female, age 66 years, D02]). Participants' rating for the level of personalization of the spoken animated video messages ranged from 6 to 10 (median 8).

Relevance (Added Value)

Most participants (9/15) indicated that, due to the absence of new information or their current level of PA, the spoken animated video messages did not meet their needs. However, other interviewees (6/15) confirmed that the spoken animated video messages fit their needs, since watching them raised their awareness of the importance of PA in relation to type 2 diabetes and made them understand they were at risk. Some (2/15) stated that the video messages provided a good insight into type 2 diabetes and, therefore, are especially relevant for people who have just had type 2 diabetes diagnosed. Some participants (3/15) confirmed that the spoken animated video messages fit their needs, pointing out the power of repeating the message. Participants' rating of the relevance (added value) of the animated video messages to their health ranged from 6 to 10 (median 8).

Motivation and Persuasion

Some participants indicated that the spoken animated video messages did not motivate them to increase their level of PA, since they were either already sufficiently physically active, meeting Dutch PA recommendations (5/15), or were impeded by other health factors such as low back pain, "weak knees," or fibromyalgia, even though they were convinced of its importance (3/15). Those who did feel motivated by the video messages to change their PA behavior mentioned the severity of the possible consequences of type 2 diabetes (2/15) and the importance of PA for their health (3/15) as key reasons. Participants' rating of the extent to which the spoken animated video messages motivated them to become more physically active ranged from 1 to 10 (median 5).

Points for Improvement

We divided suggestions for improvement into 5 main topics: animation figures used, didactics, message content in general, specific video messages, and language use. Animation figures were seen as old-fashioned (2/15), too much unilaterally emphasizing the relationship between overweight and diabetes (2/15), or containing too little cultural diversity (2/15). Suggested didactic improvements were providing an overview of upcoming animated video topics at the start of the video program (1/15) and ending with a summary of the content of each animated video message (2/15). Others suggested adding examples for people who are single or do not have children (2/15) and using testimonials (1/15). Two participants suggested that by standardizing the amount of PA, people could see 30 minutes per day as a maximum instead of a minimum. One suggested adding more concrete personalized outcome measures, whereas another suggested linking the creation of an action plan to a direct concrete result. One participants suggested avoiding the use of real names in the examples and replacing them with "friend" or "partner" (1/15). One participant suggested

explaining and emphasizing the link with nutrition more in the animated video messages on PA, and explaining PA more in the video messages on nutrition, despite the fact that there is a separate module on nutrition. Suggestions for improving the content of specific animated video messages concerned the number of messages given simultaneously in 1 video and the amount of information on 1 screen. Two participants perceived the question "If you would get these complaints how disagreeable would you find that?" as strange and rhetorical. Two others occasionally experienced the language use as pushy ("It is a bit like 'you must do this, you will do that,' a little compulsive.").

Intention to Use the Spoken Animated Video Messages as a Tool to Increase Physical Activity Behavior

Several participants indicated they did not intend to watch the spoken animated video messages again, since they either were already sufficiently physically active (5/15), had sufficient knowledge concerning diabetes (4/15), or preferred the assistance of a medical professional (1/15). Of those participants who expressed the intention to use the video messages as a tool, some (6/15) thought they would use the video program as a manual or guideline (eg, "...the guidelines that it provides to handle it proactively, and an individual concretization." [male, age 66 years, D05]). Participants' rating of their intention to use the spoken animated video messages as a tool to change their PA behavior ranged from 1 to 10 (median 6).

Intention to Recommend the Spoken Animated Video Messages to Others

Close to half (6/15) would recommend the video messages to important others and thought that the messages would especially be helpful for people with a new diagnosis of type 2 diabetes, since it would provide them with important information at an early stage. Due to time pressure during a consultation, there is not always time to explain things well, they commented. Some interviewees (4/15) mentioned ignorance or unawareness of the possible consequences of type 2 diabetes, among others, as reasons to recommend the spoken animated video messages, while others (5/15) intended to recommend the video messages to motivate people to change their lifestyle, and 1 participant mentioned the concreteness and individual approach as reasons for recommending them to others. Participants' rating of their intention to recommend the spoken animated video messages to important others ranged from 5 to 10 (median 8).

Evaluation of the Behavior Change Techniques

To evaluate participants' appreciation of the different applied BCTs (aim 2), we conducted a post hoc analysis of the qualitative data.

Participants (10/15) thought that the provided personalized feedback on their answers or performance immediately made the message personal and drew their attention. All participants appreciated the positive reinforcement, mentioning the friendliness of speech, and some (5/15) mentioned the positive tone throughout. Most participants recognized the information provided but commented that they did not hear any new information. All participants were convinced of the importance of information on the consequences of behavior and type 2 diabetes mellitus, and some (4/15) considered the information provided to be quite confrontational. The majority (13/15) acknowledged the relevance of mobilizing social support, although some participants (2/15) disliked the idea of others interfering with their lives and behavior and were not interested in what other people think or do. The gain-framing arguments used to persuade people to adopt a positive attitude toward PA behavior were confirmed by all participants, as well as the arguments used to counterpersuade participants who had a negative attitude toward PA. Not all participants agreed on the affective arguments used. One did not recognize being proud after being physically active, and others (4/15) did not feel fitter or more relaxed or disliked PA. The majority of participants (11/15) thought that the provided tips on how to make an action plan and coping plan were very clear, concrete, and helpful to become more physically active and overcome barriers. Especially the provided tips and examples on concrete goal setting, the if-then concept, the SMART (specific, measurable, achievable, relevant, time-bound) principle, and the tip on graded activity were seen as necessary for successful behavior change and maintenance. Only a few participants (2/15) said that the spoken animated video messages motivated them to become more physically active and to change their PA behavior systematically. Fewer than half of the participants stated that they did not like making an action plan (2/15) or did not need a plan, since they already embedded PA in their life and were sufficiently active (5/15). All participants recognized the provided examples of possible barriers, but varied in their perceived need to overcome these barriers by making a coping plan.

Discussion

Principal Findings

The aim of this study was 2-fold: (1) to evaluate pretesting communication elements and user appreciation, and (2) to evaluate the applied BCTs of the previously designed spoken animated video messages in a tailored self-management program for people with type 2 diabetes.

To evaluate pretesting communication elements and user appreciation (aim 1), we formulated interview questions. We used the pretesting elements as defined by Salazar [28] as a framework, covering attractiveness, comprehensibility, acceptance, believability, involvement, relevance, motivation, and improvement. To evaluate the applied BCTs (aim 2), we performed a post hoc analysis. We found a need for various improvements concerning didactics, message content, animation figures, and language used. The results of this study provide insights into what key communication elements are appreciated and what BCTs should be taken into account when designing spoken animated tailored video messages.

In general, the evaluation of the pretesting communication elements and user appreciation (aim 1) showed that participants appreciated the spoken animated video messages for being to the point, short but powerful, and positive in tone.

The comprehensibility, believability, involvement (personalization), and perceived relevance of the video program

to participants' health received high mean scores. Participants noted that simplicity, concreteness, and the absence of difficult medical terminology made the spoken animated video messages comprehensible. The use of animations to enhance the spoken words was perceived as highly supportive for understanding the messages, confirming literature indicating that spoken animation is a promising way to communicate complex health information [23,24]. The fact that the content of the spoken animated video messages corresponded to the participant's knowledge of type 2 diabetes and was similar to the information provided by health professionals made the health messages credible to most participants.

The repetition of a participant's answer, the given examples, and the possibility to identify with the animation figures made participants feel personally addressed. According to the elaboration likelihood model [46], providing tailored feedback results in more thoughtful information processing via the central route of persuasion, since tailored messages are perceived as being personally relevant and thus encourage the person to pursue the desired behavior [30]. Moreover, the use of role models, enabling identification, is a common effective method to increase attention, remembrance, self-efficacy, and skills [47,48].

Although participants were convinced that the use of the spoken animated video messages as a tool to change their PA behavior could be relevant to their health, many indicated that the animated video messages did not motivate them to become more physically active and thus did not match their needs. Reasons mentioned were the presence of other impeding health factors limiting PA participation, being sufficiently active already, or the absence of new information. In line with this, a previous study showed that the presence of comorbidities was negatively associated with PA participation in patients who had had type 2 diabetes for more than 1 year, but not in those with a new diagnosis [49]. Also, the fact some participants found the content relevant, while others did not, is in line with the I-Change model [37], which states that individuals move through stages of change.

The spoken animated video program was thought to be especially relevant for people with a new diagnosis of type 2 diabetes, since it provides them with important information at an early stage. These arguments suggest the importance of formulating inclusion criteria for participation and considering when the program is offered after diagnosis in order to connect more with a patient's needs and abilities and thereby increase the perceived relevance, usage, and PA behavior.

Following the evaluation of the applied BCTs (aim 2), a few participants expressed negative opinions on the use of some BCTs in the messages. For example, 2 participants did not appreciate the video message on social influence, considering themselves responsible for their own behavior (change), and didn't want anyone to interfere, indicating the importance of a well-thought-out formulation of a video message. Also, where the majority appreciated the BCT concerning action planning, and regarded the provided tips as helpful to change their behavior in a systematic way, others did not like or did not feel a need to make a plan. These examples may suggest the importance of the selection of BCTs based on their motivational level [50] and the users' preferences [51].

Limitations

We recruited a heterogeneous study population with well-distributed baseline characteristics of sex, marital status, and educational level. Hence, the insights obtained allowed us to develop spoken animated tailored video messages for a diverse patient population. Yet there was potentially a response bias, reflected by the small cultural diversity, as 13 participants were Dutch and 1 was German. Since Moroccans and Turkish people form a large part of our target population, the results of this pretest might not be applicable to this group.

Some of the results might be biased by the likelihood that only more interested and motivated patients participated in the study, which may affect the generalizability. Other differences in characteristics between participants and patients who did not participate, such as disease severity and other impeding health factors, could also have contributed to biased results. For instance, 3 participants indicated that, due to impeding health factors, they were not able to become more physically active, although they were motivated and wanted to. Further, a large number indicated that they had not yet experienced any serious consequences of diabetes, and were sufficiently active and therefore not motivated to increase their level of PA.

Although the use of thematic analysis to analyze the data may have limited the objectivity of this study, the use of Salazar's pretesting elements enabled us to evaluate key elements of effective communication. Although we based most of the conclusions regarding our research question on qualitative data, the addition of quantitative measures (grades) may have yielded complementary information and valuable insight into the appreciation of the developed animated video messages and their compliance with the conditions of effective communication. Lessons learned concerning the strengths and limitations of our approach may help to improve the spoken animated video messages and thereby the electronic health self-management program and its integration into diabetes care.

Conclusion

We conducted this study (1) to evaluate key elements of effective communication and user appreciation, and (2) to evaluate the applied BCTs of the designed spoken animated health messages in a tailored self-management program, in order to enhance the delivery of health messages, which in turn may enhance the user experience and usage of the tailored self-management program for patients with type 2 diabetes mellitus. Our evaluation using Salazar's pretesting elements identified several key elements that enhanced appreciation of the animated tailored health messages, as well as points for improvement. The high appreciation of the pretesting elements' comprehensibility, believability, and personalization was supported by participants' statements on BCTs and other elements that help the creation of animated tailored health messages. Based on the results of this study, we recommend that health messages be simple, concrete, to the point, short but powerful, and positive in tone, and lack difficult medical terminology. The use of animations to enhance the spoken words

was perceived as highly supportive for the understanding of health messages. Providing positive reinforcement, feedback on participants' answers, and examples that match user characteristics and the possibility to identify with the animation figures contributes to participants' involvement in a health message. To connect more with patient's needs and physical abilities—and thereby increase the perceived relevance and motivation to use an animated video program, and thus increase PA participation—when developing spoken animated health messages we suggest taking users' characteristics and preferences into account, basing the selection of BCTs on the stages of change, and considering when the program is offered after diagnosis.

Acknowledgments

There were no sponsors in this research and there was no financial disclosure.

Authors' Contributions

CvhS wrote the video script and designed the video animations and study in collaboration with SV, CH, and HdV. CvhS conducted the interviews, analyzed the data, and wrote the first draft of the manuscript. CvhS and KLC interpreted the findings. All authors reviewed and conceptualized various versions of the manuscript. CvhS wrote the final manuscript, which all authors have read and approved.

Conflicts of Interest

None declared.

References

1. Moreau M, Gagnon MP, Boudreau F. Development of a fully automated, web-based, tailored intervention promoting regular physical activity among insufficiently active adults with type 2 diabetes: integrating the I-change model, self-determination theory, and motivational interviewing components. JMIR Res Protoc 2015;4(1):e25 [FREE Full text] [doi: 10.2196/resprot.4099] [Medline: 25691346]
2. Gerich JE, Smith TS. Beta-cell defectsand pancreatic abnormalities in type 2 diabetes. In: Pickup JC, Williams G, editors. Textbook of Diabetes, 3rd edition. Malden, MA: Blackwell Science; 2003:23-21.
3. Fischer PH, Hilderink H, Harbers MM, Bree LVP. Rijksinstituut voor Volksgezondheid en Milieu. Bilthoven, Netherlands: Rijksinstituut voor Volksgezondheid en Milieu; 2014. URL: https://www.volksgezondheidenzorg.info/onderwerp/diabetes-mellitus/cijfers-context/trends#node-toekomstige-trend-diabetes-door-demografische-ontwikkelingen [accessed 2020-03-11]
4. Ali MK, McKeever BK, Imperatore G, Barker L, Gregg EW, Centers for Disease Control and Prevention (CDC). Characteristics associated with poor glycemic control among adults with self-reported diagnosed diabetes--National Health and Nutrition Examination Survey, United States, 2007-2010. MMWR Suppl 2012 Jun 15;61(2):32-37. [Medline: 22695461]
5. van Vugt M, de Wit M, Cleijne WHJJ, Snoek FJ. Use of behavioral change techniques in web-based self-management programs for type 2 diabetes patients: systematic review. J Med Internet Res 2013;15(12):e279 [FREE Full text] [doi: 10.2196/jmir.2800] [Medline: 24334230]
6. Pickup J, Williams G. Management of diabetes: diet and lifestyle modification. In: Textbook of Diabetes. Oxford United Kingdom: Blackwell Pub; 2003:31-36.
7. Cheung BMY, Ma RCW. Drug therapy: special considerations in diabetes. In: Holt RIG, Cockram C, Flyvbjerg A, Goldstein BJ, editors. Textbook of Diabetes. New York, NY: John Wiley & Sons, Ltd; 2017:385-397.
8. Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes: a meta-analysis of the effect on glycemic control. Diabetes Care 2002 Jul;25(7):1159-1171. [Medline: 12087014]
9. Avery L, Flynn D, van Wersch A, Sniehotta FF, Trenell MI. Changing physical activity behavior in type 2 diabetes: a systematic review and meta-analysis of behavioral interventions. Diabetes Care 2012 Dec;35(12):2681-2689 [FREE Full text] [doi: 10.2337/dc11-2452] [Medline: 23173137]
10. Kodama S, Tanaka S, Heianza Y, Fujihara K, Horikawa C, Shimano H, et al. Association between physical activity and risk of all-cause mortality and cardiovascular disease in patients with diabetes: a meta-analysis. Diabetes Care 2013 Feb;36(2):471-479 [FREE Full text] [doi: 10.2337/dc12-0783] [Medline: 23349151]
11. Onafhankelijk wetenschappelijk adviesorgaan voor regering en parlement. Beweegrichtlijnen 2017. Den Haag, Netherlands: De Gezondheidsraad; 2017. URL: https://www.gezondheidsraad.nl/documenten/adviezen/2017/08/22/beweegrichtlijnen-2017 [accessed 2020-03-02]
12. Rijksinstituut voor Volksgezondheid en Milieu. Bewegen: cijfers & context: huidige situatie. Bilthoven, Netherlands: RIVM; 2017. URL: https://www.volksgezondheidenzorg.info/onderwerp/sport-en-bewegen/cijfers-context/huidige-situatie [accessed 2017-07-04]

13. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
14. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
15. Dijkstra A, De Vries H. The development of computer-generated tailored interventions. Patient Educ Couns 1999 Feb;36(2):193-203. [Medline: 10223023]
16. Glasgow RE, Bull SS, Piette JD, Steiner JF. Interactive behavior change technology. A partial solution to the competing demands of primary care. Am J Prev Med 2004 Aug;27(2 Suppl):80-87. [doi: 10.1016/j.amepre.2004.04.026] [Medline: 15275676]
17. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [doi: 10.1207/s15324796abm3103_2] [Medline: 16700634]
18. Walthouwer MJL, Oenema A, Lechner L, de Vries H. Comparing a video and text version of a web-based computer-tailored intervention for obesity prevention: a randomized controlled trial. J Med Internet Res 2015;17(10):e236 [FREE Full text] [doi: 10.2196/jmir.4083] [Medline: 26481772]
19. Vandelanotte C, Mummery WK. Qualitative and quantitative research into the development and feasibility of a video-tailored physical activity intervention. Int J Behav Nutr Phys Act 2011;8(1):70 [FREE Full text] [doi: 10.1186/1479-5868-8-70] [Medline: 21718546]
20. Smit ES, Linn AJ, Van Weert JCM. Taking online computer-tailoring forward: the potential of tailoring the message frame and delivery mode of online health behaviour change interventions. Eur Health Psychol 2015;17(1):25-31.
21. Stanczyk N, Bolman C, van Adrichem M, Candel M, Muris J, de Vries H. Comparison of text and video computer-tailored interventions for smoking cessation: randomized controlled trial. J Med Internet Res 2014;16(3):e69 [FREE Full text] [doi: 10.2196/jmir.3016] [Medline: 24589938]
22. Cheung KL, Schwabe I, Walthouwer MJL, Oenema A, Lechner L, de Vries H. Effectiveness of a video-versus text-based computer-tailored intervention for obesity prevention after one year: a randomized controlled trial. Int J Environ Res Public Health 2017 Oct 23;14(10) [FREE Full text] [doi: 10.3390/ijerph14101275] [Medline: 29065545]
23. Leiner M, Handal G, Williams D. Patient communication: a multidisciplinary approach using animated cartoons. Health Educ Res 2004 Oct;19(5):591-595 [FREE Full text] [doi: 10.1093/her/cyg079] [Medline: 15150139]
24. Meppelink CS, van Weert JCM, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
25. World Health Organization. Health promotion: health literacy. Geneva, Switzerland: WHO; 2019 Feb 08. URL: https://www.who.int/healthpromotion/health-literacy/en/ [accessed 2019-12-03]
26. Höffler TN, Leutner D. Instructional animation versus static pictures: A meta-analysis. Learning Instruction 2007 Dec;17(6):722-738. [doi: 10.1016/j.learninstruc.2007.09.013]
27. Lang A. The limited capacity model of mediated message processing. J Commun 2000 Mar 01;50(1):46-70. [doi: 10.1111/j.1460-2466.2000.tb02833.x]
28. Salazar BP. Practical applications of pretesting health education concepts and materials. Health Educ Monogr Ser 2004;21(1):6-12.
29. Vluggen S, Hoving C, Schaper NC, de Vries H. A web-based program to improve treatment adherence in patients with type 2 diabetes: development and study protocol. Contemp Clin Trials 2018 Nov;74:38-45. [doi: 10.1016/j.cct.2018.10.002] [Medline: 30290275]
30. Boudreau F, Walthouwer MJL, de Vries VH, Dagenais GR, Turbide G, Bourlaud A, et al. Rationale, design and baseline characteristics of a randomized controlled trial of a web-based computer-tailored physical activity intervention for adults from Quebec City. BMC Public Health 2015 Oct 09;15:1038 [FREE Full text] [doi: 10.1186/s12889-015-2364-3] [Medline: 26453041]
31. Schulz DN. Using the Internet to Promote a Healthy Lifestyle: Testing a Multiple Behaviour Change Intervention Regarding Reach, Use, Appreciation, Effectiveness and Cost-Effectiveness [doctoral thesis]. Maastricht, Netherlands: Maastricht University; 2014.
32. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;3:CD008776. [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
33. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

34. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. The long-term efficacy of two computer-tailored physical activity interventions for older adults: main effects and mediators. Health Psychol 2011 Jul;30(4):442-452. [doi: 10.1037/a0023579] [Medline: 21639638]
35. de Vries H. An integrated approach for understanding health behavior; the I-Change model as an example. Psychol Behav Sci Int J 2017 Mar;2(2):555585. [doi: 10.19080/PBSIJ.2017.02.555585]
36. Schulz DN, Kremers SPJ, Vandelanotte C, van Adrichem MJG, Schneider F, Candel MJ, et al. Effects of a web-based tailored multiple-lifestyle intervention for adults: a two-year randomized controlled trial comparing sequential and simultaneous delivery modes. J Med Internet Res 2014 Jan 27;16(1):e26 [FREE Full text] [doi: 10.2196/jmir.3094] [Medline: 24472854]
37. de Vries H, Kremers SPJ, Smeets T, Brug J, Eijmael K. The effectiveness of tailored feedback and action plans in an intervention addressing multiple health behaviors. Am J Health Promot 2008;22(6):417-425. [doi: 10.4278/ajhp.22.6.417] [Medline: 18677882]
38. Vyond. Make your own animated videos. San Mateo, CA: GoAnimate, Inc; 2019. URL: https://goanimate.com/videomaker [accessed 2019-04-04]
39. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184. [doi: 10.1111/j.1525-1497.2005.40245.x]
40. Skivington K, Matthews L, Craig P, Simpson S, Moore L. Developing and evaluating complex interventions: updating Medical Research Council guidance to take account of new methodological and theoretical approaches. Lancet 2018 Nov 01;392(suppl 2). [doi: 10.1016/s0140-6736(18)32865-4]
41. Boeije H. Analysis in Qualitative Research. Thousand Oaks, CA: Sage Publications; 2009.
42. Wendel-Vos G, Schuit A. Short Questionnaire to Assess Health Enhancing Physical Activity. Bilthoven, Netherlands: Center for Chronic Diseases Epidemiology, National Institute of Public Health and the Environment; 2002. URL: http://smile-research.unimaas.nl/SMILE-CMS/1260-SMILE/version/1/part/6/data/handleiding%20%20squash%20juli%2020041.doc?branch=main&language=default [accessed 2020-03-11]
43. Baarda B, Bakker E, Julsing M, Fischer T. Basisboek Kwalitatief Onderzoek: Handleiding voor het Opzetten en Uitvoeren van Kwalitatief Onderzoek. Groningen, Netherlands: Noordhoff Uitgevers bv; 2013.
44. Clarke V, Braun V. Thematic analysis. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht, Netherlands: Springer Netherlands; 2014:6626-6628.
45. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
46. Cacioppo JT, Petty RE, Kao CF, Rodriguez R. Central and peripheral routes to persuasion: an individual difference perspective. J Pers Soc Psychol 1986;51(5):1032-1043. [doi: 10.1037/0022-3514.51.5.1032]
47. Ashford S, Edmunds J, French DP. What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. Br J Health Psychol 2010 May;15(Pt 2):265-288. [doi: 10.1348/135910709X461752] [Medline: 19586583]
48. Kay Bartholomew Eldredge L, Markham CM, Ruiter RAC, Kok G, Fernandez ME, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. New York, NY: John Wiley & Sons; 2016.
49. Plotnikoff RC, Lippke S, Karunamuni N, Eves N, Courneya KS, Sigal R, et al. Co-morbidity, functionality and time since diagnosis as predictors of physical activity in individuals with type 1 or type 2 diabetes. Diabetes Res Clin Pract 2007 Oct;78(1):115-122. [doi: 10.1016/j.diabres.2007.02.016] [Medline: 17379349]
50. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589]
51. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]

Abbreviations

BCT: behavior change technique
I-Change: integrated change
MET: metabolic equivalent task
PA: physical activity
SMART: specific, measurable, achievable, relevant, time-bound
SMS: short message service

Edited by G Eysenbach; submitted 08.07.19; peer-reviewed by J Papadakos, V Gregory; comments to author 22.10.19; revised version received 03.02.20; accepted 03.02.20; published 23.04.20.

Please cite as:
van het Schip C, Cheung KL, Vluggen S, Hoving C, Schaper NC, de Vries H
Spoken Animated Self-Management Video Messages Aimed at Improving Physical Activity in People With Type 2 Diabetes: Development and Interview Study
J Med Internet Res 2020;22(4):e15397
URL: http://www.jmir.org/2020/4/e15397/
doi:10.2196/15397
PMID:32324138

Original Paper

Rethinking Social Interaction: Empirical Model Development

Jone Bjornestad[1,2], PsyD, PhD; Christian Moltu[2], PsyD, PhD; Marius Veseth[3], PsyD, PhD; Tore Tjora[1], PsyD, PhD

[1]Department of Social Studies, Faculty of Social Sciences, University of Stavanger, Stavanger, Norway

[2]Department of Psychiatry, District General Hospital of Førde, Førde, Norway

[3]Department of Clinical Psychology, University of Bergen, Bergen, Norway

Corresponding Author:
Jone Bjornestad, PsyD, PhD
Department of Social Studies
Faculty of Social Sciences
University of Stavanger
PO Box 8600 FORUS
Stavanger, 4036
Norway
Phone: 47 97141599
Email: jone.r.bjornestad@uis.no

Abstract

Background: Social media is an integral part of human social life. More than 90% of young people use social media daily. Current theories, models, and measures are primarily based on face-to-face conceptions, leaving research out of sync with current social trends. This may lead to imprecise diagnoses and predictions.

Objective: To develop a theoretically based empirical model of current social interfaces to inform relevant measures.

Methods: A three-stage, qualitative, data-collection approach included anonymous individual Post-it notes, three full-class discussions, and 10 focus groups to explore 82 adolescents' relational practices. Data analysis followed a meaning-condensation procedure and a field-correspondence technique.

Results: We developed an empirical model that categorizes adolescents' social interactions into five experiential positions. Four positions result from trajectories relating to social media and face-to-face social interaction. Positions are described by match or mismatch dynamics between preferred and actual social platforms used. In matched positions, individuals prefer and use both face-to-face and social media platforms (position 1), prefer and use face-to-face platforms (position 2), or prefer and use social media platforms (position 3). In mismatched positions, individuals prefer face-to-face interactions but use social media platforms (position 4) or prefer social media but use face-to-face platforms (position 5). We propose that matched positions indicate good social functioning while mismatched positions indicate serious social challenges.

Conclusions: We propose a model that will expand previous unidimensional social interaction constructs, and we hypothesize that the described match and mismatch analyses provide conceptual clarity for research and practical application. We discuss prediction value, implications, and model validation procedures.

(J Med Internet Res 2020;22(4):e18558) doi:10.2196/18558

KEYWORDS

social interaction; social functioning; social media; model; empirical; adolescence; health science

Introduction

Social media platforms are technology-mediated tools that enable individuals to create, share, and exchange ideas, images, and information through online communities and networks [1,2]. Social media has become an integral part of current social life and provides new opportunities for accommodating the core human need of being emotionally affiliated with a community [3]. Globally, there are approximately three billion registered social media profiles, and the number of social media platforms is growing exponentially [4]. Young persons are the heaviest consumers, with more than 90% using social media on a daily basis [5]. Despite a minimum access age of 13 years for several platforms, reports suggest that 75% of 10-12-year-olds have a social media account [6,7].

Despite burgeoning social media use, current social interaction theories [8,9], models, and measures are largely based on face-to-face conceptions, resulting in an outdated understanding

of how social media phenomena materialize in social life [10-14]. This may lead to lack of analytic precision across a range of settings. For example, within the field of mental health, individuals with a rich online but limited face-to-face social life could currently be assessed as having poor social functioning. This may have monumental consequences, as a false low social-functioning score could lead to a false positive psychiatric diagnosis with subsequent incorrect or excessive treatment. Furthermore, imprecise understanding of social media functioning could deprive the helper of the opportunity to map and facilitate central online coping areas [14].

Core social capacities, such as emotion regulation, attachment, language, mentalization, and agency, develop from a starting point of physical, time-synchronized, face-to-face interactions with caregivers [8,15,16]. Mature relationships gradually manifest throughout adolescence [17,18]. Use of social media as part of human social life accelerates during the same time period. This seems significant, as adolescence, due to rapid bodily, cognitive, and emotional changes, is considered a period of both great vulnerability and great potential [19]. Youths display general limitations in reflexivity, emotion regulation, and ability to consider consequences before acting, making them more easily affected by influences, both through social media and in face-to-face situations. Negative social comparisons, social exclusion, social media addictions [18], bullying, cyberbullying [20,21], and cybervictimization [22] increase chances for poor development and psychopathology. Moreover, adolescents are formed prosaically by positive grown-up role models or peers validating and teaching flexible strategies. These factors may act protectively when youths fluctuate between the group norms and identities of their peers, both on social media and in face-to-face settings [12,17,23]. Consequently, social media may be understood as adding a further layer of complexity that adolescents must master during an already complex life period.

Technological innovation affects society and human behavior in fundamental ways and changes the interfaces between individuals [24]. Specific technologies, such as social media, do not merely add to the possibilities for communication, but also change the nature of communication. At face value, when compared with face-to-face social interaction, social media platforms represent radical changes, creating possibilities for asynchronistic and multicast communication and an unlimited number of possible contacts [2,25]. Further, by removing boundaries of time, space, and language, and by adding artificial intelligence, social media makes human relationships digital.

Online social technology may raise challenges [2,12]; for example, does social media require extra social flexibility or is it adaptable to facilitate communication for persons who may experience social limitations face-to-face, such as persons with social anxiety or severe mental illness [14,26]? Among other things, distance in both time and space means that fewer bodily senses are used and gives one the ability to pause information flow [27]. Compared to face-to-face communication, these features may reduce social withdrawal but, at the same time, may also be obstacles to precise communication by decreasing information accuracy. Social media also makes misinformation and use of several or false identities easier and more common [28].

Valid research into human social behavior can only be achieved with clear operationalizations integrating contemporary social processes. To achieve this, incorporating and investigating the added complexity that social media brings to sociality is imperative. Theories, models, and measures have largely ignored this integration [14], and the need for modification is precarious.

In this study, we implemented a large-scale, qualitative, in-depth investigation of 82 young individuals' experiences with current social life, aiming to develop an empirically informed theoretical model of face-to-face and social media interaction.

Our research question is as follows: How do adolescents experience and practice social interaction after the added complexity brought by social media?

Methods

Overview

A three-stage, qualitative, data-collection approach involved anonymous individual Post-it notes, three full-class discussions, and 10 focus groups to explore 82 adolescents' relational practices on social media. We used a reflexive thematic approach [29,30] for implementation and analyses. This study was approved by the Regional Ethical Committee (2018/2273/REK nord). Participants gave their informed written consent.

Sample and Recruitment

Most people participate in the public school system in Norway, including people with, for example, psychological problems and disabilities [31]. Special care is free in Norway, and the result is an overall representative student population and a high number of graduating students [31]. The study sample (N=82) was recruited between February and April 2019 and was based on strategic sampling from three different high schools in Rogaland county, Norway. To approximate population representativeness, we invited schools with different socioeconomic profiles, admission requirements, and geographical localizations. The sample consisted of six school classes: three classes of students in a general higher education preparation program, two classes of students in a health and social work training program, and one class of students in an electrical craft training program.

Sample size was reviewed after four and eight focus group interviews. We stopped recruiting after 10 focus groups because we considered the last two not to have contributed additional information [32]. Participants were 71% (58/82) female and participant ages ranged from 17 to 19 years. They were admitted in either programs for university admissions or vocational training. Several students from three classes had psychology as an elective subject.

Procedure and Data Collection

First, we used a design-thinking approach aiming to reveal individual perspectives [33-35]. We applied a "silent" Post-it note technique with the following instructions: (1) "Please make

as many statements as possible, positive and negative, about what social interaction is for you, including interaction on social media" and (2) "What questions are relevant to ask about current social interaction, including social media interaction?" Then, together with the participants, we categorized Post-it notes thematically, in vivo, based on similarity in content. Using this approach had two purposes: (1) to form a starting point for full-class and focus group discussions and (2) to ensure field correspondence, give voice to outliers, and compensate for limitations in the focus group approach. These full-class introduction sessions lasted approximately 45 minutes.

Second, we organized full-class discussions. We encouraged participants to elaborate on their Post-it notes and comment on their claims on social interaction. The full-class discussions lasted approximately 60 minutes each.

Third, we divided classes into focus groups for in-depth interviews: total number of participants was 82, total number of focus groups was 10, and number of participants in each focus group ranged from 5 to 12. The participants' primary teachers created the focus groups, as we considered them to be best suited for the task given their familiarity with the individual students and relationships among them. Focus groups were used to further elaborate on the themes raised in the Post-it notes and the full-class sessions. The researchers used this opportunity to raise in-depth questions about themes emerging from steps 1 and 2. As expected, the focus group format was more manageable for several of the study participants, resulting in increased engagement and new topics.

At the end of each focus group, participants were invited to provide any relevant information that had not yet been thematized. All steps of the investigation were implemented in classrooms at the participants' respective high schools. Post-it notes were photographed, and full-class discussions and focus group interviews were audio recorded and transcribed verbatim for the purposes of analysis.

Analysis

For inductive analysis, we employed a six-step reflexive thematic approach [29,30] concretized in Textbox 1. To strengthen the credibility of the study, the four researchers conducted the analytic procedure independently. During collaborative meetings, the researchers compared their interpretations, agreed on themes with accompanying quotes and model content, and validated the findings by consensus decision [36], dedicating special attention to steps 4 to 6 presented in Textbox 1. To overcome possible disagreement in the collaborative analytic meetings, we agreed on the following decision rules in the preparatory phases of the study: (1) to resolve minor disagreements by the principle of parsimoniousness and (2) to resolve major disagreements by (a) an inductive principle using the raw data as a compass, aiming to select the descriptions most closely reflecting the experience of the phenomena at issue, and (b) further applying the principle of best argument as described above.

Textbox 1. Steps of thematic analysis.

Step 1. Becoming familiar with the data through thorough reading of the Post-it notes, transcribed full-class discussions, and focus group interviews, thereby forming a main impression of the experiences of the participants and identifying potentially important themes. A *theme* was defined as a verbalization capturing an important element of the data in relation to the research question, representing a patterned response in the dataset.

Step 2. Generating initial codes, which were defined as the most basic segments of the raw data that could be assessed in a meaningful way regarding the specific phenomenon. For example, the participants' descriptions of flexibly using both face-to-face and social media platforms were given the tentative code *flexible use*.

Step 3. Searching for and developing candidate themes and subthemes. Remaining codes were set aside at this phase in a separate category for the purpose of being further analyzed and incorporated when appropriate. For example, the theme *mentalization* was initially referred to as *empathy*. However, during analysis we perceived this theme to also include the participants' reflections of their personal contribution to the social interactions they engaged in. Hence, we altered the theme name to *mentalization*.

Step 4. Reviewing themes to develop a coherent thematic map and considering the validity of individual themes in relation to the dataset.

Step 5. Defining and naming themes: further refining and defining themes, identifying the essence of themes, identifying subthemes, and summarizing the contents of the main themes into what each researcher considered to best represent participants' experiences. When refinements no longer added substantially to the themes, the analytic process was closed.

Step 6. Determining the relevance of a particular theme by counting the frequency of the relevant meaning units and combining this with our interpretation of how central the theme was perceived to be to the social life of the participants. It was at this point that we agreed that themes could be transferred to model positions.

Results

Our analysis resulted in a theoretically informed empirical model for the integration of social media and face-to-face interaction (see Figures 1 and 2).

Figure 1. Youth social interaction positions.

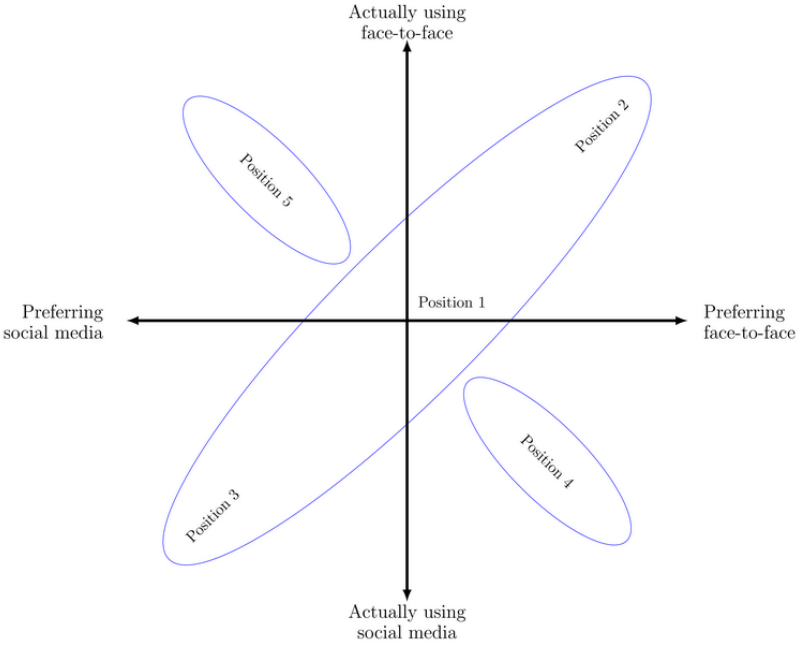
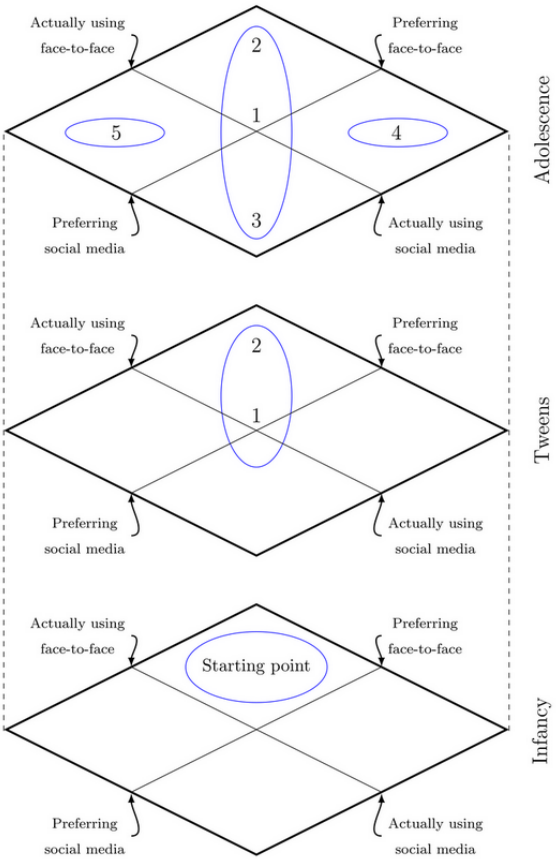


Figure 2. Social interaction positions: complete model.



Model: Suggested Social Positions for Adolescent Social Interaction

Infancy denotes the model's starting point in childhood's physical, time-synchronized, face-to-face experiences with caregivers. The second layer, *Tweens*, concerns the introduction of social media experiences, reflecting the approximate period when social media becomes a significant part of human social life [6]. The third layer, *Adolescence*, reflects the empirical analyses in this study, which suggest that adolescents' social interactions can be understood through five social positions resulting from multiple developmental trajectories. The x-axis presents the preferred type of social interaction, while the y-axis presents the actual type of social interaction. In the participants'

experiences, face-to-face interaction and social media interaction are overlapping social formats. Hence, both x- and y-axes reflect a continuum of social interactions, not binary oppositions. For transparency, illustrations of the empirical basis for the model are presented in Table 1 and Table 2. The five social positions (see Figure 2) are detailed in the following sections.

The Concept of Matching

In the model presentation, *match* implies correspondence in supply and demand between personal resources—social skills, social behaviors, personal values, interests, etc—and social context demands underlying social networks and relationship maintenance. All individuals in matched positions reported that they have communities, digital or analog, in which they have a sense of emotional affiliation and where they can act as independent agents. Hence, *match* refers to the individual's ability to gain access to available social benefits; master the current social norms, game rules, and contexts; and display social functioning fulfilling their social needs. Our empirical data suggest that some individuals satisfy their social needs, and hence achieve match, solely in face-to-face settings, whereas some achieve match solely in social media settings, and some move flexibly and interchangeably between face-to-face and social media settings. These three matched positions constitute the *cigar shape* in the model (see Figure 2).

Conversely, *mismatch* between preferred and actual social platform implies lack of correspondence between personal social resources and contextual demands. Individuals experiencing mismatch experience challenges in accessing available social benefits and accommodating to social norms and game rules, and display a social functioning profile not satisfying their social needs. Our empirical data and the current literature indicate two mismatching positions: individuals who prefer face-to-face interactions but are unwillingly using social media, and individuals who prefer social media interaction but are unwillingly using face-to-face interactions.

Position 1: Flexible Match—Preferred and Actual Interactions Are Both Face-to-Face and Via Social Media

Individuals speaking from position 1 were characterized as flexible and well-functioning individuals who are using social media and face-to-face formats in continuous adaption to contextual changes. Participants typically describe these individuals as living rich face-to-face social lives but, at the same time, actively enriching their social life with social media, for example, by initiating new relationships, seeking information and entertainment, and preserving already-established relationships. They were also perceived as good at critically assessing information quality and using their established face-to-face and online social networks to protect themselves against hazards, for example, individuals trying to exploit them or fake profiles (ie, catfishing). Position 1 appears to be the most flexible of the model positions.

Position 2: Match—Preferred and Actual Interaction Is Face-to-Face

Individuals with a current affiliation with position 2 consider face-to-face relationships to be the only authentic relationship format and therefore seek this form of contact. Participants described individuals remaining within this position as largely respecting and upholding social norms regulating face-to-face interactions, such as honesty and respecting individual differences and personal boundaries. Norm violations result in sanctioning and ultimately social exclusion. Data suggested a value-laden conflict between positions 2 and 3 on what constitutes authenticity. This conflict was mostly addressed by individuals in position 2, who suggest that purely online social life is inferior, as expressed in statements such as "a pure online life is not a full social life" or "the goal of online contact is always physical meetings."

Position 3: Match—Preferred and Actual Interaction Is Via Social Media

Individuals with a sense of belonging in position 3 reported preferring social media relationships, including gaming, over face-to-face relationships. Although they used face-to-face interactions early in life, the significance of face-to-face relationships gradually decreased with age and primarily serves to meet practical and societal demands, such as attending school and family meetings; when choosing based on their own preferences, they live their lives mainly on social media. Within their social media communities, they described it as less important whether they use a nickname or real name. Nicknames were, for some, described as just as important and real as their given names. Although real names and personal information are often gradually revealed, the goal of relationships is not necessarily to evolve into physical meetings, as is the case for individuals in position 2. Position 3 social norms seem similar to norms associated with position 2, but are typically sanctioned through platform moderators.

Answering the authenticity criticism promoted by individuals in position 2, individuals speaking from position 3 were consistent in valuing social media interactions as authentic and as, for them, more appreciated than face-to-face relationships, and they described long-lasting social media relationships. These types of relationships were mainly described in the context of flexible social media platforms that allow for complex online interactions.

Position 4: Mismatch—Preferred Interaction Is Face-to-Face; Actual Interaction Is Via Social Media

Data suggested that position 4 was taken by individuals preferring face-to-face social interaction but using social media social interaction. No participants confirmed affiliation with this position themselves. Rather, findings are based on participants' descriptions of other individuals and their first-hand experiences with them through both face-to-face and social media interactions. These third-person descriptions involved increased distance from the phenomenon compared to the first-person descriptions of position 1, 2, and 3; that is, this position is described from an outside perspective rather than as lived experiences. Consequently, the descriptions may be affected by distance in perspective, and the risk of fundamental attribution errors increases accordingly. Nevertheless, individuals in position 4 are still judged based on their behaviors in the data material, and the consequences are negative characterizations and high risk for social exclusion.

Participants described individuals in position 4 as having limited social networks, poor social skills, and poor compliance with face-to-face and social media norms and social game play. Examples describe odd behaviors, such as contacting strangers face-to-face or online in ways more suitable for close friends, and cross-border behaviors, like communicating through fake social media profiles (ie, catfishing). Poor social skills were also described as making these individuals vulnerable for exploitation, including exploitation by individuals seeking economic, sexual, or other personal gains.

Participants described a collective effort, using each other's face-to-face and online social networks, to identify and protect against hazards from individuals in position 4. Their main strategy was social exclusion, for example, through profile blocking. Despite describing these individuals as displaying similar behaviors in face-to-face and social media settings, face-to-face exclusion was described as more brute and absolute than social media exclusion. The nearly unlimited opportunities for new encounters were described as the main reason why position 4 individuals remained social media users, despite preferring face-to-face interaction. Nonetheless, the descriptions suggest that widespread social exclusion is a problem for this group and that they have few opportunities for realizing protracted social relationships.

Position 5: Mismatch—Preferred Interaction Is Via Social Media; Actual Interaction Is Face-to-Face

Position 5 is descriptive of individuals who possibly prefer social media but use or are forced to engage in face-to-face social interaction. Participants, only to a limited degree, gave third-person descriptions with position 5 characteristics, suggesting that this is an outlier position. Hence, the attribution error risk is also relevant for this position. Nevertheless, based on empirical evidence for this position from other studies [37-39], we include it in the proposed model. We suggest that this position might apply to two groups of individuals: (1) those who lack social media awareness and possibilities due to, for example, limited internet access, and (2) those who, regardless of awareness, otherwise lack social media interest, have negative attitudes toward social media, or simply resist social media participation [37,38]. Individuals with limited social media skills or in disadvantaged social economic situations also plausibly belong to this group [39]. These individuals would perhaps have a more satisfactory social life if social media were integrated in their daily routines. This may be the case for several marginalized groups, such as prison inmates or individuals in other facilities where people are compulsively placed together.

Table 1. Experiences of social interaction across current social interfaces, by theme and position.

Theme and positions	Exemplary quotes
Establishing and maintaining social relationships	
3	I use Discord a lot [complex social media platform] and for me it's not just something that can replace Skype, but also a place they can find new communities or support groups where you can meet new friends. And it's a place where you can, for example, if you meet a friend in a game and you want to play with them more, you can contact them on Discord and then just talk to them there or join like a server or like communities with them. And in that way keep in touch and when you use Discord as much as I do and many others do, you meet a lot of new people and you can have a lot of good friendships, even if they are just on Discord. For example, I have people I know better than my brothers who I only talk to on Discord.
1	Because I can't think of many friends I've met on social media. It's more to keep in touch. People I don't see that often live in other places and stuff.
1 and 3	Yeah, for me at least, if I meet new people, then first it's Facebook, then become friends on Facebook, and then maybe follow up on Instagram, and then when you've either met again or chatted on Facebook, or something like that, then you can like... You go onto SnapChat. But it's just like... Well, that's how I do it, in that order, kind of. And SnapChat is more, for me, when I know someone. While Facebook can, is kind of more, like general: "Ok, we're friends on Facebook, at least, so then we can contact each other on Facebook," if we... From there, and then develop it further to other social media, sort of.
4	All these friend requests from people in different countries and you have no idea who they are. And they have closed information. Then you don't accept them.
4	When it says you have no friends in common you don't often think "oh, I think maybe I know who they are!"
1	So the more friends you have in common, the better the chances are.
2	Yeah, cause the whole social network has just been moved over to social media. It's reality moved over onto the internet. The same things that happen in reality happen there, just a bit differently.
2	Because now I kind of have the base I need to meet people in reality, face-to-face. I don't, like, need to go online to meet people online and then meet them in reality, kind of.
1	Back to the thing about getting new friends. I feel like say you're working out at the gym and you see someone and you know who they are, but you're not brave enough to talk to them. Then social media is kind of a really good, like, ice breaker, because you can start by following them, or adding them somewhere and then that kind of opens up for a conversation next time you see each other.
3	Very true, if you both are just sniping from behind, for example, instead of just going up and being aggressive in a way, then you make a connection there.
Living on social media	
3	Because there are, like, some people who just live on social media. It's us who make the bloggers. We're the ones saying what we want. And then they have to find a way to give it to us, sort of.
3	I think maybe it has to do with... Maybe you're a bit conflict avoidant and you're worried about seeing, or, worried about how the other person will react and if you do it on social media, then you kind of don't have to see the reaction.
1	No, I mean that's hard, but I think that... For me at least, I'll admit that sometimes it can be easier to send a message than to say it face-to-face. That says something about me, too. But that the reaction, I mean, that the recipient can interpret it in another way, or in a bit of a different way, than it was meant is better than seeing the immediate reaction of the recipient, which might be different to what you hoped for, kind of.
1	But in a way... When you say something wrong in reality then it gets taken more personally than online. I can't explain why.
1	From [when] you're little you learn that you need to be careful about talking to people you don't know. So, if you've met someone first, you know them in a way, and you feel safer. And then you chat on social media. Because then you have an idea of who the person is and you're not tricked.
Generation gap	
5	Yeah, when I communicate with a person who uses them, I kind of think that I don't analyze it in the same way, because they're a different generation, so I like adapt. But, I still don't use smileys, I just do what I do, mostly, but I kind of adapt it, like I don't write slang when I chat with mum, but when I chat with others and they use it I would either think it's a typo or that it was, that there was something more behind it. So you wonder if it was on purpose.
5	We've left Facebook because of parents. And now granny's on there, as well.
1	It's good to keep things a bit separate. So, you've got different places for different relationships.
Social norms and cross-border behaviors	
2	I guess the social norms we follow are less strict online.
1	I would search for the person on Facebook. Check if we have friends in common. That's the first thing I look at. Or I look at who follows the person on Instagram. If there's no one...

Theme and positions	Exemplary quotes
4	If they're fake, they're always fake online.
4	Not to be rude, but I think it's probably easier if you find it hard to talk to people face-to-face... Then this is something that will help you. But for most people real relations are worth more. If that's ok to say.
4	Something else I've noticed is that it's people who have social problems who post their SnapChats on Facebook and say, "can you add me?" or "can I add you?" and ask for other people's SnapChats. That's understandable if people have social problems and want to find friends online. That's when you post that on Facebook.
4	There was a girl who suddenly started sending me messages saying "I've seen you at school"... That was a bit weird, I thought. And then she never stopped sending me messages. She asked me so many strange things. I don't know. It was a bit scary.
4	I feel like people ask like "hi, can we meet?" and you've never had anything to do with them before. Mostly you just try to ignore it, because you don't know who they are or what it will be like.
4	But maybe that's why it's kind of like... That's why Tinder is looked down on maybe. As that's a place you can be fooled much easier. Because some people go there to find comfort maybe.
Self-agency and perceived control	
1	You kind of need to be in direct contact with the person. You can pretend to be anybody on SnapChat—the way you talk about everything. You can seem a lot more confident, etc.
3	So, er, if you use the platforms Messenger and SnapChat, on Messenger you can choose when you want to reply, right. So if you get... Now he got a message. Of all things..., yeah. So, if you get a message on Messenger it's, like, maybe a reply you need to think about a bit, I do that myself a lot. So, it's like, you choose a bit like: "I don't want to answer this yet, I know I've got a message, but I don't need to answer yet" and then when you're like kind of ready to answer, sounds a bit weird, but then yeah, you can answer. But in real life, it's like, you kind of have to have a topic of conversation that you need to come up with then and there. But if you're like online, and like Messenger, you can like come up with stuff bit by bit, right. Because then you have a lot of time to think about different things, kind of, and if you use SnapChat, then there's that thing of having to answer right away. It's normal if you snap someone, then you don't like answer right after you get it, like. You wait maybe like the same amount of time as that person, the other person, waited to open and answer your snap, right? That's like roughly how it goes, normally...
1	But if you're not sure, you have all these other social media you can check with. If there's anyone on, for example, Tinder you think there's something strange about, you can check on Instagram and see if you find anything. Or you can go on Instagram and see if there's something there.
1	No, I mean that's hard, but I think that... For me at least, I'll admit that sometimes it can be easier to send a message than to say it face-to-face. That says something about me, too. But that the reaction, I mean, that the recipient can interpret it in another way, or in a bit of a different way, than it was meant is better than seeing the immediate reaction of the recipient, which might be different to what you hoped for, kind of.
1	When we started adding each other on SnapChat. I got her SnapChat and she got mine. And I didn't start... The first thing I sent wasn't, like, a picture of me. It was what I was doing and stuff. And then she did it, and I saw she was in high school, sat with her books and stuff.
3	It's easier to talk to people when you don't get the signals from their faces and body... It's kind of easier to talk to them when you don't see how they react, like. And easier to talk also because you get a chance to think about what you're talking about longer before you say it.
2	Yeah, true. I feel like "friend" in real life, that's more someone you can contact and be with. But friend on social media that doesn't need to be people you even want to hang out with, it's just like... Just to show that you know people!
Mentalization	
1	I think both are just as genuine, just the process of messaging goes faster, because it's easier to share since we kind of don't see the response... We always want approval and that type of thing, and see that what we say and do is right.

Table 2. Field correspondence: themes included on Post-it notes, but not included in full-class or focus group discussions.

Theme	Post-it note content
Inspiration and shopping or fashion	Identify the latest in fashion
	A place where I find inspiration
News and knowledge acquisition	Find out what's going on in the world on a daily basis
	Source of useful information
Social media as a sanctuary	A place where my parents can't interfere so much
	A place where you can hide behind anonymity
	It is my identity without being judged
	A place where you hide how you really feel
New life	A place where you can create a new and different life
	Get to know people all over the world
Display differences in socioeconomic status	Display those who cannot afford an iPhone
Fake news	Information can be fake
Misunderstandings	Misunderstandings can easily arise due to lack of body language
Monitoring others	Easy way to monitor others
	A place where I study others
Stress and psychiatric problems	Can lead to loneliness, anxiety, and depression
	Social media kept me awake at night
	Social media leads to stress
	Feeling of being an outsider. Shows what your friends are doing without you
Social media dependence	May be all-consuming and lead to poor self-image
	It is a prison that makes us slaves because it is addictive
	Social media creates FOMO (fear of missing out)
Stops face-to-face communication	I spend a lot of time on Instagram. I stop talking to my friends. I'm just chatting
Cyberbullying and harassment	Social media leads to bullying
	Consequence of too few likes
	Easier to use nasty language online
	Nude pictures can be spread
Escalate body image issues and negative social comparison	Social media creates body image issues
	I compare myself with picture-perfect people, although I know the pictures are retouched

Discussion

Principal Findings

Social media provides distinct platforms for accommodating the core human need for being emotionally affiliated with a community [3]. Venturing beyond earlier models, which are solely concerned with face-to-face positions [10-14], four of the five suggested model positions are based on trajectories resulting from introducing social media to face-to-face social interaction. Hence, findings that primarily rely on face-to-face interactions are deficient when it comes to reflecting current social interfaces.

Reflecting previous research [40,41], similarities in behaviors, values, and interests seem to be consistent triggers for relationship formation and a perceived match between the preferred and actual social platform. This is particularly the case with regard to positions 1, 2, and 3. Also in line with previous research [15], mentalization or *the ability to understand the mental state, of oneself or others, that underlies overt behavior* [42], particularly during early phases of contact, seems to be a decisive capacity when it comes to validly evaluating social contexts, while social skills seem to be catalysts for contact attainment, both on social media and in face-to-face interactions. Another similarity between social media and face-to-face interactions is that perceived agency, which *encompasses the belief in the power of one's own ability to affect outcome* [8], seems consistently correlated with social mastery, network development, and psychological well-being [8,43,44]. Thus, mastery of social media resembles mastery of face-to-face interactions and failings on social media appear similar to failings in face-to-face social settings. On this basis, we will argue that the perhaps most obvious point of departure for analysis, namely, to consider social media as the *new* (ie, the figure) and face-to-face interaction as the *point of reference* (ie, the background) is not the most fruitful approach. We propose that analyzing match and mismatch between preferred and actual social interaction platforms will result in a more valid analysis.

Match Between Preferred and Actual Social Interaction Platforms

Individuals in positions 1 and 2 seem to achieve a match after a short and cost-effective social trajectory, reflecting a model starting point of face-to-face social interactions. These positions appear to involve well-established rituals and structures [45,46]. *Match* seems determined by whether the individuals seek the available social gains, master the current social norms, and provide a social functioning type that allows their needs to be fulfilled.

The Vulnerable Position 3

Findings indicate a vulnerable transition for individuals in position 3. Although experiencing a fruitful match between preferred and actual social media platforms in late adolescence, these individuals seem to have had a social trajectory characterized by more fundamental contextual and social change. Transition value from earlier social life also seems lower compared to individuals in positions 1 and 2. Expected attendance on face-to-face arenas during adolescence [18,47], combined with accommodation of new social media game rules, seems to add additional stress for many. The aforementioned value-laden conflict between positions 2 and 3 on what authentic social interactions are may also cause stress. Stress associated with social interaction may in turn lead to increased social withdrawal from face-to-face interactions [48], yet without having established a robust social media network.

Social withdrawal is frequently used as a strategy to reduce acute social discomfort. However, as a long-term solution, withdrawal reinforces rather than solves social problems [48]. Anonymity, distance, and the emerging possibilities for complex emotional affiliation with an online community [2,25] may accelerate and cement social withdrawal from face-to-face interactions for individuals vulnerable to social distress. Hence, for individuals with match potential within positions 1 or 2, such a trajectory may result in a social *no-man's land*, position 4, and mismatch. Early position 3 affiliation may thus predict increased risk for social challenges and psychological problems [1,49]. Individuals suffering from early face-to-face social defeats or adverse events, such as bullying [20,50], seem particularly vulnerable to this route.

Conversely, individuals that are well-established in position 3 indicate that social media provides new and appealing social affiliation opportunities not available pre-social media [2,12,24,51].

The Matthew Effect

Establishing social skills and associated capacities, such as self-agency and mentalization, may in simple terms be understood through a paraphrase on the Matthew effect: those who master become better; those who fail, fail again. Individuals who, in early age, experience mastering social contexts will have a higher chance of establishing a feeling of self-agency, better social skills, and, in time, the ability to mentalize [8,15]. Findings indicate that this momentum created by mastering and building social skills may predict mastering either face-to-face, on social media, or both. The associated increase in mentalization corresponds to the finding that mentalization ability is particularly important for valid early relational evaluation [15] and increases the likelihood of picking a social platform corresponding to personal characteristics and needs. Echoing previous research, such development is associated with flexible psychological strategies, well-functioning relationships, and psychological well-being [3,15,52]. Findings indicate that these positive effects are prominent in all established matched positions.

Conversely, findings suggest that early failure in mastering face-to-face social skills lowers the chance of establishing self-agency, social skills, and mentalization. Individuals to whom this applies also seem to lack social triggers, including behaviors, values, or interests that they share with others [40,41], thus reducing the chances of relationship formation and social mobility. Findings indicate that this lack of social momentum is associated with poor mastering of face-to-face and, in time, social media interaction. This pattern also seems associated with decreased chances for achieving a match between preferred and actual type of social platform and, ultimately, social rejection from others [53]. Thus, instead of psychological well-being, these individuals may experience increased stress [54], anxiety, and *learned helplessness* [48,55,56], leading to passivity and withdrawal from social encounters. Findings indicate that this pattern is self-reinforcing, continuously increasing the distance between the person and the social mainstream. These continuous social defeats and unpleasant psychological conditions, combined with the core hunt for social affiliation, may partly explain why some individuals over time seek different online echo chambers rather than a matched position.

Implications

Model position status has the potential to predict human behavior and is, therefore, relevant for disciplines such as internet and gaming research, youth research, and mental health research, as well as for practitioners. Individuals in all established matched positions seem robust, except individuals in early position 3, who seem vulnerable.

Position 4 may predict psychopathology [2,57,58]. General limitations in social skills, self-agency, and mentalization, but plausibly also poor insight into how their own behaviors affect others, make these individuals particularly vulnerable for long-term social exclusion. Further, social media's unlimited number of possible contacts and few boundaries related to time and space [12,27,59] may reduce social correction and instead facilitate a continuation of a negative social trajectory. Position 5 may also predict psychopathology. Individuals in this position do not necessarily have poor insight, but have problems transferring from actual to preferred social platform. This may apply to individuals living in institutions and forced settings, for example, prison populations.

A detailed model of social interaction including new media platforms provide resources for measure development within health science. For instance, items targeting social functioning in widely used measures for psychiatric diagnoses do not include social media functioning [14], even if people withdrawing from social contact as part of mental illness development lead active social media lives [60]. This paradox points to a latency in

measure development in a social reality marked by rapid technology developments. Research-based operationalizations of social media functions may inform necessary development.

Limitations

The general limitations of qualitative research apply to this investigation: findings are context dependent and pertain to the participants and setting in which the study was conducted. There was an overrepresentation of females (71%), the population was selected from individuals attending school, and interviews were conducted face-to-face and only in Norway. Also, data were not used to analyze how specific demographic or other characteristics might influence the participants' social interactions, differences in their perceptions, and overall attitude. All these factors affect generalizability.

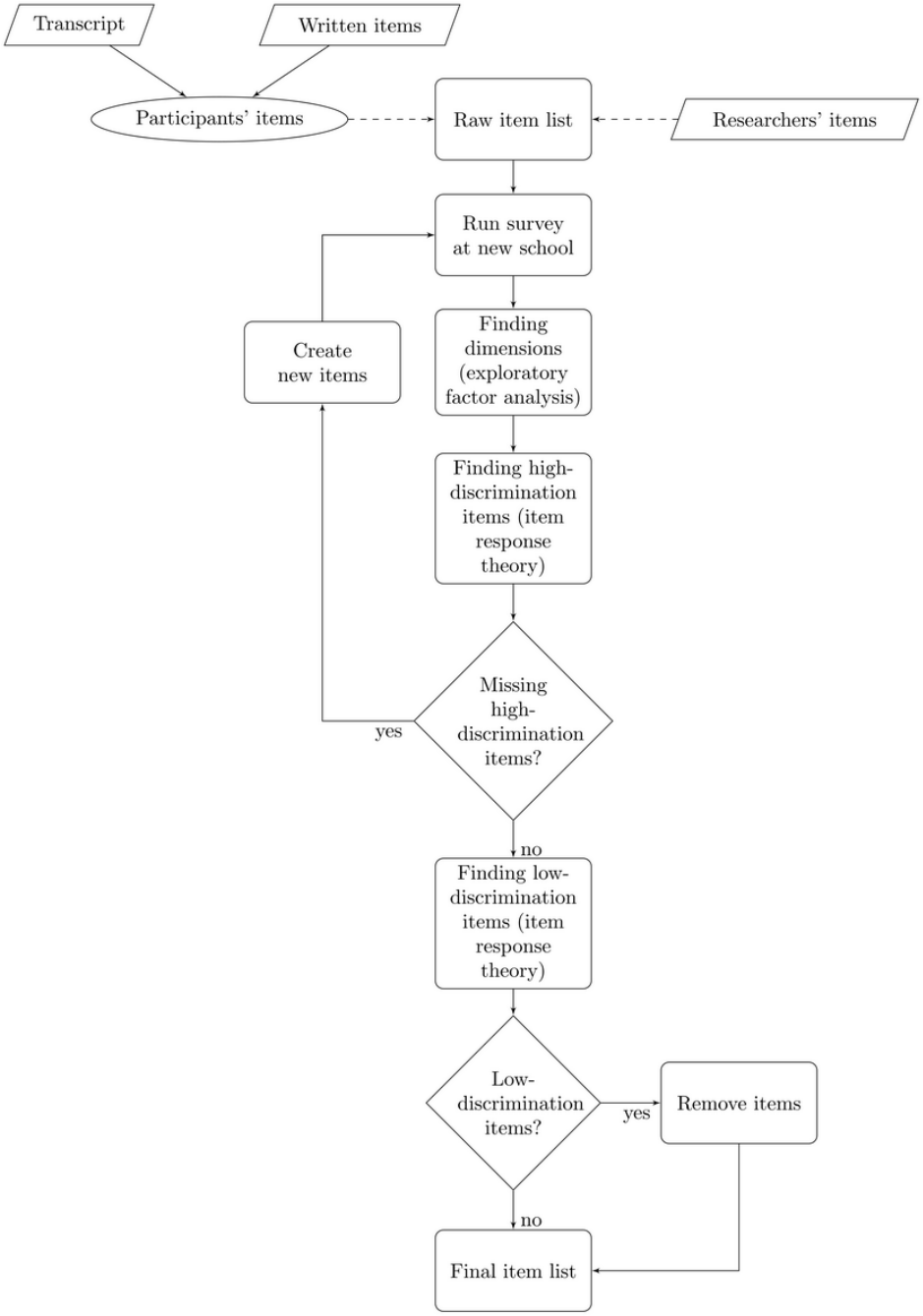
To accommodate some of the limitations associated with qualitative methodology, we performed a field-correspondence technique using Post-it notes. Although expected themes, such as pornography habits, were omitted, participants reported more personally sensitive themes in this format (see Table 2). These dealt, in particular, with the negative sides of using social media.

Future Research

The overarching goal for future research is to explore and develop the validity of the proposed model, and to investigate to what degree falling into one position is associated with positive (ie, mastery and protection) and negative health and well-being.

We will test this model with a large-scale quantitative survey design, as presented in Figure 3. The initial survey items will be drawn from suggestions written by participants in this study. These items will be supplemented with questions derived from the focus groups, class sessions, and items made by the researchers. This initial large pool of items will be used in the initial large-scale survey. Using exploratory factor analysis or similar, the questions will be divided into different dimensions. Using item response theory, items with low discrimination will be excluded. Item response theory will also be used to highlight ranges in the different dimensions where the survey is missing high-discrimination items. This framework can be used to generate testable hypotheses and items on dimensions of social interaction. To study context-specificity and generalizability, we propose that this research procedure could be implemented across different countries and contexts, and we welcome collaboration.

Figure 3. Model validation procedure.



Acknowledgments

A special thanks to the staff at the Medical Library of Stavanger University Hospital for assistance with the literature search.

Authors' Contributions

JB and TT had a main role in model development, data extraction, analysis, and in writing the first draft. CM and MV contributed in the conceptual development, data extraction, and data analysis. All authors were involved in study design, provided scientific oversight throughout the project, provided detailed comments regarding the paper across several drafts, and edited the paper.

Conflicts of Interest

None declared.

References

1. Riehm KE, Feder KA, Tormohlen KN, Crum RM, Young AS, Green KM, et al. Associations between time spent using social media and internalizing and externalizing problems among US youth. JAMA Psychiatry 2019 Sep 11;76(12):1-9. [doi: 10.1001/jamapsychiatry.2019.2325] [Medline: 31509167]
2. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
3. Thoits PA. Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav 2011 Jun;52(2):145-161. [doi: 10.1177/0022146510395592] [Medline: 21673143]
4. Statista. 2019. Number of social media users worldwide from 2010 to 2023 (in billions) URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2020-02-13]
5. Lenhart A, Duggan M, Perrin A, Stepler R, Rainie L, Parker K. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2015 Apr 09. Teens, social media & technology overview 2015: Smartphones facilitate shifts in communication landscape for teens URL: https://tinyurl.com/y9xstjeq [accessed 2019-06-23]
6. Children's Commissioner's Office, Revealing Reality. Life in 'Likes': Children's Commissioner Report Into Social Media Use Among 8-12 Year Olds. London, UK: Children's Commissioner; 2017. URL: https://www.basw.co.uk/system/files/resources/basw_94916-10.pdf [accessed 2020-02-21]
7. The Lancet. Children and social media. Lancet 2018 Jan 13;391(10116):95. [doi: 10.1016/S0140-6736(18)30049-7] [Medline: 29353623]
8. Bandura A. Self-efficacy mechanism in human agency. Am Psychologist 1982;37(2):122-147 [FREE Full text] [doi: 10.1037/0003-066X.37.2.122]
9. Ajzen I. From intentions to actions: A theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition to Behavior. Berlin, Germany: Springer-Verlag; 1985:11-40.
10. Koteyko N, Hunt D, Gunter B. Expectations in the field of the internet and health: An analysis of claims about social networking sites in clinical literature. Sociol Health Illn 2015 Mar;37(3):468-484 [FREE Full text] [doi: 10.1111/1467-9566.12203] [Medline: 25847533]
11. Hercheui MD. A literature review of virtual communities. Inf Commun Soc 2011 Feb;14(1):1-23 [FREE Full text] [doi: 10.1080/13691181003663593]
12. Lee EWJ, Ho SS, Lwin MO. Explicating problematic social network sites use: A review of concepts, theoretical frameworks, and future directions for communication theorizing. New Media Soc 2016 Oct 06;19(2):308-326 [FREE Full text] [doi: 10.1177/1461444816671891]
13. Litt E. Measuring users' internet skills: A review of past assessments and a look toward the future. New Media Soc 2013 May 24;15(4):612-630. [doi: 10.1177/1461444813475424]
14. Bjornestad J, Hegelstad WTV, Berg H, Davidson L, Joa I, Johannessen JO, et al. Social media and social functioning in psychosis: A systematic review. J Med Internet Res 2019 Jun 28;21(6):e13957 [FREE Full text] [doi: 10.2196/13957] [Medline: 31254338]
15. Fonagy P, Gergely G, Jurist EL. Affect Regulation, Mentalization, and the Development of the Self. London, UK: Routledge; 2018.
16. Vygotsky LS. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press; 1980.
17. Muuss RE. Theories of Adolescence. 5th edition. New York, NY: Random House; 1988.
18. Blakemore SJ. Inventing Ourselves: The Secret Life of the Teenage Brain. New York, NY: PublicAffairs, Hatchett Book Group; 2018.
19. Griffiths MD. Internet sex addiction: A review of empirical research. Addict Res Theory 2011 Jun 15;20(2):111-124. [doi: 10.3109/16066359.2011.588351]
20. Ehman AC, Gross AM. Sexual cyberbullying: Review, critique, & future directions. Aggress Violent Behav 2019 Jan;44:80-87 [FREE Full text] [doi: 10.1016/j.avb.2018.11.001]
21. Jenaro C, Flores N, Frías CP. Systematic review of empirical studies on cyberbullying in adults: What we know and what we should investigate. Aggress Violent Behav 2018 Jan;38:113-122 [FREE Full text] [doi: 10.1016/j.avb.2017.12.003]
22. Fisher BW, Gardella JH, Teurbe-Tolon AR. Peer cybervictimization among adolescents and the associated internalizing and externalizing problems: A meta-analysis. J Youth Adolesc 2016 Sep;45(9):1727-1743. [doi: 10.1007/s10964-016-0541-z] [Medline: 27447707]
23. Allen JP, Land D. Attachment in adolescence. In: Cassidy J, Shaver PR, editors. Handbook of Attachment: Theory, Research, and Clinical Applications. New York,NY: The Guilford Press; 1999:319-335.
24. Jasanoff S. The Ethics of Invention: Technology and the Human Future. New York, NY: WW Norton & Company; 2016.
25. Kuss DJ, Griffiths MD. Online social networking and addiction: A review of the psychological literature. Int J Environ Res Public Health 2011 Sep;8(9):3528-3552 [FREE Full text] [doi: 10.3390/ijerph8093528] [Medline: 22016701]
26. Spinzy Y, Nitzan U, Becker G, Bloch Y, Fennig S. Does the internet offer social opportunities for individuals with schizophrenia? A cross-sectional pilot study. Psychiatry Res 2012 Jul 30;198(2):319-320. [doi: 10.1016/j.psychres.2012.02.022] [Medline: 22440545]

27. Troiano G, Nante N. Emoji: What does the scientific literature say about them? A new way to communicate in the 21st century. J Hum Behav Soc Environ 2018 Feb 16;28(4):528-533 [FREE Full text] [doi: 10.1080/10911359.2018.1437103]
28. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
29. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597 [FREE Full text] [doi: 10.1080/2159676x.2019.1628806]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Fasting RB. Adapted education: The Norwegian pathway to inclusive and efficient education. Int J Inclusive Educ 2013 Mar;17(3):263-276 [FREE Full text] [doi: 10.1080/13603116.2012.676083]
32. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2019 Dec 26:1-16. [doi: 10.1080/2159676x.2019.1704846]
33. Yeager DS, Romero C, Paunesku D, Hulleman CS, Schneider B, Hinojosa C, et al. Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. J Educ Psychol 2016 Apr;108(3):374-391 [FREE Full text] [doi: 10.1037/edu0000098] [Medline: 27524832]
34. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
35. Brown T, Wyatt J. Design thinking for social innovation. Dev Outreach 2010 Jul;12(1):29-43. [doi: 10.1596/1020-797x_12_1_29]
36. Hill CE, Knox S, Thompson BJ, Williams EN, Hess SA, Ladany N. Consensual qualitative research: An update. J Couns Psychol 2005 Apr;52(2):196-205. [doi: 10.1037/0022-0167.52.2.196]
37. Reisdorf BC, Groselj D. Internet (non-)use types and motivational access: Implications for digital inequalities research. New Media Soc 2015 Dec 27;19(8):1157-1176. [doi: 10.1177/1461444815621539]
38. Brody N. Opting out of social media: Online communication attitudes mediate the relationship between personality factors and Facebook non-use. South Commun J 2018 Jan 04;83(2):75-88 [FREE Full text] [doi: 10.1080/1041794x.2017.1413415]
39. Hargittai E, Piper AM, Morris MR. From internet access to internet skills: Digital inequality among older adults. Univers Access Inf Soc 2018 May 3;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
40. Bornstein RF. Exposure and affect: Overview and meta-analysis of research, 1968–1987. Psychol Bull 1989 Sep;106(2):265-289. [doi: 10.1037/0033-2909.106.2.265]
41. Zajonc RB. Attitudinal effects of mere exposure. J Pers Soc Psychol 1968;9(2, Pt.2):1-27. [doi: 10.1037/h0025848]
42. Fonagy P, Bateman AW. Mechanisms of change in mentalization-based treatment of BPD. J Clin Psychol 2006 Apr;62(4):411-430. [doi: 10.1002/jclp.20241] [Medline: 16470710]
43. Jeannerod M. The sense of agency and its disturbances in schizophrenia: A reappraisal. Exp Brain Res 2009 Jan;192(3):527-532. [doi: 10.1007/s00221-008-1533-3] [Medline: 18709365]
44. McAdams DP. Psychopathology and the self: Human actors, agents, and authors. J Pers 2020 Feb;88(1):146-155. [doi: 10.1111/jopy.12496] [Medline: 31206660]
45. Bourdieu P. Cultural reproduction and social reproduction. In: Grusky DB, Szelenyi S, editors. Inequality: Classic Readings in Race, Class, and Gender. New York, NY: Routledge; 2018:257-272.
46. Goffman E. The Presentation of Self in Everyday Life. London, UK: Penguin Books; 1990.
47. Willoughby M. A review of the risks associated with children and young people's social media use and the implications for social work practice. J Soc Work Pract 2018 May 08;33(2):127-140 [FREE Full text] [doi: 10.1080/02650533.2018.1460587]
48. Gazelle H, Rubin KH. Social withdrawal and anxiety in childhood and adolescence: Interaction between individual tendencies and interpersonal learning mechanisms in development: Introduction to the special issue. J Abnorm Child Psychol 2019 Jul;47(7):1101-1106. [doi: 10.1007/s10802-019-00557-y] [Medline: 31089980]
49. McCrae N, Gettings S, Purssell E. Social media and depressive symptoms in childhood and adolescence: A systematic review. Adolesc Res Rev 2017 Mar 2;2(4):315-330. [doi: 10.1007/s40894-017-0053-4]
50. Franklin P, Hossain R, Coren E. Social media and young people's involvement in social work education. Soc Work Educ 2016 Mar 11;35(3):344-356 [FREE Full text] [doi: 10.1080/02615479.2016.1154710]
51. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. Soc Issues Policy Rev 2017 Jan 13;11(1):274-302. [doi: 10.1111/sipr.12033]
52. Zeegers MAJ, Colonnesi C, Stams GJM, Meins E. Mind matters: A meta-analysis on parental mentalization and sensitivity as predictors of infant-parent attachment. Psychol Bull 2017 Dec;143(12):1245-1272. [doi: 10.1037/bul0000114] [Medline: 28805399]
53. Becker HS. Outsiders: Studies in the Sociology of Deviance. New York, NY: The Free Press; 1963.
54. Beyens I, Frison E, Eggermont S. "I don't want to miss a thing": Adolescents' fear of missing out and its relationship to adolescents' social needs, Facebook use, and Facebook related stress. Comput Human Behav 2016 Nov;64:1-8. [doi: 10.1016/j.chb.2016.05.083]

55. Seligman ME. Learned helplessness. Annu Rev Med 1972;23:407-412. [doi: 10.1146/annurev.me.23.020172.002203] [Medline: 4566487]
56. Seligman MEP. Learned Optimism: How to Change Your Mind and Your Life. New York, NY: Vintage Books; 2006.
57. van Os J, Kenis G, Rutten BPF. The environment and schizophrenia. Nature 2010 Nov 11;468(7321):203-212. [doi: 10.1038/nature09563] [Medline: 21068828]
58. Gunderson JG, Herpertz SC, Skodol AE, Torgersen S, Zanarini MC. Borderline personality disorder. Nat Rev Dis Primers 2018 May 24;4:18029. [doi: 10.1038/nrdp.2018.29] [Medline: 29795363]
59. Lips AMB, Eppel EA. Understanding and explaining online personal information-sharing behaviours of New Zealanders: A new taxonomy. Inf Commun Soc 2016 May 09;20(3):428-443 [FREE Full text] [doi: 10.1080/1369118x.2016.1184697]
60. Seabrook EM, Kern ML, Rickard NS. Social networking sites, depression, and anxiety: A systematic review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]

Edited by G Eysenbach; submitted 04.03.20; peer-reviewed by O Ness, A Economides, AJ Nagarajan; comments to author 25.03.20; revised version received 05.04.20; accepted 08.04.20; published 23.04.20.

Please cite as:
Bjornestad J, Moltu C, Veseth M, Tjora T
Rethinking Social Interaction: Empirical Model Development
J Med Internet Res 2020;22(4):e18558
URL: http://www.jmir.org/2020/4/e18558/
doi:10.2196/18558
PMID:32324144

Original Paper

Patient-Reported Outcomes During Immunotherapy for Metastatic Melanoma: Mixed Methods Study of Patients' and Clinicians' Experiences

Lærke K Tolstrup[1,2*], MA, RNC; Helle Pappot[3*], MD, PhD, Prof Dr; Lars Bastholt[1,2*], MD; Ann-Dorthe Zwisler[1,4*], MD, PhD, Prof Dr; Karin B Dieperink[1,2*], RNC, PhD

[1]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

[2]Department of Oncology, Odense University Hospital, Odense, Denmark

[3]Department of Oncology, Copenhagen University Hospital, Copenhagen, Denmark

[4]The Danish Knowledge Centre for Rehabilitation and Palliative Care, Nyborg, Denmark

[*]all authors contributed equally

Corresponding Author:
Lærke K Tolstrup, MA, RNC
Department of Clinical Research
University of Southern Denmark
Sønder Boulevard 29
Odense, 5000
Denmark
Phone: 45 40295129
Email: laerke.tolstrup@rsyd.dk

Abstract

Background: The benefits of electronic patient reported outcomes (PRO) questionnaires have been demonstrated in many settings, including in hospitals and patient homes. However, it remains to be investigated how melanoma patients and their treating clinicians experience the electronic self-reporting of side effects and the derived communication.

Objective: The primary objective of this study was to examine patients' and clinicians' experiences with an eHealth intervention for weekly monitoring of side effects during treatment with immunotherapy.

Methods: An eHealth intervention based on questions from the PRO-Common Terminology Criteria for Adverse Events (CTCAE) library was used and tested in a randomized clinical trial with patients receiving immunotherapy for malignant melanoma and clinicians at a university hospital in Denmark. On a weekly basis, patients reported their symptoms from home during the treatment via a provided tablet. The electronic patient reports were available to clinicians in the outpatient clinic. A mixed methods approach was applied to investigate the patients' and clinicians' experiences with the intervention. Data from patient experiences were collected in a short survey, the Patient Feedback Form. Moreover, a subset of the patients participating in the survey was interviewed about their experience. Furthermore, one focus group interview with clinicians was carried out to elucidate their views.

Results: A total of 57 patients completed the Patient Feedback Form, and 14 patients were interviewed. The focus group interview included 5 clinicians. Overall, patients and clinicians were satisfied with the tool. They believed it enhanced patients' awareness of side effects and increased their feeling of involvement. The patients reported that it was easy to fill out the questionnaire and that it made sense to do so. However, a minority of the patients expressed in the interviews that they did not believe that the health care professionals had seen their reports when they came to the clinic, and that the reporting did not lead to increased contact with the department.

Conclusions: Overall, satisfaction with the eHealth intervention was high among patients and their treating clinicians. The tool was easy to use and contributed to greater symptom awareness and patient involvement. Thus, in terms of patient and clinician satisfaction with the tool, it makes sense to continue using the tool beyond the project period.

Trial Registration: ClinicalTrials.gov NCT03073031; https://tinyurl.com/tjx3gtu

(J Med Internet Res 2020;22(4):e14896) doi:10.2196/14896

KEYWORDS

side effects, adverse events, patient-reported outcomes, PRO-CTCAE, melanoma, eHealth, immunotherapy, patient satisfaction; CPIs, interviews

Introduction

Underreporting of symptoms by clinicians in connection with cancer therapy, particularly chemotherapy and radiotherapy, is well established [1-4]. However, over the last few decades, new therapies have been developed and various kinds of immunotherapies now play an important role in fighting cancer [5]. In particular, immunotherapy has significantly improved survival in patients with melanoma [6]. However, the side effects that patients experience when treated with immunotherapy can be severe and unpredictable, and they differ immensely from the side effects experienced by patients who receive chemotherapy [5]. Furthermore, untreated toxicities may progress and become potentially life threatening [7]. Thus, toxicity monitoring may advantageously be optimized to meet the need for early detection of symptoms. Studies have demonstrated the value in using patient-reported outcomes (PROs) to detect and monitor symptoms, and to improve communication in routine care [8], and their implementation has been encouraged [9,10]. Moreover, increased inclusion of patients in their treatment has become a priority in many health care settings worldwide [11]. Similarly, there has been an increasing awareness within the Danish health care system of patients not being sufficiently involved with their treatment and care [12], despite the fact that the Danish regions recommend planning treatment *with* the patient rather than *for* the patient [13].

The use of electronic PRO questions (ePROs) to monitor symptoms has proven to be feasible in connection with scheduled consultations (ie, in the waiting area in various oncology settings) [14], and recent evidence suggests that the method is also useful at home (ie, via a link to a webpage) [15]. Studies also demonstrate that including cancer patients in the reporting of symptoms may increase their quality of life [16], and that the general acceptability of routine data collection is high [8]. With regard to immunotherapy, previous studies have examined the quality of life during treatment [17,18], but no study has examined whether patient reporting of side effects also results in improved toxicity monitoring. Therefore, we designed a randomized controlled trial (RCT), PROMelanoma (ClinicalTrials.gov NCT03073031), with the primary aim of investigating whether the severity and frequency of severe side effects can be reduced by including the patients in the reporting of symptoms on a frequent basis. Enrollment for this study has just completed.

An exploratory endpoint of PROMelanoma was to examine whether our setup of including an eHealth intervention on symptom management is implementable in clinical practice and makes the patients feel more involved in their treatment and care. Patient and clinician satisfaction with various eHealth interventions has been measured in other studies within an oncology setting to support clinical decision making and improve patient self-management [15,19]. However, many outcome measures are not sufficiently tested in clinical practice, which is imperative before implementation. To ensure the success of PRO interventions, it is vital that they are approved by the patients [20]. Thus, there is a need for more precise measures [21] that fit the patient population under investigation [22] to make sure that the PRO intervention is feasible and easy for the patient to adopt. In this regard, studies that elucidate the usefulness of a given PRO from the perspectives of patients and their treating clinicians must be carried out. To our knowledge, no study has explored how melanoma patients treated with immunotherapy experience the electronic self-reporting of symptoms using an eHealth intervention specifically designed for this patient group [23], which makes this type of study highly relevant.

However, there is no recipe for measuring the patient experience, and measurement is not routinely conducted in a standardized manner [24]. Thus, the patient experience can be captured in different ways. To acquire a broad perspective on the topic, a mixed methods approach may be most suitable. For example, a short survey can help provide feedback about the general trends, whereas in-depth interviews may provide a more detailed understanding of both the patient and clinician perspective [25]. Similarly, Hudak et al [26] suggested that it is preferable to combine a standardized quantitative measure with a qualitative method when measuring patient satisfaction. Girgis et al [15] used a similar method when they evaluated the feasibility and acceptability of real-time reporting in a cancer population.

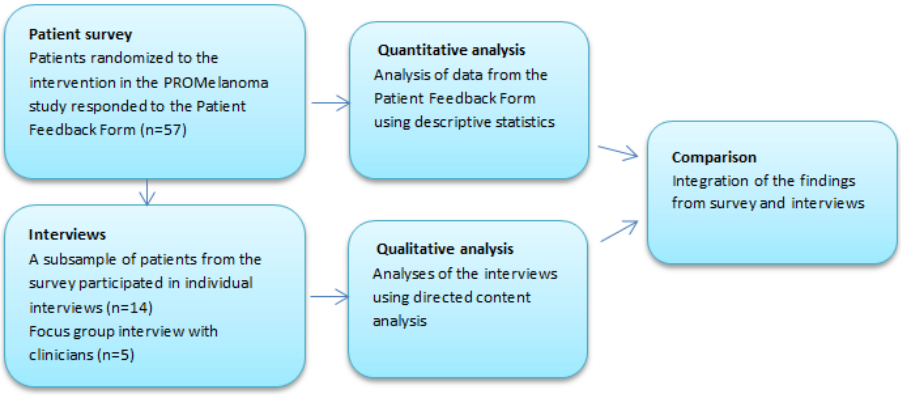
Thus, the primary objective of this study was to examine, using both qualitative and quantitative data, patients' and clinicians' experiences with an eHealth intervention to monitor the side effects during treatment with immunotherapy in routine clinical practice.

Methods

Overall Design

A mixed methods approach was employed to gain deeper insight into the feasibility of the PRO intervention for melanoma patients and their treating clinicians. For quantitative assessment, a questionnaire to measure patient satisfaction, the Patient Feedback Form [19,27], was provided to patients who experienced the PROMelanoma eHealth intervention. In addition to the questionnaire, qualitative interviews with a subsample of these patients and one focus group interview with clinicians were conducted using a deductive approach [28] to evaluate the intervention. The Consolidated Criteria for Reporting Qualitative Research (COREQ) Checklist [29] was applied to ensure that all important aspects were included. A convergent design was selected [30], in which the survey data and interview data were collected in parallel over the same period of time (February 2017 to March 2019). Data were analyzed separately and compared to determine similarities and differences. Using the triangulation technique, cross verification of data from the interviews and survey was achieved (Figure 1).

Figure 1. Overview of the mixed methods study design, including a survey, individual interviews, and one focus group interview.



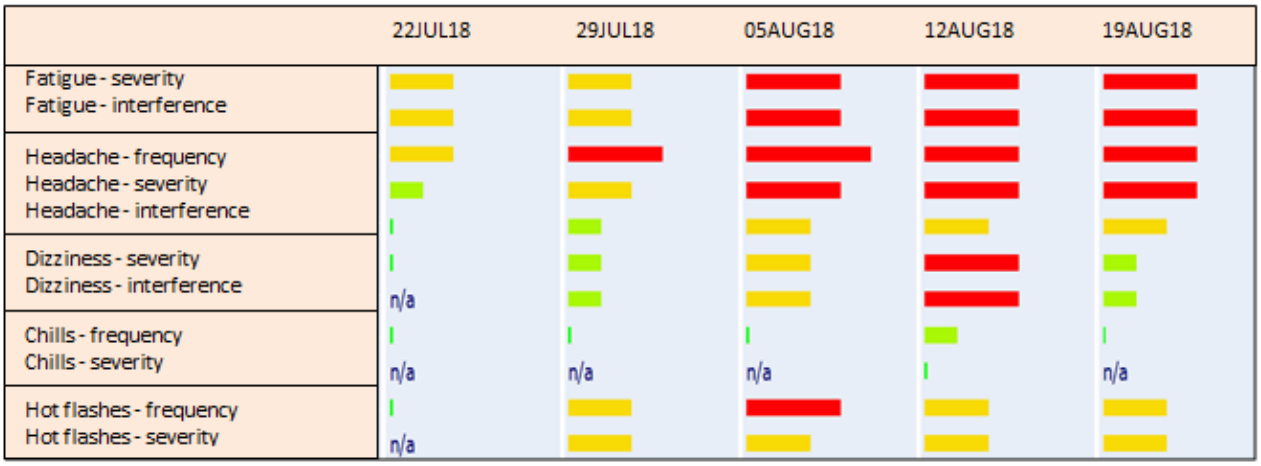
Setting

The survey and interviews took place at the Department of Oncology, Odense University Hospital, Denmark. The patients completed the Patient Feedback Form when they came to the outpatient clinic to receive their treatment for metastatic melanoma. The interviews also took place in the outpatient clinic in a separate room.

The eHealth Intervention

Common Terminology Criteria for Adverse Events (CTCAE), developed by the National Cancer Institute (NCI) for patient self-reporting [31], was chosen as the PRO tool (PRO-CTCAE), since the grading scale [32] is well known within the oncology field [33] and is used by oncologists worldwide. Through a careful selection process, the relevant items were selected from the PRO-CTCAE library [34]. The software platform AmbuFlex [35] was used, which was specifically developed for ePROs. The patients received a tablet with a SIM card to ensure internet access. The reporting took place on the tablet, at home once a week, which is the preferred recall period for PRO-CTCAE items [36], and continued for 24 weeks to ensure that the majority of symptoms could be detected. The patients did not receive a weekly reminder in the form of a text message or telephone call, but they were asked to choose a fixed weekday for reporting their symptoms when they were first introduced to the intervention so that reporting would be easier to remember. If the patients experienced a symptom, an alert would tell them to contact the department. The alert function was triggered for side effects that could potentially become severe. Accordingly, side effects such as alopecia or fatigue did not trigger an alert. As soon as the patients responded to the questionnaire, the report was visible to the health care professionals at the hospital. However, clinicians did not receive a notification when an alert was triggered by a patient report; rather, it was left up to the patients to react to the alert. The clinicians only logged onto the electronic system to see the patient's report after the patient came to the outpatient clinic. A bar attached to each symptom appeared green, yellow, or red depending on the severity of the symptom reported (Figure 2).

Figure 2. Example of part of a patient report available to clinicians.



Patients

Patients were eligible for the qualitative part of the study if they had been enrolled in the RCT PROMelanoma. The inclusion criteria were melanoma patients, >18 years old, randomized to the intervention in PROMelanoma, and had received at least one cycle of immunotherapy. Exclusion criteria were not able or willing to comply with the study procedure (eg, fill out the electronic questionnaire) or if they did not speak Danish.

Survey

All patients in the PROMelanoma intervention group of the trial were asked to fill out the Patient Feedback Form between January 2017 and April 2019, which addressed patient satisfaction relating to the eHealth intervention. The Patient Feedback Form was developed by Basch et al [27] to measure patient satisfaction with the online self-reporting of toxicity symptoms, and was subsequently adapted by Snyder et al [19] who also used it to measure patient satisfaction with PRO interventions. Thus, it is an established tool to measure quantitative feedback and was considered to be appropriate for evaluating the usefulness and acceptability of our eHealth intervention. The adapted version consists of 13 items [19]. Respondents evaluate their level of agreement or disagreement on a scale with four options. Some representative questions included were: "Was it easy to use?," "Did the questions make sense?," and "Were the patient reports included in the patient-clinician consultation?" To apply the questionnaire for evaluating the eHealth intervention, we translated it into Danish and validated it in a Danish setting according to existing guidelines, including psychometric testing [37]. The patients had carried out the weekly PROMelanoma reporting at least three times and had the opportunity to discuss their report with a physician at least once before filling out the Feedback Form. Data were analyzed using descriptive statistics when enrollment in the PROMelanoma study closed in April 2019, and 70 patients had been enrolled in the intervention group.

Interviews With Patients

Patients enrolled in the PROMelanoma study were contacted over the phone by the project manager and informed about this study between November 2017 and June 2018. The patients provided verbal consent and signed the written consent form in connection with the interview. We decided to use a convenience sample at the same time, taking into account the patients' gender and age to ensure that the group was representative. The patients already had several visits scheduled in the outpatient clinic; therefore, the interviews were planned to take place on days when they were already at the hospital so as to not burden them further. If the patients were accompanied by relatives, the relatives were invited to participate in the interview. A semistructured interview guide was prepared based on the research questions, in collaboration with an expert. The interviews were carried out by the same interviewer (LT) who also carried out audio recording and transcription. The interviewer had already talked to the majority of the patients during the inclusion screening for the PROMelanoma study, but had no contact otherwise. Given that we had some knowledge about the research area in question (ie, the interviewer had worked with this patient group for more than 10 years), there were four major categories that we wished to explore: the usefulness of the eHealth solution, the questionnaire, physician-patient communication, and involvement of relatives. Thus, a directed content analysis as suggested by Hsieh and Shannon [38] was applied, using a deductive approach [39]. The fact that the level of interpretive complexity was expected to be relatively low further contributed to our choice of content analysis as the preferred method [40]. Any text that could not be categorized within the initial categories would be given a new code during the analysis [38]. Recruitment continued until data saturation was reached.

Focus Group Interview With Clinicians

A focus group interview was chosen as the preferred method for clinicians, because the number of physicians and nurses caring for these patients was limited to a selected group, which made a questionnaire pointless. For the same reason, only one interview was conducted. The physicians and nurses who had the most experience with the intervention were chosen for the interviews. One author (KD) carried out the interview, who is a qualified researcher experienced in conducting focus group interviews. The interview was conducted in a semistructured manner [41]. Data were generated through group interaction about the specific topic predetermined by the research group. The purpose of the focus group was to explore the perspectives of the clinicians regarding the implementation and acceptability of the eHealth intervention in routine cancer care. The interview was transcribed by LT. The same content analysis approach was applied in relation to the group interview as described above [38].

Results

Survey

All patients who were randomized to the intervention arm in the PROMelanoma study (N=70, median age 65 years, 33 men and 37 women) were expected to evaluate the eHealth intervention by filling out the Patient Feedback Form (Table 1). However, 2 patients who had been randomized to the intervention group did not wish to proceed with the electronic reporting, and 2 patients were hospitalized due to side effects and never received the second series. For 9 patients, the melanoma progressed quickly and their conditions deteriorated, making it unethical to ask them to participate. Thus, a total of 57 patients evaluated the intervention. As summarized in Table 1, none of the patients found the eHealth intervention to be too time-consuming (item 1). In fact, one patient thought that it was too short. Similarly, almost all of the patients found the frequency with which the eHealth intervention was administered (item 2) to be just right. The general satisfaction was high. The lowest satisfaction ratings were obtained for items 8, 9, and 10, dealing with inclusion of the patient response in treatment and care. Overall, the majority of patients agreed/strongly agreed that the doctor used the information for their care, that the questionnaire improved the quality of care (item 9), and that the questionnaire improved communication with the doctor (item 10). The proportion of patients who responded "strongly agree," "agree," or "just right" was over 90% for 8 of the 13 questions. All of the patients (100%) recommended filling out the questionnaire to other patients and stated that they would like to continue responding to the questionnaire (items 12 and 13).

Table 1. Evaluation of the eHealth intervention PROMelanoma in a Danish study with patients with melanoma cancer (N=57).

Patient feedback form item	Response, n (%)		
	Category 1	Category 2	Category 3
1. Time it took to complete	1 (2)[a]	54 (94)[b] and 0 (0)[c]	2 (4)[d]
2. Number of times completing	1 (2)[e]	54 (94)[b] and 1 (2)[f]	1 (2) [d]
3. Easy to complete	56 (98)[g]	1 (2)[h]	0 (0)[d]
4. Completing was useful	55 (96)[g]	2 (4)[h]	0 (0)[d]
5. Easy to understand	53 (93)[g]	4 (7)[h]	0 (0)[d]
6. Easier to remember symptoms and side effects	52 (91)[g]	4 (7)[h]	1 (2)[d]
7. Improved discussions with clinician	51 (89)[g]	4 (7)[h]	2 (4)[d]
8. Clinician used information for my care	48 (84)[g]	6 (11)[h]	3 (5)[d]
9. The quality of care improved because of the questionnaire	43 (75)[g]	8 (14)[h]	6 (11)[d]
10. Communication with clinician improved	45 (78)[g]	6 (11)[h]	6 (11)[d]
11. Made me more in control of care	50 (87)[g]	6 (11)[h]	1 (2)[d]
12. Recommend to other patients	57 (100)[g]	0 (0)[h]	0 (0)[d]
13. Would like to continue responding	57 (100)[g]	0 (0)[h]	0 (0)[d]

[a]Too short.
[b]Just right.
[c]Too long.
[d]Missing.
[e]Not often enough.
[f]Too often.
[g]Strongly agree/agree.
[h]Disagree/strongly disagree.

Patient Interviews

General Characteristics of the Interviews

In addition to filling out the Patient Feedback Form, 16 of the patients were invited to participate in an in-depth interview about their experience. One patient declined and one patient who had agreed to participate was hospitalized due to deteriorating disease before the interview was conducted. Thus, 14 interviews were conducted. The median age of the patients was 67 years (range 41-79 years), including 6 men and 8 women. Apart from one patient who had only self-reported their symptoms 3 times, the patients had reported between 6 and 24 times (weeks), and the majority (10, 71%) had reported more than 15 times. Relatives were present during 10 of the interviews. The interviews lasted on average 20 minutes (range 9-33 minutes). Nine interviews lasted for more than 20 minutes. A total of 280 minutes of interview data were available for analysis. The three themes identified from the transcripts aligned with three of the predetermined categories. However, a fourth theme (involvement of relatives) did not become a theme when the final analysis was carried out.

Usefulness of the eHealth Solution

Overall, the patients reported that accessing and filling out the eHealth questionnaire was easy. Only two patients were not used to electronic devices upon entering the study. One of them stated “I’m pleasantly surprised. I think it is really easy to deal with” (man, 79 years old), and his wife (73 years old) added, “I did not think he could do it because he is a clown when it comes to computers…” Some of the patients, particularly the elderly, had a hard time using the touchscreen function with their fingers because they either pressed too hard or for too long. However, when they were given a touchscreen pen, which is more accurate than the fingertip, they did not have any problems. Only one patient could not do it and asked his wife to do the reporting following his instructions. Almost all of the patients experienced a request to update the operating system of the tablet while using it, but they were able to close the message easily and continued their reporting. Otherwise, there were only minor technical challenges, and the patients were very compliant and contacted the department in case of any technical problems. The majority of patients were pleased with the tablet. Only a few patients would have preferred a link instead of having to take home the tablet. As mentioned above, it was not possible to send a text message reminding the patient to fill out the questionnaire on the relevant days. However, this did not constitute a problem for the patients, who found it easy to remember because they were doing it on a fixed weekday. Two patients mentioned that a reminder text message would have been advantageous.

Questionnaire

The patients reported that the number of items and the length of the questionnaire were appropriate and that reporting on a weekly basis was fitting. A few of the patients would have liked a free text field where they could write a comment or elaborate if the questionnaire did not adequately cover existing symptoms: "It is as if you (health care professionals) don't get enough information" (woman, 71 years old). The patients were divided when asked if responding to the PROMelanoma questionnaire was reassuring; half of the patients confirmed that this was the case, while the other half rejected this notion: "I feel reassured enough as it is" (woman, 52 years old). The majority reported that their attention to side effects was heightened due to the intervention (eg, "Your focus is increased because you have to remember to write it" [woman, 62 years old]), and that responding to the questionnaire was useful. More of the patients also found that filling out the questionnaire made it easier to remember symptoms when they came to the clinic. One patient reported that she was reminded of her disease every time she responded. A majority of the patients reported that the alert reminding them to contact the department popped up too frequently. As one interviewee put it: "If I were to call every time it pops up, I would have to call very often" (woman, 67 years old). However, if the patients decided that it was not a new symptom or worsening of an already existing symptom, they were able to reject the alert.

Patient-Physician Communication

When the patients came to the outpatient clinic, two out of three of the patients who were interviewed felt that the health care professionals had in fact seen their reports and included them in their consultation: "It is like having an agenda for a meeting" (man, 66 years old), "It makes you feel as if you are not just a number in the system" (woman, 49 years old) . A minority did not know if their reports had been seen by the clinician: "I think they have seen it (the report), but it is not something we have discussed" (woman, 69 years old). A few believed that the clinician had in fact not seen it at all, which was of course frustrating due to the fact that they had spent time filling out the questionnaire. One third had the feeling that they contacted the department more as a result of the reporting. Thus, the majority of patients did not think that they were more in touch with the hospital due to the reporting. Overall, the reporting made the patients feel more involved in their treatment and care: "It is nice that we have something common to talk about" (man, 66 years old).

Other Themes

Many of the patients explained that a strong motivation for entering the study was that they would be able to help future patients. Of course, they believed that they themselves would benefit, but being able to help others was also important. Including relatives in the reporting was not a theme. The patients did the reporting alone, apart from one patient, and it did not prompt any discussions within the family.

Focus Group Interview With Clinicians

The participants in the focus group consisted of three doctors and two nurses. They were all women with a median age of 43 years. All of them had broad experience working with cancer patients and dealing with symptoms or side effects (6-11 years). They were also accustomed to caring for melanoma patients receiving immunotherapy. They had all seen the patient reports several times and had included them in the clinician-patient communication.

There was some discrepancy between how the patient and the clinician graded a given symptom. In some cases, the clinician did not find the symptom to be as severe as the patient. In other instances, the clinician felt that the patient had in fact neglected a symptom that they believed should have been reported: "sometimes there's a discrepancy between what you find out when you talk to the patient and what has been reported … the two things supplement each other" (physician).

Furthermore, the inclusion of patient reporting was seen as being more time-consuming than a typical consultation due to the fact that the clinicians had to log into another system to see the report. Having the reports integrated in the electronic health records (EHRs) was stated not only to save time but also make it much easier to remember to include them in the consultation.

The clinicians agreed that the patients were better prepared when they came to the outpatient clinic, and that the patients had increased focus on their symptoms and were more alert: "I think it is an advantage that the patients become more aware of the side effects that can occur" (nurse). Moreover, the information on toxicity that had been given to the patients prior to treatment start was repeated when the patients responded to the electronic questionnaire at home. Accordingly, there was a better chance that the patients would react appropriately by contacting the department in time instead of waiting for the next scheduled consultation, which might be days or weeks ahead. Thus, having the patients call more often was seen as an advantage because it might enable earlier detection. Moreover, it was an advantage to be able to use the patient reporting as the basis of the consultation by starting with the symptoms that had bothered the patient the most: "…then I scroll down to see where it is red or yellow and that is typically where we start…" (physician). In this way, the patients took part in setting the agenda. However, according to the health care professionals, the patient reporting should be seen as a supplement and not something that could replace the clinician-patient consultation. In addition, the clinicians reported that the eHealth intervention was a valuable tool, particularly for patients who are normally slightly reluctant to contact the department unscheduled: "…it may be precisely the group of patients who are not good at self-care or at least some of them…the weakest patients who…will benefit most from self-reporting by being guided into becoming more aware of when to react to symptoms" (physician). Because the patients were encouraged to make contact if they experienced a new or worsened symptom, they might have felt that it was more legitimate to call the outpatient clinic. All of the clinicians believed that the patients with the best social resources would benefit the least from the intervention because they were sure to contact the department in agreement with the given instructions.

Overall, the clinicians had a positive attitude toward the intervention using an eHealth tool, even though there was also room for improvement in some areas.

Comparison Between Survey and Patient Interviews With the Focus Group Interview

The clinicians believed that the reporting would make the patients call the hospitals more, whereas the majority of patients did not think that they called more frequently. Some of the patients thought that their reports did not provide the clinicians with enough information; however, none of the clinicians stated this to be the case. Patients and clinicians agreed that the attention to side effects was increased and that the patients were better prepared for the consultation when they came to the outpatient clinic. The patient reports also established a shared agenda for the consultation at the outpatient clinic. Overall, the findings from the survey confirmed what had been established in the patient interviews. The patients reported that it was easy to fill out the questionnaire and that it made sense to do so. Moreover, it increased symptom awareness. Both the patients and clinicians agreed that when the report was in fact included, it helped to prioritize the problems that were most acute.

Discussion

Principal Findings

The goal of this study was to elucidate the experiences of malignant melanoma patients and their treating clinicians with an eHealth intervention. Overall, acceptance was high for both clinicians and patients, and both groups believed that it improved communication during the consultation. This is in line with previous studies showing that using PROs prompted patient-clinician dialog, streamlined consultations, and increased focus on side effects [10,42]. In addition, the potential for discrepancies between the degree of severity when clinicians and patients grade a given symptom confirms previous findings [1-4].

However, a minority of the patients in this study did not believe that the clinician had actually seen their reports when they came to the clinic. This point was primarily expressed by patients who were enrolled at the beginning of the study, when monitoring the patient reports had not yet become routine in the outpatient clinic. This improved over time as the clinicians got used to taking the reports into consideration. This finding is in line with Mooney et al [43], who argued that when the advantages of systematic PRO collection in clinical care become visible, adoption will rapidly occur. Although the use of PRO in the clinic can improve communication, it does not necessarily result in intensified symptom treatment and improved symptom management [44]. Thus, it remains to be seen if patient and clinician satisfaction with the eHealth intervention will translate into a reduction in symptom severity; this aspect is being investigated in the ongoing RCT PROMelanoma.

As for the survey, patient satisfaction was extremely high for many of the questions. The three items that had the lowest scores in satisfaction (items 8, 9, and 10) deal with the inclusion of patient response in the clinic. This response is comparable with the results of other studies using the Patient Feedback Form [19]. This suggests that one of the challenges when using PROs may be to ensure that the patients' responses from questionnaires are included in treatment and care. For many years, PROs have been collected in clinical trials, but they have not been used routinely in clinics. It will likely take some time before implementing PROs in clinical practice becomes as natural as other procedures within the health care system.

The clinicians participating in the focus group interview agreed that the least resourceful patients would benefit most from the eHealth intervention, because they were usually less inclined to contact the clinic in case of any symptoms. This notion has been confirmed in other studies, which have shown that the level of patient involvement is dependent on the degree of health literacy. For example, patients with a high level of education are more inclined to be involved in medical decision making compared to patients with a low level of education [45]. Basch et al [46] also suggested that patients who do not have any computer experience may have weaker communication skills and therefore benefit more from a structured setup. It can be argued that if this patient group becomes involved in the reporting of side effects, they may be encouraged to react appropriately when an alert is triggered, thereby potentially improving toxicity management. When data from the RCT on the number of phone contacts are analyzed, it will be revealed if patients in the intervention arm actually did call more frequently. Preliminary findings revealed that 78% of the patients adhered to the intervention by reporting their symptoms on a weekly basis.

Some of the patients also argued that the eHealth intervention was very box-like and they would have liked a space where they could write more about their symptoms instead of just checking a box. The patients in the PROMelanoma study can add other symptoms as advised by the NCI, but the patients also wished to be able to elaborate on some of the symptoms. Although this is understandable from a patient point of view, one must keep in mind that the primary aim of introducing the intervention was to increase patient awareness, hoping to reduce the number of severe side effects and improve clinical outcome. Further, it was important that it was fairly easy and not too time-consuming for the clinicians to acquire a quick outline of the reporting if it were to be implementable in the clinic. Moreover, patients had the opportunity to elaborate on the various symptoms that they experienced when they came to the clinic.

Limitations

One potential limitation of the study is that a deductive approach was used by having the coding framework decided in advance, which may limit the development of new themes [28]. By using a deductive approach, and thus imposing our own structure on the data, the analysis may have been biased. However, the fact that we had some knowledge about the subject made the deductive approach an obvious choice. The fact that one of the predetermined categories, involvement of relatives, did not develop into a theme and was removed during the analysis indicates that we were not too locked in our preconception.

An obvious limitation is that we were only able to conduct one focus group interview with the clinicians. However, we aimed

at selecting participants with vast knowledge and expertise of the subject [47], which limited the potential number of informants. Of course, other physicians and nurses had treated these patients, but the fact that they did not do so on a routine basis made them unsuitable as participants. Consequently, we settled for one focus group although it would have been preferable to have more. With respect to the number of interviewed patients, we judged that data saturation was reached at patient number 14, as data were replicated, which is why we stopped including patients in the study at this point. According to Francis et al [48], data saturation may very well be reached after 14 interviews when diversity sampling is appropriate. We believed this was the case in this study.

Another potential limitation is that the alert function was triggered too frequently according to the majority of patients. This may be changed when designing future studies or implementing the intervention beyond the study period to avoid alert fatigue. Nevertheless, having an alert function is a good idea, as studies have shown that patients value advice on when it is appropriate to contact the hospital [49]. In addition, it is vital that the clinicians log into the system and see the patients' reports prior to every consultation. Otherwise, patients may lose the incentive to continue to fill out the reports. PROs must be implemented in such a way that the process is embedded as part of routine care [21] so that clinicians do not have to be reminded to view the patient report (eg, by project managers or study coordinators). In this regard, it is important that PROs are easily accessible to clinicians (ie, integrated in the EHR) to be successful, as recommended by the clinicians in the focus group interview. Recommendations on how to integrate PROs into the EHR have already been developed by the PRO-EHR Users' Guide Steering and Working Groups [50].

Conclusion

We found a high acceptance of the eHealth intervention tool among clinicians and melanoma patients being treated with immunotherapy. The tool was easy to use and contributed to greater symptom awareness and patient involvement. Thus, in terms of patient and clinician satisfaction, it makes sense to continue using the tool beyond the project period. However, it remains to be investigated whether the predominantly positive perceptions of the intervention by patients and clinicians will also be followed by a reduction in the number of severe side effects. Our RCT PROMelanoma will shed light on this aspect.

Acknowledgments

We would like to thank the patients who participated in interviews and the survey, and–despite difficult circumstances–contributed to this study. We are also grateful to the physicians and nurses for sharing their experiences and views. We also thank Niels Henrik Hjøllund, WestChronic, Occupational Medicine, University Research Clinic, Aarhus University, who provided technical support, making it possible to carry out the study. Finally, we acknowledge William Frost, Last Word Consultancy, for language editing. The study is funded by the Danish Cancer Society and the Region of Southern Denmark. None of the funding organizations was involved in reviewing or approving the manuscript.

Conflicts of Interest

None declared.

References

1. Atkinson TM, Ryan SJ, Bennett AV, Stover AM, Saracino RM, Rogak LJ, et al. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. Support Care Cancer 2016 Aug;24(8):3669-3676. [doi: 10.1007/s00520-016-3297-9] [Medline: 27260018]
2. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst 2014 Sep;106(9):dju244 [FREE Full text] [doi: 10.1093/jnci/dju244] [Medline: 25265940]
3. Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. J Clin Oncol 2015 Mar 10;33(8):910-915. [doi: 10.1200/JCO.2014.57.9334] [Medline: 25624439]
4. Bruner DW, Movsas B, Basch E. Capturing the patient perspective: patient-reported outcomes as clinical trial endpoints. Am Soc Clin Oncol Educ Book 2012:139-144 [FREE Full text] [doi: 10.14694/EdBook_AM.2012.32.139] [Medline: 24451723]
5. Hall ET, Singhal S, Dickerson J, Gabster B, Wong H, Aslakson RA, AAHPM Research Committee Writing Group. Patient-Reported Outcomes for Cancer Patients Receiving Checkpoint Inhibitors: Opportunities for Palliative Care-A Systematic Review. J Pain Symptom Manage 2019 Jul;58(1):137-156. [doi: 10.1016/j.jpainsymman.2019.03.015] [Medline: 30905677]
6. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. N Engl J Med 2015 Jul 02;373(1):23-34 [FREE Full text] [doi: 10.1056/NEJMoa1504030] [Medline: 26027431]

7. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010 Aug 19;363(8):711-723. [doi: 10.1056/NEJMoa1003466] [Medline: 20525992]
8. Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. Ann Oncol 2015 Sep;26(9):1846-1858. [doi: 10.1093/annonc/mdv181] [Medline: 25888610]
9. Beijers AJM, Mols F, van den Hurk CJ, Vreugdenhil A. Are chemotherapy-associated symptoms underestimated? A view beyond common toxicity criteria. Acta Oncol 2016;55(4):516-518. [doi: 10.3109/0284186X.2015.1073352] [Medline: 26359553]
10. Yang LY, Manhas DS, Howard AF, Olson RA. Patient-reported outcome use in oncology: a systematic review of the impact on patient-clinician communication. Support Care Cancer 2018 Jan;26(1):41-60. [doi: 10.1007/s00520-017-3865-7] [Medline: 28849277]
11. Boivin A, Lehoux P, Lacombe R, Burgers J, Grol R. Involving patients in setting priorities for healthcare improvement: a cluster randomized trial. Implement Sci 2014 Feb 20;9:24 [FREE Full text] [doi: 10.1186/1748-5908-9-24] [Medline: 24555508]
12. Danske Regioner. LUP - Den Landsdækkende Undersøgelse af Patientoplevelser. Frederiksberg: Enhed for Evaluering og Brugerinddragelse; 2015 Apr. URL: https://www.regioner.dk/media/2172/lup-2014.pdf [accessed 2020-02-27]
13. Danske Regioner. Plan for Borgernes Sundhedsvæsen - vores sundhedsvæsen. 2015. URL: https://www.regioner.dk/media/3047/borgernes-sundhedv-final-let.pdf [accessed 2020-02-27]
14. Basch E, Pugh SL, Dueck AC, Mitchell SA, Berk L, Fogh S, et al. Feasibility of Patient Reporting of Symptomatic Adverse Events via the Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) in a Chemoradiotherapy Cooperative Group Multicenter Clinical Trial. Int J Radiat Oncol Biol Phys 2017 Dec 01;98(2):409-418. [doi: 10.1016/j.ijrobp.2017.02.002] [Medline: 28463161]
15. Girgis A, Durcinoska I, Levesque JV, Gerges M, Sandell T, Arnold A, PROMPT-Care Program Group. eHealth System for Collecting and Utilizing Patient Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care) Among Cancer Patients: Mixed Methods Approach to Evaluate Feasibility and Acceptability. J Med Internet Res 2017 Dec 02;19(10):e330 [FREE Full text] [doi: 10.2196/jmir.8360] [Medline: 28970188]
16. Basch E, Abernethy AP, Mullins CD, Reeve BB, Smith ML, Coons SJ, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. J Clin Oncol 2012 Dec 01;30(34):4249-4255. [doi: 10.1200/JCO.2012.42.5967] [Medline: 23071244]
17. Petrella TM, Robert C, Richtig E, Miller WH, Masucci GV, Walpole E, et al. Patient-reported outcomes in KEYNOTE-006, a randomised study of pembrolizumab versus ipilimumab in patients with advanced melanoma. Eur J Cancer 2017 Dec;86:115-124 [FREE Full text] [doi: 10.1016/j.ejca.2017.08.032] [Medline: 28987768]
18. Abdel-Rahman O, Oweira H, Giryes A. Health-related quality of life in cancer patients treated with PD-(L)1 inhibitors: a systematic review. Expert Rev Anticancer Ther 2018 Dec;18(12):1231-1239. [doi: 10.1080/14737140.2018.1528146] [Medline: 30253680]
19. Snyder CF, Herman JM, White SM, Luber BS, Blackford AL, Carducci MA, et al. When using patient-reported outcomes in clinical practice, the measure matters: a randomized controlled trial. J Oncol Pract 2014 Sep;10(5):e299-e306 [FREE Full text] [doi: 10.1200/JOP.2014.001413] [Medline: 24986113]
20. Recinos PF, Dunphy CJ, Thompson N, Schuschu J, Urchek JL, Katzan IL. Patient Satisfaction with Collection of Patient-Reported Outcome Measures in Routine Care. Adv Ther 2017 Dec;34(2):452-465. [doi: 10.1007/s12325-016-0463-x] [Medline: 28000165]
21. Howell D, Fitch M, Bakker D, Green E, Sussman J, Mayo S, et al. Core domains for a person-focused outcome measurement system in cancer (PROMS-Cancer Core) for routine care: a scoping review and Canadian Delphi Consensus. Value Health 2013;16(1):76-87 [FREE Full text] [doi: 10.1016/j.jval.2012.10.017] [Medline: 23337218]
22. Kluetz PG, Slagle A, Papadopoulos EJ, Johnson LL, Donoghue M, Kwitkowski VE, et al. Focusing on Core Patient-Reported Outcomes in Cancer Clinical Trials: Symptomatic Adverse Events, Physical Function, and Disease-Related Symptoms. Clin Cancer Res 2016 Apr 01;22(7):1553-1558 [FREE Full text] [doi: 10.1158/1078-0432.CCR-15-2035] [Medline: 26758559]
23. Howell D, Liu G. Can routine collection of patient reported outcome data actually improve person-centered health? Healthc Pap 2011;11(4):42-7; discussion 55. [doi: 10.12927/hcpap.2012.22703] [Medline: 22543292]
24. LaVela SL, Gallan AS. Evaluation and measurement of patient experience. Patient Exp J 2014 Apr;1(1):28-36.
25. De Silva D. Evidence Scan: Measuring Patient Experience. London: The Health Foundation; 2013 Jun. Measuring patient experience URL: https://www.health.org.uk/sites/default/files/MeasuringPatientExperience.pdf [accessed 2020-02-27]
26. Hudak PL, Wright JG. The characteristics of patient satisfaction measures. Spine 2000 Dec 15;25(24):3167-3177. [doi: 10.1097/00007632-200012150-00012] [Medline: 11124733]
27. Basch E, Artz D, Dulko D, Scher K, Sabbatini P, Hensley M, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. J Clin Oncol 2005 May 20;23(15):3552-3561 [FREE Full text] [doi: 10.1200/JCO.2005.04.275] [Medline: 15908666]

28. Burnard P, Gill P, Stewart K, Treasure E, Chadwick B. Analysing and presenting qualitative data. Br Dent J 2008 Apr 26;204(8):429-432. [doi: 10.1038/sj.bdj.2008.292] [Medline: 18438371]
29. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec 23;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
31. National Cancer Institute. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE).: National Cancer Insitute, Division of Cancer Control & Population Sciences; 2017. URL: https://healthcaredelivery.cancer.gov/pro-ctcae/ [accessed 2020-02-27]
32. National Cancer Institute. Common Terminology Criteria for Adverse Events v4.: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute; 2009 May 28. URL: https://www.eortc.be/services/doc/ctc/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf [accessed 2020-02-27]
33. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, National Cancer Institute PRO-CTCAE Study Group. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncol 2015 Nov;1(8):1051-1059 [FREE Full text] [doi: 10.1001/jamaoncol.2015.2639] [Medline: 26270597]
34. Tolstrup LK, Bastholt L, Zwisler A, Dieperink KB, Pappot H. Selection of patient reported outcomes questions reflecting symptoms for patients with metastatic melanoma receiving immunotherapy. J Patient Rep Outcomes 2019 Mar 21;3(1):19 [FREE Full text] [doi: 10.1186/s41687-019-0111-8] [Medline: 30900035]
35. Schougaard LMV, Larsen LP, Jessen A, Sidenius P, Dorflinger L, de TA, 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]
36. Mendoza TR, Dueck AC, Bennett AV, Mitchell SA, Reeve BB, Atkinson TM, et al. Evaluation of different recall periods for the US National Cancer Institute's PRO-CTCAE. Clin Trials 2017 Jun;14(3):255-263. [doi: 10.1177/1740774517698645] [Medline: 28545337]
37. Tolstrup LK, Pappot H, Zangger G, Bastholt L, Zwisler A, Dieperink KB. Danish translation, cultural adaption and initial psychometric evaluation of the patient feedback form. Health Qual Life Outcomes 2018 Apr 27;16(1):77 [FREE Full text] [doi: 10.1186/s12955-018-0900-4] [Medline: 29703191]
38. 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]
39. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open 2016;2:8-14 [FREE Full text] [doi: 10.1016/j.npls.2016.01.001]
40. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
41. Doody O, Slevin E, Taggart L. Preparing for and conducting focus groups in nursing research: part 2. Br J Nurs 2013;22(3):170-173. [doi: 10.12968/bjon.2013.22.3.170] [Medline: 23411826]
42. Baeksted C, Pappot H, Nissen A, Hjollund NH, Mitchell SA, Basch E, et al. Feasibility and acceptability of electronic symptom surveillance with clinician feedback using the Patient-Reported Outcomes version of Common Terminology Criteria for Adverse Events (PRO-CTCAE) in Danish prostate cancer patients. J Patient Rep Outcomes 2017 Sep 12;1(1):1 [FREE Full text] [doi: 10.1186/s41687-017-0005-6] [Medline: 29757324]
43. Mooney K, Berry DL, Whisenant M, Sjoberg D. Improving Cancer Care Through the Patient Experience: How to Use Patient-Reported Outcomes in Clinical Practice. Am Soc Clin Oncol Educ Book 2017;37:695-704 [FREE Full text] [doi: 10.14694/EDBK_175418] [Medline: 28561689]
44. Mooney KH, Beck SL, Friedman RH, Farzanfar R, Wong B. Automated monitoring of symptoms during ambulatory chemotherapy and oncology providers' use of the information: a randomized controlled clinical trial. Support Care Cancer 2014 Sep;22(9):2343-2350 [FREE Full text] [doi: 10.1007/s00520-014-2216-1] [Medline: 24687538]
45. Brabers AEM, Rademakers JJDJM, Groenewegen PP, van Dijk L, de Jong JD. What role does health literacy play in patients' involvement in medical decision-making? PLoS One 2017;12(3):e0173316 [FREE Full text] [doi: 10.1371/journal.pone.0173316] [Medline: 28257472]
46. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
47. Doody O, Slevin E, Taggart L. Focus group interviews in nursing research: part 1. Br J Nurs 2013;22(1):16-19. [doi: 10.12968/bjon.2013.22.1.16] [Medline: 23299206]
48. Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychol Health 2010 Dec;25(10):1229-1245. [doi: 10.1080/08870440903194015] [Medline: 20204937]

49. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of Patient-Reported Outcomes in Routine Medical Care. Am Soc Clin Oncol Educ Book 2018 May 23;38:122-134. [doi: 10.1200/EDBK_200383] [Medline: 30231381]
50. Gensheimer SG, Wu AW, Snyder CF, PRO-EHR Users' Guide Steering Group, PRO-EHR Users' Guide Working Group. Oh, the Places We'll Go: Patient-Reported Outcomes and Electronic Health Records. Patient 2018 Dec;11(6):591-598. [doi: 10.1007/s40271-018-0321-9] [Medline: 29968179]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
CTCAE: Common Terminology Criteria for Adverse Events
EHR: electronic health record
ePROs: electronic patient reported outcomes
NCI: National Cancer Institute
PROs: patient reported outcomes
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 01.06.19; peer-reviewed by E Hall, H Yang; comments to author 03.10.19; revised version received 16.10.19; accepted 24.01.20; published 09.04.20.

Please cite as:
Tolstrup LK, Pappot H, Bastholt L, Zwisler AD, Dieperink KB
Patient-Reported Outcomes During Immunotherapy for Metastatic Melanoma: Mixed Methods Study of Patients' and Clinicians' Experiences
J Med Internet Res 2020;22(4):e14896
URL: https://www.jmir.org/2020/4/e14896
doi:10.2196/14896
PMID:32271150

Original Paper

Identification of the Facial Features of Patients With Cancer: A Deep Learning–Based Pilot Study

Bin Liang[1*], PhD; Na Yang[2*], MM; Guosheng He[3], MM; Peng Huang[1], MSc; Yong Yang[1], MD

[1]Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]South Building #2 Division, The 3rd Medical Center of the People's Liberation Army General Hospital, Beijing, China

[3]People's Hospital of Beijing Daxing District, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Bin Liang, PhD
Department of Radiation Oncology
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital
Chinese Academy of Medical Sciences and Peking Union Medical College
17 Panjiayuannanli Rd
Chaoyang District
Beijing, 100021
China
Phone: 86 1087788663
Email: leangbin@gmail.com

Abstract

Background: Cancer has become the second leading cause of death globally. Most cancer cases are due to genetic mutations, which affect metabolism and result in facial changes.

Objective: In this study, we aimed to identify the facial features of patients with cancer using the deep learning technique.

Methods: Images of faces of patients with cancer were collected to build the cancer face image data set. A face image data set of people without cancer was built by randomly selecting images from the publicly available MegaAge data set according to the sex and age distribution of the cancer face image data set. Each face image was preprocessed to obtain an upright centered face chip, following which the background was filtered out to exclude the effects of nonrelative factors. A residual neural network was constructed to classify cancer and noncancer cases. Transfer learning, minibatches, few epochs, L2 regulation, and random dropout training strategies were used to prevent overfitting. Moreover, guided gradient-weighted class activation mapping was used to reveal the relevant features.

Results: A total of 8124 face images of patients with cancer (men: n=3851, 47.4%; women: n=4273, 52.6%) were collected from January 2018 to January 2019. The ages of the patients ranged from 1 year to 70 years (median age 52 years). The average faces of both male and female patients with cancer displayed more obvious facial adiposity than the average faces of people without cancer, which was supported by a landmark comparison. When testing the data set, the training process was terminated after 5 epochs. The area under the receiver operating characteristic curve was 0.94, and the accuracy rate was 0.82. The main relative feature of cancer cases was facial skin, while the relative features of noncancer cases were extracted from the complementary face region.

Conclusions: In this study, we built a face data set of patients with cancer and constructed a deep learning model to classify the faces of people with and those without cancer. We found that facial skin and adiposity were closely related to the presence of cancer.

(J Med Internet Res 2020;22(4):e17234) doi:10.2196/17234

KEYWORDS

convolutional neural network; facial features; cancer patient; deep learning; cancer

Introduction

In daily social activities, people tend to judge the health status of others based on their facial appearance. Facial appearance is associated with other social judgments, such as leadership ability and attractiveness. These judgments may affect real-life outcomes such as employment and dating outcomes. Numerous studies have indicated that healthy people tend to be perceived as more attractive [1-4]. This may be explained by Darwin's theory of evolution: This preference has evolved because these "attractive" traits provide signals of biological quality, particularly physical health, and keeping away from individuals who are perceived to be less attractive (unhealthy) may prevent the contraction of a contagious disease.

Most published research focuses on the association of health status with perceived features such as skin color, facial adiposity, and symmetry. Skin color is malleable and may change rapidly in response to health status. Thus, it provides more relevant information regarding an individual's current physical condition [5-10]. Facial adiposity is positively correlated with BMI and body fat percentage, and it conveys important information about one's physical condition [1,11]. Facial symmetry is considered to be an indicator of health because it corresponds with the ability to maintain developmental stability and resist asymmetric growth (eg, pathogens or mutation rate) [4,12-15]. These studies attempt to reveal the association of perceptual facial features with physical condition, which is helpful in understanding the underlying psychological and biological reasons for an individual's behavior and social activities.

The recently developed deep learning (DL) technique [16] provides a novel method to study the correlation of facial features with health status. The facial features are extracted via a large-scale cascaded conventional neural network (CNN), which is more subjective than perceptual features. Gurovich et al [17] adopted the DL technique to identify facial phenotypes of genetic disorders. Kong et al [18] reported detection of acromegaly using the DL technique based on facial images; they achieved satisfactory accuracy and sensitivity. Wang et al [19] found that the DL technique outperformed humans at detecting sexual orientation based on facial images.

Cancer caused approximately 9.6 million deaths worldwide in 2018 [20]. It has become the second leading cause of death globally. Approximately 90%-95% of cancer cases are due to genetic mutations caused by environmental (radiation, pollution, etc.) and lifestyle (tobacco, alcohol, etc.) factors, while the remaining cases are due to inherited genetics. In either case, genetic mutation may result in facial features that are different from those of people without cancer. In this study, we aimed to reveal the facial features of patients with cancer using DL techniques and to use these features to distinguish the faces of people with cancer from those of people without cancer.

Methods

In this section, the face image data sets for people with and those without cancer are briefly introduced. Subsequently, the preprocessing procedure and network architecture are presented in detail. Finally, the transfer learning strategy and training details are described.

Ethics Statement

This study was carried out in accordance with the Declaration of Helsinki and approved and exempted from requiring informed consent by the Independent Ethics Committee of the Cancer Hospital of the Chinese Academy of Medical Sciences.

Image Data Sets

Images of the faces of patients with cancer at our institution before radiotherapy were collected over 1 year to build the cancer face image data set. The ages of the patients with cancer ranged from 1 year to 70 years, and the median age was 52. The histogram of the age distribution is shown in Table 1. The noncancer face image data set was built by randomly sampling the MegaAge data set [21]. The MegaAge data set consists of 41,941 faces of Asian people with annotated ages. Considering data balance, 8124 face images with the same sex ratio and age distribution as the cancer face data set were randomly selected from the MegaAge data set.

Table 1. Age distribution of the patients in the cancer face data set.

Age category	Patients	
	Male, n (%)	Female, n (%)
<10 years	31 (0.80)	28 (0.66)
10-20 years	41 (1.06)	23 (0.54)
20-30 years	81 (2.10)	163 (3.81)
30-40 years	258 (6.70)	738 (17.27)
40-50 years	689 (17.89)	1542 (36.09)
50-60 years	1558 (40.46)	1375 (32.18)
60-70 years	1193 (30.98)	404 (9.45)

Preprocessing

To exclude the influence of nonrelevant factors such as face pose, body profile, and image background, only upright, centered, and frontal face "chip" images were used as inputs of the convolutional neutral network (CNN). The preprocessing workflow and the intermediate results are shown in Figure 1 [22]. The human face in each image was first detected based

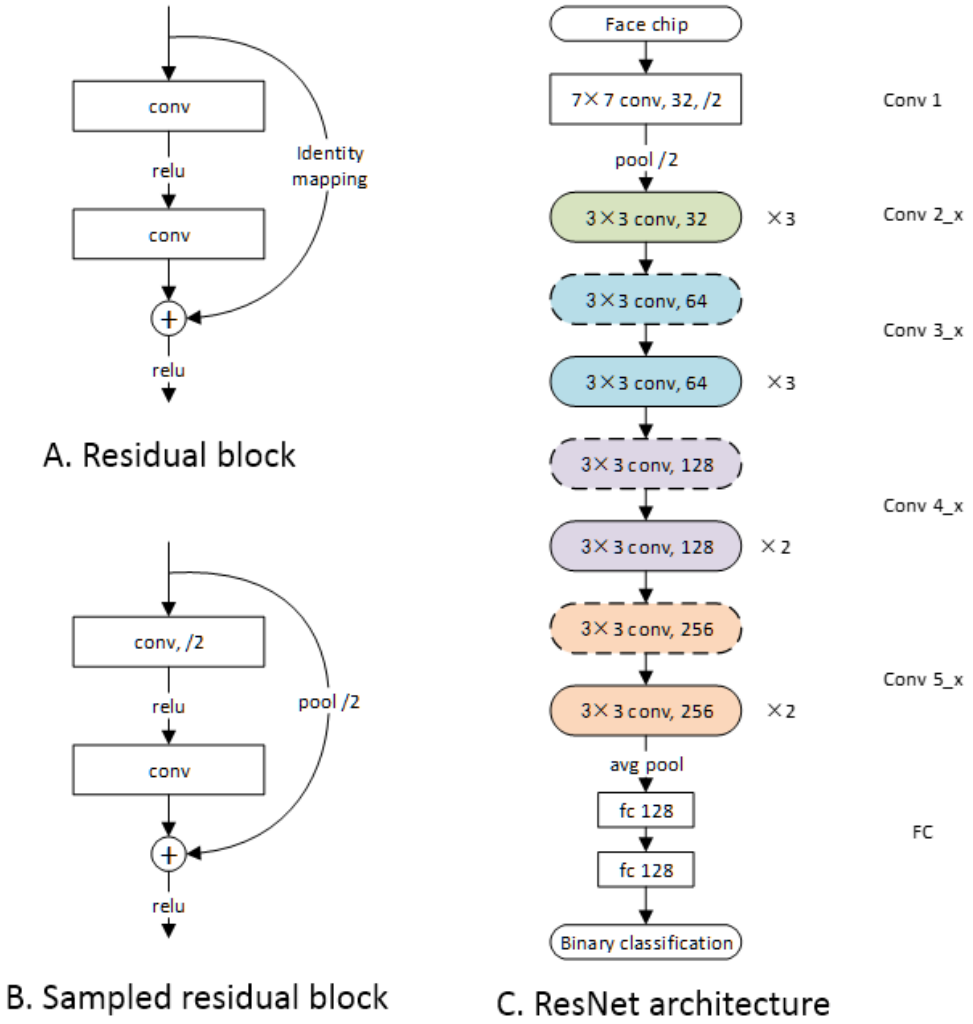
on the histogram of oriented gradients (HOG) [23]. The positions of 68 face landmarks such as the corners of the mouth and the centers of the eyes were then detected [24]. Using these landmarks, the face pose was determined. The image was rotated and aligned accordingly to obtain an upright, frontal, centered face image. The aligned image was cropped to obtain the face chip (ie, only the face part was retained and any other parts of the image were removed). Finally, the background in the resulting chip was further filtered out. The human face in the image was segmented using a fully convolutional network (FCN) [25,26], and the background was masked.

Figure 1. Pre-processing workflow to obtain chip images to use as inputs in the convolutional neutral network, showing the intermediate results. Image courtesy Barack Obama Presidential Library [22].

Network Architecture

In this work, a residual neural network (ResNet) was used to extract facial features. ResNet was proposed in 2015 by He et al [27]. Because a CNN is trained based on the gradient of the loss function, the gradient of the cascaded network becomes smaller as the network deepens. This issue was solved by adding identity shortcut connections to skip the convolutional layers in between, as shown in Figure 2A and B. The network of this study contains 27 convolutional layers and 2 full connection (FC) layers. The architecture is shown in Figure 2C, and the number and sizes of the filters in each layer are also annotated. Essentially, the model is a simplified version of ResNet-34 implemented proposed in the paper by He et al [27], with 7 layers removed and the number of filters in each layer reduced by half.

Figure 2. Network architecture. A) The basic cell of ResNet: the residual block. B) The sampled residual block. The difference is that the convolution of the first layer is performed with a 2-pixel stride. Correspondingly, 2-pixel stride sampling was performed before identity mapping. C) The ResNet architecture. The blocks shown in A) and B) are annotated with solid and dashed round-ended boxes, respectively.



Transfer Learning and Training Strategy

For the convolution layers (layers 1-5), the weights of a pretrained face encoding network [28] were directly utilized. The face-encoding network contains the same convolution layers but only one FC layer. With this structure, the features extracted by the convolution layers are flattened into a one-dimensional vector by the FC layer. In other words, the face image is encoded into a one-dimensional vector. The Euclidean distance between two vectors directly measures the similarity of the two corresponding faces. With the triplet loss function, the encoded vectors of the face images of the same individual are mapped close to each other in the Euclidean space, while the vectors of different individuals are mapped far from each other. Through this process, the encoding network captures the features that are unique to an individual.

This study aimed to determine what the faces of patients with cancer have in common from these individually distinguishable features. For our model, the weights of the convolution layers were fixed, and only the weights of the FC layers were trained for classification. This is because the weights of the convolution layers were trained on a combined data set with about 3 million faces [29,30], which is several orders of magnitude larger than the size of our data set. Addition of our data set would not introduce any substantial improvements.

In the data set, 50% of the face chips were randomly selected for training and 50% were selected for testing. To reduce the risk of overfitting, the following strategies were adopted: The minibatch size was set to 10, only 5 epochs were used, an L2 regulation penalty was added to the weights of the FC layers, and a random dropout was added on the FC1 layer. Both the learning rate and weight of the L2 regulation terms were set to 0.001. The dropout rate was set to 0.25; therefore, 25% of the extracted features were randomly dropped. The guided gradient-weighted class activation mapping (grad-CAM) method was utilized in the trained model to illustrate the discriminative features [31] of the cancer and non-cancer data sets. The guided grad-CAM method is the Hadamard production of global backpropagation and CAM. The global backpropagation is the gradient of the network, and the CAM is the activated map of each class.

This study was implemented with Python programming language. All data preprocessing procedures except FCN face segmentation were implemented using the Dlib package (v19.17.0). FCN face segmentation and ResNet were implemented using TensorFlow (v1.4.0) [32] and Keras (v2.2.4).

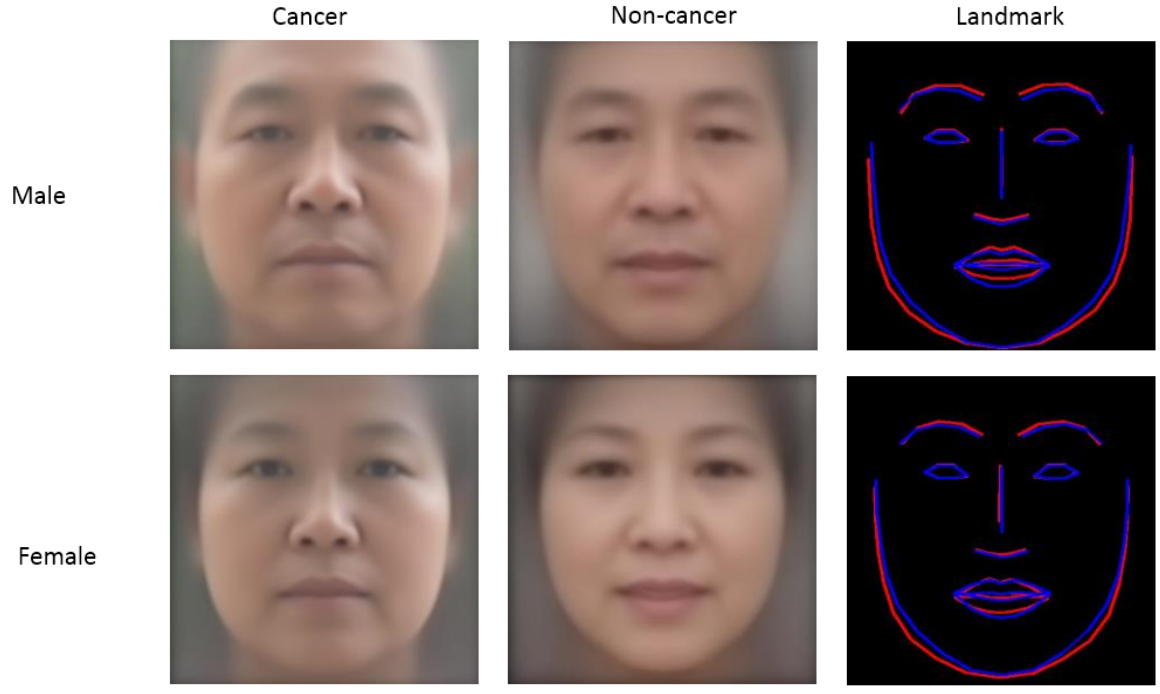
Results

We collected images of the faces of 8124 patients with cancer before radiotherapy from January 2018 to January 2019, including 3851 (47.4%) men and 4273 (52.6%) women. We compared the average faces of the cancer and noncancer data sets and presented the training and validation results; finally, we showed the distinguishable facial features of people with and those without cancer, captured by the grad-CAM method.

Average Faces

The intermediate results of the preprocessing procedure are shown in Figure 1. The average faces are shown in Figure 3. The facial landmarks of the average cancer face and the average noncancer face were detected and are depicted in the same figure for comparison. Intuitively, both the male and female average cancer faces display more obvious facial adiposity than the average noncancer faces, which was supported by the landmark comparison.

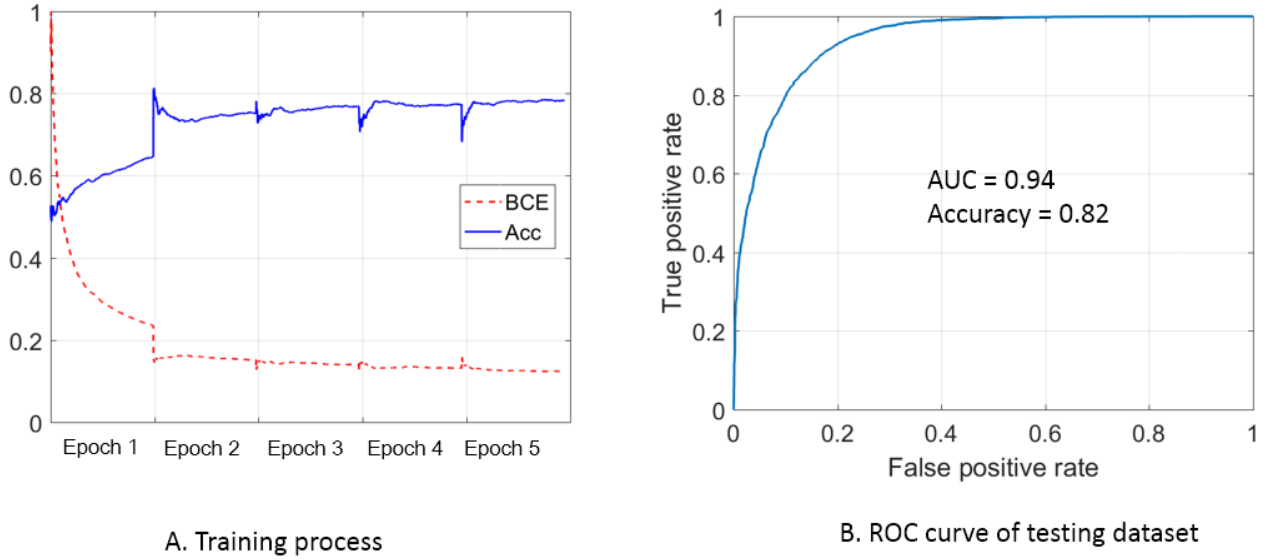
Figure 3. Average faces and landmark comparison. Top row: male, bottom row: female. First column: cancer, second column: noncancer, and third column: landmark comparison (from left to right). In the third column, the landmark of the average cancer face is depicted in red and that of the average noncancer face is depicted in blue.



Training and Testing Results

The binary cross-entropy (BCE) and accuracy during the training process are shown in Figure 4A. The fluctuation at the beginning of each epoch is due to the random shuffle of the training batch. For both BCE and accuracy, the most significant improvement was observed in epoch 1, and the improvement decreased with each epoch. No obvious improvement was observed between epoch 4 and epoch 5. The training process was terminated after 5 epochs to reduce the risk of overfitting. The receiver operating characteristic (ROC) curve of the testing data set is shown in Figure 4B. The accuracy value is 0.82, which is consistent with the training results (approximately 0.80). The area under the curve (AUC) is 0.94. The ROC curve is calculated with various "passing thresholds," and the corresponding AUC measures the discriminative ability of the model. The accuracy evaluates the discriminative ability with the "passing threshold" fixed to 0.5.

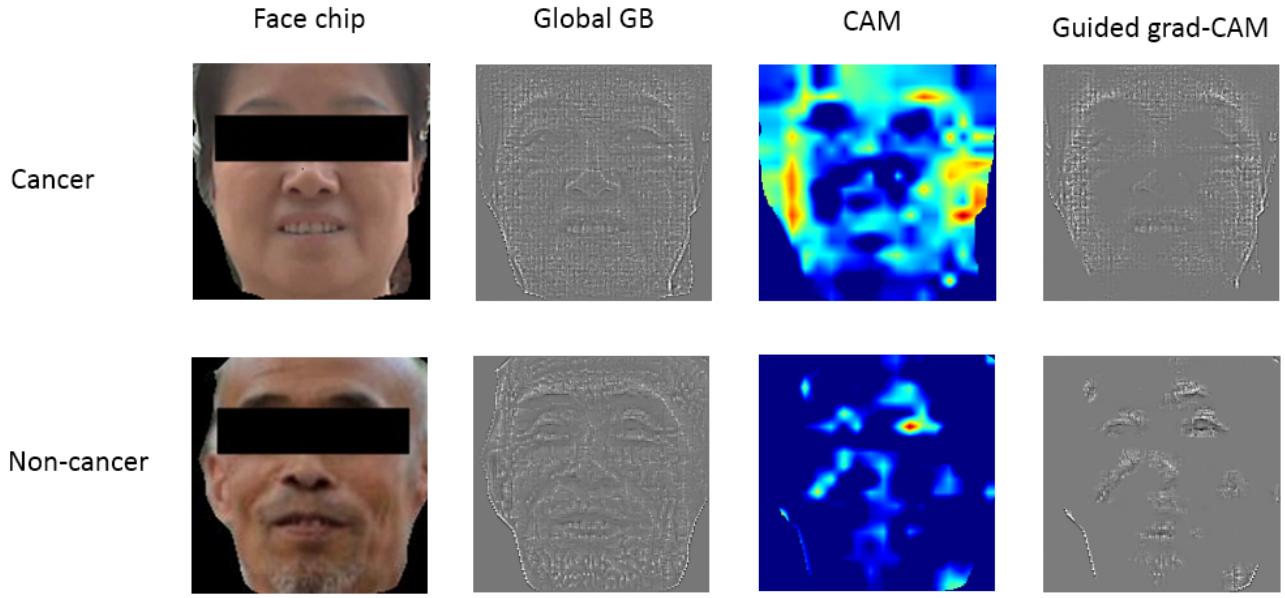
Figure 4. Training and testing results: (A) Binary cross-entropy (BCE) and accuracy (Acc) during the training process. The loss value is normalized for clarity. (B) Receiver operating characteristic (ROC) curve of the testing dataset (AUC: area under the curve).



Guided Grad-CAM

Figure 5 shows the results of the guided grad-CAM method for a typical cancer case and a typical noncancer case. The global backpropagation can be interpreted as the features extracted by the convolutional network. As shown in Figure 5, the global backpropagations of the two cases show the same pattern. This is because during the convolution process, "complete" features were extracted and the relative features of each class were activated by the following FC layers. The relative features of each class are shown by the guided grad-CAM, which is the entry-wise product of the global backpropagation and CAM. The main relative feature of people with cancer is the facial skin, that is, the facial region except for the eyes, eyebrows, lips, and nose. In contrast, the relative features of patients without cancer only include features extracted from the eyes, lips, and end-of-nose regions. This finding supports the findings from the average faces, where the average cancer face showed more obvious facial adiposity.

Figure 5. Grad-CAM analysis results for face images of people with and without cancer.



Discussion

In this study, we built a face data set of patients with cancer and constructed a deep learning model to classify the faces of people with and those without cancer. We found that facial skin and facial adiposity were closely related to the presence of cancer.

The AUC and accuracy results indicate that the network can discriminate the faces of patients with cancer from those of people without cancer. External validation must be performed to improve the presented network. Although facial features are closely related to an individual's health status, it is not solid to predict cancer incidence based on facial features. The motivation of this study was to find the facial features of cancer patients. The MegaAge data set was used as a noncancer reference data set. The possibility of people with cancer in the MegaAge data set cannot be excluded. However, the incidence of cancer cases in China is only 0.3%-0.4% [33], which would not affect the results.

We mainly attribute the satisfactory AUC and accuracy results to the transfer learning strategy. The parameters of the deep network were pretrained well with a large data set. We would like to point out that the convolution network was designed and pretrained for face image–based individual recognition, and only the FC layers were trained for classification in this study. The results demonstrate that cancer-discriminative facial features were included in the individual-discriminative features. The preprocessing procedure, which excludes the effects of factors other than the face chip, is another reason for the satisfactory results.

This study reveals that facial skin and adiposity are closely related to the presence of cancer. It has been proven that facial skin and adiposity are associated with health status. Among all facial features, these two respond most rapidly to the metabolic status. This finding can possibly be explained by the fact that malignant tumors affect metabolism and thus further affect facial skin and adiposity.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (11875320, 81801799, and 81502649).

Authors' Contributions

LB and YN conceived the project and wrote the paper; HP and YY collected and analyzed the data. All authors edited the manuscript.

Conflicts of Interest

None declared.

References

1. Henderson AJ, Holzleitner IJ, Talamas SN, Perrett DI. Perception of health from facial cues. Phil Trans R Soc B 2016 May 05;371(1693):20150380. [doi: 10.1098/rstb.2015.0380]

2. Coetzee V, Re D, Perrett DI, Tiddeman BP, Xiao D. Judging the health and attractiveness of female faces: is the most attractive level of facial adiposity also considered the healthiest? Body Image 2011 Mar;8(2):190-193. [doi: 10.1016/j.bodyim.2010.12.003] [Medline: 21354874]
3. de Jager S, Coetzee N, Coetzee V. Facial Adiposity, Attractiveness, and Health: A Review. Front Psychol 2018;9:2562 [FREE Full text] [doi: 10.3389/fpsyg.2018.02562] [Medline: 30622491]
4. Jones B, Little A, Penton-Voak I, Tiddeman B, Burt D, Perrett D. Facial symmetry and judgements of apparent health. Evol Hum Behav 2001 Nov;22(6):417-429. [doi: 10.1016/s1090-5138(01)00083-6]
5. Tan KW, Tiddeman B, Stephen ID. Skin texture and colour predict perceived health in Asian faces. Evol Hum Behav 2018 May;39(3):320-335. [doi: 10.1016/j.evolhumbehav.2018.02.003]
6. Stephen ID, Hiew V, Coetzee V, Tiddeman BP, Perrett DI. Facial Shape Analysis Identifies Valid Cues to Aspects of Physiological Health in Caucasian, Asian, and African Populations. Front Psychol 2017;8:1883 [FREE Full text] [doi: 10.3389/fpsyg.2017.01883] [Medline: 29163270]
7. Jones AL, Porcheron A, Sweda JR, Morizot F, Russell R. Coloration in different areas of facial skin is a cue to health: The role of cheek redness and periorbital luminance in health perception. Body Image 2016 Jun;17:57-66. [doi: 10.1016/j.bodyim.2016.02.001] [Medline: 26967010]
8. Tan KW, Graf BA, Mitra SR, Stephen ID. Daily Consumption of a Fruit and Vegetable Smoothie Alters Facial Skin Color. PLoS One 2015;10(7):e0133445 [FREE Full text] [doi: 10.1371/journal.pone.0133445] [Medline: 26186449]
9. Whitehead RD, Ozakinci G, Perrett DI. A randomized controlled trial of an appearance-based dietary intervention. Health Psychol 2014 Jan;33(1):99-102. [doi: 10.1037/a0032322] [Medline: 23527517]
10. Whitehead RD, Re D, Xiao D, Ozakinci G, Perrett DI. You are what you eat: within-subject increases in fruit and vegetable consumption confer beneficial skin-color changes. PLoS One 2012;7(3):e32988 [FREE Full text] [doi: 10.1371/journal.pone.0032988] [Medline: 22412966]
11. Phalane KG, Tribe C, Steel HC, Cholo MC, Coetzee V. Facial appearance reveals immunity in African men. Sci Rep 2017 Aug 07;7(1):7443 [FREE Full text] [doi: 10.1038/s41598-017-08015-9] [Medline: 28785075]
12. Marcinkowska UM, Galbarczyk A, Jasienska G. La donna è mobile? Lack of cyclical shifts in facial symmetry, and facial and body masculinity preferences-A hormone based study. Psychoneuroendocrinology 2018 Feb;88:47-53. [doi: 10.1016/j.psyneuen.2017.11.007] [Medline: 29161637]
13. Jones AL. The influence of shape and colour cue classes on facial health perception. Evolution and Human Behavior 2018 Jan;39(1):19-29. [doi: 10.1016/j.evolhumbehav.2017.09.005]
14. Mogilski JK, Welling LLM. The Relative Importance of Sexual Dimorphism, Fluctuating Asymmetry, and Color Cues to Health during Evaluation of Potential Partners' Facial Photographs : A Conjoint Analysis Study. Hum Nat 2017 Mar;28(1):53-75. [doi: 10.1007/s12110-016-9277-4] [Medline: 27752965]
15. Rhodes G, Zebrowitz LA, Clark A, Kalick S, Hightower A, McKay R. Do facial averageness and symmetry signal health? Evol Hum Behav 2001 Jan;22(1):31-46. [doi: 10.1016/s1090-5138(00)00060-x]
16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
17. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med 2019 Jan;25(1):60-64. [doi: 10.1038/s41591-018-0279-0] [Medline: 30617323]
18. Kong X, Gong S, Su L, Howard N, Kong Y. Automatic Detection of Acromegaly From Facial Photographs Using Machine Learning Methods. EBioMedicine 2018 Jan;27:94-102 [FREE Full text] [doi: 10.1016/j.ebiom.2017.12.015] [Medline: 29269039]
19. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. J Pers Soc Psychol 2018 Feb;114(2):246-257. [doi: 10.1037/pspa0000098] [Medline: 29389215]
20. 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]
21. Zhang Y, Li L, Cheng L, Chen C. Quantifying Facial Age by Posterior of Age Comparisons. 2017 Sep 15 Presented at: British Machine Vision Conference; September 4-7, 2017; London, UK p. 108.1-108.12. [doi: 10.5244/C.31.108]
22. Barack Obama Presidential Library. URL: https://www.obamalibrary.gov/research/ordering-photos-videos [accessed 2020-03-24]
23. Dalal N, Triggs B. Histograms of oriented gradients for human detection. 2005 Jun 1 Presented at: IEEE conference on computer vision and pattern recognition; 20-26 June 2005; San Diego, US p. 886-893. [doi: 10.1109/cvpr.2005.177]
24. Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. 2014 Jun 1 Presented at: IEEE conference on computer vision and pattern recognition; June 24-27, 2014; Columbus, US p. 1867-1874. [doi: 10.1109/cvpr.2014.241]
25. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. 2105 Jun 1 Presented at: IEEE conference on computer vision and pattern recognition; Jun 7-12, 2015; Boston, US p. 3431-3440. [doi: 10.1109/cvpr.2015.7298965]
26. Shen X, Hertzmann A, Jia J, Paris S, Price B, Shechtman E, et al. Automatic Portrait Segmentation for Image Stylization. Comput Graph Forum 2016 May 27;35(2):93-102. [doi: 10.1111/cgf.12814]

27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Jun 1 Presented at: IEEE conference on computer vision and pattern recognition; June 26-July 1, 2016; Las Vegas, US p. 770-778. [doi: 10.1109/cvpr.2016.90]
28. Schroff F, Kalenichenko D, Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering. 2015 Jun 1 Presented at: IEEE conference on computer vision and pattern recognition; Jun 7-12, 2015; Boston, US p. 815-823. [doi: 10.1109/cvpr.2015.7298682]
29. Ng H, Winkler S. A data-driven approach to cleaning large face datasets. 2014 Oct 1 Presented at: IEEE International Conference on Image Processing; Oct 27-30, 2014; Paris, France p. 343-347. [doi: 10.1109/icip.2014.7025068]
30. Parkhi O, Vedaldi A, Zisserman A. Deep Face Recognition. 2015 Sep 1 Presented at: British Machine Vision Conference; September 7-10, 2015; Swansea, UK p. 41.1-41.12. [doi: 10.5244/C.29.41]
31. Selvaraju R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. 2017 Oct 1 Presented at: IEEE international conference on computer vision; October 22-29, 2017; Venice, Italy p. 618-626. [doi: 10.1109/iccv.2017.74]
32. Abadi M, Agarwal A, Barham P, Brevdo E, Zheng X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arxiv 2016 Mar 16:1603.04467 [FREE Full text]
33. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]

Abbreviations

AUC: area under the curve
BCE: binary cross-entropy
CAM: class activation mapping
CNN: conventional neural network
DL: deep learning
FC: full connection
FCN: fully convolutional network
grad-CAM: gradient-weighted class activation mapping
HOG: histogram of oriented gradients
ResNet: residual neural network
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 27.11.19; peer-reviewed by W Sun, R Wei; comments to author 06.02.20; revised version received 12.02.20; accepted 05.03.20; published 29.04.20.

Please cite as:
Liang B, Yang N, He G, Huang P, Yang Y
Identification of the Facial Features of Patients With Cancer: A Deep Learning–Based Pilot Study
J Med Internet Res 2020;22(4):e17234
URL: http://www.jmir.org/2020/4/e17234/
doi:10.2196/17234
PMID:32347802

Original Paper

The Kids Obesity Prevention Program: Cluster Randomized Controlled Trial to Evaluate a Serious Game for the Prevention and Treatment of Childhood Obesity

Isabelle Mack[1], PhD; Nadine Reiband[1,2], MSc; Carolin Etges[1], MD; Sabrina Eichhorn[1]; Norbert Schaeffeler[1], PhD; Guido Zurstiege[3], Prof Dr; Caterina Gawrilow[2], Prof Dr; Katja Weimer[4], PhD; Riyad Peeraully[5], PhD, MD; Martin Teufel[6], Prof Dr; Gunnar Blumenstock[7], MD, PhD; Katrin Elisabeth Giel[1], Prof Dr; Florian Junne[1], MD, MSc; Stephan Zipfel[1], Prof Dr

[1]Department of Psychosomatic Medicine and Psychotherapy, University Medical Hospital, Tübingen, Germany
[2]Department of School Psychology, University of Tübingen, Tübingen, Germany
[3]Department of Media Studies Tübingen, University of Tübingen, Tübingen, Germany
[4]Department of Psychosomatic Medicine and Psychotherapy, Ulm University Medical Center, Ulm, Germany
[5]Department of Paediatric Surgery, Nottingham Children's Hospital, Nottingham, United Kingdom
[6]LVR-Clinic for Psychosomatic Medicine and Psychotherapy, University of Duisburg-Essen, Essen, Germany
[7]Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Tübingen, Germany

Corresponding Author:
Isabelle Mack, PhD
Department of Psychosomatic Medicine and Psychotherapy
University Medical Hospital
Osianderstraße 5
Tübingen, 72076
Germany
Phone: 49 70712985614
Email: isabelle.mack@uni-tuebingen.de

Abstract

Background: Health games provide opportunities for the treatment and prevention of childhood obesity. We developed a motion-controlled serious game for children that addresses 3 core topics of nutrition, physical activity, and stress coping. It is the first serious game that extensively targets the dietary energy density principle (DED-P) in relation to nutrition. The game is intended to provide an additional educational component for the prevention and treatment of obesity in children.

Objective: The Kids Obesity Prevention study aimed to evaluate the newly developed game and to evaluate how well children are able to understand and apply the DED-P.

Methods: This cluster randomized controlled trial collected data from 82 primary school children aged 9 to 12 years and their parents at baseline (T0), at 2 weeks after study commencement (T1), and at the 4-week follow-up (T2). The dropout rate was 3.6%. The intervention group (IG) played the game within 2 weeks (2 sessions with different game modules). One part of the game involves selection of food with the lower energy density when presented with a pair of foods. This allows assessment of whether the children have understood the DED-P and whether they can apply it to unknown foods under time pressure. The control group (CG) received a brochure about the food pyramid concept and physical activity. The primary outcome was the gain in knowledge (nutrition and stress coping) and measured with a pretested questionnaire. The secondary outcomes were the maintenance of knowledge, application of the DED-P, feelings during game play, game acceptance, and behavioral measures (physical activity, media consumption, and dietary intake).

Results: The knowledge score ranging from 0 to 100 increased from T0 (IG: 53 [SD 10], CG: 50 [SD 11]) to T1 (IG: 69 [SD 11], CG: 52 [SD 12]) in IG versus CG (P<.001). At T2, the knowledge score of IG remained at the same level as that of T1. Game data showed that after DED-P education, the classification under time pressure of unknown versus known food pairs according to their DED category was similar (hit rate around 70%). Overall, 95% of the children liked the game very much or much. No group changes were observed at the behavioral level.

Conclusions: The Kids Obesity Prevention program sustainably increased knowledge in the areas of nutrition and stress coping, and children were able to apply the DED-P.

Trial Registration: ClinicalTrials.gov NCT02551978; https://clinicaltrials.gov/ct2/show/NCT02551978

(J Med Internet Res 2020;22(4):e15725) doi:10.2196/15725

KEYWORDS

children; serious game; nutrition; stress; energy density

Introduction

Obesity and Weight Management Principles

Obesity and its associated comorbidities such as type 2 diabetes and hypertension are a severe public health problem. Unfortunately, many school-aged children are already overweight or obese [1], and most of them will remain overweight and obese in adulthood [2]. Obese children and adolescents often differ in several physiological [3-6] and psychological parameters [6-8] when compared with their peers. Some children even suffer from comorbidities [6] and are victims of stigmatization [9,10]. Therefore, it is imperative to establish effective weight gain prevention [11] and treatment programs [12] during childhood.

Weight management principles are aimed at reducing energy intake by improving diet and eating behavior and at increasing energy expenditure by increasing physical activity and avoiding sedentary behaviors [13]. At the dietary level, portion size and the dietary energy density (DED) of foods [13,14] are critical factors. Knowledge of appropriate food portion sizes is important in view of the increase in portion sizes over recent decades [15]. In adults, a recent short-term study reported that it is possible to recalibrate the perception of food portion sizes. In this study, participants were served small food portions, which affected their perceptions of what constitutes a normal-sized serving and resulted in choosing to eat less food in the future [16]. The food pyramid or my plate concept has been applied for many years to guide portion sizes for the different food groups [17,18], although the flexibility of adapting to personal preferences is limited when using this concept.

Obese adults, adolescents, and children have increased gastric capacities and might need to eat larger portions to perceive satiety in comparison with normal-weight participants, which is problematic if foods high in DED are consumed [19-21]. Adults eat consistent food amounts over several days even if the DED is varied, and this may also be true for children aged 3 to 5 years [14]. Thus, consideration of the DED of foods may be an important factor for weight regulation.

The success of lifestyle interventions is critically dependent on psychosocial and psychological aspects, such as stress [22]. Learning how to cope with stress may be an important factor in weight regulation, as stress can be associated with an unfavorable lifestyle, including low physical activity and high energy intake by consuming more foods with high DED and probably fewer foods with low DED [23]. For interventions to be successful in childhood, family support and goal commitment are also necessary [22].

Digital Media

Digital media offer a ubiquitous and highly accepted tool, especially in younger age groups. In the United Kingdom, for instance, 3- to 7-year-old children watch TV for about 2 hours a day, whereas 8- to 15-year-old children spend 2.5 hours in the Web, watch TV for 2 hours a day, and play games for 2 hours a day [24]. Similar results have been reported in Germany [25]. Although screen time is considered a risk factor for childhood obesity [26], there appear to be differences between passive and active screen time, with active being less problematic [27]. For lifestyle interventions, the well-considered and appropriate use of digital media offers enormous advantages [28-30]. These digital media include serious games, which are games designed to fulfill a *serious* purpose by providing education from health professionals via a digital device [31]. Serious games can be applied to convey health-related information and can also use interactive components to reinforce the information and train behaviors, for example, by incorporating cognitive bias therapy into games [32,33].

In a recent systematic review, the abilities and limitations of video games, including exergames (focusing on increasing activity) and serious games, to combat and prevent childhood obesity were investigated. It was found that the children like playing the games, and on a qualitative level, most studies report positive effects on obesity-related outcomes (improvement of weight-related parameters, physical activity, or dietary behavior and knowledge). However, at the quantitative level, the observed effects were small [29]. When analyzing the teaching of concepts regarding nutrition and eating habits by games, most studies focus on the concepts of food pyramid and nutritional balance [34-43], fruit-vegetable or 5-a-day (aimed at increasing fruit and vegetables to 5 portions per day [44-47]), or the Mediterranean diet [48]. Intake of dietary fat, excess sugar [42,49], liquids, packed snacks, and fast food [39,47,48] have been addressed in other studies. None of the games investigated have addressed the DED-P in detail or considered psychosocial or psychological aspects of nutrition and eating habits, such as stress coping [29].

Objective and Hypothesis

We developed a motion-controlled serious game for children, the Kids Obesity Prevention (KOP) program, addressing the areas of nutrition, physical activity, and stress coping. In addition to a motion control interface, a tablet is used for knowledge-based and cognitive tasks. Unlike in most other games, the nutritional episodes in the game presented here relate to the nutrition pyramid and, beyond that, to the DED of food and drink, as well as stress and stress-coping strategies, such as relaxation exercises. The game was developed to be applied

in obesity prevention and as an additional educational component of inpatient and outpatient settings for obesity treatment in children.

The aim of the KOP study was to evaluate the game in a cluster randomized controlled trial with two parallel groups in a primary school setting in children aged 9 to 12 years. The primary outcome was the gain in knowledge about important lifestyle factors with the focus on nutrition, especially the DED-P. The secondary outcomes were the maintenance of knowledge, the application of the DED-P, game acceptance, dietary habits, physical activity, and media consumption. Our primary hypothesis was that after intervention, (1) the intervention group (IG) would have higher knowledge scores compared with the control group (CG). We further hypothesized that (2) after the 4-week follow-up, the knowledge scores of IG would be higher than that at baseline but lower than the scores immediately after game play; (3) children would understand and apply the DED-P; (4) children would enjoy playing the game, and game acceptance would be high; and (5) no changes at the behavioral level would be observed because of the primary study aim and design.

Methods

Study Design and Participants

The KOP study was a cluster randomized controlled trial involving school children aged 9 to 12 years; it was conducted from September to November 2015 in Germany. The study period was 2 weeks, and the follow-up was conducted 4 weeks after study termination. The inclusion criterion was an age between 9 and 12 years (all children who were in the fourth grade of primary school). The exclusion criterion was children with major linguistic difficulties.

At baseline, participants were assessed for knowledge about a healthy lifestyle, with a focus on nutrition issues, weight and height, physical activity, and dietary behavior. Parents provided information about their child regarding physical activity, dietary behavior, and media consumption.

The detailed study procedure for the intervention program was set up before participant enrollment in arrangement with the teachers. The recruitment of children took place at a parents' evening of the school classes. Parents and children provided written informed consent. The study protocol was approved by the ethics committee of the medical faculty of the Eberhard Karls University Tübingen and the University Hospital Tübingen, Germany (050/2014BO1). The KOP-1 study was registered before the study commencement at ClinicalTrials.gov (NCT02551978).

Randomization and Masking

Five fourth grade classes in a single school were randomly allocated to an IG or a CG. To ensure balanced age and sex distributions and to avoid interactions between children in different groups within one class, classes were randomized en bloc to one of the groups by a coin toss. The randomization took place before study commencement. Owing to the nature of the intervention, participants and outcome assessors could not be blinded to treatment assignments.

Treatments

Intervention

The IG played the game twice over a 2-week period, with a different selection of game modules played at each of the two sessions (Section Game Modules and Multimedia Appendix 1). Each session took about 45 min. An investigator was present during the whole time. The game [50] starts with a framing story of a child living in a medieval fantasy world presented by a storyteller showing pictures in an ancient book on a screen. The story is about the competition between two villages to regain knowledge about nutrition and a healthy lifestyle. Competitive elements of the story and elements where children had supportive function of the character aimed at motivating the children to engage in the game. After the presentation of the story in a video, the player has to move an avatar in a 3D medieval world to walk from the site of one task to another. The player's physical movement is necessary to make the avatar move in the game. Therefore, the player walks on the spot, lifting the knees upward (gaming PC: ARLT Silent Gamer GTX 650 [Windows 7], Intel Core i3 3240 [2× 3.4 GHz], 8 GB RAM, 2000 GB hard disk drive, NVIDIA GeForce GTX 650; Kamera Xbox One Microsoft Kinect Sensor; and BenQ MW851UST DLP-Beamer). When the avatar reaches certain locations for tasks to be completed, a switch to touch screens occurs (Acer Iconia Tab W510, NT.L0KEG.001). Depending on the gamer, the locations are reached within 1 to 2 min. The topics addressed by the game belong to the categories of nutrition, physical activity, and stress coping. The nutrition segment of the game deals with the food pyramid and the sugar content of liquids, and focuses on factors that are important for satiety, therefore teaching the concept of DED-P extensively. A self-reflective diagnostic tool to analyze daily food intake is also offered. Other aspects including eustress, distress, and stress-coping strategies are addressed by relaxation exercises and a tool for reflecting and planning everyday activities. An overview and description of the game modules is given below in the section Game Modules, screenshots are provided as Multimedia Appendix 2, and a video is provided as Multimedia Appendix 3.

Control Condition

The CG received basic information about a healthy lifestyle via a brochure from the informational service of the Federal Ministry of Food and Agriculture entitled "So macht Essen Spaß (How to make food fun)" and was handed out to the children at the beginning of the study phase. The CG received the intervention of the IG after study completion because of ethical considerations (the ethics committee wished the CG to play the game too).

Game Modules

Pack Your Backpack With Food

Both game sessions start with this self-reflective diagnostic tool to analyze the individual daily food intake. In this task, the children must pack as many foods into a backpack as they think they need to consume in a normal day to be energy balanced. Before the children start the game, the sex, age, height, weight, and activity level (high, medium, and low) are requested to calculate (1) BMI z-score, (2) energy expenditure, and (3)

energy intake by the program. In this study, the children's basic information is entered in advance by the investigators. A medium activity level is assumed for all children. BMI is calculated according to the standard formula: BMI=weight in kilogram/height in m^2 and is subsequently transformed into the sex- and age-specific BMI z-scores according to German reference values [51]. Energy intake requirements are calculated based on the recommendations of the Research Institute of Child Nutrition, Dortmund, Germany, which, in turn, are based on the Food-Based Dietary Guidelines in Europe [52]. Individual energy expenditure is calculated according to the document *Human energy requirements*, a report of a joint Food and Agriculture Organization and World Health Organization and United Nations Organization expert consultation [53]. For children with a BMI z-score of less than −0.4, the program sets the energy requirements to the values calculated for children of the same age and sex and a BMI z-score of −0.4. Similarly, for children with a BMI z-score greater than 0.4, the program sets the energy requirements to the values calculated for children of the same age and sex and a BMI z-score of 0.4. Setting these cutoffs, we assure that the children do not pack portions that are connected with undereating and overeating. In the task, the children can choose from a large variety of foods from the food groups for breakfast, lunch, dinner, and two mid-meals. They can also choose to skip meals or mid-meals. The foods are presented as pictures for which the exact amount and energy contents and their assignment to the appropriate food group are entered in a database. Finally, individual visual and verbal feedback is given about the child's energy balance and the distribution of the selected foods according to the food groups. The feedback is provided in a positive way. Besides the *food pyramid* recommendations [54,55], other state-of-the-art information retrieved from large epidemiological cohort studies, for example, the European Prospective Investigation into Cancer and Nutrition study [56] and from the World Cancer Research Fund International is included in this tool, for example, if a child exceeds the amount of food in the meat plus fish plus egg group, the recommendation will be to focus on fish and poultry and to avoid red meat.

Balloon Game

This game deals with the food pyramid and its food groups. At the beginning of the balloon game, the children are introduced to the food pyramid and nutrition circle [54,55]. To make the food groups clear, we invented animals representing the food groups: (1) fruit and vegetable group with a monkey with a banana, (2) grain food group with a hamster with an ear of corn, (3) dairy group with a cow with a milk bottle, (4) meat and fish group with a cat with a fish, (5) sweets and fats group with a bear with honey, and (6) liquids with a dolphin with a bottle of water. Next, a fun game starts where foods from the five food groups (without liquids) hanging from a balloon fly across the screen. On the bottom are five boxes labeled with the pictures of the food-group animals. The children then have to touch and pop the balloons to make the foods fall into the appropriate box. If the food falls into the wrong box, the food jumps out again and will appear with a balloon again. The game is over when all foods are in their correct food group box. In this module, we do not include the *dolphin group*. This last group is introduced in the module *Liquid rankings on the sugar scale*.

Foods Under the Microscope

In the next part of the game, the children learn about the main factor that induces satiety, which is the volume of food [57,58]. Next, the dietary energy density (DED) concept [59,60] is introduced, which organizes foods into 3 groups: (1) green foods with a DED up to 1.5 kcal/g, (2) yellow foods with a DED of 1.5 to 2.5 kcal/g, and (3) red foods with a DED above 2.5 kcal/g. According to the DED concept, green foods can be eaten in relatively large amounts and help to induce satiety, whereas red foods should be only eaten in small amounts and are mainly for enjoyment. Next, on the bottom of the screen, five box-like rectangles appear, labeled with the food-group animals. After clicking on one of these rectangles, five selected foods from the specific food group appear. The player can freely choose the foods to investigate them *under the microscope*. For every food, the proportion of water, fat, carbohydrates, protein, and fiber are displayed visually in test tubes in addition to the DED color information. The child also has the opportunity to obtain the information via additional audio texts. All foods must be investigated before the player is allowed to proceed. Next, a similar setup is used, but this time the child has to decide which food belongs to which DED color while the information about the test tubes is provided. In this game activity, the children collect points. They receive 3 points if the answer is correct the first time, 2 points if the answer is correct the second time, and 1 point if the answer is finally correct. After completion of every food group, a summary about the specific foods is given as audio text.

Liquid Rankings on the Sugar Scale

This game deals with the questions of (1) which liquids should be generally used for water intake (ie, water or sugar-free herbal and fruit teas), (2) how much liquid should be consumed daily (ie, as much as necessary to keep the urine transparent to very light yellow), (3) how much sugar is in liquids, (4) the impact of liquids on satiety [58,61]), and (5) how much exercise has to be done to burn the calories of 1 L coke or fruit juice. This game starts with the food pyramid concept, now introducing the dolphin group. Next, the game starts where a horizontal number line of lumps of sugar is provided on the top of the screen. On the left side of the screen, the following 1 L bottles of liquids are provided: water, sugar-free herbal tea, orange juice, orange lemonade, coke, apple spritzer, and ice tea. Each liquid appears in the middle of the screen when it is selected. Next, the player has to decide how many sugar lumps are *hidden* in the selected liquid. The player has three chances and receives verbal feedback about misestimation after every trial, for example, "there is much more sugar in this liquid," "there is a little bit more sugar in the liquid," "there is less sugar in the liquid," and "the answer is correct." After the third try (or before if the estimation was correct), the selected liquid is placed alongside its appropriate place on the number line of lumps of sugar. At the end of the game, the player has an overview of the sugar content in the liquids in relation to each other.

Kangaroo-Turtle Race

In this game, the player has to apply the knowledge about DED without having much time to think. A kangaroo (the computer) races against a turtle (the player). The kangaroo hops continuously, but the turtle is only fast if the rocket on its back is activated. In the middle of the screen, pairs of foods or liquids are presented in random order, with one food or liquid always having a higher DED than the other. The task is to choose the food or liquid with the lower DED. If the answer is correct, the rocket on the turtle is activated, and if the answer is false, a penalty of 3 seconds occurs. Some food or liquid pairs are already known from the previous games, and others are new to test the *transfer* of the DED concept to unknown foods. The winner is whoever passes the finishing line first. This is a fast game and is played three times in succession. Many food and liquid pairs are provided for this game so that during one game run, not all pairs will be displayed.

Bursting Bubble Game

This game provides information about eustress, distress, and coping strategies. The children then have the chance to burst bubbles on their tablet. The instruction is to breathe in slowly and breathe out slowly and loudly. The longer they breathe out, the larger the bubble will be. The children can burst the bubbles by tapping their finger on them. The basis for the functioning of the game is the microphone of the tablet. There should be no background noise in the room while the game is played.

Relaxation Story

This task begins with an introduction about eustress and distress and strategies to cope with stress. The idea of balancing activities in life is also introduced in this task. Next, the relaxation story starts. The children receive the instruction to sit conformably in a chair and to relax and simply listen. The story is about 2 min long.

Everyday Activities Task

The relaxation story is followed by a task where we measure the activities the children remember to do in their leisure time. The children are then motivated to reflect on their behavior. A set of typical activities is presented to the children (eg, playing soccer, riding a bike, and watching television). Next, they have the task to sort the activities they do into the categories “I do this every day,” “every week,” and “seldom.” The children also have the opportunity to add their own favorite activities. Finally, they are asked to plan their activities for the next week and to try to balance active and relaxing activities.

Outcome Measures

Measurements were taken at baseline (T0), at 2 weeks after study commencement (T1), and at the 4-week follow-up (T2). The primary outcome of the study was the gain in knowledge (eg, nutrition and stress coping) measured by a self-constructed, pilot-pretested questionnaire tailored specifically for the serious game and to be analyzed between IG and CG at T0 versus T1. In a pilot study with 17 children, the game and the questionnaire were evaluated and further developed and adapted. It consists of 13 questions. The answers for the items were transformed to scales ranging from 0 to 100. Subsequently, sum scores were calculated and divided by the number of items resulting again in values between 0 and 100. The total knowledge test score includes all items; the DED score includes the items 3, 5, 7, and 9; the food pyramid score includes the items 1, 2, and 4, and the stress score includes the items 10, 11, 12, and 13. The questionnaire is provided in Multimedia Appendix 4.

Secondary outcomes were the maintenance of knowledge at T2, perception of knowledge gain, acceptance of the game, game module data dealing with the DED-P to support the results of the knowledge questionnaire, emotions during game play [62], game data regarding changes of dietary behavior [63,64], physical activity [65], and media consumption [66] measured by validated questionnaires. A detailed overview of the applied measurements for the secondary outcomes is provided below.

Maintenance of knowledge was measured by applying the knowledge questionnaire at the 4-week follow-up after T1. At this time, the CG had also just completed the intervention in view of ethical considerations. Maintenance of knowledge of IG was compared with the knowledge of CG, which had newly acquired the knowledge.

Acceptance of the game (IG only at T1) was measured by the three items “Overall, I like the game,” “Playing the game was fun,” and “Playing the game does not make me feel bored” answered on a 4-point scale, and the 2 items “Would you recommend the game to a friend?” and “Would you also like to play the game at home?” answered on a 3-point scale.

Emotions during game play (IG only at T1) was measured using the validated self-assessment manikin [62] as a nonverbal pictorial assessment (corresponding to a 5-point scale), which is applied to measure pleasure, arousal, and dominance associated with the children’s affective reaction during the game. The items were (1) “When I play the game I feel good (1)—bad (5),” (2) “When I play the game I feel relaxed (1)—aroused or excited (5),” and (3) “When I play the game I feel tall and strong (1)—small and weak (5).”

Changes in dietary behavior [63,64] *(IG and CG at T0 and T2)* were measured using the “Ernährungsmusterindex” (an index for healthy nutrition) [63]. This index consists of food items, which are considered to be indicators for healthy and unhealthy eating behaviors. The indicator food items used for this index are vegetables and fruits (cooked, raw, frozen, and tinned), whole-grain bread, soft drinks, fast food, chocolate, and snacks such as crisps or pretzel sticks. The questions were taken from the corresponding validated Food Frequency Questionnaire (FFQ) [64]. The children completed the questionnaire by themselves, with the parents also completing the same questionnaire for their children.

Physical activity (IG and CG at T0 and T2) was measured by a validated questionnaire [65], which was completed by the children themselves and by the parents on behalf of their children. The questionnaire consists of seven items, each with six answer options. Finally, a score was calculated, allowing categorizing the activity level into low, medium, and high.

Media consumption (IG and CG at T0 and T2) was measured by questions from the German Health Interview and Examination Survey for Children and Adolescents (KIGGS) questionnaire [66]. Four questions seek information on the

average time the child spends watching television or playing video games or time spent on the computer during the week and at weekends. Each question has five response options.

Game data that assessed the application of the DED concept under time pressure (IG and CG, Kangaroo-Turtle race) were analyzed. The hit rate and reaction time for known foods were calculated using data from three consecutively conducted independent runs.

Game data of the Kangaroo-Turtle race (Section Game Modules and Multimedia Appendices 2-4), which tests the application of the DED-P under time pressure, were analyzed to support the results of the DED-P subscore from the knowledge questionnaire (IG and CG). The hit rate and reaction time for foods were calculated using data from three consecutively conducted independent runs.

Procedure of Measurements

Data were collected using standard operating procedures. All questionnaires were completed by all participating children simultaneously in a classroom setting with separated tables using paper and pencil. A teacher, trained school psychologist or medical student were present at all times during this process. No specific instructions were given to individual children. The parents completed the questionnaires at home using paper and pencil. The log data of the educational game were saved for additional analysis. The BMI z-score was calculated by calculating the BMI, on the basis of body weight measured by a calibrated scale and body height measured by a stadiometer and with reference to age- and sex-specific norms [51].

Sample Size

On the basis of pilot data, the sample size was calculated regarding the primary outcome, that is, the group (IG vs CG)×time (T0 vs T1) interaction of knowledge. A univariate two-group repeated measures analysis of variance (ANOVA) will have 80% power to detect a variance among the group marginal means of 1.266, will have 99% power to detect a variance of 1.891 among the means of the 2 levels, and will have 99% power to detect an interaction between groups and levels with a variance of 0.766, assuming that the between-group error term is 3.92, the within-group error term is 1.77, the measure of *sphericity* of the covariance matrix, epsilon, is 1.00 when the significance level is 0.05, and the sample size in each of the two groups is 25. The sample size was adjusted for cluster randomization by assuming an interclass correlation coefficient of 0.035 and a mean of 15 pupils per cluster, resulting in a variance inflation factor of 1.49. This results in a sample size of 37 per study arm [67,68]. We assumed an additional dropout rate of 5% because of illness or other specifics at the school setting. Therefore, the final sample size for each group should be 39, resulting in 78 participants overall.

Statistical Analysis

Data were analyzed using SPSS version 25 (IBM, Ehningen, Germany). Data are presented as means (standard deviation) and frequencies along with their percentages unless stated otherwise. Before test statistics, the normality distribution of data was tested using the Kolmogorov-Smirnov test, and the equality of variances between groups was tested using Levene test. Baseline differences between the groups were analyzed using unpaired *t* tests; if nonparametric, the differences were analyzed using the Mann-Whitney U test; or if nonmetric, the differences were analyzed using the chi-square test or the Fisher-Freeman-Halton test [69]. The Freeman-Halton test is an extension of Fisher exact test and was applied if the chi-square test was inappropriate because the frequencies in the cross-tables were less than 5 in a cell in greater than 20% of the cells.

The primary and secondary outcomes were calculated as unadjusted and adjusted estimates.

Unadjusted Model

The sample size was calculated as follows: the calculation of group (IG vs CG)×time (T0 vs T1 for knowledge and T0 vs T2 for other variables) interaction by 2×2 ANOVA. Different baseline levels for the DED score were controlled for using analysis of covariance (dependent variable: DED change score from T0 to T1, fixed factor: group, and covariable: DED score at T0). Nonmetric data were analyzed using the chi-square test or the Fisher-Freeman-Halton test (see the above specification) using the difference between the values at T0 and T2.

Adjusted Model Accounting for Cluster Randomization

To account for cluster randomization, an ANOVA with contrasts was performed using the school classes as factor. Planned contrasts were calculated not only for the IG versus CG school classes but also for a set of different class combinations to check the specificity and plausibility of observed contrast effects.

Further Analysis

The data of the Kangaroo-Turtle race was analyzed across both groups (IG and CG) after game play using the Wilcoxon signed-rank test because of nonparametric data distribution.

Handling of Missing Data of Single Items in Questionnaires at Baseline (Secondary Outcomes)

At baseline, all parents and children filled in questionnaires as secondary outcome variables, but there were instances of single questions of questionnaires not being answered. The predictive mean matching (PMM) method [70-72] with five cases in each match set was used for missing data (mis-d) imputation of the FFQ, parent version (mis-d: 1.7%) and child version (mis-d: 0.9%); the activity questionnaires, parent version (mis-d: 0.9%) and child version (mis-d: 1.2%); and questions about media consumption (mis-d: 5.8%).

Intention-to-Treat Analysis and Handling of Missing Data of Single Items in Questionnaires at the End of Study Period

We analyzed the primary and secondary outcomes by intention-to-treat (ITT) analysis using the PMM with five cases in each match set. Mis-d of single items in the questionnaires was imputed as described for the handling of missing data at baseline. The percentages of mis-d of the questionnaires were as follows: FFQ parent version 10.4% and child version 3.7%, activity questionnaire parent version 8.9% and child version 3.8%, and questions about media consumption 12.5%. No mis-d

imputation was performed for game acceptance where only descriptive data were reported.

Results

Participant Characteristics

The flow of the participants in the study is shown in Figure 1. Of 122 children approached for the trial, 82 participated. After allocation to groups, 3 children dropped out because of illness (dropout rate: 3.6%). In Table 1, the baseline characteristics of the 82 children are shown, which were included in the ITT analysis. Overall, the study groups did not differ in sex, age, BMI z-score, and media consumption reported by parents, physical activity level and healthy eating index reported by children and parents, and the primary outcome knowledge about a healthy lifestyle with the focus on nutrition (total score). Regarding the knowledge subscores, the food pyramid score and the stress score did not differ between the groups. The knowledge subscore DED was higher in IG than in CG (U=587, *P*=.02).

Outcomes

The data for the primary and secondary outcomes are presented in Tables 2 and 3. In addition, score changes from baseline are presented in Multimedia Appendix 5.

Figure 1. Consolidated Standards of Reporting Trials 2010 flow diagram for the Kids Obesity Prevention Study. ITT: intention-to-treat.

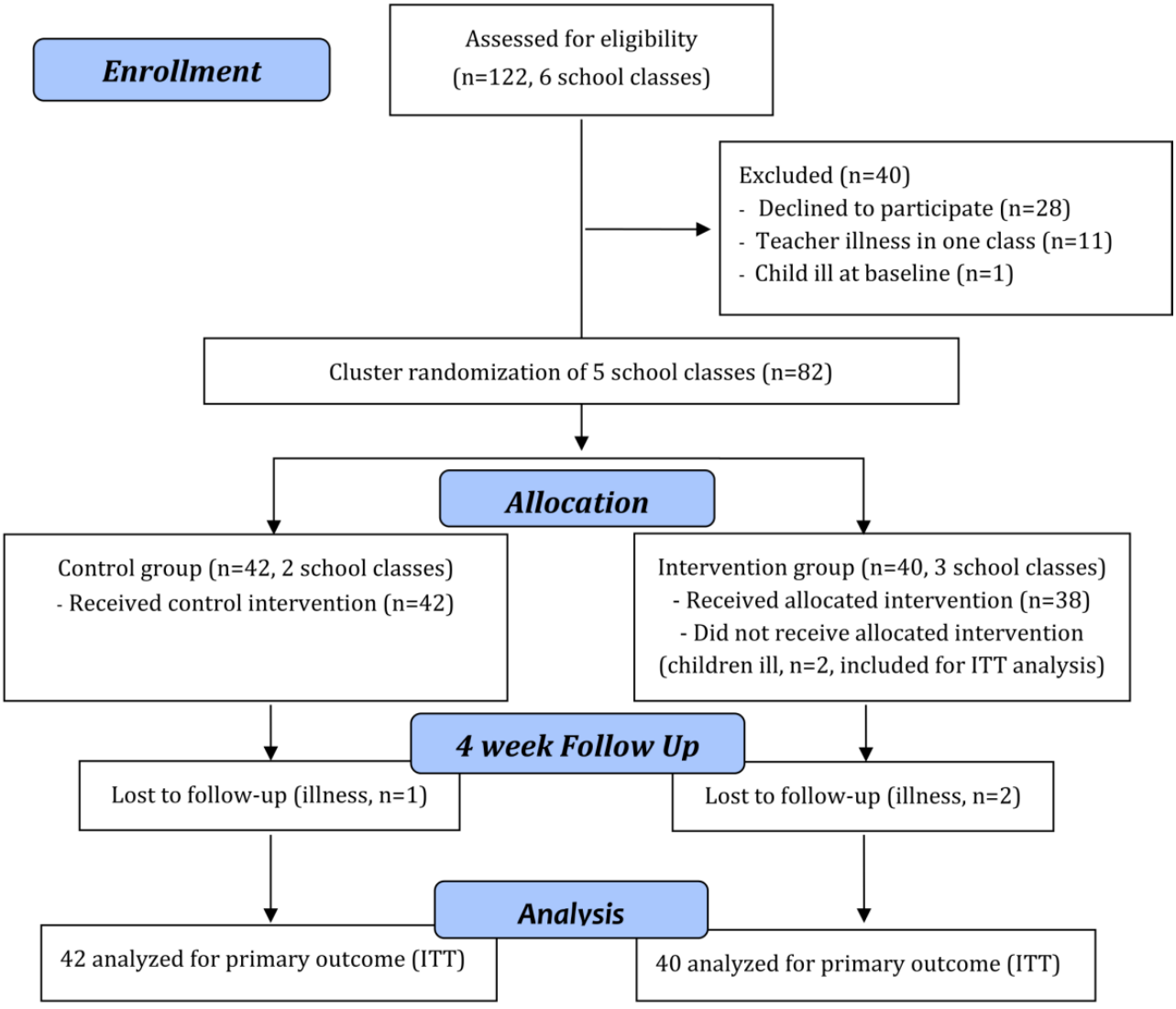


Table 1. Baseline characteristics of the study population.

Characteristics	Intervention group (n=40)	Control group (n=42)	*P* value
Age (years), mean (SD, range)	9.6 (4.9, 9-11)	9.7 (0.5, 9-11)	.38
Sex (n)			**.66**
Male	19	24	
Female	21	18	
Weight (kg), mean (SD, range)	37.3 (10.4, 24-65)	34.4 (9.2, 24-70)	.16
Height (cm), mean (SD, range)	143 (9, 129-166)	141 (6, 129-162)	.18
BMI z-score, mean (SD, range)	**0.21 (1.26, −2.44 to 2.65)**	**−0.02 (1.08, −2.07 to 2.57)**	**.34**
Normal weight, n (%)	27 (68)	35 (83)	
Thinness, n (%)	2 (5)	1 (2)	
Overweight, n (%)	7 (17)	4 (10)	
Obesity, n (%)	4 (10)	2 (5)	
Language spoken at home (n)			**.80**
German	17	23	
German and foreign language	16	14	
Foreign language	7	5	
Knowledge score, mean (SD)			
Percentage of total score (primary outcome)	53 (10)	50 (11)	.13
Percentage of food pyramid score	50 (13)	49 (14)	.70
Percentage of dietary energy density score	51 (18)	41 (19)	.02
Percentage of stress score	56 (17)	59 (20)	.52
Healthy Nutrition Index			
Reported by parents			
Score, mean (SD)	9.2 (2.0)	9.8 (2.1)	.91
Unfavorable, n (%)	1 (3)	2 (5)	
Neutral, n (%)	29 (72)	23 (55)	
Favorable, n (%)	10 (25)	17 (40)	
Reported by children			
Score, mean (SD)	8.8 (2.1)	8.9 (2.2)	.13
Unfavorable, n (%)	1 (3)	4 (10)	
Neutral, n (%)	29 (72)	30 (71)	
Favorable, n (%)	10 (25)	8 (19)	
Physical activity level			
Reported by parents			
Score, mean (SD)	−0.9 (2.8)	−0.9 (3.2)	.97
Low, n (%)	15 (38)	17 (40)	
Medium, n (%)	19 (47)	17 (41)	
High, n (%)	6 (15)	8 (19)	
Reported by children			
Score, mean (SD)	−1.6 (2.7)	−1.5 (3.2)	.86
Low, n (%)	17 (42)	18 (43)	
Medium, n (%)	20 (50)	19 (45)	
High, n (%)	3 (8)	5 (12)	

Characteristics	Intervention group (n=40)	Control group (n=42)	*P* value
Media consumption			
Watching TV or video films per day reported by parents, n (%)			**.63**
≤0.5 hours	13 (33)	19 (45)	
1-2 hours	26 (65)	22 (53)	
≥3 hours	1 (2)	1 (2)	
Doing gaming activities at a computer per day reported by parents, n (%)			**>.99**
≤0.5 hours	37 (92)	38 (90)	
1-2 hours	3 (8)	4 (10)	
≥3 hours	0 (0)	0 (0)	

Table 2. Outcomes from the intention-to-treat analysis.

Outcomes[a]	IG[b] (n=40)	CG[c] (n=42)	IG versus CG	Contrast IG versus CG *P* value	Effect size Eta^2/Cramers V[d]
Knowledge, mean (SD); 95% CI					
Percentage of total score (primary outcome)	69 (11); 65 to 72	52 (12); 49 to 56	17 (12 to 22)	<.001	0.235
Percentage of food pyramid score	77 (12); 74 to 81	54 (12); 50 to 57	23 (18 to 28)	<.001	0.337
Percentage dietary energy density score	64 (17); 58 to 69	46 (22); 40 to 53	18 (1 to 27)	.097	0.131
Percentage of stress score	68 (18); 63 to 74	58 (24); 51 to 66	10 (1 to 19)	<.01	0.086
Physical activity level					
Reported by parents					
Score, mean (SD); 95% CI	−2.1 (2.8); 3.0 to −1.2	−2.7 (2.7); −3.6 to −1.9	−0.6 (−0.6 to 1.8)	.48	0.012
Low, n (%)	21 (53)	27 (64)	−6 (−11)		
Medium, n (%)	17 (42)	12 (29)	5 (13)		
High, n (%)	2 (5)	3 (7)	−1 (−2)		
Reported by children					
Score, mean (SD); 95% CI	−2.7 (3.1); −3.6 to −1.7	−2.6 (3.4); −3.7 to −1.6	0.1 (−1.3 to 1.5)	.87	0.000
Low, n (%)	25 (62)	25 (60)	0 (0)		
Medium, n (%)	13 (33)	12 (28)	1 (5)		
High, n (%)	2 (5)	5(12)	−3 (−7)		
Healthy Nutrition Index					
Reported by parents					
Score, mean (SD); 95% CI	10.0 (2.7); 9.1 to 10.8	10.2 (2.0); 9.5 to 10.7	0.2 (−0.8 to 1.2)	.31	0.014
Unfavorable, n (%)	2 (5)	0 (0)	2 (5)		
Neutral, n (%)	20 (50)	26 (62)	−6 (−12)		
Favorable, n (%)	18 (45)	16 (38)	−2 (−7)		
Reported by children					
Score, mean (SD); 95% CI	9.5 (2.2); 8.8 to 10.1	9.3 (2.5); 8.5 to 10.1	0.2 (−0.9 to 1.2)	.68	0.002
Unfavorable, n (%)	0 (0)	3 (7)	−3 (−7)		
Neutral, n (%)	25 (62)	27 (64)	−2 (−2)		
Favorable, n (%)	15 (38)	12 (29)	3 (9)		
Media consumption					
Watching TV or video films per day reported by parents, n (%)					**0.045[d]**
<0.5 hours	14 (35)	21 (50)	−7 (−15)	N/A[e]	
1-2 hours	24 (60)	16 (38)	8 (22)	N/A	
≥3 hours	2 (5)	5 (12)	−3 (−7)	N/A	
Doing gaming activities at a computer per day reported by parents, n (%)					**0.141[d]**
≤0.5 hours	37 (93)	34 (81)	3 (12)	N/A	
1-2 hours	2 (5)	7 (17)	−5 (−12)	N/A	
≥3 hours	1 (2)	1 2)	0 (0)	N/A	

[a]For all outcomes, except for media consumption, the data are presented as mean (standard deviation) and 95% CIs for the IG and CG and for the

changes between the groups (IG vs CG). For the physical activity level, the Healthy Nutrition Index, and media consumption, the data are also presented as sample size and percentage of the corresponding category. The *P* values of the contrasts IG versus CG of the ANOVA of the adjusted model along with the effect sizes are presented.

[b]IG: intervention group.

[c]CG: control group.

[d]The effect size according to Cramers V.

[e]Not applicable.

Table 3. Complete overview of statistics from the intention-to-treat analysis.

Outcome[a]	Unadjusted group effect, *P* value	Overall class effect, *P* value	Contrast IG[b] versus CG[c] (95% CI of contrast)	Contrast IG versus CG, *P* value	Effect size (Eta^2/Cramers V[d])
Knowledge					
Percentage of total score (primary outcome)	<.001	<.001	0.39 (0.238 to 0.542)	<.001	0.235
Percentage of food pyramid score	<.001	<.001	0.686 (0.494 to 0.878)	<.001	0.337
Percentage of dietary energy density score	.001	.19	0.241 (−0.045-0.527)	.097	0.131
Percentage of stress score	.008	.05	0.379 (0.109 to 0.648)	<.01	0.086
Physical activity level					
Reported by parents					
Score	.320	.48	1.270	.48	0.012
Categories	.951		(−2.325 to 4.866)		0.048[d]
Reported by children					
Score, n	.883	.96	0.278	.87	0.000
Categories	.843		−3.121 to 3.678		0.079[d]
Healthy Nutrition Index					
Reported by parents					
Score	.297	.06	−1.453	.31	0.014
Categories	.353		(−4.304 to 1.398)		0.163[d]
Reported by children					
Score	.681	.58	0.673	.68	0.002
Categories	.819		(−2.537 to 3.884)		0.072[d]
Media consumption					
Watching TV or video films per day in hours reported by parents	.948 (categories)	N/A[e]	N/A	N/A	0.045[d]
Doing gaming activities at a computer per day in hours reported by parents	.676 (categories)	N/A	N/A	N/A	0.141[d]

[a]The complete statistics for the outcomes of the study are presented for the unadjusted and adjusted models, the latter being an analysis of variance with contrasts.

[b]IG: intervention group.

[c]CG: control group.

[d]Cramers V effect size.

[e]Not applicable.

Primary Outcome: Knowledge of the Game

The primary outcome was the gain in knowledge (total knowledge test score). Knowledge increased significantly from T0 to T1 when comparing IG with CG in the unadjusted and adjusted model. All subscores increased significantly in the unadjusted model from T0 to T1. In the adjusted model, the subscores *food pyramid* and *stress score* increased, and a clear trend was shown for the DED score from T0 to T1. In addition, all scores improved slightly from T0 to T1 regardless of the group allocation (total knowledge test score: $F_{1,80}$=49.597,

P>.001; food pyramid score: $F_{1,80}$=98.600, P<.001; DED score: $F_{1,79}$=11.921, P=.001; and stress score: $F_{1,80}$=6.891, P=.01).

Secondary Outcomes

Maintenance of knowledge was tested at the 4-week follow-up from T1. The gain in knowledge was maintained in IG over a 4-week period (from T1 to T2) and was comparable with the knowledge of CG, which had just completed the intervention (see the subsection Outcome Measures under the Methods section), and no differences between the groups were evident (total knowledge test score for IG: 69.7% [SD 11.0] vs CG: 68.3% [SD 18.7]; food pyramid score: IG 78.9% [SD 7.9] vs CG: 77.4 [SD 11.2]; DED score: IG: 68.5% [SD 14.1] vs CG: 65.2 [SD 16.0]; and stress score: IG: 69.3 [SD 22.2] vs CG: 68.3 [SD 18.7

The Dietary Energy Density Principle Under Time Pressure (Intervention Group and Control Group After Game Play)

To support the data of the DED score, we analyzed the game data of both groups where the DED-P had to be applied under time pressure (Kangaroo-Turtle race) after DED education. Under these conditions, the children were able to classify unknown foods equally to known foods according to their DED (hit rate: 72.5% [SD 14.9] vs 70.5% [SD 14.2]; T=−1127; P=.260). However, it took the children longer to make a decision when unknown foods were presented in comparison with known foods (1542 ms [SD 346] vs 1495 ms [SD 465]; T=−3484; P<.001).

Acceptance of the Game and Feelings During Game Play (Intervention Group Only)

Overall, 92% (37/40) of participants in IG responded that they liked the game very much or much, 3% (1/40) found it okay and 5% (2/40) were mis-d. Furthermore, 90% (36/40) of participants in IG agreed that playing the game was fun, whereas 5% (2/40) agreed partly and 3% (1/40) disagreed and 3% (1/40) were mis-d. In addition, 85% (34/40) of participants in IG agreed that it did not make them feel bored, whereas 10% (4/40) agreed partly and 5% (2/40) disagreed or were mis-d. Furthermore, 80% (32/40) of participants in IG also would like to play the game at home, 7.5% were not sure (n=3), and 13% (5/40) would not or were mis-d. More than two third of participants in IG (31/40, 78%) would recommend the game to a friend, 15% (6/40) were not sure, and 8% (3/40) would not or were mis-d. Using a nonverbal pictorial assessment, the children reported that overall they felt good, strong, and relaxed when playing the game. For question 1, the mean was 1.1 (SD 0.3); for question 2, it was 1.8 (SD 1.2); and for question 3, it was 1.8 (SD 1.0).

Dietary Behavior

Dietary behavior was assessed using a Healthy Nutrition Index (HNI). At baseline, HNI was predominantly neutral or favorable as reported by the children and their parents. From T0 to T2, no significant changes were observed between the groups neither in HNI reported by the children themselves nor by their parents (Tables 2 and 3 and Multimedia Appendix 5). However, regardless of the group allocation, HNI improved from T0 to T2 as reported by parents ($F_{1,80}$=4.916; P=.029) and children ($F_{1,80}$=10.863; P trend=.057).

Physical Activity

At baseline, most of the children had a low or moderate physical activity level. No changes were observed in the time course between the different group allocations as reported by children and their parents (Tables 2 and 3 and Multimedia Appendix 5). However, from T0 to T2, the physical activity level decreased across all groups according to children ($F_{1,80}$=15.924; P<.001) and parents ($F_{1,80}$=22.153; P<.001).

Media Consumption

The duration of watching TV or video films on weekdays was between 0 and 30 min or 1 to 2 hours for most children at baseline. Gaming activities duration at a computer on weekdays was between 0 and 30 min for most children at baseline. No significant changes were found between the groups in the time course. However, changes in media consumption independent of group allocation were found for both investigated items (each P<.001, Fisher exact test, has no test value). The most prominent changes were found for watching TV or video films, where almost balanced bidirectional changes in consumption occurred in 35% to 40% of children (Tables 2 and 3 and Multimedia Appendix 5).

Discussion

Principal Findings

Knowledge

To our knowledge, KOP is the first health game that addresses three core areas of obesity prevention and treatment: nutrition, physical activity, and stress coping. In the nutrition section, it is the first that extensively focuses on the DED concept.

The main hypothesis was verified, showing that children gained sustainable knowledge about the food pyramid concept, the DED concept (including the topic of liquids), and about stress and stress-coping strategies after game play. Second, we hypothesized that after the 4-week follow-up, the knowledge scores would be higher than at baseline but lower in comparison with the scores directly after the game. Interestingly, 4 weeks after the intervention, the knowledge level was similar to the level directly after intervention. This was unexpected considering that knowledge is quickly lost if not consciously reviewed from time to time [73,74]. This is a promising result, which may be explained by the high level of challenging interactions, repetitions, and self-reflective tools applied in the game. It also shows how powerful games can be for standardized knowledge transfer when correctly designed and when topics are presented appealingly to children (and adolescents and adults) [75]. It is possible that the knowledge transfer was only sustainable because the children might have been more receptive to the topic and solidified the knowledge in thought and practice after the intervention. This was not measurable at the HNI level, however, as this did not differ between the groups but improved in the course of time in both groups, as will be discussed below. The maintenance of knowledge has not been tested by other serious games targeting nutrition and obesity [29], except in a

recent report in which a 2-week follow-up nutrition knowledge test was performed where at the follow-up, similar results were observed for IG and CG. This was explained by (1) high baseline knowledge levels of both groups not leaving much room for knowledge gain and (2) CG having an extremely good learning curve by repeated exposure to the same tests [47]. In concordance, we also saw a time-dependent knowledge effect independent of the group, although this was marginal.

In contrast to the abovementioned study, in this study, the knowledge testing after baseline was after 4 weeks, and the knowledge questionnaire was designed using a wide range from simple to difficult questions, thereby leaving room for grading of knowledge and making it difficult to score highly if certain content was not explicitly taught.

Understanding and Applying the Dietary Energy Density Principle

The third hypothesis stated that the children could understand and apply the DED-P after game play. At knowledge level, the subscore DED did not reach significance between the groups in the adjusted model, but a clear trend was observed ($P<.1$). This observation may be because of (1) that the questionnaire was designed for the total score and not explicitly for the subscores (eg, fewer points can be collected with this score when compared with the food pyramid score) and (2) that the IG already scored significantly higher at baseline regarding the knowledge of DED-P. More importantly, the children were able to apply their DED knowledge by transferring it to unknown foods when tested under time pressure, as 70% were able to correctly categorize the food pairs according to their DED. No reference values from the literature could be retrieved, either for children or adults. Hermmans et al [47] included a game on healthy and unhealthy food pairs (without defining what healthy and unhealthy are) where some pairs may have indirectly presented pairs of different DEDs. Here, the hit rate for *healthy food choices* was around 75%. The observed 70% correct categorization rate for food pairs under time pressure appears substantial. The finding that the correct categorization rate was similar between known and unknown food pairs is also promising. Although it took slightly longer to make the decision when unknown food pairs were presented, the principle was shown to have been understood, and a transfer of that knowledge to other foods was possible. This may be important when making subconscious decisions regarding food selection [76]. We have no baseline data for this DED-P game module (Kangaroo-Turtle race) because the learning effect of the module itself is assumed to be high and, therefore, could not have been applied in CG or IG before intervention.

Game Acceptance and Behavior

Hypothesis four was verified by showing that children felt good during game play and game acceptance was high. Finally, we tested behavioral aspects, although we hypothesized no changes at the behavioral level would occur because of the study focus and design. Behavioral measures were assessed at baseline and T2. In the course of the study period, we observed bidirectional changes of media consumption in both groups but no differences between the groups. We observed behavioral changes at the physical and dietary levels across both groups, reported independently by the children and their parents. Physical activity decreased in roughly 40% or 30% of children depending on the reporting source, whereas in 10%, it increased. Thus, 20% to 30% of children decreased their activity level, which may be explained by the fact that the study commenced in autumn after 6 weeks of summer holidays. The decreasing physical activity levels observed may be related to more inactivity during school time, shorter daylight periods, and a temperature drop. The intervention itself was not associated with a change in physical activity. This is in keeping with the short study period and the game design being unlikely to effect a change in physical activity level. The motion control was used to avoid long times of inactivity during game play but did not focus on increasing physical activity as classical exergames do [29].

As mentioned previously, HNI improved slightly over the course of the study. Approximately 20% of the children improved in category, whereas around 10% showed a deterioration in category. It must be noted that the baseline HNI in about 95% of children was already either neutral or favorable, and category changes appear to be rather moderate because the proportion of children in the unfavorable category did not increase. The changes observed over time may be spontaneous because of real or perceived dietary changes occurring in the context of the switch between holidays and school routine together with seasonal factors, may reflect participation in this health study, or may involve a combination of these factors. CG also received material about a healthy lifestyle, which may have motivated some families to maintain a healthier diet. The HNI data here are in line with the findings of a national health survey (KIGGS), which used the same instrument for dietary assessment [63].

Overall, it is a positive finding that taking part in this health game study improved dietary behavior measured with HNI in families, but for sustainable lifestyle changes in children, it is important to promote child-parent interactions. Parents are role models for their children and have the power to influence their behavior and thinking or help to solidify new patterns of activity in children and adolescents [77,78].

Limitations and Outlook

To date, there is a lack of health game interventions, which include parents [29].

This is a limitation of this study, as we only involved parents for data collection. Second, our follow-up period was short, and we did not include behavioral tests, for example, for measuring eating behavior. We also have to consider that knowledge improvements about health issues do not necessarily result in changes at the behavioral level. Owing to the nature of this intervention, no double-blind and placebo-controlled study design was possible to evaluate the game, and this may have influenced the results.

To address some of these limitations, we are in the process of undertaking a further randomized controlled trial to evaluate the game regarding (1) its acceptance and efficacy in parents and their children and (2) whether the children can benefit by the involvement of their parents. A 6-month follow-up is also included in this study. We also plan to conduct trials where the efficacy and acceptance of the game are tested in established

inpatient and outpatient treatment settings for children. By involving not only the children but also their parents, these studies aim to improve parental support for their children and compliance of the children. Making the topics of *nutrition* and *stress and stress regulation* more appealing using modern media should improve the outcome of the children in these programs. Finally, the approach of supporting individual lifestyle changes should be complemented by community-based and environment-oriented approaches to combat obesity [79-81].

Conclusions

Taken together, KOP is the first serious game for children addressing the three topics of nutrition, physical activity, and stress coping. To our knowledge, it is also the first game, which extensively targets the DED concept in the area of nutrition, in addition to playfully dealing with stress coping. The game was highly accepted by children, sustainably increased their knowledge of the topics addressed, and could be a useful tool for further studies and education.

Acknowledgments

This study was funded by grants from the Leibniz-ScienceCampus *Informational Environments*. IM received a grant from the Ministry of Science Baden-Württemberg and the European Social Fund. The authors acknowledge support from the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of Tübingen University. The authors thank the group members from Cluster 7 (Overcoming barriers to prevention and treatment of obesity in childhood and adolescence—a digital media approach) of the Leibniz-ScienceCampus Tübingen, Elena Kounadis, and Olaf Schliesing for their valuable discussions and Jessica Dugmore for correcting the manuscript as native English speaker. Permissions from the persons named in the Acknowledgments have been received.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of topics and modules.
[PDF File (Adobe PDF File), 34 KB - jmir_v22i4e15725_app1.pdf]

Multimedia Appendix 2
Screeenshots of the game.
[PDF File (Adobe PDF File), 2071 KB - jmir_v22i4e15725_app2.pdf]

Multimedia Appendix 3
Video of the game.
[MP4 File (MP4 Video), 69490 KB - jmir_v22i4e15725_app3.mp4]

Multimedia Appendix 4
Questionnaire.
[PDF File (Adobe PDF File), 251 KB - jmir_v22i4e15725_app4.pdf]

Multimedia Appendix 5
Results-changes from baseline.
[PDF File (Adobe PDF File), 65 KB - jmir_v22i4e15725_app5.pdf]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2398 KB - jmir_v22i4e15725_app6.pdf]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
2. Geserick M, Vogel M, Gausche R, Lipek T, Spielau U, Keller E, et al. Acceleration of BMI in early childhood and risk of sustained obesity. N Engl J Med 2018 Oct 4;379(14):1303-1312. [doi: 10.1056/NEJMoa1803527] [Medline: 30281992]

3. Sauer H, Ohla K, Dammann D, Teufel M, Zipfel S, Enck P, et al. Changes in gustatory function and taste preference following weight loss. J Pediatr 2017 Mar;182:120-126. [doi: 10.1016/j.jpeds.2016.11.055] [Medline: 27989411]
4. Mazurak N, Sauer H, Weimer K, Dammann D, Zipfel S, Horing B, et al. Effect of a weight reduction program on baseline and stress-induced heart rate variability in children with obesity. Obesity (Silver Spring) 2016 Feb;24(2):439-445 [FREE Full text] [doi: 10.1002/oby.21355] [Medline: 26704529]
5. Weimer K, Sauer H, Horing B, Valitutti F, Mazurak N, Zipfel S, et al. Impaired gastric myoelectrical reactivity in children and adolescents with obesity compared to normal-weight controls. Nutrients 2018 May 31;10(6):pii: E699 [FREE Full text] [doi: 10.3390/nu10060699] [Medline: 29857470]
6. Pulgarón ER. Childhood obesity: a review of increased risk for physical and psychological comorbidities. Clin Ther 2013 Jan;35(1):A18-A32 [FREE Full text] [doi: 10.1016/j.clinthera.2012.12.014] [Medline: 23328273]
7. Rankin J, Matthews L, Cobley S, Han A, Sanders R, Wiltshire HD, et al. Psychological consequences of childhood obesity: psychiatric comorbidity and prevention. Adolesc Health Med Ther 2016;7:125-146 [FREE Full text] [doi: 10.2147/AHMT.S101631] [Medline: 27881930]
8. Mölbert SC, Sauer H, Dammann D, Zipfel S, Teufel M, Junne F, et al. Multimodal body representation of obese children and adolescents before and after weight-loss treatment in comparison to normal-weight children. PLoS One 2016;11(11):e0166826 [FREE Full text] [doi: 10.1371/journal.pone.0166826] [Medline: 27875563]
9. Harrison S, Rowlinson M, Hill AJ. 'No fat friend of mine': young children's responses to overweight and disability. Body Image 2016 Sep;18:65-73. [doi: 10.1016/j.bodyim.2016.05.002] [Medline: 27322671]
10. Thiel A, Alizadeh M, Giel K, Zipfel S. [Stereotyping of overweight children by their contemporaries]. Psychother Psychosom Med Psychol 2008 Dec;58(12):e16-e24. [doi: 10.1055/s-2008-1067340] [Medline: 19043828]
11. Bleich SN, Vercammen KA, Zatz LY, Frelier JM, Ebbeling CB, Peeters A. Interventions to prevent global childhood overweight and obesity: a systematic review. Lancet Diabetes Endocrinol 2018 Apr;6(4):332-346. [doi: 10.1016/S2213-8587(17)30358-3] [Medline: 29066096]
12. Foster BA, Farragher J, Parker P, Sosa ET. Treatment interventions for early childhood obesity: a systematic review. Acad Pediatr 2015;15(4):353-361 [FREE Full text] [doi: 10.1016/j.acap.2015.04.037] [Medline: 26142067]
13. Raynor HA, Champagne CM. Position of the academy of nutrition and dietetics: interventions for the treatment of overweight and obesity in adults. J Acad Nutr Diet 2016 Jan;116(1):129-147. [doi: 10.1016/j.jand.2015.10.031] [Medline: 26718656]
14. Rolls BJ. Dietary energy density: applying behavioural science to weight management. Nutr Bull 2017 Sep;42(3):246-253 [FREE Full text] [doi: 10.1111/nbu.12280] [Medline: 29151813]
15. Benton D. Portion size: what we know and what we need to know. Crit Rev Food Sci Nutr 2015;55(7):988-1004 [FREE Full text] [doi: 10.1080/10408398.2012.679980] [Medline: 24915353]
16. Robinson E, Kersbergen I. Portion size and later food intake: evidence on the 'normalizing' effect of reducing food portion sizes. Am J Clin Nutr 2018 Apr 1;107(4):640-646. [doi: 10.1093/ajcn/nqy013] [Medline: 29635503]
17. Haack SA, Byker CJ. Recent population adherence to and knowledge of United States federal nutrition guides, 1992-2013: a systematic review. Nutr Rev 2014 Oct;72(10):613-626. [doi: 10.1111/nure.12140] [Medline: 25209465]
18. Montagnese C, Santarpia L, Buonifacio M, Nardelli A, Caldara AR, Silvestri E, et al. European food-based dietary guidelines: a comparison and update. Nutrition 2015;31(7-8):908-915. [doi: 10.1016/j.nut.2015.01.002] [Medline: 26015390]
19. Park M, Camilleri M. Gastric motor and sensory functions in obesity. Obes Res 2005 Mar;13(3):491-500 [FREE Full text] [doi: 10.1038/oby.2005.51] [Medline: 15833933]
20. Mejía-Rivas M, Remes-Troche JM, Montaño-Loza A, Herrera M, Valdovinos-Díaz MA. Gastric capacity is related to body mass index in obese patients. A study using the water load test. Rev Gastroenterol Mex 2009;74(1):71-73. [Medline: 19666328]
21. Mack I, Sauer H, Weimer K, Dammann D, Zipfel S, Enck P, et al. Obese children and adolescents need increased gastric volumes in order to perceive satiety. Obesity (Silver Spring) 2014 Oct;22(10):2123-2125 [FREE Full text] [doi: 10.1002/oby.20850] [Medline: 25070787]
22. Barlow SE, Dietz WH. Obesity evaluation and treatment: Expert Committee recommendations. The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services. Pediatrics 1998 Sep;102(3):E29. [doi: 10.1542/peds.102.3.e29] [Medline: 9724677]
23. Tajik E, Zulkefli NA, Baharom A, Minhat HS, Latiff LA. Contributing factors of obesity among stressed adolescents. Electron Physician 2014;6(1):771-778 [FREE Full text] [doi: 10.14661/2014.771-778] [Medline: 25763144]
24. Ofcom. 2017. Children and Parents: Media Use and Attitudes Report URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/108182/children-parents-media-use-attitudes-2017.pdf [accessed 2020-01-20]
25. mpfs. Startseite | mpfs. 2017. Jugend, Informartion, (Multi-) Media, JIM 2017 URL: https://www.mpfs.de/fileadmin/files/Studien/JIM/2017/JIM_2017.pdf [accessed 2020-01-20]
26. Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput J, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):S240-S265. [doi: 10.1139/apnm-2015-0630] [Medline: 27306432]

27. Rey-López JP, Vicente-Rodríguez G, Biosca M, Moreno LA. Sedentary behaviour and obesity development in children and adolescents. Nutr Metab Cardiovasc Dis 2008 Mar;18(3):242-251. [doi: 10.1016/j.numecd.2007.07.008] [Medline: 18083016]
28. Dias JD, Domingues AN, Tibes CM, Zem-Mascarenhas SH, Fonseca LM. Serious games as an educational strategy to control childhood obesity: a systematic literature review1. Rev Lat Am Enfermagem 2018 Sep 3;26:e3036 [FREE Full text] [doi: 10.1590/1518-8345.2509.3036] [Medline: 30183872]
29. Mack I, Bayer C, Schäffeler N, Reiband N, Brölz E, Zurstiege G, et al. Chances and limitations of video games in the fight against childhood obesity-a systematic review. Eur Eat Disord Rev 2017 Jul;25(4):237-267. [doi: 10.1002/erv.2514] [Medline: 28467004]
30. Holzmann SL, Dischl F, Schäfer H, Groh G, Hauner H, Holzapfel C. Digital gaming for nutritional education: a survey on preferences, motives, and needs of children and adolescents. JMIR Form Res 2019 Feb 13;3(1):e10284 [FREE Full text] [doi: 10.2196/10284] [Medline: 30758290]
31. Gentry SV, Gauthier A, Ehrstrom BL, Wortley D, Lilienthal A, Car LT, 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]
32. Boendermaker WJ, Boffo M, Wiers RW. Exploring elements of fun to motivate youth to do cognitive bias modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]
33. Barnes S, Prescott J. Empirical evidence for the outcomes of therapeutic video games for adolescents with anxiety disorders: systematic review. JMIR Serious Games 2018 Feb 28;6(1):e3 [FREE Full text] [doi: 10.2196/games.9530] [Medline: 29490893]
34. Turnin MC, Tauber MT, Couvaras O, Jouret B, Bolzonella C, Bourgeois O, et al. Evaluation of microcomputer nutritional teaching games in 1,876 children at school. Diabetes Metab 2001 Sep;27(4 Pt 1):459-464. [Medline: 11547219]
35. Serrano EL, Anderson JE. The evaluation of food pyramid games, a bilingual computer nutrition education program for Latino youth. J Fam Consum Sci Educ 2004;22(1):1-16 [FREE Full text]
36. Schneider KL, Ferrara J, Lance B, Karetas A, Druker S, Panza E, et al. Acceptability of an online health videogame to improve diet and physical activity in elementary school students: 'Fitter Critters'. Games Health J 2012 Aug;1(4):262-268 [FREE Full text] [doi: 10.1089/g4h.2012.0009] [Medline: 24761317]
37. Robertson TP, Zalles DR. Nutrition education program Nutrition Pathfinders teaches children how to make healthful food choices. J Nutr Educ Behav 2005;37(1):41-42. [doi: 10.1016/s1499-4046(06)60258-3] [Medline: 15745655]
38. Moore JB, Pawloski LR, Goldberg P, Kyeung MO, Stoehr A, Baghi H. Childhood obesity study: a pilot study of the effect of the nutrition education program Color My Pyramid. J Sch Nurs 2009 Jun;25(3):230-239. [doi: 10.1177/1059840509333325] [Medline: 19363107]
39. Majumdar D, Koch PA, Lee H, Contento IR, Islas-Ramos AD, Fu D. 'Creature-101': a serious game to promote energy balance-related behaviors among middle school adolescents. Games Health J 2013 Oct;2(5):280-290 [FREE Full text] [doi: 10.1089/g4h.2013.0045] [Medline: 24761326]
40. Johnson-Glenberg MC, Savio-Ramos C, Henry H. 'Alien Health': a nutrition instruction Exergame using the Kinect sensor. Games Health J 2014 Aug 1;3(4):241-251 [FREE Full text] [doi: 10.1089/g4h.2013.0094] [Medline: 25083315]
41. Baños RM, Cebolla A, Oliver E, Alcañiz M, Botella C. Efficacy and acceptability of an internet platform to improve the learning of nutritional knowledge in children: the ETIOBE Mates. Health Educ Res 2013 Apr;28(2):234-248. [doi: 10.1093/her/cys044] [Medline: 22498924]
42. Sharma SV, Shegog R, Chow J, Finley C, Pomeroy M, Smith C, et al. Effects of the Quest to Lava Mountain computer game on dietary and physical activity behaviors of elementary school children: a pilot group-randomized controlled trial. J Acad Nutr Diet 2015 Aug;115(8):1260-1271. [doi: 10.1016/j.jand.2015.02.022] [Medline: 25912520]
43. Munguba MC, Valdés MT, da Silva CA. The application of an occupational therapy nutrition education programme for children who are obese. Occup Ther Int 2008;15(1):56-70. [doi: 10.1002/oti.244] [Medline: 18288771]
44. Baranowski T, Baranowski J, Thompson D, Buday R, Jago R, Griffith MJ, et al. Video game play, child diet, and physical activity behavior change a randomized clinical trial. Am J Prev Med 2011 Jan;40(1):33-38 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.029] [Medline: 21146765]
45. Thompson D, Baranowski T, Cullen K, Watson K, Liu Y, Canada A, et al. Food, fun, and fitness internet program for girls: pilot evaluation of an e-Health youth obesity prevention program examining predictors of obesity. Prev Med 2008 Nov;47(5):494-497. [doi: 10.1016/j.ypmed.2008.07.014] [Medline: 18718846]
46. Thompson D, Bhatt R, Vazquez I, Cullen KW, Baranowski J, Baranowski T, et al. Creating action plans in a serious video game increases and maintains child fruit-vegetable intake: a randomized controlled trial. Int J Behav Nutr Phys Act 2015 Mar 18;12:39 [FREE Full text] [doi: 10.1186/s12966-015-0199-z] [Medline: 25890060]
47. Hermans RCJ, van den Broek N, Nederkoorn C, Otten R, Ruiter EL, Johnson-Glenberg MC. Feed the Alien! The effects of a nutrition instruction game on children's nutritional knowledge and food intake. Games Health J 2018 Jun;7(3):164-174. [doi: 10.1089/g4h.2017.0055] [Medline: 29634366]

48. Marchetti D, Fraticelli F, Polcini F, Lato R, Pintaudi B, Nicolucci A, et al. Preventing adolescents' diabesity: design, development, and first evaluation of 'Gustavo in Gnam's Planet'. Games Health J 2015 Oct;4(5):344-351 [FREE Full text] [doi: 10.1089/g4h.2014.0107] [Medline: 26287925]
49. Haerens L, Deforche B, Maes L, Brug J, Vandelanotte C, de Bourdeaudhuij I. A computer-tailored dietary fat intake intervention for adolescents: results of a randomized controlled trial. Ann Behav Med 2007;34(3):253-262. [doi: 10.1007/bf02874550] [Medline: 18020935]
50. Zurstiege G, Zipfel S, Ort A, Mack I, Meitz TG, Schäffeler N. Managing obesity prevention using digital media: a double-sided approach. In: Buder J, Hesse F, editors. Informational Environments. Cham: Springer; 2017:97-123.
51. Kromeyer-Hauschild K, Wabitsch M, Kunze D, Geller F, Geiß HC, Hesse V, et al. Perzentile für den Body-mass-Index für das Kindes- und Jugendalter unter Heranziehung verschiedener deutscher Stichproben. Monatsschr Kinderheilkd 2001;149(8):807-818. [doi: 10.1007/s001120170107]
52. Kersting M, Lentze M. Perspektiven in der Kinderernährung. Monatsschr Kinderheilkd 2014;162(7):623-629. [doi: 10.1007/s00112-014-3088-7]
53. -. Human energy requirements: report of a joint FAO/ WHO/UNU Expert Consultation. Food Nutr Bull 2005 Mar;26(1):166. [Medline: 15810802]
54. von Ruesten A, Oberritter H. The German food pyramid. In: Preedy V, Hunter LA, Patel V, editors. Diet Quality. New York, NY: Humana Press; 2013:153-165.
55. Welsh S, Davis C, Shaw A. Development of the Food Guide Pyramid. Nutr Today 1992;26(6):12-23. [doi: 10.1097/00017285-199211000-00005]
56. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002 Dec;5(6B):1113-1124. [doi: 10.1079/PHN2002394] [Medline: 12639222]
57. Janssen P, Berghe PV, Verschueren S, Lehmann A, Depoortere I, Tack J. Review article: the role of gastric motility in the control of food intake. Aliment Pharmacol Ther 2011 Apr;33(8):880-894 [FREE Full text] [doi: 10.1111/j.1365-2036.2011.04609.x] [Medline: 21342212]
58. Welch RW. Satiety: have we neglected dietary non-nutrients? Proc Nutr Soc 2011 May;70(2):145-154. [doi: 10.1017/S0029665110004763] [Medline: 21275081]
59. Smethers AD, Rolls BJ. Dietary management of obesity: cornerstones of healthy eating patterns. Med Clin North Am 2018 Jan;102(1):107-124 [FREE Full text] [doi: 10.1016/j.mcna.2017.08.009] [Medline: 29156179]
60. Rolls BJ. The relationship between dietary energy density and energy intake. Physiol Behav 2009 Jul 14;97(5):609-615 [FREE Full text] [doi: 10.1016/j.physbeh.2009.03.011] [Medline: 19303887]
61. Almiron-Roig E, Palla L, Guest K, Ricchiuti C, Vint N, Jebb SA, et al. Factors that determine energy compensation: a systematic review of preload studies. Nutr Rev 2013 Jul;71(7):458-473 [FREE Full text] [doi: 10.1111/nure.12048] [Medline: 23815144]
62. Bradley MM, Lang PJ. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
63. Kleiser C, Mensink GB, Karth BM, Neuhauser H, Schenk L. Semantic Scholar. 2007. Ernährungsverhalten von Kindern und Jugendlichen mit Migrationshintergrund ? KiGGS-Migrantenauswertung - Endbericht (Nutrition behaviour of children and adolescents with a migration background - data from the KIGGS-study) URL: https://pdfs.semanticscholar.org/fc15/29094d13875bc4a17d14f0e92c131b67da5e.pdf [accessed 2020-02-04]
64. Truthmann J, Mensink GB, Richter A. Relative validation of the KiGGS Food Frequency Questionnaire among adolescents in Germany. Nutr J 2011 Dec 7;10:133 [FREE Full text] [doi: 10.1186/1475-2891-10-133] [Medline: 22152115]
65. Bayer O, Jarczok M, Fischer J, von Kries R, de Bock F. Validation and extension of a simple questionnaire to assess physical activity in pre-school children. Public Health Nutr 2012 Sep;15(9):1611-1619. [doi: 10.1017/S1368980012001243] [Medline: 22647318]
66. Manz K, Schlack R, Poethko-Müller C, Mensink G, Finger J, Lampert T, KiGGS Study Group. [Physical activity and electronic media use in children and adolescents: results of the KiGGS study: first follow-up (KiGGS wave 1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014 Jul;57(7):840-848. [doi: 10.1007/s00103-014-1986-4] [Medline: 24950833]
67. Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. Fam Pract 2000 Apr;17(2):192-196. [doi: 10.1093/fampra/17.2.192] [Medline: 10758085]
68. van Breukelen GJ, Candel MJ. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient!. J Clin Epidemiol 2012 Nov;65(11):1212-1218 [FREE Full text] [doi: 10.1016/j.jclinepi.2012.06.002] [Medline: 23017638]
69. Freeman GH, Halton JH. Note on an exact treatment of contingency, goodness of fit and other problems of significance. Biometrika 1951 Jun;38(1-2):141-149. [Medline: 14848119]
70. Vink G, Frank LE, Pannekoek J, van Buuren S. Predictive mean matching imputation of semicontinuous variables. Statistica Neerlandica 2014;68(1):61-90. [doi: 10.1111/stan.12023]
71. Rubin DB. Statistical matching using file concatenation with adjusted weights and multiple imputations. J Bus Econ Stat 1986;4(1):87-94. [doi: 10.1080/07350015.1986.10509497]

72. Kleinke K. Multiple imputation by predictive mean matching when sample size is small. Methodology 2018;14(1):3-15. [doi: 10.1027/1614-2241/a000141]
73. Ebbinghaus H. Über das Gedächtnis. Untersuchungen zur Experimentellen Psychologie. Leipzig: Verlag von Duncke&Humblot; 1985.
74. Murre JM, Dros J. Replication and analysis of Ebbinghaus' forgetting curve. PLoS One 2015;10(7):e0120644. [doi: 10.1371/journal.pone.0120644] [Medline: 26148023]
75. Goodman W, McFerran E, Purves R, Redpath I, Beeken RJ. The untapped potential of the gaming community: narrative review. JMIR Serious Games 2018 Sep 25;6(3):e10161 [FREE Full text] [doi: 10.2196/10161] [Medline: 30274962]
76. Jacquier C, Bonthoux F, Baciu M, Ruffieux B. Improving the effectiveness of nutritional information policies: assessment of unconscious pleasure mechanisms involved in food-choice decisions. Nutr Rev 2012 Feb;70(2):118-131. [doi: 10.1111/j.1753-4887.2011.00447.x] [Medline: 22300598]
77. Scaglioni S, de Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors influencing children's eating behaviours. Nutrients 2018 May 31;10(6):pii: E706 [FREE Full text] [doi: 10.3390/nu10060706] [Medline: 29857549]
78. Stein D, Weinberger-Litman SL, Latzer Y. Psychosocial perspectives and the issue of prevention in childhood obesity. Front Public Health 2014;2:104 [FREE Full text] [doi: 10.3389/fpubh.2014.00104] [Medline: 25133140]
79. Mendoza JA, Watson K, Cullen KW. Change in dietary energy density after implementation of the Texas Public School Nutrition Policy. J Am Diet Assoc 2010 Mar;110(3):434-440 [FREE Full text] [doi: 10.1016/j.jada.2009.11.021] [Medline: 20184994]
80. Hayes D, Dodson L. Practice paper of the Academy of Nutrition and Dietetics: comprehensive nutrition programs and services in schools. J Acad Nutr Diet 2018 May;118(5):920-931. [doi: 10.1016/j.jand.2018.02.025] [Medline: 29703343]
81. Weihrauch-Blüher S, Wiegand S. Risk factors and implications of childhood obesity. Curr Obes Rep 2018 Dec;7(4):254-259. [doi: 10.1007/s13679-018-0320-0] [Medline: 30315490]

Abbreviations

ANOVA: analysis of variance
CG: control group
DED-P: dietary energy density principle
FFQ: Food Frequency Questionnaire
HNI: Healthy Nutrition Index
IG: intervention group
ITT: intention-to-treat
KIGGS: German Health Interview and Examination Survey for Children and Adolescents
KOP: Kids Obesity Prevention
mis-d: missing data
PMM: predictive mean matching

Edited by G Eysenbach; submitted 02.08.19; peer-reviewed by J Alvarez Pitti, W Goodman, N Khalili-Mahani; comments to author 20.09.19; revised version received 02.12.19; accepted 16.12.19; published 24.04.20.

Please cite as:
Mack I, Reiband N, Etges C, Eichhorn S, Schaeffeler N, Zurstiege G, Gawrilow C, Weimer K, Peeraully R, Teufel M, Blumenstock G, Giel KE, Junne F, Zipfel S
The Kids Obesity Prevention Program: Cluster Randomized Controlled Trial to Evaluate a Serious Game for the Prevention and Treatment of Childhood Obesity
J Med Internet Res 2020;22(4):e15725
URL: https://www.jmir.org/2020/4/e15725
doi:10.2196/15725
PMID:32329742

Original Paper

Discrepancies in Demand of Internet of Things Services Among Older People and People With Disabilities, Their Caregivers, and Health Care Providers: Face-to-Face Survey Study

Heayon Lee[1*], MD, PhD; Yu Rang Park[2*], PhD; Hae-Reong Kim[2], MS; Na Young Kang[1], BS; Gahee Oh[3], MD, MPH; Il-Young Jang[1], MD; Eunju Lee[1], MD, PhD

[1]Division of Geriatrics, Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[2]Department of Biomedical System Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[3]Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Il-Young Jang, MD
Division of Geriatrics, Department of Internal Medicine
Asan Medical Center
University of Ulsan College of Medicine
88, Olympic-ro 43-gil, Songpa-gu
Seoul, 05505
Republic of Korea
Phone: 82 2 3010 1658
Email: onezero2@gmail.com

Abstract

Background: Home Internet of Things (IoT) services and devices have the potential to aid older adults and people with disabilities in their living environments. IoT services and devices can also aid caregivers and health care providers in conveniently providing care to those in need. However, real-world data on the IoT needs of vulnerable people are lacking.

Objective: The objective of this study is to conduct a face-to-face survey on the demand for IoT services among older people and people with disabilities, their caregivers, and health care providers in a real-world setting and to see if there are any differences in the aspects of need.

Methods: We conducted a face-to-face survey with 500 participants between January 2019 and March 2019. A total of 300 vulnerable people (200 older adults aged ≥65 years and 100 physically disabled people aged 30-64 years) were randomly sampled from either a population-based, prospective cohort study of aging—the Aging Study of Pyeongchang Rural Area (ASPRA)—or from the outpatient clinics at the Asan Medical Center, Seoul, South Korea. Simultaneously, their caregivers (n=150) and health care providers (n=50) participated in the survey. Detailed socioeconomic status, digital literacy, health and physical function, and home IoT service needs were determined. Among all commercially available IoT services, 27 services were classified into five categories: emergency and security, safety, health care, convenience (information), and convenience (operation). The weighted-ranking method was used to rank the IoT needs in different groups.

Results: There were discrepancies in the demand of IoT services among the vulnerable groups, their caregivers, and health care providers. The home IoT service category that was required the most by the vulnerable groups and their caregivers was emergency and security. However, health care providers indicated that the safety category was most needed by the older adults and disabled people. Home IoT service requirements differed according to the different types of disabilities among the vulnerable groups. Participants with fewer disabilities were more willing to use IoT services than those with more disabilities.

Conclusions: Our survey study shows that there were discrepancies in the demand of IoT services among the vulnerable groups, their caregivers, and health care providers. IoT service requirements differed according to the various types of disabilities. Home IoT technology should be established by combining patients' priorities and individualized functional assessments among vulnerable people.

Trial Registration: Clinical Research Information Service (CRIS; KCT0004157); https://tinyurl.com/r83eyva

(J Med Internet Res 2020;22(4):e16614) doi:10.2196/16614

KEYWORDS

Internet of Things; older adults; disability; health care; mobile phone

Introduction

The Internet of Things (IoT) is an emerging technology that connects a variety of everyday devices and systems, such as sensors, appliances, actuators, computers, and cellular phones, leading toward a highly distributed intelligent system capable of communicating with other devices and human beings [1]. Applications of IoT services and devices have the potential to aid older people and the physically disabled in their homes. IoT can help vulnerable groups to *age in place*, which is a concept whereby older people are able to continue living in their own homes as they age despite changes to their health and mobility [2,3]. In recent years, progress in wearable devices and sensor technologies have started to improve the prospects of services for assisting older people and physically disabled people [4-7]. It is agreed that helping the elderly and disabled people to live independently instead of in health care facilities provides cost savings and has significant potential to enhance quality of life [8-10]. Since some of the elderly and disabled people rely entirely on their family members for their assisted-care needs, their family caregivers may have emotional stress. Therefore, IoT can also aid their family caregivers and health care providers to conveniently give care and monitor those in need, as well as provide relief for family members [11-13].

Although a lot of work is going on regarding IoT-based care of older adults and disabled people, the adoption of these services is still quite low [14-16]. Among older adults and people with disabilities and their family members, older adults especially are generally slow to adopt emerging technologies, as they often have difficulties using electronic devices and they are concerned about their privacy [14]. Also, the concept of smart homes and IoT is relatively new; therefore, poor understanding about new innovative solutions can also be a factor of slow adoption, as well as cost [16-18].

Understanding the needs of vulnerable groups and their specific requirements is the key to success of the smart homes meant for their well-being. Yet, there is a lack of research on the needs of older adults and disabled people regarding smart homes and IoT [10]. The objective of this study was to investigate the demand for IoT services and devices among older people and disabled people, their caregivers, and health care providers in the real-world setting, and to see if there are any differences in the aspects of need according to the different types of disabilities among the groups.

Methods

Study Design and Recruitment of Study Population

This study was designed to investigate the needs for IoT in everyday life from the perspectives of older adults or people with disabilities. For this, we conducted a face-to-face survey with 500 participants between January 2019 and March 2019. We randomly selected 300 participants who were physically vulnerable (200 older adults aged ≥65 years and 100 disabled participants aged 30-64 years), 150 participants who were their caregivers, and 50 health care providers to participate in the survey.

We randomly selected 200 older adults with scores of 4-9 (ie, mild-to-moderate limitations in physical performance) on the Short Physical Performance Battery (SPPB) [19,20], with or without disability: 120 participants were from a population-based, prospective cohort study of aging—the Aging Study of Pyeongchang Rural Area (ASPRA)—and 80 participants were from the outpatient clinics at the Asan Medical Center, Seoul, South Korea. The ASPRA cohort was established in the Pyeongchang rural area, located 180 kilometers east of Seoul, South Korea, and has been described elsewhere [21,22].

In the disabled people's group, we randomly selected 100 participants with (1) three or more comorbidities and/or (2) physical disabilities or mobility disability (ie, inability to walk more than 400 meters) [23]. Among them, 60 subjects were from the outpatient clinics at the Asan Medical Center and 40 participants were from the Pyeongchang rural area [24].

We also conducted a companion face-to-face survey on 150 caregivers, who were either trained professionals (n=5) or family members (n=145), and 50 health care providers, who were described as physicians, nurses, visiting nurses, and community care workers who regularly interact with older adults or people with disabilities. The caregivers and health care providers were asked to answer, from the perspectives of the vulnerable groups, what IoT services and devices they needed. All participants provided written informed consent. This study was approved by the Institutional Review Board of the Asan Medical Center (Institutional Review Board No. 2019-0041).

Questionnaire Items

We developed paper questionnaires based on several publications [24-27] and conducted a face-to-face survey for about 30 minutes. In order to ensure the validity of the questionnaire, the questionnaire items were developed through several revisions with five experts. The questionnaire was composed of four parts and consisted of three versions, with respect to respondents: vulnerable groups (see Multimedia Appendix 1), caregivers, and health care providers. We researched all commercially available IoT services in the Korean market as of October 1, 2018. There were 84 services that were available. We removed IoT services and devices that had duplicate functions and grouped similar services, resulting in 27 services that were included in the questionnaire. We classified these services into five categories: security and emergency, safety, health care, convenience (information), and convenience (operation; see Table 1).

Table 1. Available Internet of Things (IoT) services by category.

Category	Services
Security	Home security and closed-circuit television (CCTV); Smart band and mobile *SOS* bell; Smart home *SOS* bell; Front door smart sensor; Voice-recognition front door lock
Safety	IoT-based power-system protection device; IoT-based smart gas monitoring; GPS trackers; Gas-valve remote control; Track use of smart water purifier
Health care	Smart home air purifier; Doctor's appointments; Fitness program using smart television (TV); House temperature and humidity control; Smart IoT chair
Convenience (information)	Schedule appointments and alarms; Bus arrival time notification; Weather forecast; Traffic information
Convenience (operating)	Taxi call; Food order and delivery; Robot vacuum cleaner; Smart washing machine mobile app; Voice-recognition radio remote system; Voice-recognition TV remote system; Smart light switch; Voice-recognition alarm setting

Vulnerable groups were asked questions about digital literacy [28] (ie, using the internet or smartphones and their willingness to use these devices). Questions about perceptions on knowledge of IoT, actual use of IoT, and willingness to use IoT services were then asked. Lastly, participants were asked to choose which IoT services they needed the most on a scale of 1 (*the most needed*) to 10 (*the tenth-most needed*). Caregivers and health care providers were asked the same questions from the points of view of the vulnerable groups.

Questions about the following characteristics were asked: baseline socioeconomic status, age, and sex of vulnerable group members; caregiver status (ie, age, sex, number of visits per week, and average time of stay per visit); living area (ie, rural or urban); type of dwelling (ie, apartment, house, or semibasement house); monthly income (ie, <US $1000, US $1000-$2000, or >US $2000); literacy level (ie, proficient, basic-to-intermediate, or below basic); and television-watching behavior (ie, all the time vs as needed). To evaluate disability, we used validated scales for Koreans to assess dependence in six activities of daily living (ADL)—toileting, feeding, dressing, grooming, physical ambulation, and bathing—and in eight instrumental activities of daily living (IADL)—using the telephone, shopping, food preparation, housekeeping, laundry, using public transportation, taking personal medication, and the ability to handle finances [29]. ADL and IADL scores were rated on a scale of 0-100: 0 (*total dependence*), 25 (*extensive assistance*), 50 (*limited assistance*), 75 (*supervision only*), and 100 (*independent*). Disability was defined if assistance was required from another person in performing any of the above activities in ADL and/or IADL [29].

Information was obtained on comorbidities (ie, hypertension, hyperlipidemia, diabetes, cardiovascular disease, cerebrovascular disease, thyroid disease, biliary disease, osteoporosis, chronic liver disease, renal or urinary tract stone, asthma, tuberculosis, gastric or duodenal ulcer, gout, osteoarthritis, rheumatoid arthritis, depression, dementia, Parkinson disease, and cancer) and clinical symptoms (ie, fatigue, respiratory symptoms, constipation, poor oral intake, palpitation or chest discomfort, headache, dizziness, falls, weight loss, edema, depression, anxiety, insomnia or sleep disorders, memory loss, wandering, destructive behavior, and hallucinations). Levels in hearing (*bad*, *average*, or *good*), vision (*bad*, *average*, or *good*), and speech (*bad*, *average*, or *good*) were self-reported. Sensory disability was defined if the level of either hearing, vision, or speech was *bad*.

When comparing the IoT needs regarding underlying conditions and disability, we grouped the number of total ADL disabilities, number of total IADL disabilities, number of comorbidities, number of clinical symptoms, and number of triple (ie, hearing, visual, and speech) impairments into three groups: 0 (*no deficit*), 1-X (*mild-to-moderate deficit*), and >X (*severe deficit*), where the value of X is different for each impairment.

Statistical Analysis

Analyses were conducted using R software, version 3.5.3 (The R Foundation), and Microsoft Excel, version 2016. Descriptive statistics for proportions of respondents, work profiles (eg, work experience, area of expertise, and institution), and responses regarding data demand, data linking, and deidentification were explored. The descriptive analysis examined differences in terms of five categories of IoT needs and deidentification processes. In the case of continuous data, we used the one-way analysis of variance (ANOVA) test to identify the differences in IoT needs between the groups. In the case of categorical data, we looked at the differences between the groups using the chi-square test. In general, when the expectation frequency was small, the Fisher exact test was used, except when there were more than three categories. The weighted-ranking method was used to rank the IoT needs in different groups. We asked for up to 10 responses per person for IoT needs. The higher the rank number, the lower the weight value, and vice versa: for example, the most important IoT needs have the lowest rank number (ie, 1), so are given the highest weight (ie, 10); on the other hand, the least important IoT needs have the highest rank number (ie, 10), so are given the lowest weight (ie, 1). Also, the weight was multiplied by the number of people who actually selected the services, and the value of the weighted sum was ranked.

Results

Socioeconomic Characteristics of Vulnerable Groups

The socioeconomic characteristics of the older adults and disabled people are seen in Table 2. The number of visits per week (mean 6.4, SD 1.6, vs mean 5.8, SD 2.2) and average hours per stay of the main caregivers (mean 19.7, SD 7.7, vs mean 17.8, SD 9.7) were significantly higher in the disabled people's group compared to the older adults' group. A total of 64.0% (128/200) of older adults lived in rural areas compared

to 50.0% (50/100) of disabled people. The majority of people lived in houses in both groups. The majority (144/200, 72.0%) of older adults had monthly incomes of less than US $1000, whereas in the disabled people's group, 35.0% (35/100) had monthly incomes of more than US $2000 and 25.0% (25/100) had monthly incomes of less than US $1000. Literacy levels were higher among the disabled people's group compared to the older adults' group (93/100, 93.0%, vs 131/200, 65.5%). Older adults tended to watch more television than disabled people. All of the results were statistically significant ($P<.05$).

Table 2. Socioeconomic characteristics of the vulnerable groups.

Characteristic	Older adults (n=200)	Disabled people (n=100)	*P* value
Total group members (n=300), n (%)	200 (66.7)	100 (33.3)	
Caregiver, mean (SD)			
Number of visits per week	5.8 (2.2)	6.4 (1.6)	.008
Average time of stay per visit (hours)	17.8 (9.7)	19.7 (7.7)	.07
Living area (rural), n (%)	128 (64.0)	50 (50.0)	.03
Type of dwelling, n (%)			**<.001**
Apartment	39 (19.5)	35 (35.0)	
House	157 (78.5)	52 (52.0)	
Semibasement house	3 (1.5)	6 (6.0)	
Monthly income (US $), n (%)			**<.001**
<1000	144 (72.0)	25 (25.0)	
1000-2000	14 (7.0)	18 (18.0)	
>2000	23 (11.5)	35 (35.0)	
Literacy level, n (%)			**<.001**
Proficient	131 (65.5)	93 (93.0)	
Basic-to-intermediate	34 (17.0)	4 (4.0)	
Below basic	32 (16.0)	0 (0)	
Television-watching behavior, n (%)			**.002**
All the time	95 (47.5)	26 (26.0)	
As needed	86 (43.0)	57 (57.0)	

Perception of Digital Literacy

The difference in perception of digital literacy based on the internet, smartphone use, and IoT is shown among the two vulnerable groups, their caregivers, and health care providers (see Table 3). The mean ages of older adults and disabled people were 78.13 years (SD 6.00) and 52.65 years (SD 10.19), respectively. A total of 33.5% (67/200) of older adults were male compared to 44.0% (44/100) of disabled people. Older adults tended to use the internet (22/200, 11.0%, vs 62/100, 62.0%) and smartphones (78/200, 39.0%, vs 88/100, 88.0%) less than did the disabled people. However, the proportion of older adults and disabled people willing to use the internet (81/200, 40.5%, vs 72/100, 72.0%) or smartphones (115/200, 57.5%, vs 62/100, 62.0%) was relatively high. Each of their caregivers replied similarly, showing matching trends.

In terms of IoT, the proportion of older adults that had heard of IoT was much lower compared to disabled people (25/200, 12.5%, vs 66/100, 66.0%). Only 1.0% (2/200) of older adults were currently using IoT services, whereas 20.0% (20/100) of disabled people were using them. The proportion of older adults willing to use IoT services in the future increased up to 57.5% (115/200) compared to 90.0% (90/100) of the disabled people. Each of their caregivers replied similarly, showing resembling trends.

The IoT needs in total and in five categories were presented as multiple-choice questions. The older adults indicated that security (125/384, 32.6%) was most needed, followed by safety (112/384, 29.2%) and then convenience (operation; 99/384, 25.8%). Their caregivers replied in the same order. The disabled group members indicated that security (76/254, 29.9%) was most needed, followed by convenience (operation; 61/254, 24.0%) and then safety (52/254, 20.5%). Their caregivers replied slightly differently. The health care providers also indicated that security (46/191, 24.1%) was most needed, followed by safety (45/191, 23.6%) and then convenience (operation; 37/191, 19.4%).

Table 3. Digital literacy and Internet of Things (IoT) needs of the three groups.

Category	Vulnerable group (n=300)		Caregivers (n=150)		Health care providers (n=50)	Total participants (N=500)
	Older adults (n=200)	Disabled people (n=100)	Older adults (n=100)	Disabled people (n=50)		
Age (years), mean (SD)	78.13 (6.00)	52.65 (10.19)	65.39 (13.45)	51.40 (15.68)	38.88 (9.34)	63.89 (17.00)
Gender (male), n (%)	67 (33.5)	44 (44.0)	51 (51.0)	16 (32)	5 (10)	183 (36.6)
ICT[a] perception, n (%)						
Able to use internet	22 (11.0)	62 (62.0)	14 (14.0)	34 (68)	N/A[b]	132 (29.3)[c]
Willing to use internet	81 (40.5)	72 (72.0)	46 (46.0)	37 (74)	N/A	236 (52.4)[c]
Able to use smartphone	78 (39.0)	88 (88.0)	42 (42.0)	44 (88)	N/A	252 (56.0)[c]
Willing to use smartphone	115 (57.5)	67 (67.0)	62 (62.0)	43 (86)	N/A	287 (63.8)[c]
Heard of IoT	25 (12.5)	66 (66.0)	7 (7.0)	26 (52)	N/A	124 (27.6)[c]
Currently using IoT	2 (1.0)	20 (20.0)	4 (4.0)	13 (26)	N/A	39 (8.7)[c]
Willing to use IoT	115 (57.5)	90 (90.0)	63 (63.0)	45 (90)	N/A	313 (69.6)[c]
IoT needs by service category, n (%)						
Total[d]	384 (100)	254 (100)	307 (100)	166 (100)	191 (100)	1302 (100)
Security	125 (32.6)	76 (29.9)	82 (26.7)	45 (27.1)	46 (24.1)	374 (28.73)
Safety	112 (29.2)	52 (20.5)	81 (26.4)	41 (24.7)	45 (23.6)	331 (25.42)
Health care	26 (6.8)	44 (17.3)	42 (13.7)	29 (17.5)	31 (16.2)	172 (13.21)
Convenience (information)	22 (5.7)	21 (8.3)	26 (8.5)	17 (10.2)	32 (16.8)	118 (9.06)
Convenience (operation)	99 (25.8)	61 (24.0)	76 (24.8)	34 (20.5)	37 (19.4)	307 (23.58)

[a]ICT: information and communication technology.

[b]N/A: not applicable.

[c]Total number of respondents was 450.

[d]Respondents could select multiple needs; individual category percentages are based on the total IoT needs in each column.

Internet of Things Needs by Category Depending on Underlying Conditions and Disability Types in Vulnerable Groups

Tables 4 and 5 show the difference in underlying characteristics and conditions in terms of IoT category needs in vulnerable groups. In older adults, people with lower incomes paid more attention to security and safety, whereas people with higher incomes paid more attention to health care or convenience (information), and these findings were statistically significant (see Table 4). Among the people who were physically disabled, those with lower incomes paid more attention to safety and convenience (operating), whereas people with higher incomes paid more attention to health care, however, these results were not significant (see Table 5). In older adults, variations in smartphone use, ADL, and IADL caused significant differences in IoT service category needs. However, among physically disabled people, there were no significant differences.

Figures 1 and 2 show IoT needs—multiple replies were possible—depending on disability type among both vulnerable groups. Among both older adults (see Figure 1) and disabled people (see Figure 2), those with mild-to-moderate disabilities were more willing to use IoT services than those with severe disabilities. People with no sensory disabilities (ie, triple impairment in hearing, vision, and speech) were more willing to use IoT services than people with mild-to-moderate sensory disabilities in both groups.

Table 4. Differences in underlying characteristics and conditions in terms of Internet of Things (IoT) category needs in older adults.

Variables	IoT needs by service category (n=384 total selections), n (%)					*P* value
	Security (n=125)	Safety (n=112)	Health care (n=26)	Convenience (information) (n=22)	Convenience (operating) (n=99)	
Age (years), mean (SD)	78.26 (5.76)	78.33 (5.78)	73.54 (5.36)	76.59 (6.77)	78.88 (5.75)	.001
Gender (male)	41 (32.8)	34 (30.4)	15 (58)	13 (59)	31 (31)	.009
Living area						**<.001**
Rural	81 (64.8)	82 (73.2)	6 (23)	1 (4)	72 (73)	
Urban	44 (35.2)	30 (26.8)	20 (77)	21 (95)	27 (27)	
Type of dwelling						**<.001**
Apartment	22 (17.6)	17 (15.2)	11 (42)	12 (54)	15 (15)	
House	103 (82.4)	95 (84.8)	15 (58)	10 (45)	84 (85)	
Monthly income (US $)						**<.001**
<1000	102 (81.6)	94 (83.9)	15 (58)	11 (50)	87 (88)	
1000-2000	8 (6.4)	5 (4.5)	3 (11)	6 (27)	5 (5)	
>2000	15 (12.0)	13 (11.6)	8 (31)	5 (23)	7 (7)	
Literacy level						**.28**
Below basic	23 (18.4)	23 (20.5)	2 (8)	1 (4)	18 (18)	
Basic-to-intermediate	22 (17.6)	23 (20.5)	2 (8)	2 (9)	18 (18)	
Proficient	80 (64.0)	66 (58.9)	22 (85)	19 (86)	63 (64)	
Television-watching behavior						**.07**
As needed	53 (42.4)	42 (37.5)	14 (54)	12 (54)	38 (38)	
Do not watch	12 (9.6)	8 (7.1)	4 (15)	5 (23)	8 (8)	
All the time	60 (48.0)	62 (55.4)	8 (31)	5 (23)	53 (53)	
ICT[a] perception						
Able to use internet	12 (9.6)	8 (7.1)	4 (15)	4 (18)	6 (6)	.23
Willing to use internet	56 (44.8)	52 (46.4)	17 (65)	12 (54)	43 (43)	.27
Able to use smartphone	50 (40.0)	40 (35.7)	22 (85)	11 (50)	34 (34)	<.001
Willing to use smartphone	75 (60.0)	69 (61.6)	15 (58)	8 (36)	58 (59)	.29
Able to use IoT	1 (0.8)	1 (0.9)	0 (0)	0 (0)	1 (1)	.98
Willing to use IoT	73 (58.4)	75 (67.0)	16 (61)	10 (45)	61 (62)	.37
ADL[b] deficits						**.04**
None (0)	20 (16.0)	15 (13.4)	10 (38)	7 (32)	11 (11)	
Mild-to-moderate (1-3)	76 (60.8)	72 (64.3)	11 (42)	11 (50)	62 (63)	
Severe (4-6)	29 (23.2)	25 (22.3)	5 (19)	4 (18)	26 (26)	
IADL[c] deficits						**.004**
None (0)	24 (19.2)	16 (14.3)	8 (31)	8 (36)	12 (12)	
Mild-to-moderate (1-4)	74 (59.2)	75 (67.0)	11 (42)	5 (23)	68 (69)	
Severe (5-8)	27 (21.6)	21 (18.8)	7 (27)	9 (41)	19 (19)	
Clinical symptoms						**.11**
None (0)	12 (9.6)	9 (8.0)	4 (15)	5 (23)	7 (7)	
Mild-to-moderate (1-9)	82 (65.6)	74 (66.1)	20 (77)	15 (68)	62 (63)	
Severe (>10)	31 (24.8)	29 (25.9)	2 (8)	2 (9)	30 (30)	

Variables	IoT needs by service category (n=384 total selections), n (%)					*P* value
	Security (n=125)	Safety (n=112)	Health care (n=26)	Convenience (information) (n=22)	Convenience (operating) (n=99)	
Comorbidity						**.07**
None (0)	5 (4.0)	3 (2.7)	1 (4)	1 (4)	3 (3)	
Mild-to-moderate (1-3)	49 (39.2)	42 (37.5)	16 (61)	15 (68)	35 (35)	
Severe (4-6)	71 (56.8)	67 (59.8)	9 (35)	6 (27)	61 (62)	
Triple impairment[d]						**.13**
None (0)	72 (57.6)	68 (60.7)	20 (77)	9 (41)	61 (62)	
Mild-to-moderate (1-2)	52 (41.6)	43 (38.4)	5 (19)	12 (54)	38 (38)	
Severe (3)	1 (0.8)	1 (0.9)	1 (4)	1 (4)	0 (0)	

[a]ICT: information communication and technology.

[b]ADL: activities of daily living.

[c]IADL: instrumental activities of daily living.

[d]Hearing, visual, and speech impairments.

Table 5. Differences in underlying characteristics and conditions in terms of Internet of Things (IoT) category needs in disabled people.

Variables	IoT needs by service category (n=254 total selections), n (%)					*P* value
	Security (n=76)	Safety (n=52)	Health care (n=44)	Convenience (information) (n=21)	Convenience (operating) (n=61)	
Age (years), mean (SD)	53.11 (10.10)	52.90 (10.37)	51.77 (9.81)	53.52 (9.02)	51.75 (10.85)	.89
Gender (male)	34 (45)	23 (44)	10 (23)	7 (33)	28 (46)	.10
Living area						**.26**
Rural	42 (55)	33 (63)	20 (45)	8 (38)	33 (54)	
Urban	34 (45)	19 (36)	24 (54)	13 (62)	28 (46)	
Type of dwelling						**<.001**
Apartment	28 (37)	19 (36)	22 (50)	12 (57)	23 (38)	
House	46 (60)	33 (63)	21 (48)	8 (38)	37 (61)	
Income (US $)						**.63**
<1000	33 (43)	26 (50)	14 (32)	8 (38)	29 (47)	
1000-2000	15 (20)	11 (21)	9 (20)	6 (29)	14 (23)	
>2000	28 (37)	15 (29)	21 (48)	7 (33)	18 (29)	
Literacy level						**.98**
Below basic	1 (1)	1 (2)	1 (2)	0 (0)	1 (2)	
Basic-to-intermediate	4 (5)	4 (8)	2 (4)	0 (0)	4 (7)	
Proficient	71 (93)	47 (90)	41 (93)	21 (100)	56 (92)	
Television-watching behavior						**.56**
As needed	43 (57)	32 (61)	31 (70)	15 (71)	34 (56)	
Do not watch	12 (16)	5 (10)	7 (16)	3 (14)	11 (18)	
All the time	21 (28)	15 (29)	6 (14)	3 (14)	16 (26)	
ICT[a] perception						
Able to use internet	54 (71)	36 (69)	30 (68)	16 (76)	43 (70)	.97
Willing to use internet	56 (74)	40 (77)	28 (64)	14 (67)	43 (70)	.80
Able to use smartphone	67 (88)	43 (83)	40 (91)	20 (95)	52 (85)	.56
Willing to use smartphone	52 (68)	39 (75)	27 (61)	13 (62)	42 (69)	.85
Able to use IoT	12 (16)	7 (13)	8 (18)	4 (19)	14 (23)	.76
Willing to use IoT	68 (89)	46 (88)	41 (93)	20 (95)	57 (93)	.82
ADL[b] deficits						**.45**
None (0)	30 (39)	22 (42)	22 (50)	12 (57)	22 (36)	
Mild-to-moderate (1-3)	28 (37)	21 (40)	13 (29)	3 (14)	26 (43)	
Severe (4-6)	18 (24)	9 (17)	9 (20)	6 (29)	13 (21)	
IADL[c] deficits						**.68**
None (0)	24 (32)	18 (35)	19 (43)	10 (48)	20 (33)	
Mild-to-moderate (1-4)	32 (42)	23 (44)	14 (32)	5 (24)	28 (46)	
Severe (5-8)	20 (26)	11 (21)	11 (25)	6 (29)	13 (21)	
Clinical symptoms						**.81**
None (0)	7 (9)	5 (10)	6 (14)	0 (0)	5 (8)	
Mild-to-moderate (1-9)	63 (83)	44 (85)	34 (77)	20 (95)	50 (82)	
Severe (>10)	6 (8)	3 (6)	4 (9)	1 (5)	6 (10)	

Variables	IoT needs by service category (n=254 total selections), n (%)					*P* value
	Security (n=76)	Safety (n=52)	Health care (n=44)	Convenience (information) (n=21)	Convenience (operating) (n=61)	
Comorbidity						**.68**
None (0)	11 (14)	7 (13)	7 (16)	4 (19)	10 (16)	
Mild-to-moderate (1-3)	54 (71)	36 (69)	31 (70)	17 (81)	46 (75)	
Severe (4-6)	11 (14)	9 (17)	6 (14)	0 (0)	5 (8)	
Triple impairment[d]						**.95**
None (0)	48 (63)	34 (65)	31 (70)	14 (67)	39 (64)	
Mild-to-moderate (1-2)	28 (37)	18 (35)	13 (29)	7 (33)	22 (36)	
Severe (3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	

[a]ICT: information communication and technology.

[b]ADL: activities of daily living.

[c]IADL: instrumental activities of daily living.

[d]Hearing, visual, and speech impairments.

Figure 1. The total number of Internet of Things (IoT) needs by category, as a function of underlying conditions in older adults. ADL: activities of daily living; IADL: instrumental activities of daily living; Triple impairment: hearing, vision, and speech impairments.

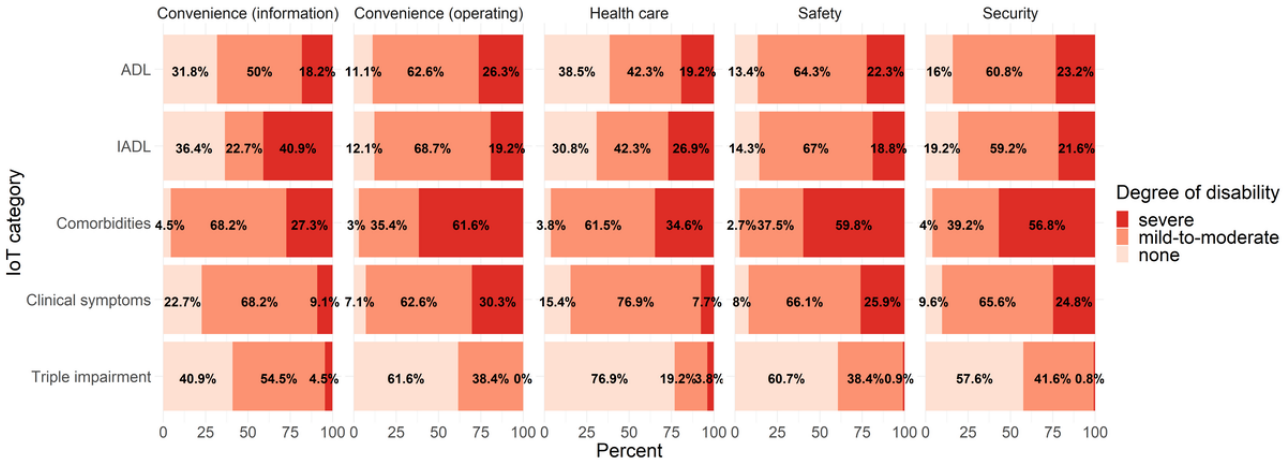
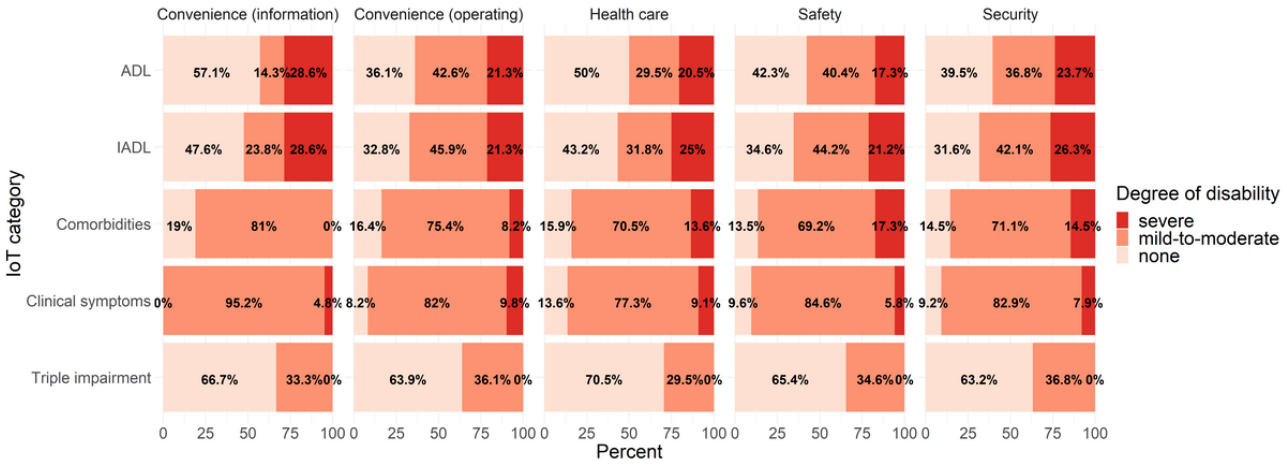


Figure 2. The total number of Internet of Things (IoT) needs by category, as a function of underlying conditions in disabled people. ADL: activities of daily living; IADL: instrumental activities of daily living; Triple impairment: hearing, vision, and speech impairments.



Weighted Rankings of Internet of Things Needs Depending on Different Groups

Figure 3 shows weighted rankings of IoT needs depending on various groups. The top-three IoT services (ie, ranked 1-3) for older adults were *smart home SOS bell*, *home security and closed-circuit television (CCTV)*, and *smart band and mobile SOS bell services*; these were within the emergency and security category. Although the order of the service rankings were slightly different, their caregivers replied similarly in both groups. The services ranked 4-6 were also similar in both groups. The services they selected were all in the safety categories: *IoT-based power-system protection device*, *IoT-based smart gas monitoring*, and *GPS tracker service*.

The top-two IoT services (ie, ranked 1 and 2), for disabled people were *smart band and mobile SOS bell* and *home security and CCTV*, which were both in the emergency and security category. The service ranked as number 3 was *IoT-based smart gas monitoring*, which was in the safety category. Their caregivers replied slightly differently. The *home security and CCTV* service topped the ranks in the emergency and security category; the *IoT-based smart gas monitoring* service was ranked as number 2 and *IoT-based power-system protection device* was ranked as number 3, both of which were in the safety category.

In the health care provider group, participants' top-two IoT services (ie, ranked 1 and 2) were *smart gas monitoring* and *GPS tracker*, both in the safety category, which was different from the other groups. The service ranked as number 3 was *smart home SOS bell* in the emergency and security category.

Figure 3. Weighted rankings of Internet of Things (IoT) services and needs depending on different groups. CCTV: closed-circuit television.

Definition of each color for IoT service categories					
Color	Emergency and security	Safety	Health care	Convenience [operating]	Convenience [information]
Pattern of IoT needs and service categories					
Rank	**Older adults (n=100)**	**Caregiver of older adults (n=100)**	**Disabled people (n=100)**	**Caregiver of disabled people (n=50)**	**Health care provider (n=50)**
1	Smart home *SOS* bell	Smart home *SOS* bell	Smart band and mobile *SOS* bell	Home security and CCTV	IoT-based smart gas monitoring
2	Home security and CCTV	Smart band and mobile *SOS* bell	Home security and CCTV	IoT-based smart gas monitoring	GPS trackers
3	Smart band and mobile *SOS* bell	Home security and CCTV	IoT-based smart gas monitoring	IoT-based power system protection device	Smart home *SOS* bell
4	IoT-based power-system protection device	IoT-based smart gas monitoring	Smart home *SOS* bell	Smart band and mobile *SOS* bell	Taxi call
5	IoT-based smart gas monitoring	IoT-based power-system protection device	Robot vacuum cleaner	Smart home *SOS* bell	Home security and CCTV
6	GPS trackers	GPS trackers	IoT-based power-system protection device	Front door smart sensor	Doctor's appointments
7	Smart light switch	IoT-based smart gas monitoring	Fitness on television	Robot vacuum cleaner	IoT-based power-system protection device
8	Robot vacuum cleaner	Taxi call	Front door smart sensor	IoT-based smart gas monitoring	Smart band and mobile *SOS* bell
9	IoT-based smart gas monitoring	Doctor's appointments	Voice-recognition door lock	Voice-recognition door lock	Weather forecast
10	Front door smart sensor	Smart light switch	Smart light switch	Smart light switch	IoT-based smart gas monitoring

Discussion

Principal Findings

The purpose of this study was to conduct a face-to-face survey on the demand for IoT among older people and people with disabilities, their caregivers, and health care providers in the real-world setting and to see if there are any differences in the aspects of need. The primary finding of this study was that IoT service needs were different among the vulnerable groups, their caregivers, and health care providers. The most required IoT service category selected by the vulnerable groups and their caregivers was security. Meanwhile, health care providers decided that IoT services in the safety category were most needed by the vulnerable groups. In addition, IoT service preferences differed according to various types of disabilities in the vulnerable groups.

Smart home technology has been anticipated to be at the front line of individualized health care, yet there are still hurdles that prevent home IoT technology from spreading among vulnerable populations. Various studies have addressed learning and adherence issues as the main problems [30-32]. Even with well-targeted populations, wearable-device studies have reported only 10% use within one year of incentive shrinkage [31]. Other studies also showed that the adoption rates of IoT services and devices are still very low and were reported to be around 5%-15% in older adults and physically disabled people [30]. However, previous studies have been focused on the technical aspects or digital literacy only in determining the causes of low

adoption. In our study, we showed that the percentage of IoT use in older adults and physically disabled people was still very low regardless of underlying literacy or digital literacy levels. The reason for this was mainly due to lack of knowledge and exposure to the IoT devices compared with internet or smartphone exposure, especially in older adults. Therefore, education on using and experiencing these devices is needed to improve adoption, as our results showed that there is an interest and willingness to use IoT devices among the vulnerable groups.

One of our study's strengths is that we explored the different perspectives of IoT demands that have been subject to stereotypes and insufficiently studied in real-world settings [17,33]. We showed the actual needed IoT service categories and rankings that the users, their caregivers, and health care providers chose in person. Vulnerable people are very different physically, mentally, and environmentally from the general population; therefore, patient-centered approaches should be emphasized as they may improve outcomes for people with multiple chronic conditions [34]. We showed the users' priorities based on their socioeconomic status, literacy levels, digital literacy levels, underlying disabilities, and remnant physical performances, which provide a comprehensive view for practical implementation of IoTs in older adults and disabled people.

Furthermore, our study results showed the IoT service needs from the perspective of the caregivers and health care providers, which is also important. Older adults and physically disabled people often rely on caregivers and health care providers for validation of behaviors, including purchase and use of technology [35]. Also, IoT-based systems can share information with them so they can intervene in case of emergency and provide support [36]. Our results were different in that the vulnerable groups were willing to try new technology, quite contrary to the social perception. We also showed that the IoT service needs of the caregivers who were mainly family were in agreement with the vulnerable group members, which proves that the caregivers were sensitive to the specific needs of their care recipients. Our results may have differed if most of the caregivers had been paid. Paid caregivers may have a different financial perspective than family members and, therefore, may have replied differently. Interestingly, the health care providers replied differently, which shows that their demands deviate from the needs of the vulnerable people and their families. We need to be aware of these discrepancies when recommending and applying IoT devices to vulnerable users.

Internet of Things Service Categories in Each Group

The most required IoT services chosen by the vulnerable groups and their caregivers in our study were within the security category. It has been reported that vulnerable groups value independence, privacy, and social interactions, while they have negative impressions about personal emergency alarms because they are obtrusive and even shameful and they dislike being watched [37-40]. Our study showed otherwise. There was a strong need for security among vulnerable people and their family caregivers. Vulnerable people are physically less mobile and their activities mostly take place within the home environment. They can fear loneliness and isolation; however, they wish to remain independent as long as possible [41]. Therefore, the security services they selected, especially *SOS* alarms and wearables, can act as a backup plan for self-management and can be used in emergency situations to notify family members or caregivers and providers [42-45].

The health care providers were focused more on the safety category. The reason the vulnerable groups did not choose this service as their top priority in our study could be that they are still relatively independent and so it did not meet their demands [38,39]. This shows that health care providers did not fully comprehend the needs of the vulnerable groups and may have stereotyped them as all needing support in their daily activities. Health care providers should be aware of these discrepancies and, therefore, consider a patient-centered approach when considering IoT services and devices.

In all of the groups, services related to health care were the least popular. Previous IoT solutions in these vulnerable groups had been mainly designed for health monitoring, such as monitoring medical parameters, activity level, medical compliance, nutrition, fitness, and sleep [36,46-48]. However, health care–related services in our list were not favored by the users. Unfortunately, due to the Personal Information Protection Act and the Medical Service Law enacted in 2011 in South Korea, we are unable to use health care devices for remote monitoring regarding medical information security. Therefore, we could only select commercially available IoT devices that were related to health care and somewhat less common. Our results may have differed if the devices had been directly monitoring health. IoT regulation depends on the country and different domestic circumstances need to be acknowledged.

Internet of Things Needs Vary Depending on Different Types of Disabilities

In our study, IoT needs varied due to the different combinations of disabilities. People with mild-to-moderate disabilities tended to respond more to needing IoT services compared to people with either no disabilities or severe disabilities. This indicates that, perhaps for people with severe disabilities, IoT services were too difficult to use or participants were too frail and needed continued care [49]. On the other hand, it is possible that people with no disabilities did not necessarily need the IoT services. Many of the IoT services designed to be used do not usually consider that the functional limitations of each user are different [42]. Therefore, the key challenge is IoT customization for older adults and people with disabilities. Individualized, comprehensive, functional assessment among vulnerable people to analyze their underlying conditions, functional status, and disabilities are recommended. We are currently undergoing trials to apply IoT devices for use among vulnerable people based on these assessments.

Limitations

The main limitation of this study is that the respondents were mostly people with mild functional disabilities. However, the purpose of IoT services is to enhance usability among vulnerable people with mild functional disabilities before their conditions deteriorate. Although our study may not directly represent the opinions of the aging or disabled population, it does illuminate, through a cross-sectional approach, the present status of home

IoT needs from the perspectives of users and their families in the real world. Further investigation into groups of people with severe functional disabilities is needed to represent the overall opinions of the population.

Conclusions

Our survey study shows that there were inconsistencies in the demand of IoT services among vulnerable groups, their caregivers, and health care providers. IoT service requirements differed according to the various types of disabilities. Home IoT technology should be established by combining patients' priorities and individualized functional assessments of vulnerable people in an environment where patient-centered approaches and collaborative decision making are emphasized. This information and future trial data can inform public health professionals and industry workers in designing home IoT services for vulnerable populations.

Acknowledgments

This study was supported by the Social Service R&D Project funded by the Ministry of Health and Welfare, Republic of Korea, through the Korea Health Industry Institute (KHIDI; HI18C2400).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire for older adults and physically disabled people.
[DOCX File , 27 KB - jmir_v22i4e16614_app1.docx]

References

1. Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Comput Netw 2010 Oct;54(15):2787-2805. [doi: 10.1016/j.comnet.2010.05.010]
2. Iecovich E. Aging in place: From theory to practice. Anthropol Noteb 2014;20(1):21-33 ISSN 1408-032X [FREE Full text]
3. Carnemolla P. Ageing in place and the Internet of Things: How smart home technologies, the built environment and caregiving intersect. Vis Eng 2018 Dec 6;6:1-16 [FREE Full text] [doi: 10.1186/s40327-018-0066-5]
4. Majumder S, Aghayi E, Noferesti M, Memarzadeh-Tehran H, Mondal T, Pang Z, et al. Smart homes for elderly healthcare: Recent advances and research challenges. Sensors (Basel) 2017 Oct 31;17(11):2496 [FREE Full text] [doi: 10.3390/s17112496] [Medline: 29088123]
5. Majumder S, Mondal T, Deen M. Wearable sensors for remote health monitoring. Sensors (Basel) 2017 Jan 12;17(1):130 [FREE Full text] [doi: 10.3390/s17010130] [Medline: 28085085]
6. Domingo MC. An overview of the Internet of Things for people with disabilities. J Netw Comput Appl 2012 Mar;35(2):584-596. [doi: 10.1016/j.jnca.2011.10.015]
7. Das R, Tuna A, Demirel S, Yurdakul MK. A survey on the Internet of Things solutions for the elderly and disabled: Applications, prospects, and challenges. Int J Comput Netw Appl 2017 Jun 30;4(3):84-92 ISSN: 2395-0455 [FREE Full text] [doi: 10.22247/ijcna/2017/49023]
8. World Health Organization. Health Education Service. Health education in self-care: Possibilities and limitations. Geneva, Switzerland: World Health Organization; 1984. URL: https://apps.who.int/iris/bitstream/handle/10665/70092/HED_84.1.pdf?sequence=1&isAllowed=y [accessed 2020-04-05]
9. Ray PP. Home Health Hub Internet of Things (H3IoT): An architectural framework for monitoring health of elderly people. In: Proceedings of the International Conference on Science Engineering and Management Research (ICSEMR). New York, NY: IEEE; 2014 Presented at: International Conference on Science Engineering and Management Research (ICSEMR); November 27-29, 2014; Chennai, India p. 1-3 URL: https://ieeexplore.ieee.org/document/7043542 [doi: 10.1109/ICSEMR.2014.7043542]
10. Siegel C, Dorner TE. Information technologies for active and assisted living: Influences to the quality of life of an ageing society. Int J Med Inform 2017 Apr;100:32-45. [doi: 10.1016/j.ijmedinf.2017.01.012] [Medline: 28241936]
11. 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]
12. AbdElnapi NMM, Omran NF, Ali AA, Omara FA. A survey of Internet of Things technologies and projects for healthcare services. In: Proceedings of the International Conference on Innovative Trends in Computer Engineering (ITCE). New York, NY: IEEE; 2018 Presented at: International Conference on Innovative Trends in Computer Engineering (ITCE); February 19-21, 2018; Aswan, Egypt p. 48-55 URL: https://ieeexplore.ieee.org/abstract/document/8316599 [doi: 10.1109/itce.2018.8316599]
13. Chi N, Demiris G. The roles of telehealth tools in supporting family caregivers: Current evidence, opportunities, and limitations. J Gerontol Nurs 2017 Feb 01;43(2):3-5. [doi: 10.3928/00989134-20170111-04] [Medline: 28128391]

14. Coughlin JF, Pope J. Innovations in health, wellness, and aging-in-place: Development of a consumer-centered approach to intelligent home services. IEEE Eng Med Biol Mag 2008 Jul;27(4):47-52 [FREE Full text] [doi: 10.1109/memb.2008.919493]
15. Wilson C, Hargreaves T, Hauxwell-Baldwin R. Benefits and risks of smart home technologies. Energy Policy 2017 Apr;103:72-83. [doi: 10.1016/j.enpol.2016.12.047]
16. Balta-Ozkan N, Davidson R, Bicket M, Whitmarsh L. Social barriers to the adoption of smart homes. Energy Policy 2013 Dec;63:363-374. [doi: 10.1016/j.enpol.2013.08.043]
17. Demiris G, Rantz MJ, Aud MA, Marek KD, Tyrer HW, Skubic M, et al. Older adults' attitudes towards and perceptions of "smart home" technologies: A pilot study. Med Inform Internet Med 2004 Jun;29(2):87-94. [doi: 10.1080/14639230410001684387] [Medline: 15370989]
18. Lee C, Coughlin JF. PERSPECTIVE: Older adults' adoption of technology: An integrated approach to identifying determinants and barriers. J Prod Innov Manage 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
19. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med 1995 Mar 02;332(9):556-561. [doi: 10.1056/NEJM199503023320902] [Medline: 7838189]
20. Pavasini R, Guralnik J, Brown JC, di Bari M, Cesari M, Landi F, et al. Short Physical Performance Battery and all-cause mortality: Systematic review and meta-analysis. BMC Med 2016 Dec 22;14(1):215 [FREE Full text] [doi: 10.1186/s12916-016-0763-7] [Medline: 28003033]
21. Jung H, Jang I, Lee YS, Lee CK, Cho E, Kang WY, et al. Prevalence of frailty and aging-related health conditions in older Koreans in rural communities: A cross-sectional analysis of the aging study of Pyeongchang rural area. J Korean Med Sci 2016 Mar;31(3):345-352 [FREE Full text] [doi: 10.3346/jkms.2016.31.3.345] [Medline: 26952571]
22. Jang I, Jung H, Lee CK, Lee YS, Kim K, Kim KW, et al. Rural and urban disparities in frailty and aging-related health conditions in Korea. J Am Geriatr Soc 2016 Apr;64(4):908-911 [FREE Full text] [doi: 10.1111/jgs.14074] [Medline: 27100601]
23. Chang M, Cohen-Mansfield J, Ferrucci L, Leveille S, Volpato S, de Rekeneire N, et al. Incidence of loss of ability to walk 400 meters in a functionally limited older population. J Am Geriatr Soc 2004 Dec;52(12):2094-2098. [doi: 10.1111/j.1532-5415.2004.52570.x] [Medline: 15571549]
24. Korean Longitudinal Study of Aging (KLoSA): Employment Survey. Seoul, Korea: Korea Employment Information Service; 2016. URL: http://survey.keis.or.kr [accessed 2017-12-01]
25. Korea Health Statistics 2011: Korea National Health and Nutrition Examination Survey (KNHANES V-2). Cheongju, South Korea: Ministry of Health of Welfare, Korea Centers for Disease Control and Prevention; 2012. URL: http://knhanes.cdc.go.kr/ [accessed 2012-12-31]
26. Korean Institute for Health and Social Affairs. Current Status of the Elderly in 2017. Sejong, South Korea: Korean Institute for Health and Social Affairs; 2017. URL: http://wish.welfare.seoul.kr/upload/data/_20180601142141142.pdf [accessed 2020-04-05]
27. 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 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
28. Eshet-Alkalai Y. Digital literacy: A conceptual framework for survival skills in the digital era. J Educ Multimed Hypermedia 2004;13(1):93-106 ISSN: ISSN-1055-8896 [FREE Full text]
29. Won C, Yang K, Rho Y, Kim S, Lee E, Yoon J. The development of Korean Activities of Daily Living (K-ADL) and Korean Instrumental Activities of Daily Living (K-IADL) scale. J Korean Geriatr Soc 2002;6:107-120 [FREE Full text]
30. Lau J. Building a National Technology and Innovation Infrastructure for an Aging Society [master's thesis]. Boston, MA: Massachusetts Institute of Technology; 2006 Sep. URL: https://dspace.mit.edu/bitstream/handle/1721.1/38566/154715153-MIT.pdf?sequence=2&isAllowed=y [accessed 2020-04-05]
31. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): A randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
32. Jang I, Kim HR, Lee E, Jung H, Park H, Cheon S, et al. Impact of a wearable device-based walking program in rural older adults on physical activity and health outcomes: Cohort study. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e11335 [FREE Full text] [doi: 10.2196/11335] [Medline: 30463838]
33. Niemelä‐Nyrhinen J. Baby boom consumers and technology: Shooting down stereotypes. J Consum Mark 2007 Aug 07;24(5):305-312. [doi: 10.1108/07363760710773120]
34. Tinetti ME, Naik AD, Dindo L, Costello DM, Esterson J, Geda M, et al. Association of patient priorities-aligned decision-making with patient outcomes and ambulatory health care burden among older adults with multiple chronic conditions: A nonrandomized clinical trial. JAMA Intern Med 2019 Oct 07;179(12):1688-1697. [doi: 10.1001/jamainternmed.2019.4235] [Medline: 31589281]
35. Wang A, Redington L, Steinmetz V, Lindeman D. The ADOPT Model: Accelerating Diffusion of Proven Technologies for older adults. Ageing Int 2011 Mar;36:29-45. [doi: 10.1007/s12126-010-9072-1]

36. Azimi I, Rahmani AM, Liljeberg P, Tenhunen H. Internet of Things for remote elderly monitoring: A study from user-centered perspective. J Ambient Intell Humaniz Comput 2017 Apr;8:273-289. [doi: 10.1007/s12652-016-0387-y]
37. Peek ST, Wouters EJ, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJ. Factors influencing acceptance of technology for aging in place: A systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
38. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
39. Walsh K, Callan A. Perceptions, preferences, and acceptance of information and communication technologies in older-adult community care settings in Ireland: A case-study and ranked-care program analysis. Ageing Int 2011 Mar;36:102-122. [doi: 10.1007/s12126-010-9075-y]
40. Stefanov D, Bien Z, Bang W. The smart house for older persons and persons with physical disabilities: Structure, technology arrangements, and perspectives. IEEE Trans Neural Syst Rehabil Eng 2004 Jun;12(2):228-250. [doi: 10.1109/TNSRE.2004.828423] [Medline: 15218937]
41. Boldy D, Grenade L, Lewin G, Karol E, Burton E. Older people's decisions regarding 'ageing in place': A Western Australian case study. Australas J Ageing 2011 Sep;30(3):136-142. [doi: 10.1111/j.1741-6612.2010.00469.x] [Medline: 21923707]
42. Celler BG, Lovell NH, Chan DK. The potential impact of home telecare on clinical practice. Med J Aust 1999 Nov 15;171(10):518-521. [Medline: 10816699]
43. Deen MJ. Information and communications technologies for elderly ubiquitous healthcare in a smart home. Pers Ubiquitous Comput 2015 Jul;19:573-599. [doi: 10.1007/s00779-015-0856-x]
44. Bottazzi D, Corradi A, Montanari R. Context-aware middleware solutions for anytime and anywhere emergency assistance to elderly people. IEEE Commun Mag 2006 Apr;44(4):82-90. [doi: 10.1109/mcom.2006.1632653]
45. van Kasteren TLM, Englebienne G, Kröse BJA. An activity monitoring system for elderly care using generative and discriminative models. Pers Ubiquitous Comput 2010 Sep;14:489-498 [FREE Full text] [doi: 10.1007/s00779-009-0277-9]
46. Ahmed M, Banaee H, Rafael-Palou X, Loutfi A. Intelligent healthcare services to support health monitoring of elderly. In: Proceedings of the 4th International Internet of Things Summit (IoT 2014). Cham, Switzerland: Springer; 2014 Presented at: 4th International Internet of Things Summit (IoT 2014); October 6-8, 2014; Cambridge, MA p. 178-186. [doi: 10.1007/978-3-319-19656-5_26]
47. Rivero-Espinosa J, Iglesias-Pérez A, Gutiérrez-Dueñas J, Rafael-Palou X. SAAPHO: An AAL architecture to provide accessible and usable active aging services for the elderly. ACM SIGACCESS Access Comput 2013 Sep(107):17-24 [FREE Full text] [doi: 10.1145/2535803.2535806]
48. Carús JL, García S, García R, Waterworth J, Erdt S. The ELF@Home project: Elderly sELF-care based on sELF-check of health conditions and sELF-fitness at home. Stud Health Technol Inform 2014;200:164-166. [Medline: 24851984]
49. Williams J, Hughes G, Blackwell S. Attitudes Towards Funding of Long-Term Care of the Elderly. Dublin, Ireland: Economic & Social Research Institute (ESRI); 2006 Mar 02. URL: https://www.esri.ie/system/files/media/file-uploads/2015-07/BKMNINT185_Main%20Text.pdf [accessed 2020-04-05]

Abbreviations

ADL: activities of daily living
ANOVA: analysis of variance
ASPRA: Aging Study of Pyeongchang Rural Area
CCTV: closed-circuit television
IADL: instrumental activities of daily living
IoT: Internet of Things
KHIDI: Korea Health Industry Institute
SPPB: Short Physical Performance Battery

Edited by E Borycki, G Eysenbach; submitted 09.10.19; peer-reviewed by L Luo, V Vennu, J Wei; comments to author 25.10.19; revised version received 12.01.20; accepted 23.03.20; published 15.04.20.

Please cite as:
Lee H, Park YR, Kim HR, Kang NY, Oh G, Jang IY, Lee E
Discrepancies in Demand of Internet of Things Services Among Older People and People With Disabilities, Their Caregivers, and Health Care Providers: Face-to-Face Survey Study
J Med Internet Res 2020;22(4):e16614
URL: http://www.jmir.org/2020/4/e16614/
doi:10.2196/16614
PMID:32293575

Original Paper

Deploying Patient-Facing Application Programming Interfaces: Thematic Analysis of Health System Experiences

Aaron Neinstein[1,2*], MD; Crishyashi Thao[3*], MPH; Mark Savage[2*], JD; Julia Adler-Milstein[1,3*], PhD

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

[2]Center for Digital Health Innovation, University of California, San Francisco, San Francisco, CA, United States

[3]Center for Clinical Informatics and Improvement Research, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[*]all authors contributed equally

Corresponding Author:
Aaron Neinstein, MD
Center for Digital Health Innovation
University of California, San Francisco
1700 Owens Street
Suite 541
San Francisco, CA, 94158
United States
Phone: 1 4154765397
Email: aaron.neinstein@ucsf.edu

Abstract

Background: Health systems have recently started to activate patient-facing application programming interfaces (APIs) to facilitate patient access to health data and other interactions.

Objective: This study sought to ascertain health systems' understanding, strategies, governance, and organizational infrastructure around patient-facing APIs, as well as their business drivers and barriers, to facilitate national learning, policy, and progress toward adoption.

Methods: We performed a content analysis of semistructured interviews with a convenience sample of 10 health systems known to be leading adopters of health technology, having either implemented or planning to implement patient-facing APIs.

Results: Of the 10 health systems, eight had operational patient-facing APIs, with organizational strategy driven most by federal policy, the emergence of Health Records on iPhone, and feelings of ethical obligation. The two priority use cases identified were enablement of a patient's longitudinal health record and digital interactions with the health system. The themes most frequently cited as barriers to the increased use of patient-facing APIs were security concerns, an immature app ecosystem that does not currently offer superior functionality compared with widely adopted electronic health record (EHR)–tethered portals, a lack of business drivers, EHR vendor hesitation toward data sharing, and immature technology and standards.

Conclusions: Our findings reveal heterogeneity in health system understanding and approaches to the implementation and use of patient-facing APIs. Ongoing study, targeted policy interventions, and sharing of best practices appear necessary to achieve successful national implementation.

(J Med Internet Res 2020;22(4):e16813) doi:10.2196/16813

KEYWORDS

application programming interface; consumer health information; electronic health record; health information exchange; health policy; patient engagement

Introduction

Background

An array of federal policy efforts seeks to create a more patient-centered, consumer-empowered health care system by improving patients' access to their electronic health information (EHI) [1,2]. Previous efforts to create patient-controlled health records (most famously, Google Health and Microsoft HealthVault) never gained traction, due in part to the lack of easily available, digital patient data [3]. Since then, the health

care system has digitized health data significantly, with 96% of hospitals and 78% of physician offices now using certified electronic health record (EHR) technology [4]. A major national survey of consumers shows that consumers want and use electronic access to their health information, including mobile access through smartphones and use of health apps [5]. In parallel, regulations now require that patients get timely (within 48 hours) [6] access to their health records to view, download, and transmit them electronically (as opposed to only being able to receive a paper copy). Beginning January 1, 2019, with stage 3 of the federal Medicare and Medicaid EHR Incentive Programs (now known as the Promoting Interoperability Program) and bolstered by provisions of the 21st Century Cures Act, health systems must give patients electronic access to their EHI via patient-facing application programming interfaces (APIs), "without special effort" [1,7].

These patient-facing APIs have now begun to go live, enabling patients to access their EHI as well as to direct access by third-party software apps [8]. This presents a potentially pivotal moment in the evolution of the US health care system because success could break open long-siloed patient data [9], leading to better progress in interoperability, enabling patients to create their own longitudinal health record, and realizing the potential of patient engagement via an emerging ecosystem of third-party apps. However, success is not a foregone conclusion, with health systems starting to pursue patient-facing APIs amidst uncertainty and complex issues spanning policy, reimbursement, technical implementation, governance, privacy, and security [10]. As health systems wrestle with numerous decisions about how to deploy patient-facing APIs, they risk doing so in silos, without the benefit of interorganizational learning [11]. This would limit the identification and sharing of emerging best practices, as well as challenges encountered, increasing the likelihood of early failures that could deter the next wave of adopters. Capturing and sharing early experiences with patient-facing APIs is, therefore, critically important to inform ongoing policy and implementation efforts seeking to ensure that this national transition is safe and secure and has the intended impact of accelerating patient data access, improving quality of care, and increasing options for tools available to digitally engaged patients.

Objectives

We sought to identify and summarize how early adopter health systems are planning for and implementing patient-facing APIs. We specifically wanted to ascertain their understanding and their overall strategies and key use cases, including the associated organizational infrastructure, governance, policies, and processes, and how their approaches are shaped by policies and other external drivers. Finally, we aimed to identify barriers encountered in these early experiences that could inform policy and practice efforts and help to promote broad adoption and effective use of patient-facing APIs.

Methods

Recruitment

We created a convenience sample of health systems known to be early adopters of health technology, including health systems engaged in the Health Records on iPhone patient-facing API pilot (a list of health systems included in Apple's press release) and others known to be the users of API technology (a list from authors' professional networks), and then also selecting for regional variation by ensuring at least one health system per census region (Northeast, West, Midwest, and South). Although APIs can be used by patients, providers, payers, third-party apps, and others, we used "patient-facing APIs" to describe use cases where *patients* are the actors requesting EHI via the API, or directing third-party apps to access the information on their behalf. In business-to-business API use cases, health systems or other covered entities connect via API with other covered entities or business associates to exchange health data. Although the API itself might be technically identical in both use cases, the federal policies, data authorization workflows, and ultimately many of the specific use cases are different, which led us to examine patient-facing APIs separately.

Interview Guides and Data Collection

We developed a semistructured interview guide that covered current patient-facing API efforts and capabilities, perceived barriers and risks, policy issues, organizational governance, and lessons learned (Multimedia Appendix 1). We reached out to the systems' chief information officers (CIOs) by email, as these are the leaders who typically possess the greatest breadth of understanding of the topics and invited them or a designee with subject matter expertise on APIs to participate in the interview. We then shared the interview questions in advance and scheduled and conducted individual hour-long interviews. We included the first ten health systems that agreed to participate in interviews. We conducted interviews (by phone or videoconference) between September and December 2018. All interviews were recorded and transcribed. Our study was approved by the University of California, San Francisco, institutional review board (18-25416).

Analysis

We performed a content analysis of interview transcripts using analytic matrices to extract relevant statements in three core topic categories related to our research domains: strategy and implementation, technical issues, and barriers and costs.

Specific topic categories within the strategy and implementation domain included (1) overall API strategy (both provider and patient-facing), (2) patient-facing API use cases, (3) status of implementation plan, (4) factors that shaped API strategy (organizational and policy), and (5) specific decisions and rationales for read and write APIs, white-listing or black-listing apps, publicity and marketing of APIs to patients and developers, and participation in Health Records on iPhone.

Specific topic categories within the technical issues domain included (1) level of effort required to implement patient-facing APIs, (2) level of EHR vendor engagement and support, (3) using APIs directly or with middleware, (4) staffing for API management, (5) one-time versus persistent access tokens, (6) ability to monitor the use of patient-facing APIs, and (7) technical lessons learned.

Specific topic categories within the barriers and costs domain included (1) barriers to and costs of implementing and using

patient-facing APIs, (2) fees that the system assesses for the use of patient-facing APIs, and (3) fees that the EHR vendor assesses for the use of patient-facing APIs.

To minimize bias, we gave each respondent their own row in the matrices. We then identified themes within and across topic categories through discussion and consensus among the four authors. In some cases, we summarized the theme in a more concise, structured way and then reported the associated number of respondents with content supporting that theme. In other cases, we summarized the theme drawing more heavily on narrative. Under either approach, we supported the theme with example quotes. In the Results section below, we coded quotes with a unique but anonymous identifier per respondent, based on the respondent's organizational role, for example, CIO1, CIO2, CTO1, and CTO2.

Results

Sample Characteristics

The health systems interviewed were located across the United States, with most in the Northeast or Midwest. Health system size was heterogeneous, with annual clinical revenues ranging from below US $10 billion to more than US $20 billion. Most health systems had an EHR from Epic Systems as their primary EHR, with a subset of these systems using additional EHRs in different locations because of affiliations, acquisitions, and regional consolidations. The leaders we interviewed covered a range of roles, including CIO, executive vice president, chief technology officer, senior medical director, and associate chief transformation officer (Table 1).

Table 1. Sample characteristics (n=10 organizations; 13 interviewees).

Characteristic	Value, n
Geographic location	
Northeast	5
Midwest	3
West	1
Southeast	1
Size (discharges per year)	
<100,000	4
100,000-200,000	3
>200,000	3
Total clinical revenue, US $	
<10 billion	3
10-20 billion	3
>20 billion	4
Primary electronic health record vendor	
Epic	7
Cerner	2
Homegrown	1
Role of interviewee	
Chief information officer	4
Executive VP[a]	2
CTO[b]	2
Other (including executive director, senior medical director, chief digital officer, associate CTO, VP of information systems)	5

[a]VP: vice president.

[b]CTO: chief technology officer.

Patient-Facing Application Programming Interface Strategies

Table 2 describes the API strategies of the organizations interviewed. Of the 10 organizations, eight had patient-facing APIs live and operational, while two were in the planning phase. Of the eight organizations with operational patient-facing APIs, five had delineated specific strategies (including dedicated technical support and explicit governance) for their use. The other three of eight organizations had activated patient-facing APIs but had a general API strategy for all users, not specific patient-facing API strategies or infrastructure. When we asked about API governance in more detail, seven organizations said that API decisions were made at the executive or board level,

two organizations said that API decisions were made by technical teams (below the executive level), and one organization had created a dedicated API committee to make decisions about both patient-facing and business-to-business APIs.

We identified two broader themes related to organizational API strategies, which included factors and use cases shaping health systems' strategies and encouraging their adoption and use of patient-facing APIs.

Table 2. Application programming interface strategies.

Type	Organizations, n	Description and example quotes
Operational patient-facing APIs[a] with specific patient-facing API strategy	5	• Defined strategy to support the technical implementation of patient-facing APIs and governance body with explicit oversight for patient-facing APIs • Example quote: "It [API steering committee] originally started off as trying to launch the provider applications that connected to FHIR within our EHR and over time it has evolved to include some patient-facing use cases. The primary apps that it's connected to is Apple Health." [CIO1]
Operational patient-facing APIs without specific patient-facing API strategy	3	• Patient-facing API strategy (with associated technology and governance) not distinct from overall API strategy • Example quote: "We see this as part of a broader strategy of around getting the right data to the right person at the right time and don't have a distinct strategy around that relative to use of APIs for other purposes." [CIO2]
Planning for patient-facing APIs	2	• Patient-facing APIs on the radar but specific implementation and governance plans are not yet developed. • Example quote: "There is no conceptual reservation about doing this. We intend to do this. We will be opening up that consumer API and registering ourselves into the Apple and Google ecosystems as soon as we can." [CTO1]

[a]API: application programming interface.

Theme 1: Diverse Factors Shape the Patient-Facing Application Programming Interface Strategy and Facilitate the Adoption and Use of Patient-Facing Application Programming Interfaces: Federal Policy and Regulations, Health Records on iPhone Availability, and Ethical Obligation

When asked about what shaped their strategy, three distinct and diverse factors were identified. The first factor was federal policy and regulations requiring patient-facing APIs, specifically the requirement that patients have access to and are able to use their health data through patient-facing APIs. As one interviewee explained:

> *I would say that CMS [Centers for Medicare & Medicaid Services] regulations is one of the top factors or rather what we feel like CMS regulation will become. That's both in Meaningful Use and other kinds of regulations around interoperability []. I have to say that regulations do play a large influence.* [CTO2]

The second factor was Apple's decision to include the option for patients to download their health record data to their Apple smartphone via Health Records on iPhone. A total of six organizations in our sample chose to partner with Apple to deploy this functionality, which required use of patient-facing APIs. The third factor was a value assessment or perceived ethical obligation to deploy functionality giving patients better access to their health data. As one interviewee stated:

> *We believe that patients should have their data in a format that is easy for them to share with who they want to share it with.* [EVP1]

Theme 2: Two Priority Use Cases Are Driving Patient-Facing Application Programming Interface Efforts

The first use case (cited by six organizations) includes use of patient-facing APIs to enable patients' access to health information to facilitate a longitudinal health record combining information from multiple providers and integration of provider and payer data. Patient-facing APIs are uniquely capable of facilitating this use case (beyond an EHR-tethered patient portal) because they enable data integration from multiple sources. The second use case (cited by four organizations) includes the use of patient-facing APIs to enhance patients' digital experience with the health system, including their ability to use convenience features such as telehealth, for example, transmitting health data to and from virtual visits and facilitating self-scheduling. For these systems, this use case aligns with consumers' increasing expectations that health care, like other industries such as travel, food services, and financial services, provides data access and convenient self-service functionality that are widely available (primarily via mobile apps) [12,13].

When asked whether their organization's primary use cases might change over the next 2 or 3 years, most interviewees thought that they would not. However, they did share specific ideas for how the functionality might evolve to deliver new value, in particular, by starting to enable better data-driven patient self-management. For example:

> *I think we're going to see an explosion in more analytics and intelligence pushed out to the patient-facing APIs for self-triage management and escalation, as opposed to today where it comes across to the server side and we [health systems] have to*

> *manage it and triage it...I think it will be more proactive than reactive.* [SMD1]

Said another:

> *[In the future] if you check my data and you're able to combine it with lifestyle data and other pieces of information that will add insights and value to my information beyond what my provider portal can do, then I am intrigued.* [VP1]

Operational and Technical Features of Patient-Facing Application Programming Interfaces

Among the eight health systems with operational patient-facing APIs, we summarize key features of their operational and technical approaches and implementation decisions (Table 3). APIs can allow patients to access data (read) or provide data (write). All eight of these health systems had read APIs in operation. Although seven had at least some write capability, mostly focused on patients' home device data, these implementations of write API functionality were variable, with some offering it only to a subset of invited patients rather than to all patients. There was also variation in whether patients could write data directly to the EHR or whether health system review and approval were required to authorize the data to be written into the EHR.

Table 3. Current status of patient-facing application programming interfaces (among eight organizations with operational application programming interfaces).

Category and status	Organizations, n
Types of APIs[a] that are operational	
Read and write	4
Read with limited access to write	3
Read only	1
App authorization	
White List: Health-System managed	4
White List: EHR[b] vendor managed	4
Availability of API documentation	
From EHR vendor	4
From health system	2
No documentation available	2
Patient communication about APIs	
Patient-facing APIs explained on health system website	2
Fees[c]	
System charges or plans fees to patients for use	0
EHR vendor charges system fees for patient-facing API use	0

[a]API: application programming interface.

[b]EHR: electronic health record.

[c]Although covered entities and their business associates under the Health Insurance Portability and Accountability Act may not charge patients a fee for electronic *access* to their health information through certified EHR technology—for example, to view, download, transmit, or access the health information through a patient-facing API—we still asked about actual practices in the field given the public reports and concerns about the fact and amount of fees being charged. Unlike this access, covered entities may charge patients a reasonable, cost-based fee for an electronic *copy* of their health information [14].

A second API implementation decision relates to how third-party apps are approved for API access. All eight health systems required prior approval of the app, or whitelisting, before an app could access the API. For four of these, the health system maintained the whitelist; for the other four, the EHR vendor maintained it. For third-party apps to connect to a patient-facing API, a developer requires documentation about the API configuration. A total of four health systems stated that this documentation was available to developers through their EHR vendor, with two health systems saying they would make it available to developers upon request and two health systems not having documentation available for developers. Of note, none of the health systems had publicly available API documentation for any developer to access. In all, two health systems noted that they were actively publicizing to their patient populations the availability of patient data via their patient-facing APIs, with one specifically noting its focus on how to utilize patient-facing API functionality to help "market to them the experience [rather than the underlying technology]" (SMD1).

In terms of more technical dimensions of patient-facing APIs, two health systems were relying on their EHR vendors to

manage APIs, while six were using a middleware vendor for API management. API management refers to an array of capabilities, including API access control, API creation and design, API protection, business value reporting, and API traffic control, performance, and throttling [15]. Reliance on the EHR vendor rather than middleware appeared to be driven at least in part by EHR vendor capabilities, whether real or perceived, and possible differences between vendors. As one interviewee stated:

> *Some of the EHRs have made it easy, I would say [vendor 1] is one of them. Others are more difficult. I think its complex because it's based on their internal resources, focus, structure, and capabilities.* [CTO2]

Another interviewee added:

> *You must have API middleware. [Otherwise], there is no way that you are going to be able to monitor the traffic flows or constrain traffic flows, or identify patterns. The EHR vendors are never going to figure this out.* [CIO3]

We also asked about health systems' approaches to access tokens, which can be persistent (eg, log-in credentials not required each time the API attempts to read or write patient data) or one-time (ie, log-in required for each session the APIs are used). Overall, interviewees conveyed uncertainty about what types of tokens were in use, with two organizations explicitly not sure, two that thought they used persistent tokens, and three that thought they used one-time tokens.

Finally, when asked about fees associated with patient-facing APIs, none of the health systems were charging or planning to charge their patients a fee for use of patient-facing APIs—which was appropriate, given the regulatory prohibition against charging fees for patient access to their health information using patient-facing APIs in certified EHR technology. In addition, none of the health systems were paying specific fees to their EHR vendor for the use of patient-facing APIs.

Perceived Barriers to the Current and Future Implementation and Use of Patient-Facing Application Programming Interfaces

We identified six items cited as barriers to implementation and use of patient-facing APIs (Table 4, Multimedia Appendix 2). The most common, cited by five health systems, was concerns about security or privacy introduced by enabling patient-facing APIs, such as unauthorized access or potential attacks by malicious actors. For example:

> *We worry a little bit that the spirit of the [regulation] is basically, well if any app comes knocking on your door, and the patient really wants to do this, then who are you to judge...[but] what if we can't deliver it securely because the app is badly created or truly doing things like forking data off to a pharma company?* [CIO3]

Health systems felt unprepared to identify and differentiate the trustworthy from the untrustworthy app vendors, expressing additional uncertainty about whose responsibility this should become as the app ecosystem develops. Among our interviewees, some suggested an ongoing role for the EHR vendors, whereas others suggested instead a third-party rating consortium, with one suggesting "the creation of an expertise driven team to curate that portfolio would make the most sense" (CIO2). One element frequently tied to security concerns was a health system's financial exposure, such as:

> *When something bad happens, they are not going to go after some little startup vendor. They are going to go after [health system]. So that's why I think we have to be as clear as possible about what this decision means and where your information is going. You are taking this responsibility for this decision and we are not.* [VP1]

One health system interviewed had already taken a step toward trying to mitigate this, saying:

> *We do have a list of these recommendations and kind of a primer for patients for their digital health privacy...If we make the [health data] available, people should know how to unauthorize an application and know what to do if they suspect some app is doing something wrong with their data.* [CIO1]

Four health systems cited insufficient functionality in apps using patient-facing APIs as a barrier. Health systems believed that the current third-party app ecosystem did not yet offer a compelling alternative to EHR-tethered portals. For example:

> *First there need to be consumer apps that offer some added compelling value more than what the vendor mobile apps are doing.* [ED1]

Another interviewee said:

> *[We have] invested so much into the portal. [We are] concerned about the possibility of drawing some people away from it.* [CTO1]

A related barrier, cited by three health systems, was the lack of a perceived business value or financial return on investment, with one person saying:

> *[There is a] lack of external stimulus. It's not a revenue driver. It's not a cost reduction.* [CTO1]

Another said:

> *At some point if you have 100,000 different apps connecting with people, at some point someone is going to have to pay for those. And I don't know exactly how that is going to get sorted out.* [CIO1]

Another barrier cited by three health systems was their EHR vendors' wariness about data sharing. Said one executive:

> *Many vendors see data acquisition as a strategy for them so they don't let you share.* [EVP1]

Said another:

> *[they think] patients will download malware that will take their patient data and do sinister things with it.* [ED1]

One system complained as follows:

> *[Y]ou [EHR vendor] need to open up your data so that we can do stuff in addition to what you are doing,*

> *and so that other people can do stuff to augment what you are doing.* [CIO2]

Finally, systems noted less evolved technology and standards in health care as a barrier. For example,

> *[T]he speed at which they [FHIR] are adding [data elements] is glacial. There are 4,000 data elements in [our EHR] and I think they've got maybe 150 done through FHIR [Fast Healthcare Interoperability Resources].* [CIO2]

Another interviewee added:

> *It's sort of interesting that APIs really require that you build a sort of robustness and scalability that you've never really engineered for in the past because this all used to be done by a fairly finite population of doctors with fairly understandable behaviors, not thousands if not tens of thousands of people who...could update their medical records every six seconds...[FHIR] is so rapid in its evolution...and so suddenly you're engineering a set of apps on shifting sands.* [CIO3]

There were concerns about available technical settings, such as:

> *We don't have the ability to tell the patient for how long this application will have access to their data. That's a pretty big factor.* [VP1]

Another specific concern related to less evolved technical capabilities was the fact that certain settings created in the EHR might not be transmitted across a patient-facing API. For example:

> *In the portal you can put in rules like don't show cancer diagnosis until released by the doctor, but I hear in the HealthKit API all those filters are gone, so you may discover your cancer diagnosis on your phone before the doctor even knows.* [CIO3]

Table 4. Perceived barriers to the current/future implementation/use of patient-facing application programming interfaces.

Category	Organizations, n
Security or privacy concerns	5
Insufficient functionality available in third-party apps	4
Lack of perceived return-on-investment: costs vs business value	3
Vendor unwillingness to share data outside of their system	3
Technology not as evolved in health care vs other industries	3
Lack of application programming interface and semantic standards	2

Discussion

Principal Findings

In this study, we interviewed 10 leading health systems about their approach to patient-facing APIs to understand early experiences and identify insights for policy and practice. Our results suggest a reason for optimism about the prospects for patient-facing APIs and their impact on the US health care system. Building upon the progress made in the use of patient portals, patient-facing APIs further expand technical capabilities, leading us further toward the hopes inscribed nearly 2 decades ago in the Health Insurance Portability and Accountability Act Privacy Rule, which gave patients a right to access and use their full designated health record set.

The health systems we interviewed all planned to increase the use of patient-facing APIs, and many stated that this was "the right thing to do." Two use cases emerged as the strategic driving forces for health systems: the ability for the patient to create an aggregated longitudinal health record and better digital patient engagement. However, we also found substantial heterogeneity in the strategy as well as operational and technical approaches. This suggests that there are not yet best practices to guide health systems on how to work with patient-facing APIs, nor clear business models driving a common strategic roadmap for their implementation. We also identified that there were commonly perceived barriers spanning a range of domains, suggesting that additional work is required, at both policy and practice levels, to ensure that patient-facing APIs successfully fulfill their intended purposes and use cases.

The implementation challenges and perceived risks we identified are leading organizations to act carefully and slowly around implementation and to avoid aggressively marketing the new functionality. Health systems commonly cited privacy and security concerns—a noteworthy perceived barrier as the regulations compel health systems to share electronic health data with patients via API to any third party and via any format requested by the patient. Health systems had associated concern about financial exposure. Specifically, even where patients have transmitted their data from the system to unaffiliated third-party apps and the system is not legally liable for the patient's subsequent choices about disclosure and use, a patient might still choose to sue and may still garner a settlement given health systems' "deep pockets." Technical topics often brought out confusion about policy and implementation details, for example, with access tokens and whitelisting of apps, as well as sometimes opposing or divergent views, such as whether to use a middleware vendor rather than EHR vendor for API management. This is likely because there is not yet sufficient longitudinal experience with using patient-facing APIs for crisp, confident understanding of the technical nuances. The fact that some CIOs and other interviewees from large health systems with substantial technical experience are still on the learning curve about patient-facing APIs is an important finding from our work.

Health systems were also cautious about the market demand for the use of patient-facing APIs. Although they noted that patients expect more convenient tools for managing their care, the health systems we interviewed generally thought that their EHRs' tethered portals still deliver greater patient convenience and value than the currently available third-party apps utilizing patient-facing APIs. From some health systems' perspectives, a controlled, integrated system using an EHR-tethered portal seemed preferable to the potentially piecemeal, heterogeneous functionality of modular systems using patient-facing APIs. It is an open question, however, whether patients agree or would have an alternative assessment. For example, patients may want to choose their own appointment scheduling or cost-estimator app, or integrate their data across health systems, none of which is currently possible with a tethered portal.

On the basis of our findings, we believe additional actions are needed to spur health systems' uptake and effective use of patient-facing APIs, primarily by tipping the balances toward increased upside and decreased downside for health systems. First, given the lack of clarity we found about some technical implementation best practices, we think ongoing dialog and sharing of experiences within the industry will be critical to bridge the heterogeneous technical approaches to implementation. For example, sharing experiences around the use of API middleware platforms vs embedded EHR vendor tools for API management would likely be valuable and could be facilitated by organizations such as the Health Services Platform Consortium, the College of Healthcare Information Management Executives, the Healthcare Information and Management Systems Society, the American Medical Informatics Association, or the Association of Medical Directors of Information Systems. Second, as we found frequent concerns about security, privacy, and health system liability; a broader public awareness campaign and tools to educate and aid the public in learning how to manage their health data may be needed. One specific strategy could feature increased commitment to use the Office of the National Coordinator for Health Information Technology's (ONC) Model Privacy Notice to help consumers identify which apps may present higher privacy and security risk. Another strategy that might help convince health systems to expedite progress would be additional policy clarity and protections for health systems for data privacy breaches by apps chosen by the patient. Recent frequently asked questions published by the US Department of Health and Human Services are a good start in this direction [16]. A third strategy toward overcoming these concerns will be the emergence and development of expert consortia, making recommendations about the safety and efficacy of apps, to help health systems and patients understand which apps to use or avoid and thus aid adoption. It remains unclear whether this will be taken on by the ONC, EHR vendors, the Food and Drug Administration, hospital-specific "digital diagnostics and therapeutics committees" [17], multistakeholder coalitions, or some combination of all of these [18].

Finally, given that we found ongoing uncertainty about the relative value and cost of tethered EHR portals compared with third-party patient apps, we believe that there must be greater pressure on EHR vendors to expand the set of available APIs, even for functionality they believe might enable competitive products. This would require more rapid expansion of the set of data to be shared by standardized APIs, including the US Core Data for Interoperability (USCDI). Although ONC currently proposes to add three new data elements to the current set of 20, it could add other standardized data elements more rapidly, such as those that already exist as voluntary certified health information technology modules, for example, family history and social determinants of health, and focus on APIs that would facilitate patient-driven transactional interactions, such as appointment scheduling. Speeding these additions to the USCDI would more rapidly enable new apps to meet currently unmet needs, in turn driving patient interest and demand, leading to a virtuous cycle of increased health system prioritization of further implementation and use of patient-facing APIs.

Limitations

Given that we studied a convenience sample of 10 early adopter health systems planning for or implementing patient-facing APIs, our results may be limited in their generalizability, and we may have failed to capture all important themes. Our sample was also skewed toward larger, urban health systems, which may fail to capture differences in experiences of smaller, more rural hospitals and health systems. We interviewed one person (or sometimes two) at each organization and left it up to the organization to select the most appropriate interviewees. As a result, respondents varied in their depth of knowledge about the broad range of topics covered, meaning certain quotes may reflect respondents' perceptions.

Comparison With Prior Work

Given the recent availability of patient-facing APIs, there is limited prior work on this topic. Four recent studies have assessed the uptake and use of APIs. One study examined the use of APIs across 12 large health systems and found relatively low, but steadily increasing, uptake [19]. Relevant to our findings, this study found very low use of APIs compared with levels of use of EHR-tethered patient portals. A second study, using the 2017 American Hospital Association data, found that 33% of hospitals reported that patients could access their EHI through EHR APIs [20]. A third study from Partners' Healthcare system examined demographics of users and found that male patients and younger patients were more likely to use patient-facing APIs [21]. Finally, a study from the University of California, San Diego, surveyed the first 425 patients who had started using their personal health record API feature, with a very high percentage (75%-96%) of 132 survey respondents noting satisfaction with this new connectivity and the improved health informational benefits derived [22]. These findings complement our results in that initial results show slow uptake but significant enthusiasm for the potential benefits of patient-facing APIs.

Conclusions

In summary, we found that diverse factors drive health systems' efforts to pursue patient-facing APIs, including the benefits and opportunities for patient engagement and empowerment they bring to health care. However, we found substantial

heterogeneity in the approaches by early adopter health systems, which are proceeding with caution and remain uncertain about multiple dimensions including the long-term business drivers. To ensure that national policy goals for interoperability and patient data access are met, there will need to be an ongoing understanding of the usage of and barriers to patient-facing APIs and increasing discussion and sharing of best practices, likely with targeted ongoing interventions to support and bolster their use.

Acknowledgments

The authors would like to thank Edwin Martin and Victor Galvez at the University of California, San Francisco for providing expertise on API technology.

Conflicts of Interest

AN has received research support from Cisco Systems, Inc; has been a consultant to Steady Health, Nokia Growth Partners, WebMD, and Grand Rounds; has received speaking honoraria from Academy Health and Symposia Medicus; and is an uncompensated medical advisor for Tidepool. MS has received research support from Cisco Systems, Inc. The other authors have no disclosures.

Multimedia Appendix 1
Interview guide for “Early Insights from Health System Deployment of Patient-Facing APIs.”.
[DOCX File , 30 KB - jmir_v22i4e16813_app1.docx]

Multimedia Appendix 2
Results: barriers to implementation and use.
[PNG File , 158 KB - jmir_v22i4e16813_app2.png]

References

1. Library of Congress. 2015. Public Law 114–255 URL: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf [accessed 2019-04-01]
2. Library of Congress. 2009 Feb 17. Public Law 111–5 URL: https://www.congress.gov/111/plaws/publ5/PLAW-111publ5.pdf [accessed 2019-04-01]
3. Talbot D. MIT Technology Review. 2011 Jun 29. How a Broken Medical System Killed Google Health URL: https://www.technologyreview.com/s/424535/how-a-broken-medical-system-killed-google-health/ [accessed 2019-04-01]
4. Office of the National Coordinator for Health Information Technology (ONC). 2016. 2016 Report To Congress on Health IT Progress: Examining the HITECH Era and the Future of Health IT URL: https://www.healthit.gov/sites/default/files/2016_report_to_congress_on_healthit_progress.pdf [accessed 2019-04-01]
5. National Partnership for Women & Families. 2014 Dec. Engaging Patients and Families: How Consumers Value and Use Health IT URL: http://www.nationalpartnership.org/our-work/resources/health-care/digital-health/archive/engaging-patients-and-families.pdf [accessed 2019-08-01]
6. US Department of Health and Human Services. 2020. What is the intersection of the HIPAA right of access and the HITECH Act’s Medicare and Medicaid Electronic Health Record Incentive Program’s 'View, Download, and Transmit' provisions? URL: https://www.hhs.gov/hipaa/for-professionals/faq/2057/what-is-the-intersection-of-the-hipaa-right/index.html [accessed 2019-04-01]
7. Center for Medicare and Medicaid Services. US Government Publishing Office. 2015 Oct 16. Electronic Health Record Incentive Program-Stage 3 and Modifications to Meaningful Use in 2015 Through 2017, 80 FR 62761 URL: https://www.govinfo.gov/content/pkg/FR-2015-10-16/pdf/2015-25595.pdf [accessed 2019-04-01]
8. Holt M, Dunn O, Krueger K. MobiHealthNews. 2018 Dec 18. EMRs, APIs, App stores & all that: More data URL: https://www.mobihealthnews.com/content/emrs-apis-app-stores-and-all-more-data [accessed 2019-04-01]
9. Kim Y, Lee B, Choe EK. Investigating data accessibility of personal health apps. J Am Med Inform Assoc 2019 Mar 12;26(5):412-419 [FREE Full text] [doi: 10.1093/jamia/ocz003] [Medline: 30861531]
10. Mandl K, Kohane IS. Time for a patient-driven health information economy? N Engl J Med 2016 Jan 21;374(3):205-208. [doi: 10.1056/NEJMp1512142] [Medline: 26789868]
11. US Department of Health and Human Services (HHS). Office of the National Coordinator for Health Information Technology (ONC). 2017. Key Privacy and Security Considerations for Healthcare Application Programming Interfaces (APIs) URL: https://www.healthit.gov/sites/default/files/privacy-security-api.pdf [accessed 2019-04-01]
12. Payne TH, Corley S, Cullen TA, Gandhi TK, Harrington L, Kuperman GJ, et al. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. J Am Med Inform Assoc 2015 Sep;22(5):1102-1110 [FREE Full text] [doi: 10.1093/jamia/ocv066] [Medline: 26024883]

13. Deloitte. 2016. What matters most to the health care consumer? Insights for health care providers from the Deloitte 2016 Consumer Priorities in Health Care Survey URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-cx-survey-pov-provider-paper.pdf [accessed 2019-06-30]
14. US Department of Health & Human Services. 2016 Feb 25. Individuals' Right under HIPAA to Access their Health Information 45 CFR § 164.524 URL: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html [accessed 2019-07-01]
15. O'Neill M, Malinverno P. Gartner Research. 2018. Critical Capabilities for Full Life Cycle API Management URL: https://www.gartner.com/en/documents/3873522 [accessed 2019-04-01]
16. US Department of Health and Human Services. HIPAA FAQs for Professionals: Health Information Technology URL: https://www.hhs.gov/hipaa/for-professionals/faq/health-information-technology/index.html [accessed 2019-04-01]
17. Auerbach AD, Neinstein A, Khanna R. Balancing innovation and safety when integrating digital tools into health care. Ann Intern Med 2018 May 15;168(10):733-734. [doi: 10.7326/M17-3108] [Medline: 29582072]
18. Bates DW, Landman A, Levine DM. Health apps and health policy: what is needed? J Am Med Assoc 2018 Nov 20;320(19):1975-1976. [doi: 10.1001/jama.2018.14378] [Medline: 30326025]
19. Adler-Milstein J, Longhurst C. Assessment of patient use of a new approach to access health record data among 12 US health systems. JAMA Netw Open 2019 Aug 02;2(8):e199544 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9544] [Medline: 31411708]
20. Holmgren AJ, Apathy NC. Hospital adoption of API-enabled patient data access. Healthc (Amst) 2019 Dec;7(4):100377. [doi: 10.1016/j.hjdsi.2019.100377] [Medline: 31471262]
21. Gordon WJ, Bates DW, Fuchs D, Pappas J, Silacci S, Landman A. Comparing characteristics of patients who connect their iPhones to an electronic health records system versus patients who connect without personal devices: cohort study. J Med Internet Res 2019 Aug 22;21(8):e14871 [FREE Full text] [doi: 10.2196/14871] [Medline: 31441430]
22. Dameff C, Clay B, Longhurst CA. Personal health records: more promising in the smartphone era? J Am Med Assoc 2019 Jan 29;321(4):339-340. [doi: 10.1001/jama.2018.20434] [Medline: 30633300]

Abbreviations

API: application programming interface
CIO: chief information officer
EHI: electronic health information
EHR: electronic health record
ONC: Office of the National Coordinator for Health Information Technology
USCDI: US Core Data for Interoperability

Edited by G Eysenbach; submitted 27.10.19; peer-reviewed by A Landman, W Gordon; comments to author 08.01.20; revised version received 21.01.20; accepted 26.01.20; published 03.04.20.

Please cite as:
Neinstein A, Thao C, Savage M, Adler-Milstein J
Deploying Patient-Facing Application Programming Interfaces: Thematic Analysis of Health System Experiences
J Med Internet Res 2020;22(4):e16813
URL: https://www.jmir.org/2020/4/e16813
doi:10.2196/16813
PMID:31983680

Original Paper

Using Complexity Assessment to Inform the Development and Deployment of a Digital Dashboard for Schizophrenia Care: Case Study

Andreas Gremyr[1,2], MSc; Boel Andersson Gäre[2,3], PhD, MD; Trisha Greenhalgh[4], FMedSci, MD; Ulf Malm[5], PhD, MD; Johan Thor[2], PhD, MD; Ann-Christine Andersson[2], PhD

[1]Department of Schizophrenia Spectrum Disorders (Psykiatri Psykos), Sahlgrenska University Hospital, Mölndal, Sweden

[2]Jönköping Academy for Improvement of Health and Welfare, School of Health and Welfare, Jönköping University, Jönköping, Sweden

[3]Futurum Academy for Health and Care, Region Jönköping County, Jönköping, Sweden

[4]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[5]Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Corresponding Author:
Andreas Gremyr, MSc
Department of Schizophrenia Spectrum Disorders (Psykiatri Psykos)
Sahlgrenska University Hospital
Göteborgsvägen 31
Mölndal, 43180
Sweden
Phone: 46 733664000
Email: andreas.gremyr@vgregion.se

Abstract

Background: Health care is becoming more complex. For an increasing number of individuals, interacting with health care means addressing more than just one illness or disorder, engaging in more than one treatment, and interacting with more than one care provider. Individuals with severe mental illnesses such as schizophrenia are disproportionately affected by this complexity. Characteristic symptoms can make it harder to establish and maintain relationships. Treatment failure is common even where there is access to effective treatments, increasing suicide risk. Knowledge of complex adaptive systems has been increasingly recognized as useful in understanding and developing health care. A complex adaptive system is a collection of interconnected agents with the freedom to act based on their own internalized rules, affecting each other. In a complex health care system, relevant feedback is crucial in enabling continuous learning and improvement on all levels. New technology has potential, but the failure rate of technology projects in health care is high, arguably due to complexity. The Nonadoption, Abandonment, and challenges to Scale-up, Spread, and Sustainability (NASSS) framework and complexity assessment tool (NASSS-CAT) have been developed specifically to help identify and manage complexity in technology-related development projects in health care.

Objective: This study aimed to use a pilot version of the NASSS-CAT instrument to inform the development and deployment of a point-of-care dashboard supporting schizophrenia care in west Sweden. Specifically, we report on the complexity profile of the project, stakeholders' experiences with using NASSS-CAT, and practical implications.

Methods: We used complexity assessment to structure data collection and feedback sessions with stakeholders, thereby informing an emergent approach to the development and deployment of the point-of-care dashboard. We also performed a thematic analysis, drawing on observations and documents related to stakeholders' use of the NASSS-CAT to describe their views on its usefulness.

Results: Application of the NASSS framework revealed different types of complexity across multiple domains, including the condition, technology, value proposition, organizational tasks and pathways, and wider system. Stakeholders perceived the NASSS-CAT tool as useful in gaining perspective and new insights, covering areas that might otherwise have been neglected. Practical implications derived from feedback sessions with managers and developers are described.

Conclusions: This case study shows how stakeholders can identify and plan to address complexities during the introduction of a technological solution. Our findings suggest that NASSS-CAT can bring participants a greater understanding of complexities in digitalization projects in general.

(J Med Internet Res 2020;22(4):e15521) doi:10.2196/15521

KEYWORDS

health care; complexity; schizophrenia; coproduction; learning health systems

Introduction

Health Care Challenges

Health care is growing more complex and difficult to manage due to factors such as a rapidly expanding body of knowledge, a shift towards more people living with chronic disease and multi-morbidity [1], challenges in coordinating multiple providers and actors, and, not least, the need to include the preferences and values of the individuals seeking health care [2]. For an increasing number of individuals, interacting with health care means addressing more than just one illness or disorder, undergoing more than one treatment, and collaborating with more than one care provider [1].

Schizophrenia as an Example

Addressing and adapting to complexity might be especially challenging in health care focusing on individuals with severe mental illnesses such as schizophrenia who are disproportionately affected by comorbid medical conditions [3]. With typical onset in early adulthood and a lifelong course, schizophrenia is among the top 10 disorders in terms of disability-adjusted life years lost [4]. Characteristic symptoms are hallucinations, delusions, and disturbances of thought. These features tend to make it challenging to establish and maintain relationships [5]. Treatment failure is common, increasing the risk of suicide [6], despite access to effective treatments [7]. Health and social services for persons with schizophrenia are marked by a high level of complexity. They involve a multimodal treatment approach with a range of treatments from a multiprofessional team, often requiring coordination with other providers (eg, primary care, care for other chronic disorders, social support, housing, and vocational rehabilitation) [5].

Complex Adaptive Systems

Systems can be described either as simple (straight forward and predictable, with few components), complicated (predictable but with more interacting components), or complex (unpredictable and dynamic, where the whole is more than its constituent parts) [8]. Traditional linear cause-and-effect-thinking is not sufficient when studying systems that evolve in ways that are hard or even impossible to predict. Knowledge of complex adaptive systems can aid in understanding and studying health services [9]. A complex adaptive system is a collection of interconnected agents with the freedom to act based on their own internalized rules, affecting each other. These rules, in human-based complex adaptive systems, could be instinct and implicit mental models. Agents adapt in various ways through interactions, which causes the system to change over time [2]. Capability among individual agents in a complex system, that is "the extent to which an individual can adapt to change, generate new knowledge, and continue to improve their performance" [10], can be supported by minimum specifications (simple rules to guide behavior) and feedback loops, letting individuals gradually upgrade their internalized rules through experience. Relevant feedback on performance is crucial to enable continuous learning and improvement at all levels of a health care system, from the level of individual patients to organizational management and policy levels, potentially enabled by the use of new technologies [11-15].

New Technology: A Blessing or a Curse?

The use of new technologies, argue Pavel et al [16], is essential to achieving personalized, evidence-based, and economically viable health care. Meanwhile, experiences so far reveal significant challenges. Uptake of "disruptive" technologies in health care is slow [17,18], and fundamental quality, safety, and cost problems have not been resolved by digitalization [19]. Moreover, the failure rate of technology projects in health care is high; large and complex projects often tend to fail to deliver anticipated results [20-22]. There seems to be a gap between the development of technology and usefulness in practice within health care organizations that needs to be bridged if technologies are to support health care rather than further increase its complexity [19].

Addressing Complexity in Health Care Technology Projects

Greenhalgh et al [23] argued that adoption, scale-up, and spread of new technologies often fail due to complexity. They employed theories on complex adaptive systems and on the diffusion of innovations in health care to create a framework for using principles or rules to facilitate the development and application of technological innovations in complex contexts. The Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability (NASSS) framework and complexity assessment tool (NASSS-CAT) were developed to help stakeholders identify and manage complexity in technology innovation projects in health care [8,23].

A university hospital in west Sweden started to develop a point-of-care dashboard to support patients and health care professionals in schizophrenia care. Successful prototyping and pilot testing led to the decision to scale-up the initiative to all the department's outpatient units. Challenges started to accrue when planning for larger-scale development and deployment of the dashboard and related tools. This case study evaluates the stakeholders' use of the NASSS-CAT to inform the development and deployment of the dashboard and reports on the complexity profile of the project, stakeholders' experiences when using a pilot version of the NASSS-CAT, and practical implications. This study also aims to inform a future multinational study with multiple cases having maximum intercase variation to field-test the NASSS-CAT [24].

Methods

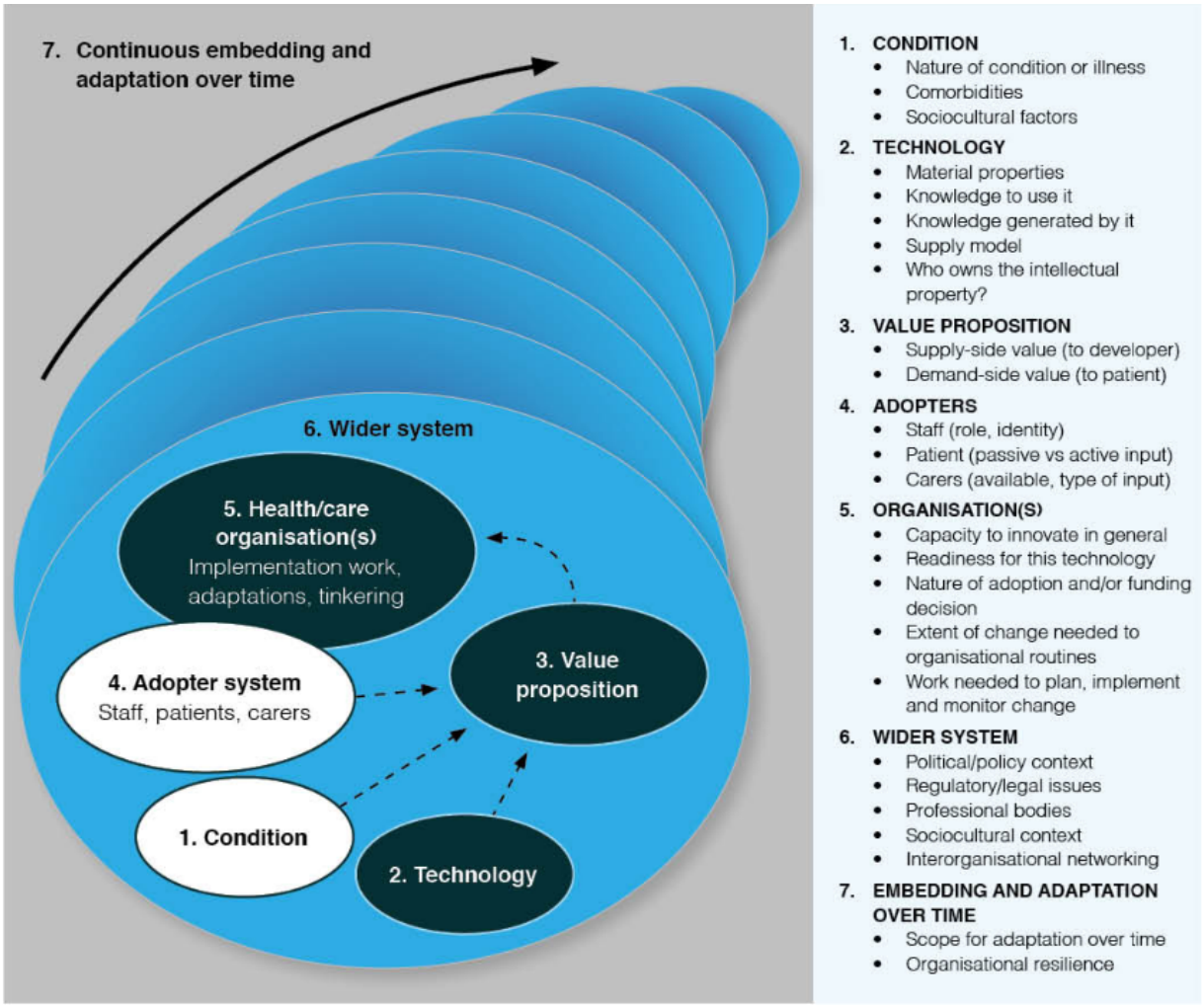
Overall Design

This case study [25,26], informed by the principles of action research [27] and action evaluation [28], involved stakeholders' use of the NASSS-CAT to inform the development and deployment of the point-of-care dashboard for patients and health care professionals in schizophrenia care. Another study will specifically evaluate patients' experiences while using the dashboard at the point of care.

Nonadoption, Abandonment, and Challenges to Scale-up, Spread, and Sustainability (NASSS) Framework

The developers of NASSS noted that it was crucial to understand the sociotechnical interaction between individuals, organizations, technology, and policy to explain why a new technology is adopted and sustained (or not) in health and social care [23,29]. The NASSS framework features 7 key domains, identified through systematic hermeneutic literature review and refined through empirical case studies of technology implementation [8]. It is intended to be used to guide and evaluate the success of technology deployment in sociotechnical systems. By addressing questions in the tool's domains, properties of the technology and adopting system are placed along a continuum ranging from simple to complicated to complex (Figure 1) [8]. Knowledge of a technology project's domain-specific complexity can aid stakeholders to respond adaptively, lessen complexity, and strengthen their capability to handle complexity [30]. Principles or "simple rules" can act as recommendations to guide further development (Textbox 1) [30].

Figure 1. The Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability (NASSS) framework by Greenhalgh et al [8]. Used with permission.



Textbox 1. Ten simple rules for managing complexity [30].

1. Strengthen program leadership, which may be distributed across the project and across contributing disciplines.
2. Codevelop an overall vision for the project and maintain dialogue around that evolving vision.
3. Nurture key relationships between individuals and organizations.
4. Develop individuals and encourage them to solve local problems creatively.
5. Make resources available for creative individuals and teams to use for generating solutions to local challenges.
6. Capture data on progress and feed it into ongoing deliberations.
7. Acknowledge and address the concerns of frontline staff.
8. Work with intended users to codesign technologies and the work routines they are intended to support, building in adaptability.
9. Control scope creep.
10. Address regulatory and policy barriers.

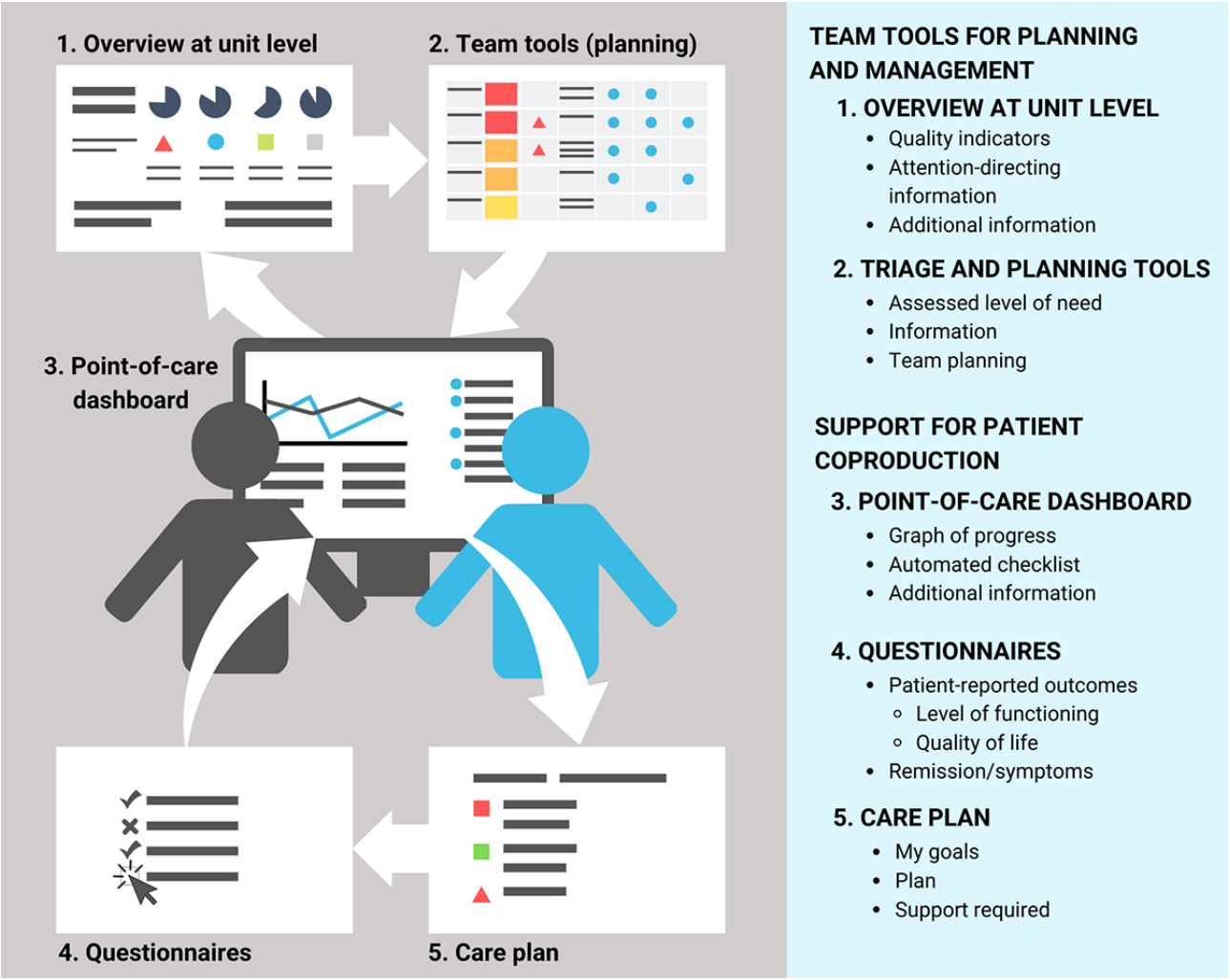
Nonadoption, Abandonment, and Challenges to Scale-up, Spread, and Sustainability Complexity Assessment Tool (NASSS-CAT)

The NASSS-CAT has two primary components. The initial component, based on the NASSS framework, supports the development of a rich narrative that surfaces key areas of uncertainty and interdependence in the project. The second component, based on an adapted version of the complexity assessment tool by Maylor et al [31], consists of a series of questions to support emergent project planning and evaluation and, in particular, to prompt project teams to consider how they might either reduce or manage complexity across the different NASSS domains.

Case Project: Developing and Deploying New Technology

The Department for Schizophrenia Spectrum Disorders at Sahlgrenska University Hospital in Gothenburg, Sweden (the Department) delivers specialized care for people with psychotic disorders in the metropolitan Gothenburg area (population of approximately 600,000 people). It serves 2600 patients, with schizophrenia as the most common diagnosis, at 8 outpatient units. About 20% of these patients need acute inpatient care at one of the Department's 5 wards each year. To support patient engagement at the point of care, the Department developed a digital dashboard to visualize key indicators of each patient's health and care status. The dashboard is one of several connected applications and displays to visualize data fed by several systems developed for several years and piloted at 2 outpatient units with some 400 patients (Figure 2). It includes team tools for care planning and management and tools to support coproduction of health and care among patients, their family members, and psychiatry staff. These tools include a unit-level overview of quality indicators identifying patients at risk, triage and planning tools including support for patient coproduction, a dashboard to be jointly reviewed at the point of care by patients and case managers/psychiatrists to support evaluation and planning, outcomes questionnaires, and patients' care plans.

Figure 2. The dashboard and related tools: applications and visualizations.



Patient-perspective Vignette: The Role of the Technology

Below is a vignette from the fictitious patient Ana's perspective [32] of a follow-up visit with her case manager at the outpatient clinic. It illustrates how the imagined dashboard-enhanced service model would differ from the traditional service model (pre-dashboard, Textbox 2).

> *Ana is a 34-year-old woman who has suffered from psychosis since her early twenties. After several diagnostic assessments, she was diagnosed with schizophrenia at age 28. Her medication treatment is combined with psychosocial interventions to help her manage her situation. On three occasions, psychotic episodes brought Ana into emergency inpatient care at the local hospital. She fears another episode and worries about how she would cope if it happened again. She is invited to annual check-ups as part of her continuous care at the outpatient unit.*

Textbox 2. Predashboard situation compared with the imagined dashboard-enhanced service model.

Predashboard situation

At her regular follow-up, it was hard for Ana to answer all the questions. Was she feeling better? Was there an increase in side effects? She struggled to remember how she felt the last time they jointly assessed her level of functioning and symptoms. She was not sure if the medication helped her. She dreaded making medication changes, for fear of new side effects. Both Ana and her case manager completed printed questionnaires. The content of the questionnaires provided some structure, covering important aspects of Ana's situation. They were possibly useful for the clinic in documenting relevant information but did not help Ana in understanding her situation or lessening her anxiety.

Dashboard-enhanced service model

Using the point-of-care dashboard not only makes it easier to complete the questionnaires and review the care plan but also shows Ana's progress and changes over time. An automated checklist signals to the case manager that it is time to update Ana's care plan and perform a general health assessment within 3 months. Ana and her case manager review the digital visualization of Ana's care plan and progress. Although Ana has not been feeling well over the past week, she is comforted by seeing how her symptoms and level of functioning have changed over time. Things are moving in the right direction. More than 2 years have passed since she had her last psychotic episode. Ana actively discusses the care plan with her case manager and psychiatrist, and they jointly update her goals.

Collecting Case Study Data

A workshop using NASSS-CAT was set up for the project stakeholders. The focus of the workshop was to increase understanding of the influence of complexity in digitalization projects, determine the complexity profile of the dashboard project, explore the usefulness of the NASSS-CAT, and reflect on ways to manage complexity. The 11 participants were line managers, department directors, organization developers, and programmers representing different professions: psychiatrist, psychologist, occupational therapist, and IT developer. Before the workshop, participants received written and oral information about the study and the voluntary nature of their participation. All invited participants gave their informed consent and agreed to join. The day included both small breakout group sessions and large group discussions. Data collection was framed according to the NASSS-CAT and included observations, field notes, notes from the participants, and audio recordings of discussions during the workshop that were subsequently transcribed. The use of NASSS-CAT yielded a complexity profile and the second part of NASSS-CAT, with questions to prompt consideration of how to handle complexity, was particularly used to identify preliminary practical implications. The complexity profile and practical implications were later presented at two feedback sessions to department directors, managers, developers, and assistants involved in planning the future deployment of the dashboard project. Their discussion of the analysis served to validate the findings as a form of member checking [33] and to deepen understanding of practical implications in relation to the project. Data related to the use and usefulness of the NASSS-CAT from the workshop were analyzed by the authors (AG, ACA) using an inductive thematic approach inspired by Braun and Clarke [34]. The resulting themes and selected illustrative quotes are reported in the Results section in the subsection Experiences With Using NASSS-CAT. The piloting of the NASSS-CAT tools in health care settings within the United Kingdom had been approved by the UK Health Research Authority, Health Research Wales, and Health Research Scotland (IRAS no. 258679; REC no. 19/LO/0550). No formal ethical review was required for piloting the tools in Sweden.

Results

The results are presented under three headings: complexity profile, experiences with using NASSS-CAT, and practical implications.

Complexity Profile

Complexity mapping of the 7 NASSS-CAT domains showed significant complexity in 6 of the 7 domains. The domain of intended adopters (ie, health care professionals and patients at the point of care) was perceived to be the least complex.

The Condition or Illness

Schizophrenia is considered to be a complex condition due to its high level of multimorbidity [3,35] and the associated need for multimodal treatment and coordination of care between multiple providers of health and social care [5]. Despite extensive research on schizophrenia and the effectiveness of multimodal treatment programs, challenges remain regarding the successful coordination of multiple providers.

Due to cognitive impairment, persons with schizophrenia have varying degrees of insight into their condition and motivation, which affects adherence to treatment including medication, sometimes resulting in involuntary care and a need for coercive measures. Access to individualized treatment, housing, and support also varies substantially.

The Technology

The dashboard was developed within the Department in collaboration with other psychiatric departments at the hospital. It has significant technical interdependencies with systems controlled by the regional information technology (IT) department. The development of the dashboard has been intertwined with older systems, making use of work processes already in place. There are uncertainties on how to adapt the technology to enable scale-up across the whole department. To what extent the technology will be obsolete within 3-5 years is unknown, but the IT department plans the broad implementation of other new health information systems within that timeframe.

The Value Proposition

The value proposition of the project is uncertain. Case managers report finding the technology useful, as do patients, according to preliminary data. Local testing and piloting have generated evidence of perceived effectiveness, although the degree of cost effectiveness remains unknown. The staff spends less time on related administration. The dashboard provides an overview of patients' progress and risks and supports collaborative planning of care.

The technology's potential value as a commercial product is uncertain and probably impossible to assess because the new technology is interwoven with older systems.

Additional uncertainties are related to the IT department's role in the maintenance and related costs.

The Intended Adopters

The domain of intended adopters is the least complex domain due to a perceived readiness within the organization. The primary users are health care professionals in the care team as well as patients during visits to the outpatient clinics. Secondary users include managers and administrators. The technology is expected to lessen the workload for administrators since more tasks are completed at the point of care by the patient and health care professional. The dashboard pilot testing at 2 outpatient units for 12 and 20 months, respectively, indicated that the innovation is useful for both health care professionals and managers. Furthermore, participating health care professionals report that most patients use the dashboard with ease at yearly follow-up visits. Most patients would prefer to have the next such visit include the dashboard.

The Organization(s)

The Department and technology have a good organization-innovation fit, as the innovation was developed in-house to support the organization's mission and ambitions. Digitalization is perceived as a quality improvement strategy

rather than as product development. Horizon scanning has increased the awareness of innovative technologies, and the organization has a tradition of supporting and trying new ways of working. In recent years, technological innovations have been a focus, enabled by recruitment of IT developers, and embedded in the organization. It has been challenging to pilot and evaluate new technical innovations due to dependencies on the regional IT department. The development of a dashboard is neither part of a regional initiative to develop the future health information system nor part of a product development plan with a clear business case. Internal support has made the pilot tests possible at the Department, but the lack of sponsors at higher organizational levels and uncertainty of the value proposition from a wider organizational perspective add challenges within this domain.

The Wider Context

Changes in the wider context may impact the organization and the introduction of the technology. In particular, implementation of a new health information system can potentially crowd out efforts to deploy the dashboard technology within the Department. There is an enormous drive for innovation and digitalization in Swedish health care, either in the form of large national or regional projects that are deemed hard to influence, or as small projects such as freestanding apps that cannot make use of available health care data. There are few opportunities to learn from other organizations; almost no other organization exists that uses similar technologies, and if they do, they mostly concern patient groups other than those within mental health services.

Project-Specific Complexity

The specific project to develop and deploy the technology across the Department brings challenges related to technical, structural, operational, and sociopolitical complexities. These include the fact that the technology does not yet exist in a robust and dependable form and that regulatory requirements related to secure authentication and access to patient-specific data are not finalized.

Structural and operational complexities include the fact that the technology depends on several other systems to access data. Lines of responsibility for tasks and deliverables are not yet defined, and there is a high dependency on key individuals in a small development team. The people managing the project are not wholly allocated to the project and do not have adequate control over resources, including project staff. Other key projects, particularly the new health information system implementation, can have a major impact on the project.

Sociopolitical complexities stem from the lack of a senior sponsor in the larger hospital and health system organization who recognizes the benefits of the dashboard initiative and can facilitate its progress. Its internal value proposition (within the Department) is clearer than a possible external business case, implying that organizational benefits, costs, and risks are largely unknown.

Experiences With Using the Nonadoption, Abandonment, and challenges to Scale-Up, Spread, and Sustainability Complexity Assessment Tool (NASSS-CAT)

The thematic analysis of data for the stakeholders' experiences of the use and usefulness of NASSS-CAT complexity mapping yielded 3 themes: new insights, threshold to start using NASSS-CAT, and inclusion of relevant stakeholders. These are presented with illustrative quotes from workshop participants. Quotes were translated from Swedish.

New Insights

Using the tool in a workshop increased awareness of the role of complexity. Participants highlighted the importance of identifying complexity and of possibilities to address it in this project in particular and in other future projects in general. Opinions varied about when in the process the tool would bring the greatest benefit or whether it ought to be used throughout the whole project.

> *(Complexity mapping) supports getting perspective and new insights that might help in addressing challenges differently.* [Participant 2, director]

> *I think this (the NASSS-domains) is helpful in (identifying) what to consider in the different digitalization projects and deployment initiatives before getting on with it.* [Participant 5, developer]

Threshold to Start Using the Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability Complexity Assessment Tool (NASSS-CAT)

Participants stated that a basic understanding of the assessment tool´s core concepts would be helpful before starting to use it. During the workshop, the participating authors needed to interpret several NASSS-CAT concepts in light of the local dashboard context for workshop participants (eg, does the "organization(s)" include the IT department or do "users" also include any family members present at a patient's visit to the outpatient unit?).

> *We discussed the meaning of different concepts and had different interpretations. Maybe preparations could have helped.* [Participant 1, manager]

> *Less time had been necessary to use to gain a mutual understanding of concepts, if they, in advance, had been defined more specifically in relation to the project.* [Participant 4, manager]

Inclusion of Relevant Stakeholders

Several challenges surfaced during the workshop that could not be addressed directly because they depended on functions or parts of the organization that were not represented at the workshop, such as the IT department. Participants wished to include them in the complexity assessment to get a better understanding and in the development and deployment of the technology to increase the chances of success. Another issue discussed during the workshop was the importance of creating shared experiences and insights for staff and leaders by allocating time to participate in such an exercise.

> *I think it could have been a very interesting discussion if other parts of the organization had been represented here with more people with other perspectives on these matters. It would have been great.* [Participant 4, manager]
>
> *I believe it would have been important for more people to attend today to share some of the experiences. It creates more power to move on, actually.* [Participant 2, director]

Practical Implications

By using the second part of the NASSS-CAT in particular, with questions to prompt consideration of how to handle complexity, the simple rules for managing technology projects in complex systems (Textbox 1), and the feedback from directors and managers on the complexity analysis through the feedback and validation sessions, the following practical implications were identified, serving as project-specific recommendations:

1. Develop a clear value proposition with information on costs, benefits, and risks. This can guide decisions to make more resources available or to halt further development.
2. Update the overall vision and maintain dialogue to keep it common and up-to-date as the initiative evolves (Rule 2).
3. Strengthen the project's leadership and support structure by clarifying how the project is governed and organized. Earmark resources for it in terms of both money and dedicated time of key individuals (Rules 1 and 5).
4. Maximize benefits and minimize complexity by focusing on parts of the technology/innovation with a low(er) threshold to deployment (Rules 4-9) related to the front-line users: case managers, psychiatrists, and patients. Set and keep the scope of the development and deployment project by using measures to monitor and understand progress and benefits such as saving time, reducing administration, and gaining a better overview.
5. Act strategically in the wider context (Rules 9 and 10) to strengthen the initiative by creating a strategy and plan for communication upwards in the organizational hierarchy and outwards to gain acceptance and sponsorship from key individuals (Rule 3), considering if "rebranding" can make the dashboard's development and its work processes better fit into the policy context, and connecting to other departments that develop, use, or evaluate similar technologies.

Discussion

Principal Findings

The dashboard initiative's complexity profile – with considerable complexity demonstrated in 6 of 7 domains – indicates that the initiative is unlikely to proceed successfully under current circumstances, reflecting the observation of Maylor et al [31]: "the greater the complexity posed by a project, the lower the chance that any successful outcome, let alone an innovative one, will be achieved." Nevertheless, using complexity as a lens when assessing the initiative was perceived as meaningful since it revealed not only challenges but also strengths. Developing the dashboard locally resulted in a high level of engagement and readiness to participate among both health care professionals and patients in focus groups, user testing, and pilots. The collaboration and codesign of technologies and work routines in the local development project might explain why participating stakeholders in the workshop perceived the domain of "intended adopters" to be the least complex when considering further deployment of the dashboard. Perhaps the greatest remaining complexity in the dashboard project concerns the challenges of connecting top-down and bottom-up initiatives and processes, whether related to the development and spread of the dashboard or the related technology, goal alignment, or governance.

Methodological Considerations

There are several limitations to this study. It is a small, single case study of an application of the NASSS-CAT to see if it could be useful in a local development and deployment project. It is restricted to use in the local project and does not test the usability and usefulness of the tool at higher levels in the organization, such as macro-level leadership or involvement from the regional IT department, even though the project has strong interdependencies to those parts of the organization. Patient involvement in the use of the complexity assessment tool could have helped gather further useful information, but the involvement of patients has been restricted to the development of the digital dashboard and not the use of NASSS-CAT. Assessing the complexity of a technology project can aid in understanding and planning but may not be enough on its own to identify what needs to be addressed to succeed with scale-up and spread of innovations [36]. Further research is needed to identify how complexity assessment, using the NASSS-CAT, at various levels of organizations from individuals to the top management can support the development and deployment of new technologies in complex health care contexts.

Conclusions

Complexity assessment of the dashboard project using the NASSS-CAT helped highlight important areas and challenges identified through rigorous research as important in the development and deployment projects of new technologies in health care settings. Experiences from the workshop and validation sessions showed that domains that otherwise might have been neglected received more attention and were brought forward to subsequent planning of the project. The assessment identified strengths of the dashboard initiative to further build upon, while also exposing wider organizational complexity that can challenge the process and spread of the initiative. The assessment helped stakeholders generate specific ideas for how to reduce complexity and strengthen the ability to manage any remaining complexity. This pilot testing of the NASSS-CAT in a real-life setting suggested that the NASSS-CAT can provide participants with a greater understanding of complexities in digitalization projects in general.

Acknowledgments

The authors would like to thank the participants in the workshop and feedback sessions from the Sahlgrenska University Hospital, Gothenburg, Sweden. TG's input to the study was funded by the UK National Institute for Health Research Oxford Biomedical Research Centre, grant no BRC-1215-20008. Partial funding for AG from Kurt and Ingrid Dahréns Foundation is gratefully acknowledged.

Authors' Contributions

All authors were involved in the design and writing of the paper. AG conducted initial data collection and analysis using an early version of the NASSS-CAT (v3.), which was developed by TG's team. A workshop was conducted by AG and ACA to collect data using NASSS-CAT (v4.), and they conducted an initial analysis and reported the results to all authors. All data were used in the final analysis using the latest available version of the NASSS-CAT (v7. Feb 2019) and discussed among all authors. AG was responsible for coordinating the writing of the paper, and all authors contributed throughout the process to both the outline and content. TG also provided the material on the NASSS-CAT. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Struckmann V, Leijten FR, van Ginneken E, Kraus M, Reiss M, Spranger A, SELFIE consortium. Relevant models and elements of integrated care for multi-morbidity: Results of a scoping review. Health Policy 2018 Jan;122(1):23-35 [FREE Full text] [doi: 10.1016/j.healthpol.2017.08.008] [Medline: 29031933]
2. Plsek PE, Greenhalgh T. Complexity science: The challenge of complexity in health care. BMJ 2001 Sep 15;323(7313):625-628 [FREE Full text] [doi: 10.1136/bmj.323.7313.625] [Medline: 11557716]
3. Druss BG, Zhao L, Von Esenwein S, Morrato EH, Marcus SC. Understanding Excess Mortality in Persons With Mental Illness. Medical Care 2011;49(6):599-604. [doi: 10.1097/mlr.0b013e31820bf86e]
4. Preedy VR, Watson R, editors. Handbook Of Disease Burdens And Quality Of Life Measures. New York, NY: Pringer; 2020.
5. Gaebel W. Schizophrenia: Current science and clinical practice. Oxford: John Wiley & Sons; 2011.
6. Forsman J, Taipale H, Masterman T, Tiihonen J, Tanskanen A. Adherence to psychotropic medication in completed suicide in Sweden 2006-2013: a forensic-toxicological matched case-control study. Eur J Clin Pharmacol 2019 Oct 19;75(10):1421-1430. [doi: 10.1007/s00228-019-02707-z] [Medline: 31218371]
7. Tiihonen J, Mittendorfer-Rutz E, Majak M, Mehtälä J, Hoti F, Jedenius E, et al. Real-World Effectiveness of Antipsychotic Treatments in a Nationwide Cohort of 29 823 Patients With Schizophrenia. JAMA Psychiatry 2017 Jul 01;74(7):686-693 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.1322] [Medline: 28593216]
8. 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):1-15. [doi: 10.1186/s12916-018-1050-6]
9. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med 2018 Jun 20;16(1):1-6. [doi: 10.1186/s12916-018-1089-4]
10. Fraser SW, Greenhalgh T. Complexity science: Coping with complexity: educating for capability. BMJ 2001 Oct 06;323(7316):799-803. [doi: 10.1136/bmj.323.7316.799]
11. Data for individual health. Rockville: Agency for Healthcare Research and Quality; 2014 Nov 01. URL: https://digital.ahrq.gov/sites/default/files/docs/publication/2014-jason-data-for-individual-health.pdf [accessed 2018-02-24]
12. Sacristán JA, Dilla T. No big data without small data: learning health care systems begin and end with the individual patient. J Eval Clin Pract 2015 Mar 31;21(6):1014-1017. [doi: 10.1111/jep.12350]
13. Friedman CP, Rubin JC, Sullivan KJ. Toward an Information Infrastructure for Global Health Improvement. Yearb Med Inform 2017 Sep 11;26(01):16-23. [doi: 10.15265/iy-2017-004]
14. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008 Nov(175):1-1422. [Medline: 19408968]
15. Ovretveit J, Nelson E, James B. Building a learning health system using clinical registers: a non-technical introduction. J of Health Org and Mgt 2016 Oct 10;30(7):1105-1118. [doi: 10.1108/jhom-06-2016-0110]
16. Pavel M, Jimison HB, Wactlar HD, Hayes TL, Barkis W, Skapik J, et al. The Role of Technology and Engineering Models in Transforming Healthcare. IEEE Rev. Biomed. Eng 2013;6:156-177. [doi: 10.1109/rbme.2012.2222636]
17. Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector. Bull. World Health Organ 2017 Nov 23;96(1):66-68. [doi: 10.2471/blt.17.197426]

18. van Limburg M, van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011 Dec 28;13(4):e124 [FREE Full text] [doi: 10.2196/jmir.1674] [Medline: 22204896]
19. Celi LA, Davidzon G, Johnson AE, Komorowski M, Marshall DC, Nair SS, et al. Bridging the Health Data Divide. J Med Internet Res 2016 Dec 20;18(12):e325 [FREE Full text] [doi: 10.2196/jmir.6400]
20. Sligo J, Gauld R, Roberts V, Villa L. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017 Jan;97:86-97. [doi: 10.1016/j.ijmedinf.2016.09.007] [Medline: 27919399]
21. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull. World Health Organ 2012 May 01;90(5):357-364. [doi: 10.2471/blt.11.099424]
22. Fitzsimons M, Power K, White M, Cavalleri G, Dunleavey B, Delanty N, et al. Information When and Where Needed for Safe and Effective Integrated Care. Int J Integr Care 2017 Oct 17;17(5):476. [doi: 10.5334/ijic.3796]
23. 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]
24. Greenhalgh T, Maylor H, Shaw S, Wherton J, Papoutsi C, Betton V, et al. The NASSS-CAT tools for supporting technology projects in health and social care: co-design of tools and protocol for further testing. JMIR Research Protocols 2020:1-33 (forthcoming). [doi: 10.2196/preprints.16861]
25. Yin RK. Case study research and applications: Design and methods. Thousand Oaks, CA: SAGE Publications, Inc; 2018.
26. Baker GR. The contribution of case study research to knowledge of how to improve quality of care. BMJ Quality & Safety 2011 Mar 30;20(Suppl 1):i30-i35. [doi: 10.1136/bmjqs.2010.046490]
27. Nielsen KA, Svensson L. Action research and interactive research: beyond practice and theory. Maastricht: Shaker Publishing; 2006.
28. Ovretveit J. Action evaluation of health programmes and changes: a handbook for a user-focused approach. Abingdon: Radcliffe Medical; 2002.
29. Greenhalgh T, Shaw S, Wherton J, Hughes G, Lynch J, A'Court C, et al. SCALS: a fourth-generation study of assisted living technologies in their organisational, social, political and policy context. BMJ Open 2016 Feb 15;6(2):e010208. [doi: 10.1136/bmjopen-2015-010208]
30. Greenhalgh T. How to improve success of technology projects in health and social care. Public Health Res Pract 2018 Sep 27;28(3):1-4 [FREE Full text] [doi: 10.17061/phrp2831815] [Medline: 30406256]
31. Maylor HR, Turner NW, Murray-Webster R. How Hard Can It Be?: Actively Managing Complexity in Technology Projects. Research-Technology Management 2015 Dec 28;56(4):45-51. [doi: 10.5437/08956308x5602125]
32. Gremyr A, Malm U, Lundin L, Andersson A. A learning health system for people with severe mental illness: a promise for continuous learning, patient coproduction and more effective care. Digital Psychiatry 2019 Jun 06;2(1):8-13. [doi: 10.1080/2575517x.2019.1622397]
33. Creswell JW, Creswell JD. Research design: Qualitative, quantitative, and mixed methods approaches. Thousand Oaks, CA: SAGE Publications, Inc; 2017.
34. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Druss BG, Walker ER. The Synthesis Project: New insights from research results. 2011 Feb. Mental disorders and medical comorbidity URL: https://www.integration.samhsa.gov/workforce/mental_disorders_and_medical_comorbidity.pdf [accessed 2020-02-18]
36. Greenhalgh T, Papoutsi C. Spreading and scaling up innovation and improvement. BMJ 2019 May 10:l2068. [doi: 10.1136/bmj.l2068]

Abbreviations

IT: information technology.
NASSS: The Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability framework.
NASSS-CAT: The Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability framework and complexity assessment tool.

Edited by G Eysenbach; submitted 17.07.19; peer-reviewed by S Hinder, S Shaw, J Lynch; comments to author 03.10.19; revised version received 21.11.19; accepted 03.02.20; published 23.04.20.

Please cite as:
Gremyr A, Andersson Gäre B, Greenhalgh T, Malm U, Thor J, Andersson AC
Using Complexity Assessment to Inform the Development and Deployment of a Digital Dashboard for Schizophrenia Care: Case Study
J Med Internet Res 2020;22(4):e15521
URL: http://www.jmir.org/2020/4/e15521/
doi:10.2196/15521
PMID:32324143

Original Paper

Smart, Remote, and Targeted Health Care Facilitation Through Connected Health: Qualitative Study

Sonia Chien-I Chen[1*], BA, MSc, PhD; Ridong Hu[1*], PhD; Rodney McAdam[2*], PhD

[1]Institute of Quantitative Economics, Huaqiao University, Xiamen, China

[2]Ulster Business School, Ulster University, Newtownabbey, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Ridong Hu, PhD
Institute of Quantitative Economics
Huaqiao University
No. 668, Jimei Avenue
Jimei dist
Xiamen, 361021
China
Phone: 86 17750619886
Email: j_rdhu@hqu.edu.cn

Abstract

Background: Societies around the world are aging. Widespread aging creates problems for social services and health care practices. In this light, research on connected health (CH) is becoming essential. CH refers to a variety of technological measures that allow health care to be provided remotely with the aim of increasing efficiency, cost-effectiveness, and satisfaction on the part of health care recipients. CH is reshaping health care's direction to be more proactive, more preventive, and more precisely targeted and, thus, more effective. CH has been demonstrated to have great value in managing and preventing chronic diseases, which create huge burdens on health care and social services. In short, CH provides promising solutions to diseases and social challenges associated with aging populations. However, there are many barriers that need to be overcome before CH can be successfully and widely implemented.

Objective: The research question of this study is as follows: How can CH facilitate smart, remote, and targeted health care? The objective is to identify how health care can be managed in more comprehensive ways, such as by providing timely, flexible, accessible, and personalized services to preserve continuity and offer high-quality seamless health care.

Methods: A qualitative approach was used based on 60 multistage, semistructured stakeholder interviews.

Results: The results can be divided into two functions of CH: ecosystem and platform. On the one hand, the interviews enabled the authors to develop a stakeholder classification and interaction diagram. These stakeholders interacted sequentially to provide technology-based content to end users. On the other hand, interviewees reflected on how CH serves as a platform to address remote monitoring and patient self-management. In the Discussion section, three innovation strategies are discussed to reflect the manner in which CH promotes smart, timely, and precise health care.

Conclusions: This study indicates that it is essential to continually revise CH business models, given the ongoing and rapid changes in technology across groups of CH stakeholders. We also found that global trends toward smart, timely, and precise health care shape what individuals expect from products and services, providing firms with unique opportunities for growth.

(J Med Internet Res 2020;22(4):e14201) doi:10.2196/14201

KEYWORDS

connected health care; smart health care; health care quality; access; remote monitoring; precision medicine; self-management

Introduction

Background

Our society is aging. This trend is projected to result in increased chronic health conditions and potential labor shortages. As traditional health care models are not fully equipped to face the unique challenges of providing health care for an aging society, innovative practices are being proposed and developed to decrease the pressures of widespread aging. Through self-management and remote monitoring platforms, connected health (CH) has been offered as a new technology-based model of health care delivery that presents a promising solution for future, aging-oriented health care [1].

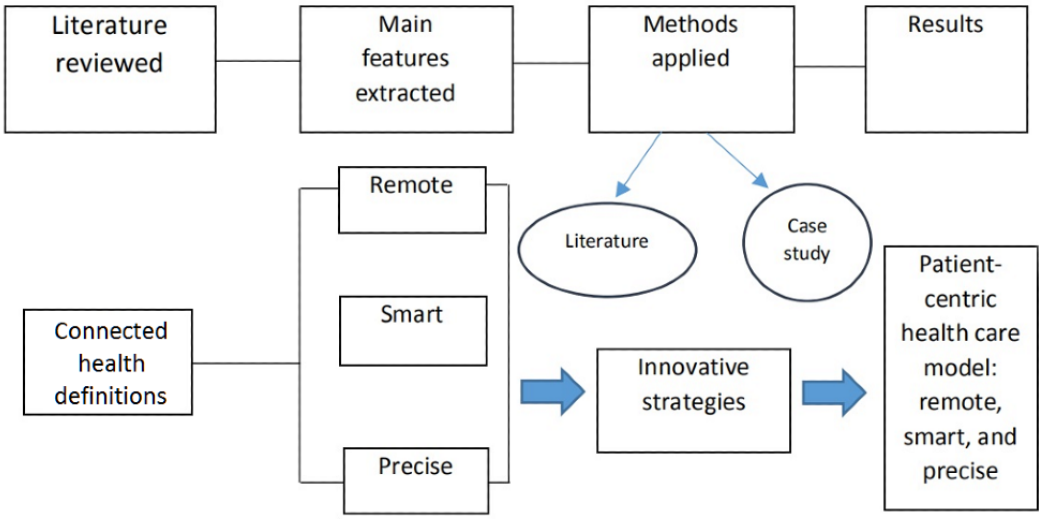
The rise of CH is attributable to four factors. The first factor is related to the widespread tendency to pursue excellence in health care through the promotion and monitoring of health care services' quality, efficiency, safety, and level of customer service. Because of this tendency, health professionals and institutions demand that health care services be more accessible, of a higher quality, and more efficient, so they often turn to CH [2]. Second, aging populations have increased health care costs due to the increase in chronic conditions, higher survival rates among patients fighting serious diseases, and longer lifespans that accompany aging [3]. Thus, the health care economy has become more complex than in the past as evidenced by the rising costs of caring for changing demographics. Third, increasingly frequent provider shortages and other relevant issues, including the geographic dispersion of families and troubling disparities in care between ethnic groups, are significant concerns for aging populations [4]. Lastly, patients in varying contexts have demanded better customer service from health care providers [5]. Some authors suggest that the development of consumerism in health care may be a catalyst for the development of patient-centric health care [6-8]. In combination, these factors have created a stronger impetus to force health care innovation from both within and outside the ecosystem.

Although CH seeks to make health care services more proactive, preventive, and precisely targeted, there are many challenges and barriers to its sustainable implementation that need to be addressed and overcome. These challenges include the cost of devices, privacy and data security concerns, health system bureaucracy, and training health professionals and patients to use new technologies. This study identifies strategies to make health care systems smarter, timelier, and more precise by exploring the interaction of stakeholders in the CH ecosystem and CH's role as a platform.

Study Design and Conceptual Framework

Figure 1 illustrates the conceptual framework that forms the basis for this study. First, we reviewed definitions of CH in the literature to determine its main features. Second, three features of timely, smart, and targeted care were determined and applied to evaluate how patient-centric health care services can be delivered through CH. Third, we conducted interviews and engaged case studies to identify effective strategies for the implementation of CH. Although CH is promising, many challenges, such as issues of cost, infrastructure, technology, and business sustainability, still remain to be overcome. Therefore, we examined extant and innovative strategies for implementing CH and/or augmenting existing health care systems and interviewed relevant stakeholders in the CH ecosystem to discover how to reach these goals. Finally, the results were analyzed with an eye to expressing the implications of addressing a comprehensive health care system that can fulfill patients' needs and suggestions on how to do so.

Figure 1. The study's design and conceptual framework.



Definition of Connected Health

Broadly speaking, CH is an *umbrella term* that covers the telemedicine family spectrum. An increasingly developed base of knowledge around CH and CH practices has generated many definitions of exactly what CH is [2,8]. For instance, the American Medical Association (AMA) defines CH as a model that utilizes technology to maximize health care resources and offer enhanced, flexible opportunities for patients to engage with clinicians and better self-manage their care [9]. In order to achieve these goals, many technologies, including telemedicine and mobile health, are used to facilitate remote,

mobile, and site-to-site medical care. To focus our research, we classified definitions of CH into three types based on their salient features: remote, smart, and precise. This study then defines CH as a platform that offers remote monitoring and self-management, helping patients to address their own health care needs in a smart, timely, and precise manner through information and communication technologies (ICTs). Below, we briefly discuss our three classifications of CH definitions.

Remote Health Care

Remote medical care is a key feature of CH [3-5,9-12]. Studies indicate that both the concept of remote health care and the practical application of information technology systems to health care include aspects of remote medical care, such as a *telecare medicine information system*, a *personally controlled health records system*, and *patient monitoring*. These studies further point out that in CH applications, "user authentication can ensure the legality of patients' care" [13]. Remote health care–focused definitions of CH are often seen as part of a new lexicon for telemedicine, even though these definitions and their proponents may exhibit a greater focus on how to best connect clients and health care professionals.

Smart Health Care

Although technology plays a significant role in CH, researchers indicate that CH is not just about technologies, but its relationship with people and the health care system [14]. Smart-oriented definitions of CH demonstrate how CH is not just about health care and technology, but about managing patients and their care [12,14]; they indicate that *CH is a new model for health management* and that CH has the potential to *put the correct information in the correct hands at the correct time*, thus improving decision making and care. These decisions can "save lives, save money and ensure a better quality of life during and after treatment." CH can help deliver better health outcomes by allocating health resources more efficiently and effectively via the better management and integration of health care systems and services, which are, in theory, typified by smart health care.

According to the literature review, CH tends to be a combination of people, processes, and technology [4,15]. Take, for example, the following definition: "connected health...refers to a conceptual model for health management where devices, services or interventions are designed around the patient's needs" [15]. Based on this definition, CH is a patient-centered care model where the patient is at the center of the processes that connect stakeholders—a process that takes place over a variety of settings, from the patient's home to an acute care setting. This definition also indicates that patient care pathways, revenue models, and data analytics can be connected through technology in the CH platform. Therefore, technology can enable the implementation of a more proactive, episodic health care model, which is starkly different from the more reactive model of conventional health care. With the assistance of technology, health care professionals, patients, and/or caregivers can be empowered to engage in more effective and efficient health care.

Precise Health Care

CH differs from conventional health care in that it collects data in a timely and seamless fashion, thus providing more information for clinicians to address patients' precise health care needs [2]. Furthermore, "stakeholders in the process are connected...using timely sharing and presentation of accurate and pertinent information regarding patient status through the smarter use of data, devices, communication platforms and people" [16]. Therefore, through CH, patients can receive care in a manner that is as efficient and precise as possible.

Value Propositions Based on Trends

As noted above, the patient-centric values of CH can be defined and classified into smart, timely, and precise care. The relevant literature reviews how such values can be proposed and captured through strategies. Values are created by a value chain comprising suppliers, firms, and buyers [17]. How much value can be captured via strategies depends on the capacity of these strategies to add value to a given service or care practice. The analysis of value - based business strategies depends on the cooperative game theory among firms, how they act toward one another as competitors, firms' willingness to pay, and the opportunity costs of acting. Although value types vary, value propositions in business models are the key to increase the attraction of firms' willingness to pay [18]. Three innovative strategies have been identified to potentially rebuild the business model: *infuse and augment*, *combine and transcend*, and *counteract and reaffirm* [19]. We review examples from prominent companies in the literature to support the idea that these strategies can influence success in service innovation [12,19]. Furthermore, innovation should be inspired by clients' motivations and innovations' appeal to clients; therefore, using these strategies may be an effective way to achieve this goal [20].

Despite the importance of value propositions and strategic planning, most research focuses on two primary performance factors: cost-effectiveness and operational efficiency [21]. Although the qualities of timely, smart, and precise health care are essential to create a patient-centric health care model, more research is required to explore how these goals can be achieved in practice. According to the literature, technology is not the sole factor driving more personalized health care services. Management and other factors that involve qualitative components may help to maintain human esteem and dignity through a CH platform [22,23]. Although CH's value has been demonstrated in the management of chronic diseases, researchers expect tremendous progress in CH's applications in disease prevention [24]. The proliferation of advanced technology enables CH to be more proactive than conventional medicine [25]. Thus, a qualitative approach is employed to achieve the goal of providing more proactive, preventative, and precise health care.

Despite the positive impacts of CH described above, barriers to implementing CH still need to be identified and overcome in order to accelerate and deliver more effective and cost-efficient health care models [15]. Health care that is proactive, preventive, and targeted is likely to reduce many of the burdens of an aging society [26,27]. Yet, there is still a gap

in the research regarding how medical interventions such as the implementation of CH can both be effective and also maintain the dignity of elderly patients [28]. To fill this gap, this study explores the relationship between technology and wellness, and how both have been optimized in certain cases. It identifies the preconditions and essential requirements for developing a CH ecosystem and provides typical, illustrative case studies to answer the research question. Therefore, this study aims to qualitatively explore and discover hitherto unknown aspects of how health care can be more comprehensive by being smarter, timelier, and more precise.

Methods

Overview

This study uses literature reviews and empirical research to investigate how timely, smart, and targeted health care can be achieved via CH platforms. Although the majority of CH research has used quantitative methods to test and validate assumptions [29-31], this study utilized a qualitative research method, conducting 60 interviews that were designed to answer the research question and explore the unknown and novel phenomena of health care in an aging society. We employed three methods to identify stakeholders in the CH ecosystem: (1) reviewing the CH organizational process assets (ie, the plans, processes, policies, procedures, and knowledge bases specific to and used by the performing organization) and their environmental factors; (2) interviewing experts in the field of CH; and (3) conducting brainstorming sessions. Regarding CH-related value propositions, the strategies based on trends in the literature are relevant. Therefore, the innovative strategies proposed in previous studies [19] are employed as a foundation for classifying and discussing the results of this study.

Research Methodology

Owing to the developmental nature of CH literature, this study used an exploratory and inductive theory-building approach based on our 60 multistage, semistructured stakeholder interviews. The target sample we used to address the study aims and research question was CH stakeholders and experts, including medical researchers, industry leaders, government officials, and end users. This qualitative methodology facilitates a better understanding of a rapidly developing discourse, such as that of CH, by exploring and using insightful input, having broader discussions, and synthesizing diverse opinions [29,31]. Thus, this approach can be used to bridge the knowledge gaps in literature.

Recruitment Criteria of the Participants

Stakeholders and experts were selected for the interviews according to the literature review and snowball sampling [32]. These participants were interviewed on a voluntary basis and gave verbal consent according to ethical guidelines. This study was approved by Ulster University's institutional review board (reference No. RG3 RMcAdam2). The study's participants included health care professionals, industry players, academic researchers, and government agents. The researchers identified participants' backgrounds, competencies, gender, etc, as they may relate to the outcomes of this research. Our semistructured interview guide comprised the following: (1) working sectors, (2) background information, (3) gender, and (4) levels of competency.

Taiwan was selected as a case study to explore the myriad phenomena accompanying aging populations and test strategies for the development and implementation of CH. We chose this case study because Taiwan has experienced both significant challenges and progress in its core health care activities. Another reason for choosing Taiwan was because of its complete CH business ecosystem, which includes mature ICT applied to health care practices and an integrated health care system. Approaching our research questions through the multi-stakeholder case study of Taiwan helps us to explore rich contextual information by asking *what* and *how* questions, rather than generalizing principles to a population [29,31]. Multistage interviews were conducted to deepen the research by iteratively collecting data [29]. The data were collected using a multistage approach with interviewees, considering the CH ecosystem's progressive and multidimensional nature. This is illustrated in Tables 1-3.

Pilot interviews were conducted (n=16) with key CH influencers across stakeholder groupings to explore the CH phenomenon and identify the primary contributing factors and stakeholders in Taiwan's CH ecosystem. Next, Stage 1 interviews (n=22) were held with CH stakeholders who were identified from the pilot stage interviews' findings and analysis. An analysis was conducted using a snowball sampling method [32]. Stage 2 interviews (n=22) were conducted 1 year after the Stage 1 interviews to explore how systemic problems develop and persist over time in CH ecosystems and to review longitudinal changes from a Business Model Innovation (BMI) perspective in terms of the sustainability of CH.

The pilot interviews were conducted in order to confirm the need for, and significance of, further CH research. This pilot study was also advantageous because it identified potential directions and perspectives for conducting the multistage interviews, which helped us collect data more effectively. The results of the pilot interviews suggested that further study of CH is meaningful. Barriers that obscured the development of CH were further investigated in the Stage 1 interviews. As mentioned above, 16 interviewees were included in the pilot interviews (see Table 1). The majority were health care providers, as they have rich experience of CH technologies and services. Their feedback was significant in reflecting the effectiveness of CH products and services. Therefore, understanding their views is important in the design of successful and user-friendly interactive systems. This pilot interview focused on medical doctors and health professionals rather than patients and their families, due to ethical concerns.

Table 1. Profiles of the pilot study interviewees.

Participant number	Sector and participants in that sector (n=16), n (%)	Organization	Gender[a]	Title
1	Industry, 2 (13)	June Sun Digicom	Male	General Manager
2	Industry, 2 (13)	DigiO2	Male	Tech Advisor
3	Government or semigovernment, 2 (13)	Taiwan Forces for Medical Travel	Male	Chief Executive Officer (CEO)
4	Government or semigovernment, 2 (13)	Taiwan Forces for Medical Travel	Female	Vice Project Manager
5	Academia, 1 (6)	Taipei Medical University, School of Gerontology Health Management, College of Nursing	Female	Assistant Professor or Attending Physician
6	Health care providers, 11 (69)	Chang Gung Memorial Health Village	Female	Staff
7	Health care providers, 11 (69)	Antai Medical Care Hospital	Male	Director or General Practitioner (GP)
8	Health care providers, 11 (69)	Antai Medical Care Hospital	Female	GP
9	Health care providers, 11 (69)	San-Chung Health Center	Female	Head Nurse
10	Health care providers, 11 (69)	YR Chinese Medicine Clinic	Male	Doctor
11	Health care providers, 11 (69)	YR Chinese Medicine Clinic	Female	Manager
12	Health care providers, 11 (69)	Tri-service General Hospital	Male	Medical Doctor
13	Health care providers, 11 (69)	Taiwan University Hospital	Male	Medical Doctor
14	Health care providers, 11 (69)	Kaohsiung Municipal Hsiaokang Hospital	Female	Pharmacist
15	Health care providers, 11 (69)	Zhang Bi Zheng Family Physicians' Clinic	Male	Director or GP
16	Health care providers, 11 (69)	Home Physician	Male	Therapist

[a]The sample was made up of 56% (9/16) males and 44% (7/16) females.

Table 2. Profiles of Stage 1 interviewees.

Participant number	Sector and participants in that sector (n=22), n (%)	Organization	Gender[a]	Title
17	Industry, 8 (36)	Netown	Male	Sales
18	Industry, 8 (36)	Far EasTone Telecommunications	Male	Senior Engineer
19	Industry, 8 (36)	Far EasTone Telecommunications	Male	Senior Engineer
20	Industry, 8 (36)	Huede Technology	Male	Chief Executive Officer (CEO)
21	Industry, 8 (36)	Huede Technology	Female	Nurse or Health Management
22	Industry, 8 (36)	Guidercare	Male	Vice President
23	Industry, 8 (36)	Acomotech	Male	Tech Advisor
24	Industry, 8 (36)	Medsense	Male	Founder
25	Government, 2 (9)	Ministry of Health & Welfare	Male	Director of ICT[b] Department
26	Government, 2 (9)	Sang Chung Health Center	Male	Administrator
27	Academia, 3 (14)	National Taiwan University	Male	Director or Professor
28	Academia, 3 (14)	National Taipei University of Technology	Female	Lecturer
29	Academia, 3 (14)	UL (Underwriters Laboratories) Life & Health	Male	Sales Manager
30	Health care providers, 9 (41)	Luo Dong Care Institute	Male	CEO
31	Health care providers, 9 (41)	En Chu Kong Hospital	Female	Senior Nurse
32	Health care providers, 9 (41)	Taoyuan Fu Hsing Township Health Station	Male	Director or General Practitioner (GP)
33	Health care providers, 9 (41)	Taoyuan Fu Hsing Township Health Station	Female	Head Nurse
34	Health care providers, 9 (41)	Taipei Medical University, Telehealth & Telecare Center	Female	Director or Health Management
35	Health care providers, 9 (41)	Mennonite Christian Hospital	Male	Management of Information Service (MIS) Director
36	Health care providers, 9 (41)	Mennonite Christian Hospital	Female	Head Nurse
37	Health care providers, 9 (41)	Taiwan University Hospital, Telehealth Center	Female	Nurse or Health Management
38	Health care providers, 9 (41)	Taiwan University Hospital	Male	GP

[a]The sample was made up of 68% (15/22) males and 32% (7/22) females.

[b]ICT: information and communication technology.

Table 3. Profiles of Stage 2 interviewees.

Participant number	Sector and participants in that sector (n=22), n (%)	Organization	Gender[a]	Title
39	Industry, 15 (68)	Agfa HealthCare Image (imaging technology)	Male	Consultant
40	Industry, 15 (68)	Asus Compute Inc (internet technology)	Male	Senior Manager
41	Industry, 15 (68)	G-cloud UK (cloud computing technology)	Male	Chief Executive Officer (CEO)
42	Industry, 15 (68)	Sheng-En Development Co (value network)	Male	Division Manager
43	Industry, 15 (68)	Asus Cloud Corporation (internet, storage, and cloud computing technology)	Male	Department Director
44	Industry, 15 (68)	Far EasTone Telecommunications	Male	Director
45	Industry, 15 (68)	Taidoc/Fora Care	Male	Sales Manager
46	Industry, 15 (68)	Smart Catch International Co, Ltd	Female	Specialist
47	Industry, 15 (68)	MitraStar	Male	Senior Director
48	Industry, 15 (68)	Gemtek Technology Co, Ltd	Female	Product Manager
49	Industry, 15 (68)	Z-Com/ZWA Inc	Male	Sales Director
50	Industry, 15 (68)	Isentek/Partnership with TXCorpration	Female	Deputy Sales Manager
51	Industry, 15 (68)	Through Tek Technology (TUTK) Co, Ltd	Female	Product Manager
52	Industry, 15 (68)	Pioneer Material Precision Tech	Female	Buyer
53	Industry, 15 (68)	Auden Group	Female	Executive Assistant
54	Health care providers, 6 (27)	CH[b] Healthy Village (value network)	Female	Division Manager
55	Health care providers, 6 (27)	Changhua Christian Hospital (CCH) (telecare health service center) (BMI[c])	Female	Head Nurse
56	Health care providers, 6 (27)	CK Memorial Hospital (health professional)	Female	Senior Pharmacist
57	Health care providers, 6 (27)	Tai Tong Health Center (CH consumer in remote area)	Male	Medical Doctor
58	Health care providers, 6 (27)	Show-Chwan Hospital	Female	Nurse
59	Health care providers, 6 (27)	Show-Chwan Hospital	Female	Nurse
60	Academia, 1 (5)	Industrial Technology Research Institute/Ministry of Economic Affairs (MOEA)	Female	Manager

[a]The sample was made up of 45% (10/22) males and 55% (12/22) females.

[b]CH: connected health.

[c]BMI: Business Model Innovation.

The Stage 1 interview asked participants to consider problems reported in the literature on CH. These questions covered the cost of implementing CH and creating the necessary infrastructure and technology, business sustainability, different CH business models, collaboration, and communication-related issues. The results were analyzed to identify further knowledge gaps and to increase the depth of current knowledge. Stage 2 interviews addressed the gaps identified in Stage 1 and sought to discover insights of sustaining CH businesses. Due to the challenges of time limits and the distance of remote areas in Taiwan during data collection, we required two phases of interviews to ensure that we adequately covered some important issues and themes.

Interviews lasted between 1 and 2 hours and were transcribed and coded using NVivo 10 (QSR International) qualitative data analysis software. Different data sources, including interviews, documents, and public records, were synthesized and coded for data triangulation to increase the credibility and validity of the study's results [29]. Data were cleaned through an integration process to merge different terms with similar meanings. For

example, CH could be called *remote health*, *telehealth*, and *telecare* in the interviews; these terms were merged because the interviewees used them to refer to the same thing. Inaccurate, incomplete, or irrelevant parts of the dataset were detected, corrected, or removed from the raw data in the data-cleaning phase. The data were then clustered and coded.

The data analysis strategy of this study included thematic and systematic approaches in a collaborative qualitative analysis. The data collected underwent a deep familiarization, and this led to the development of major themes, which in turn organized the thematic analysis. This study utilized NVivo 10 to analyze the interview data by grouping words or sentences based on how frequently they occurred in the interviews. Content analysis was conducted to objectively and systematically induce meaning on the content through coding [33]. This analysis adopted a grounded theoretical approach. Initially, we drew out coded data and themes with the potential for rapid change. The number of new and emerging themes gradually diminished, and the number of refinements to these themes decreased until theoretical saturation was met. The data collection and iterative analysis were repeated until saturation occurred. The data were summarized, confirmed, and discussed in subsequent interviews with the same interviewees from the pilot stage, Stage 1, and Stage 2. This process was intended to ensure the data's reliability and validity. This method is beneficial as the interviews' primary content can be confirmed and corrected quickly by experts in the CH ecosystem. Throughout the process, constant reference to existing literature contributed to the data analysis [29].

Results

This study employed both thematic analysis and systematic content analysis. The results are sequentially demonstrated in the following subsections.

Results of Thematic Analysis

Figure 2 shows the results of thematic analysis in terms of how qualitative data were coded and iteratively compared and integrated into themes. For instance, the description "the distance and time spent for remote residents visiting health facilities..." and the non-cost-effectiveness of "...remote health..." were coded as geographic isolation barriers and developed into the theme of *remote health* according to iterative comparison. Other interview data were clustered and merged according to a similar mechanism.

Figure 3 illustrates a map of how codes converged as themes from the qualitative data. Codes for integration, communication, bureaucracy, ecosystem, self-management, and interdisciplinary talent were merged as the theme *smart health*. Seamless, contiguity, prevention, targeted health, and personalized health were categorized into *precision medicine*. The theme *remote health* covered flexibility, accessibility, empowerment, timely manner, geographic isolation, wider coverage, and security concerns. In summary, smart health, precision medicine, and remote health consisted of the CH features.

Figure 2. Results of thematic analysis.

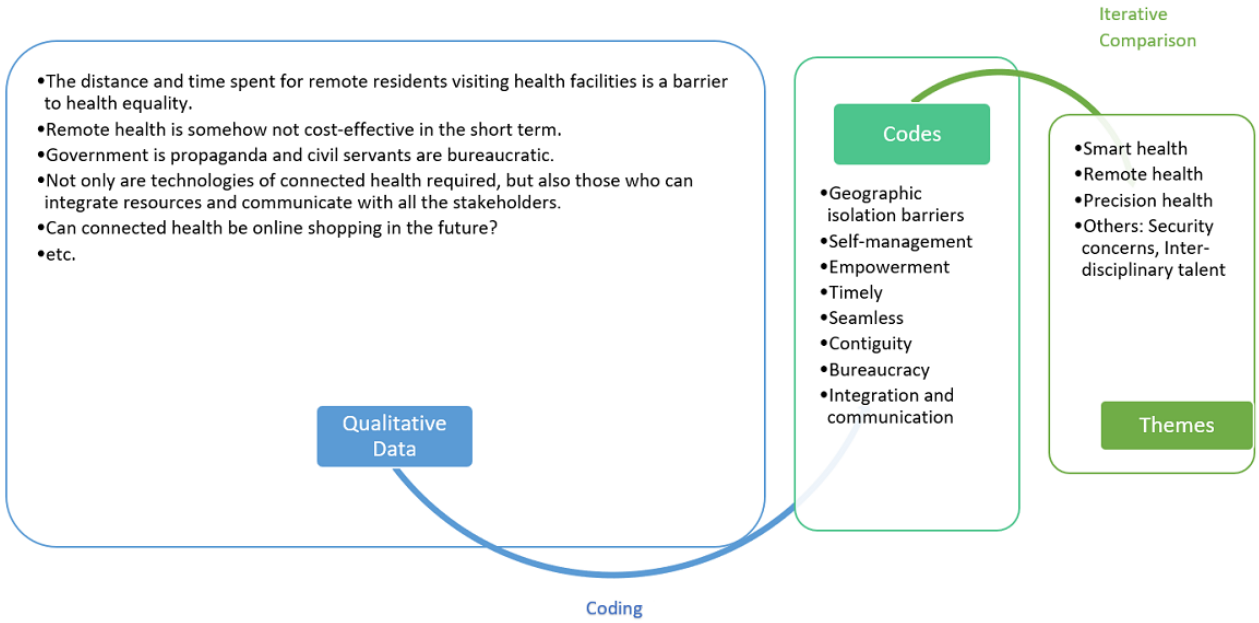
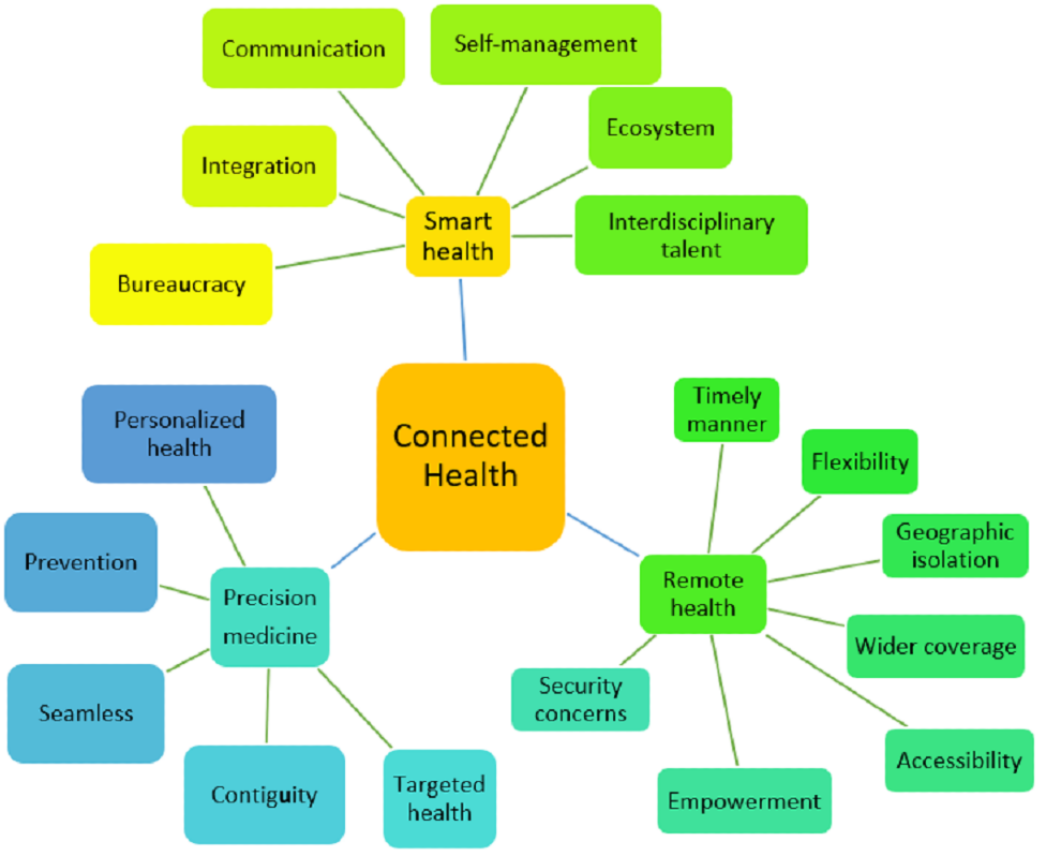


Figure 3. Thematic analysis mapping.



Results of Systematic Analysis

NVivo 10 was used to analyze the interview data by grouping words or sentences based on how frequently they occurred in the interviews. The results are shown in Table 4 and Multimedia Appendix 1. According to the word cloud of Multimedia Appendix 1 it is noticeable that keywords such as care, business, data, system, model, time, technology, government, and management were most frequently mentioned in the interviews. The word-frequency framework was used to identify meaningful words and their potential significance. The words that frequently occurred were clustered into three themes: technology, organization, and leadership and management issues (see Table 4). These words that were shown the most frequently were compared according to the original context and grouped into key themes.

Results of the analyses conducted on the interview data are summarized and presented in the following sections: (1) CH as an ecosystem, (2) CH as a platform, and (3) CH's innovative strategies based on value propositions. First, the principles adopted to identify stakeholders in the ecosystem, based on the power and interest of their ability to affect CH's objectives, were reviewed. The stakeholder interaction diagram (see Figure 4) was created based on the foundation of a previous study that identified stakeholders in the CH ecosystem [34]. In our diagram, governments take the lead and initiate the infrastructure projects needed to implement CH, and the sequential interaction that follows provides technology-based content to end-user stakeholders. Second, CH as a platform enables remote monitoring and makes patients' self-management possible. Moreover, interviewees contributed insights regarding the challenges that the implementation of CH faces and how they can be addressed. Third, CH's innovative strategies based on value propositions were referred to the value propositions from current trends, which may inspire practitioners and policy makers. Three innovation strategies, based on information emanating from Harvard Business School, are discussed to reflect how timely, smart, and targeted health care can be delivered via the CH platform [12].

Figure 4 illustrates the overall CH ecosystem, along with relationships between stakeholders and players [35]. Figure 4 identifies eight types of stakeholders and players, which can be classified into three sectors: government, industry, and academia. Policy making and infrastructure building, led by government, initiates the implementation of CH. Figure 4 demonstrates that governments are influential throughout the CH stakeholder ecosystem, as they fund many short-term projects and initiate the infrastructure required for CH's implementation.

Academia allows the research, discovery, innovation, and transfer of knowledge regarding CH processes to entities such as government stakeholders, and academic stakeholders are influential throughout the CH ecosystem. Although academia is primarily involved in the development of CH technology, it has become increasingly involved in the social, business, and marketing aspects of value-centric CH.

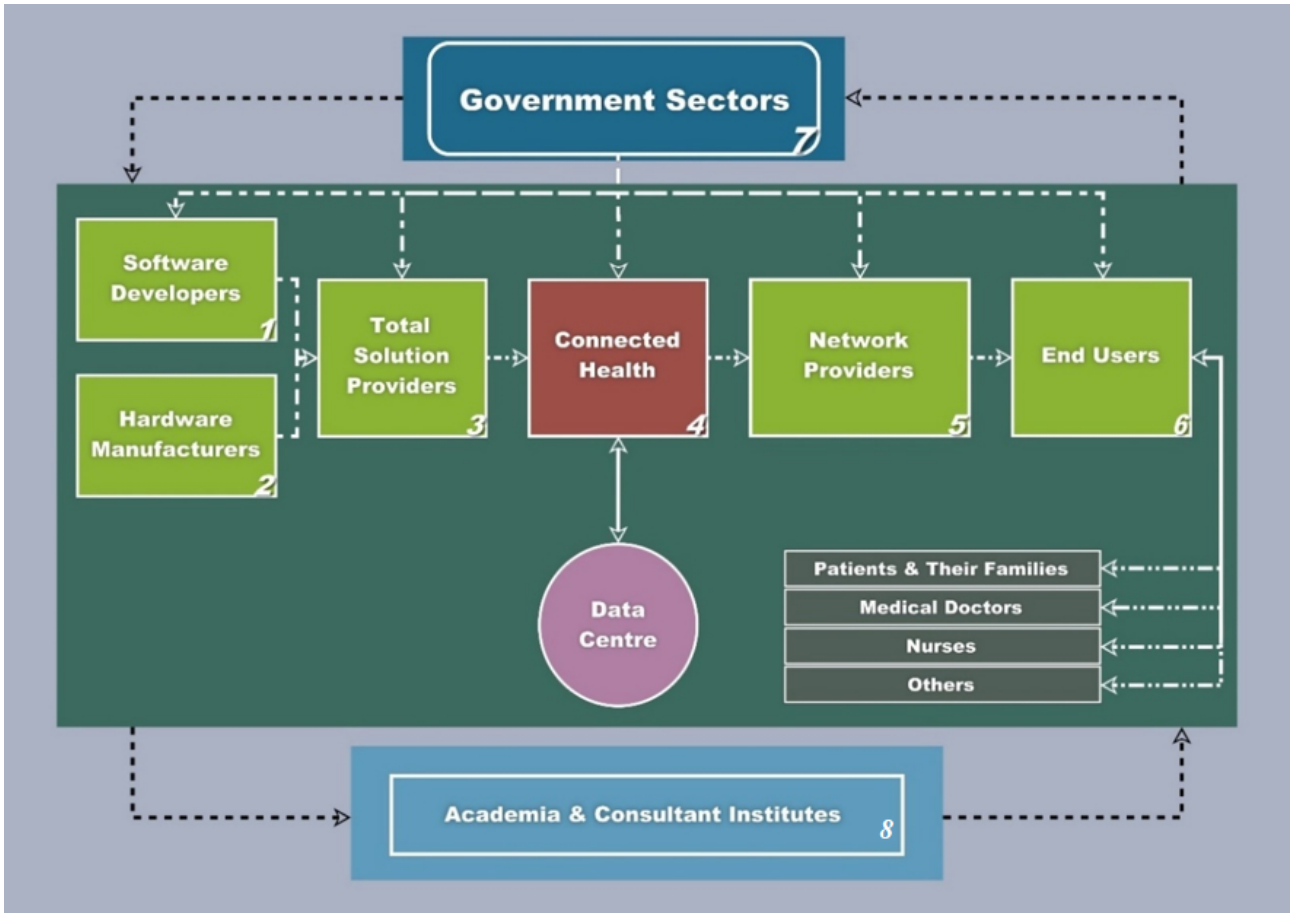
Finally, industry stakeholders are major players in the CH ecosystem who offer services and manufacturing products for use by end users. Industry covers a variety of domains: (1)

software developers, (2) hardware manufacturers, (3) total solution providers, (4) CH service providers, (5) network services, and (6) end-user stakeholders. Besides the three major stakeholders mentioned above, other actors in the CH system include patients, patients' families, doctors, nurses, caregivers, and additional persons, primarily volunteers.

Table 4. Themes and word frequencies.

Theme and words	Word frequency (N=61,489), n (%)
Technology issues	
Care	289 (0.47)
Data	152 (0.25)
System	116 (0.19)
Model	112 (0.18)
Technology	109 (0.18)
Time	108 (0.18)
Information	93 (0.15)
Remote	79 (0.13)
Distance	72 (0.12)
Concept	64 (0.10)
Support	64 (0.10)
Research	61 (0.10)
Staff	61 (0.10)
Personal	60 (0.10)
Policy	60 (0.10)
Measuring	57 (0.09)
Cost	53 (0.09)
Innovation	53 (0.09)
Quality	48 (0.08)
Mobile	40 (0.07)
Cloud	42 (0.07)
Device	39 (0.06)
Smart	34 (0.06)
Organization issues	
Government	95 (0.15)
System	116 (0.19)
Business	189 (0.31)
Model	112 (0.18)
Support	64 (0.10)
Concept	64 (0.10)
Staff	61 (0.10)
Public	68 (0.11)
Leadership and management issues	
Management	83 (0.14)
Support	64 (0.10)
Time	108 (0.18)
Research	61 (0.10)
Concept	64 (0.10)

Figure 4. Diagram of connected health players and stakeholder relationships (Chen, 2018 [34]).



Connected Health as a Coexistent Ecosystem

Smart health care via a CH platform explores how health care and technology can coexist, each dependent on the other, rather than operating as separate entities in the CH ecosystem. This coexistence not only creates a health care system that can better respond to the needs of an aging society, but has the side effect of developing smarter health care methods. Therefore, it is relevant to reconsider the conventional concepts of technology and health care. As an ecosystem, CH is surrounded by software providers, hardware manufacturers, and health care professionals (see Figure 4).

As far as the sensitive nature of health care data is concerned, cybersecurity plays an important role in ensuring the safety of patients' data. Therefore, it is critical to include the role of *information technology (IT) security specialists* in the CH ecosystem. They are incorporated within both the categories *software developers* and *hardware manufacturers*. Other relevant roles are covered in these eight categories, even though not all of them are named in Figure 4. For example, the role *stakeholders with standardization* is included within the category *academia and consultant institutes*. What is noticeable is that the role of *interoperability personnel* still remains to be developed according to many cases interviewed, although this role is essential for the success of CH's implementation.

Followed by the main stakeholders in Figure 4, the classification of organizations and participants interviewed is provided in Table 5. This table classified the 60 participants from 34 organizations into eight categories according to the categories in Figure 4. When we found that some organizations bridged two or three categories, we classified them according to their main businesses. Despite having their own businesses and business concerns, they have to coexist to activate CH's implementation.

Connected Health as a Platform of Remote Monitoring and Self-Management

This section presents the interview results regarding how CH functions as a platform of remote monitoring and self-management, based on the perspectives of participants shown in Table 6.

Table 5. The classification of organizations interviewed.

Category and numbers	Organization name
1. Software developers	
101	Far EasTone Telecommunications Smart City Division
102	Guidercare
2. Hardware manufacturers	
201	Netown Corporation
202	Far EasTone Telecommunications Technical Department
203	Huede Technology
204	Guidercare
205	Acomotech
3. Total solution providers	
301	Netown Corporation
302	Far EasTone Telecommunications
303	Huede
304	Guidercare
305	Acomotech
4. Connected health (CH) service providers	
401	Luo Dong Care Institute
402	En Chu Kong Hospital
403	Taoyuan Fu Hsing Township Health Station
404	Taipei Medical University, Telehealth & Telecare Center
405	Mennonite Christian Hospital
406	Taiwan University Hospital, Telehealth Center
407	Taiwan University Hospital
5. Network providers	
501	Far EasTone Telecommunications
6. End users	
601	National Taiwan University Hospital
602	Mennonite Christian Hospital
603	Taipei Medical University, Telehealth & Telecare Center
604	En Chu Kong Hospital
605	Luo Dong Care Institute
7. Government sectors	
701	Health & Welfare Department
702	Sang Chung Health Center
703	Luo Dong Care Institute
704	Taoyuan Fu Hsing Township Health Station
8. Academia	
801	National Taiwan University, Engineering Department
802	National Taipei University of Technology
803	Taipei University
804	UL (Underwriters Laboratories) Life & Health
805	National Taiwan University, Medicine Department

Table 6. Participants' perspectives regarding how connected health (CH) functions as a platform of remote monitoring and self-management.

Participants	Summary	Quotes
1 and 7	Most interviewees indicated that remote service is a promising solution for working adults who have to balance multiple priorities in limited time and care for aging parents or dependents. In a CH ecosystem, their parents might receive services remotely, with smart and precise features tailored to their specific course of treatment. Some respondents indicated that they were willing to receive smart services if these services were available and accessible in their areas. Additionally, smart, precise, alternative health care solutions effectively reduced the restraints placed on care through the time and cost of travelling, particularly for patients who were located outside of major metropolitan areas.	• "It will be beneficial if remote services are available to look after both parents and children at home. At the same time, sending older parents to a care institution is not our culture after all." [Participant #1] • "Remote solutions offer residents in rural areas efficient and cost-effective options to receive health care services." [Participant #7]
37, 4, 15, and 20	Many interviewees agreed that, apart from alternative solutions, remote services add value to health care centers by ensuring that stakeholders and players in the CH ecosystem remain informed and connected. These services not only enable residents in less-populated areas to receive health care services, they ameliorate patients' concerns about being isolated from essential health care services and needs.	• "Our customers are connected with hospital care, thanks to the remote monitoring." [Participant #37] • "Remote services make the scalability of health care possible." [Participant #4] • "A professional exchange of opinions is enabled through remote service." [Participant #15] • "Our remote device and service enable patients to be monitored at home after their surgery...Both patients and the hospital can benefit from this facility." [Participant #20]
43, 4, and 37	Although some industrial stakeholders believed that health care should be localized, some policy makers believed that Taiwan could attract international visitors by promoting medical tourism, which would allow a future scaling-up of health care businesses as these individuals pay for and receive comprehensive services. Whether the Taiwanese health care model and service can be expanded overseas is still debatable; however, what is certain is that the results suggest that remote services are particularly helpful for overseas businesspeople since these services offer seamless, timely, health-monitoring services. Participants made comments regarding medical tourism in the Taiwanese context (see quotes in next column).	• "I think medical services should remain localized as they are developed according to local needs and culture." [Participant #43] • "Taiwan has excellent health and medical products and services; these advantages should be attractive to international tourists...It is a good opportunity for us to develop medical tourism." [Participant #4] • "Many of our customers are Taiwanese businessmen who work in mainland China...we have expanded the telehealth service from the island of Taiwan to include mainland China..." [Participant #37]
10, 14, and 13	Whether a one-size-fits-all approach to remote services is feasible is debatable. Some conventional health care providers are conservative in their implementation of remote practices. Concerns expressed included the costs of initiating investment in infrastructure and ensuring a return on those investments. Health professionals continued to disagree as to whether remote services can actually translate to more effective care or can replace personal provider-patient interactions. Although these are future trends in health care, some doctors still held more conservative views.	• "Remote service is an ideal option; however, due to our business scale and direction, we are not yet offering this service." [Participant #10] • "In the short term, remote service can be a sponsored pilot program; however, a significant amount of time is needed to receive its return on investment." [Participant #14] • "After a decade of practicing medical service, I personally believe health care should be implemented face-to face..." [Participant #13]
7	The stakeholders that we interviewed suggested that scarce CH resources and funding should be directed to effective health care services in more remote areas, because many urban areas already offer effective health care services (see quote in the next column). Additionally, government support and incentives, such as awarding funding based on CH performance measures, may be crucial to motivate or reward health care professionals and stakeholders.	• "CH is essential in remote areas. It provides residents with access to health care services."
32	Our interviewees expressed the importance of increasingly rapid innovation within the CH business model so that services can maximize the utility of current developments in CH-related technology, including remote sensing and rapid central analysis capabilities. The doctors expressing this view were primarily general practitioners with a business model and related funding, and their work mostly centered on patients who visited the general practitioners at a specific location for care (see quote in next column by a doctor working in a rural area).	• "Although we seem to lack resources compared to urban areas, as long as we can offer unique value propositions, we can obtain essential resources from companies with CH."

Participants	Summary	Quotes
13	In some of these cases, stakeholders had cultural difficulties with CH—they argued that human contact between doctor and patient is irreplaceable and essential to health care. However, many of them did concede that contact may be flexible. These stakeholders assumed that CH is a business fad, rather than a necessary practice for patients. However, these views did not reflect the difficulties experienced in providing efficient and effective health care in remote regions.	• "I have been doing medical services for a decade, and I support personal contact in health care; for me, connected health is just one of the formats of a business fad..."

Several industry interviewees seemed more concerned about the development of technology and its potential applications for CH than end users, based on their perceptions and their need for therapeutic care. This suggests more effective stakeholder interaction and business model innovation are required to coordinate end users' needs and industrial provisions. Additionally, all industry interviewees were concerned about the sustainability of CH businesses beyond grant funding, and therefore emphasized the need to align technology development with a sustainable business model. These interviewees tended to use their connections and relationships to promote nascent supply chains for CH businesses that required more systematic development. In contrast, many academic interviewees considered CH as an emergent topic of innovation and cited a need for more interdisciplinary research and education beyond technological developments.

Discussion

Principal Findings

CH as a coexistent ecosystem includes all the stakeholders who work together to facilitate its implementation. Patients' security and privacy are crucial concerns in the implementation process [36]. However, our interviewees indicated that their main priority is to test the feasibility of remote health care services, as all such services are required to follow current regulations concerning patients' safety. This does not suggest that security and privacy issues can be compromised in exchange for ease of accessibility for health professionals and industry players [37]. On the contrary, following such standards is the precondition of practicing CH, even if some stakeholders think that there is no need to overemphasize the importance of following these standards. Below, we consider a few important issues and obstacles to implementing CH.

Security Concerns

When asked about patients' security concerns, some health professionals emphasized CH's potential convenience and high quality of care. They often asserted that they believe that CH might be like online banking: offering services anytime, anywhere. In this study, governments bear the most responsibility for controlling the risks of using CH and for setting standards that will protect patients [38]. Although successful implementation of CH requires the cooperation of all stakeholders, only those who have power and impacts in the ecosystem can make final decisions regarding CH's implementation. This explains why IT security professionals may not have the power to make decisions in the ecosystem, even though they play important roles.

Opinions regarding privacy and security differ. Some health professionals believe that patients' and institutions' security should be prioritized above the convenience of CH services, but others believe that efforts to make CH more feasible and effective should come first. In general, it is difficult to get all stakeholders to share common ground [39]. The power of decision making usually distinguishes the leaders in the ecosystem. This can be seen from the conversation between the government sectors and industry players. The latter expect government to form the standard of protocol, while the former insist on the respect of the free market.

Interdisciplinary Talent

According to our interviewees, health professionals and governments have more power in the process of implementing CH than industry or patients. Industry usually follows government policies, as their products and services often require governments to initiate funding and infrastructure for the realization of these projects [40]. Patients' needs are often interpreted through health professionals. Respondents indicated that interdisciplinary staff is crucial to enable the smooth operation of CH; however, it takes time to train and develop talents.

Connected Health as a Platform to Facilitate Smart Health Care

Although many efforts have been made to improve the quality of health care, treating some diseases remains challenging because many caregivers and patients lack the information they need for effective treatment. Although CH provides a wide range of care delivery models, it can also be used simply to connect patients with the information they need for effective treatment [41]. In this way, CH provides a pathway to smart, precise health care. CH can put power back into the hands of patients [42], and the data collected from CH platforms can enable the provision of smart health care [43]. Through big data and artificial intelligence technologies, CH offers a platform that enables smart homes, smart cities, and patient-centered, personalized health interventions [44,45].

Connected Health's Innovative Strategies Based on Value Propositions

Scholars at Harvard Business School have discussed three innovation strategies that reflect how timely, smart, and targeted health care can be met through a CH platform. These strategies are the *infuse and augment* strategy, the *combine and transcend* strategy, and the *counteract and reaffirm* strategy [19].

The *infuse and augment* strategy involves the development of new services and products under the current service structure,

without significant change to the attributes and functions of that structure. This strategy increases and emphasizes elements that deliver services that meet the needs and desires that are presently unfulfilled by major markets. A case in point is the Changhua Christian Hospital's (CCH) eHealth and Diabetes Health Management Center, which was created in 2013 to meet the needs of an aging population with an increasing level of chronic health conditions. CCH had been a symbol of public welfare and social care for over a century, and many observers thought that the most plausible reaction to an aging society would have been to reduce the cost of care. However, the CCH believed that reducing the cost of care would have jeopardized the quality of care. Instead, CCH engaged with global institutions from developed countries, from which the trend of *aging in place* emerged:

> *For over a century, CCH has been playing the role of health care hub in the mid-south of Taiwan; now we are expanding our services and business to the global context.* [Participant #55]

Furthering this argument from the patient perspective, individuals prefer to remain in a familiar location, and may prefer to manage their own health rather than increase the number of medications they take. These insights explain why the CCH opted to create the aforementioned center instead of reducing the cost of care. Such a center offers more comprehensive and cost-effective services than conventional care. Creating the center allowed the CCH to avoid an across-the-board price reduction, in contrast to many hospitals that have responded to aging populations by reducing costs. In this case, the CCH viewed global market trends as opportunities for innovation and renewal:

> *We are proactive in absorbing new practices and exploring global trends...We discovered that our customers tend to follow the concept of "aging in place," which echoes the global trend.* [Participant #55]

> *The head of the hospital is really supportive regarding both finances and resources in initiating this center so that we can offer a comprehensive service to our customers.* [Participant #55]

Taipei Medical University Hospital's (TMUH) answer to an aging population's call for seamless services represents another example of the *infuse and augment* strategy. One of the top hospitals in Taipei, TMUH introduced its telehealth-care service program in 2007, which offers consultation and over-the-phone care for patients outside standard working hours. This was achieved by integrating resources and increasing connected services. For example, customers who join the telehealth-care program are able to consult doctors online and book an appointment in advance. Connected devices, meal delivery, and laundry services are also available to ensure seamless and complete care. By augmenting its traditional health care services with these innovations, TMUH has thereby infused its value proposition with an online service:

> *Our clients can obtain access to health care information anytime they need to, even during out-of-office time.* [Participant #34]

> *We hope that our customers can use health care services in the future as if they are using online banking: 24 hours, seamless services.* [Participant #34]

The *combine and transcend* strategy offers a different, relatively radical approach to implementing CH. This strategy involves combining features of a service's existing value propositions with quality-transferring changes arising from a trend. In this way, a novel experience is created, one which may lead the institution into a new market space. At first glance, employing resources to incorporate elements of a different domain into one's core offerings sounds like it is not worthwhile. However, TaiDoc's shift to integrate medical sensing solutions into its reputation for its integrated circuit design demonstrates how the *combine and transcend* strategy can be successful. In 2002, TaiDoc joined with hospitals to develop biosensing devices that can monitor patients' health conditions remotely. By combining TaiDoc's original value propositions for electronic supplies with those of health providers, the company entered a new field of engagement with the health sector. Within a decade, TaiDoc became the top manufacturer of blood glucose medical devices and now accounts for a significant share of the global market.

When the sales manager was asked why they are determined to jump into the business of medical devices, the response was as follows:

> *We are determined to expand our business to medical devices as our founder encountered an emerging need from his family.* [Participant #31]

> *As long as we saw the needs corresponding to the trend, we took the leap...although it was challenging, we enjoyed the fruits of success within 10 years.* [Participant #45]

Another example of the *combine and transcend* strategy is the case of Far EasTone Telecommunications. After seeing the potential of a smart health care industry, Far EasTone Telecommunications combined telecommunication services with medical devices to enable telehealth-care services. In order to make this *transcendence* of existing capabilities possible, they encouraged engineers to engage in work-related training and introduced their test models in their partnered hospitals. Far EasTone Telecommunications has now expanded their businesses to over 80 hospitals and institutions within the past 5 years and has become a leading telehealth services provider in Taiwan.

> *Our manager has an insightful vision of the market; she believes that telehealth-care will become a leading industry in the future.* [Participant #18]

> *Our boss considers a telehealth-care service to be a long-term strategy. He allows exploring of possibilities for a period of time and believes that success will come soon.* [Participant #18]

The *counteract and reaffirm* strategy entails understanding disagreements with the value statements of current services or products and reclaiming a new value proposition that is considered to be superior to existing values. An example of such a service is the Nanwui Foundation, a health and social

care service established by a local physician from Tai-tong Health Center. Like other health facilities, they offer general health care services and health checks. However, the Nanwui Foundation also incorporates the fundamental needs of remote areas into its service manifesto; it holds educational workshops and has a green farm community, which allows residents to explore their values and elaborate on their talents. The case of the Nanwui Foundation supports the assertion that health care is not only about people's physical health care needs, but also their need for dignity. Our interviewees reflected that patients are happier and healthier when they have a goal to pursue and a stage to play on:

> *When I saw these patients, they appeared old, poor, and weak, but when they participated in the social workshops that I organized, they became energetic and healthy.* [Participant #57]
>
> *Physical conditions are only a partial reflection of their health conditions...offering them a stage on which to perform that allows them to exert that their value may be more relevant to their well-being.* [Participant #57]
>
> *In remote areas, residents are poor and lack medical resources. When there is no Superman, we can be our own Superman.* [Participant #57]

Connected Health's Innovative Strategies

Although CH seems to provide promising solutions for the problems posed by an aging population with many chronic health conditions, these solutions are not easy or quick—innovative implementation strategies are required to enable its practices [46]. This is especially true because the players in the CH ecosystem compete for the same organizational resources and stakeholders. Therefore, each actor in the ecosystem needs to adopt different strategies to help them differentiate their value propositions and obtain useful resources [47]. Some interviewees indicated that many CH players failed due to lack of strategic planning. Therefore, this study adapted the strategies proposed by scholars to demonstrate how organizations use strategies to succeed with CH. These experiences may inspire players in various industries.

Limitations and Future Work

This study explored how timely, smart, and targeted health care can be achieved through CH platforms. This means that some relevant content, such as a detailed stakeholder analysis, comprehensive consideration of security and privacy concerns, and detailed assessments of business models, was not described in a large amount of detail. These topics can be extended in future work to offer an inclusive picture of CH implementation. Apart from that, this study is based on the Taiwanese context and different cultural perspectives can be introduced and included in future studies.

Conclusions

This study makes a relevant contribution to the literature by exploring perspectives on how trend-associated changes in customers' opinions and behaviors influences health care and social care. Once these trends have been identified on a broad scale, appropriate innovation and intervention strategies can be developed and implemented. When innovation and entrepreneurship are correctly combined, CH businesses can be propelled and spectacular results can be expected. The *infuse and augment* strategy will permit CH businesses to emerge in successful and socially beneficial ways if the essential values of these businesses remain meaningful to consumers who engage with the service trends. If the results imply that the value statements are incongruent with consumers' new expectations, it suggests that the current innovation strategy is insufficient and needs to be transcended or combined with another strategy to create a new value statement. If the current solutions are associated with negative impacts or conflict with the mainstream system, counteracting these aspects by reaffirming the core values of your business is an ideal strategy. The findings indicate that strategic planning could be beneficial to leverage resources for the implementation of CH. Moreover, continually revisiting business models and strategies is essential because of the rapid, ongoing technological changes across CH stakeholder groupings. In short, we believe that timely, smart, and targeted health care can be created through a CH platform.

Acknowledgments

This research was partially supported by the Vice Chancellor Research Studentship from Ulster University, and SC Limited Company. The APC was funded by Huaqiao University High-level Talent Research Project and The Straits Postdoctoral Fellowship Funding Program in 2018, Fujing Financial office no [2017] 32/605 50X18137. All authors made equal contributions. The authors would like to thank the supervisors, Professor Rodney McAdam, Dr Michael Brennan, and Dr Shirley Davey from Ulster University, for sharing their pearls of wisdom during this research. The authors are also immensely grateful to Mr Ray Palmer and David Liu for their comments on an earlier version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Word frequency cloud of the interview data.
[PNG File , 585 KB - jmir_v22i4e14201_app1.png]

References

1. Barr PJ, McElnay JC, Hughes CM. Connected health care: The future of health care and the role of the pharmacist. J Eval Clin Pract 2012 Feb;18(1):56-62. [doi: 10.1111/j.1365-2753.2010.01522.x] [Medline: 20698917]
2. 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]
3. Della Mea V. What is e-health (2): The death of telemedicine? J Med Internet Res 2001;3(2):E22 [FREE Full text] [doi: 10.2196/jmir.3.2.e22] [Medline: 11720964]
4. Chouvarda IG, Goulis DG, Lambrinoudaki I, Maglaveras N. Connected health and integrated care: Toward new models for chronic disease management. Maturitas 2015 Sep;82(1):22-27. [doi: 10.1016/j.maturitas.2015.03.015] [Medline: 25891502]
5. Carroll N. Key success factors for smart and connected health software solutions. Computer 2016 Nov;49(11):22-28. [doi: 10.1109/MC.2016.340]
6. O'Leary P, Carroll N, Clarke P, Richardson I. Untangling the complexity of connected health evaluations. In: Proceedings of the 2015 International Conference on Healthcare Informatics. New York, NY: IEEE; 2015 Oct 21 Presented at: 2015 International Conference on Healthcare Informatics; October 21-23, 2015; Dallas, TX p. 272-281 URL: https://ieeexplore.ieee.org/abstract/document/7349700 [doi: 10.1109/ichi.2015.39]
7. Kellermann AL, Jones SS. What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Aff (Millwood) 2013 Jan;32(1):63-68. [doi: 10.1377/hlthaff.2012.0693] [Medline: 23297272]
8. Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B, et al. Care coordination/home telehealth: The systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008 Dec;14(10):1118-1126. [doi: 10.1089/tmj.2008.0021] [Medline: 19119835]
9. Rheuban K, Shanahan C, Willson K. Telemedicine: Innovation has outpaced policy. Virtual Mentor 2014 Dec 01;16(12):1002-1009 [FREE Full text] [doi: 10.1001/virtualmentor.2014.16.12.pfor1-1412] [Medline: 25493370]
10. Xu L, Wu F. Cryptanalysis and improvement of a user authentication scheme preserving uniqueness and anonymity for connected health care. J Med Syst 2015 Feb;39(2):10. [doi: 10.1007/s10916-014-0179-x] [Medline: 25631840]
11. Galbraith B, Mulvenna M, McAdam R, Martin S. Open innovation in connected health: An empirical study and research agenda. In: Proceedings of the XIX International Society for Professional Innovation Management (ISPIM) Conference. Manchester, UK: International Society for Professional Innovation Management; 2008 Jun 15 Presented at: XIX International Society for Professional Innovation Management (ISPIM) Conference; June 15-18, 2008; Tours, France URL: https://www.researchgate.net/profile/Maurice_Mulvenna/publication/228373568_Open_innovation_in_Connected_Health_an_empirical_study_and_research_agenda/links/542307530cf26120b7a6b92d/Open-innovation-in-Connected-Health-an-empirical-study-and-research-agenda.pdf
12. Das AK, Goswami A. A secure and efficient uniqueness-and-anonymity-preserving remote user authentication scheme for connected health care. J Med Syst 2013 Jun;37(3):9948. [doi: 10.1007/s10916-013-9948-1] [Medline: 23660745]
13. Fernandez-Luque L, Vilmarlund V, Borycki E, Schulz S, Kuziemsky C, Marschollek M, et al. Social media as catalyzer for connected health: Hype or hope? Perspectives from IMIA working groups. Stud Health Technol Inform 2016;225:602-604. [Medline: 27332272]
14. Caulfield BM, Donnelly SC. What is connected health and why will it change your practice? QJM 2013 Aug;106(8):703-707. [doi: 10.1093/qjmed/hct114] [Medline: 23676416]
15. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232]
16. Brandenburger A, Stuart J. Value‐based business strategy. J Econ Manag Strategy 1996;5(1):5-24 [FREE Full text] [doi: 10.1111/j.1430-9134.1996.00005.x]
17. Yu C. Building a value-centric e-government service framework based on a business model perspective. Int Conf Electron Gov 2008;5184:160-171. [doi: 10.1007/978-3-540-85204-9_14]
18. Ehrenhöfer C, Kreuzer E. The role of business model design in the service engineering process: A comparative case study in the field of cloud computing to join service engineering with business model design. In: Proceedings of the 2012 Annual Service Research and Innovation Institute (SRII) Global Conference. New York, NY: IEEE; 2012 Jul 24 Presented at: 2012 Annual Service Research and Innovation Institute (SRII) Global Conference; July 24-27, 2012; San Jose, CA p. 283-292. [doi: 10.1109/srii.2012.39]
19. Ofek E, Wathieu L. Are you ignoring trends that could shake up your business? In: Harvard Business Review on Rebuilding Your Business Model. Boston, MA: Harvard Business School Publishing Corporation; 2011:87-106.
20. Breuer H, Lüdeke-Freund F. Values-Based Innovation Management: Innovating by What We Care About. London, UK: Palgrave; 2017.
21. Weinstein RS, Lopez AM. Health literacy and connected health. Health Aff (Millwood) 2014 Jun;33(6):1103-1104. [doi: 10.1377/hlthaff.2014.0396] [Medline: 24889963]
22. Chesbrough H. Business model innovation: It's not just about technology anymore. Strategy Leadersh 2007 Nov 13;35(6):12-17. [doi: 10.1108/10878570710833714]

23. Barclay L. In sickness and in dignity: A philosophical account of the meaning of dignity in health care. Int J Nurs Stud 2016 Sep;61:136-141. [doi: 10.1016/j.ijnurstu.2016.06.010] [Medline: 27351830]
24. Agboola SO, Ball M, Kvedar JC, Jethwani K. The future of connected health in preventive medicine. QJM 2013 Sep;106(9):791-794. [doi: 10.1093/qjmed/hct088] [Medline: 23598385]
25. Iglehart JK. Connected health: Emerging disruptive technologies. Health Aff (Millwood) 2014 Feb;33(2):190. [doi: 10.1377/hlthaff.2014.0042] [Medline: 24493758]
26. Friedman CP, Rubin JC, Sullivan KJ. Toward an information infrastructure for global health improvement. Yearb Med Inform 2017 Aug;26(1):16-23 [FREE Full text] [doi: 10.15265/IY-2017-004] [Medline: 28480469]
27. Rogers WA, Mitzner TL. Envisioning the future for older adults: Autonomy, health, well-being, and social connectedness with technology support. Futures 2017 Mar;87:133-139 [FREE Full text] [doi: 10.1016/j.futures.2016.07.002] [Medline: 28458395]
28. Zahran Z, Tauber M, Watson HH, Coghlan P, White S, Procter S, et al. Systematic review: What interventions improve dignity for older patients in hospital? J Clin Nurs 2016 Feb;25(3-4):311-321. [doi: 10.1111/jocn.13052] [Medline: 26818360]
29. Yin RK. Applications of Case Study Research. 3rd edition. Thousand Oaks, CA: SAGE Publications; 2012.
30. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: Qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
31. Carson D, Gilmore A, Perry C, Gronhaug K. Qualitative Marketing Research. London, UK: SAGE Publications; 2001.
32. Pattison PE, Robins GL, Snijders TA, Wang P. Conditional estimation of exponential random graph models from snowball sampling designs. J Math Psychol 2013 Dec;57(6):284-296. [doi: 10.1016/j.jmp.2013.05.004]
33. Rourke L, Anderson T. Validity in quantitative content analysis. Educ Technol Res Dev 2004 Mar;52(1):5-18. [doi: 10.1007/BF02504769]
34. Chen SC. Technological health intervention in population aging to assist people to work smarter not harder: Qualitative study. J Med Internet Res 2018 Jan 04;20(1):e3 [FREE Full text] [doi: 10.2196/jmir.8977] [Medline: 29301736]
35. Carroll N, Travers M, Richardson I. Evaluating multiple perspectives of a connected health ecosystem. In: Proceedings of the 9th International Conference on Health Informatics (HEALTHINF). 2016 Feb 21 Presented at: 9th International Conference on Health Informatics (HEALTHINF); February 21-23, 2016; Rome, Italy URL: http://hdl.handle.net/10344/5442 [doi: 10.5220/0005623300170027]
36. Hale T, Kvedar J. Privacy and security concerns in telehealth. Virtual Mentor 2014 Dec 01;16(12):981-985 [FREE Full text] [doi: 10.1001/virtualmentor.2014.16.12.jdsc1-1412] [Medline: 25493367]
37. Hall JL, McGraw D. For telehealth to succeed, privacy and security risks must be identified and addressed. Health Aff (Millwood) 2014 Feb;33(2):216-221. [doi: 10.1377/hlthaff.2013.0997] [Medline: 24493763]
38. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
39. Hoffman A, Montgomery R, Aubry W, Tunis SR. How best to engage patients, doctors, and other stakeholders in designing comparative effectiveness studies. Health Aff (Millwood) 2010 Oct;29(10):1834-1841. [doi: 10.1377/hlthaff.2010.0675] [Medline: 20921483]
40. Steele R, Min K, Lo A. Personal health record architectures: Technology infrastructure implications and dependencies. Acta Anaesthesiol Scand 2012 Mar 05;63(6):1079-1091. [doi: 10.1002/asi.22635]
41. Jethwani K, Ling E, Mohammed M, Myint-U K, Pelletier A, Kvedar JC. Diabetes connect: An evaluation of patient adoption and engagement in a Web-based remote glucose monitoring program. J Diabetes Sci Technol 2012 Nov 01;6(6):1328-1336 [FREE Full text] [doi: 10.1177/193229681200600611] [Medline: 23294777]
42. Chu LF, Shah AG, Rouholiman D, Riggare S, Gamble JG. Patient-centric strategies in digital health. In: Rivas H, Wac K, editors. Digital Health: Scaling Healthcare to the World. Cham, Switzerland: Springer; 2018:43-54.
43. Pramanik MI, Lau RY, Demirkan H, Azad MA. Smart health: Big data-enabled health paradigm within smart cities. Expert Syst Appl 2017 Nov;87:370-383. [doi: 10.1016/j.eswa.2017.06.027]
44. Muhammad G, Rahman SM, Alelaiwi A, Alamri A. Smart health solution integrating IoT and cloud: A case study of voice pathology monitoring. IEEE Commun Mag 2017 Jan;55(1):69-73. [doi: 10.1109/mcom.2017.1600425cm]
45. Kvedar J, Coye MJ, Everett W. Connected health: A review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
46. Slama D, Puhlmann F, Morrish J, Bhatnagar R. Enterprise IoT: Strategies and Best Practices for Connected Products and Services. Sebastopol, CA: O'Reilly Media; 2015.
47. Vest JR, Kash BA. Differing strategies to meet information-sharing needs: Publicly supported community health information exchanges versus health systems' enterprise health information exchanges. Milbank Q 2016 Mar;94(1):77-108 [FREE Full text] [doi: 10.1111/1468-0009.12180] [Medline: 26994710]

Abbreviations

AMA: American Medical Association
BMI: Business Model Innovation
CCH: Changhua Christian Hospital
CH: connected health
ICT: information and communication technology
IT: information technology
TMUH: Taipei Medical University Hospital

Edited by G Eysenbach; submitted 29.03.19; peer-reviewed by P Yeng, M Afzal; comments to author 18.09.19; revised version received 07.11.19; accepted 23.03.20; published 28.04.20.

Please cite as:
Chen SCI, Hu R, McAdam R
Smart, Remote, and Targeted Health Care Facilitation Through Connected Health: Qualitative Study
J Med Internet Res 2020;22(4):e14201
URL: http://www.jmir.org/2020/4/e14201/
doi:10.2196/14201
PMID:32343254

Original Paper

Social Media Coverage of Scientific Articles Immediately After Publication Predicts Subsequent Citations - #SoME_Impact Score: Observational Analysis

Niranjan Jude Sathianathen[1], MBBS; Robert Lane III[1], MS; Declan G Murphy[2], MBBCHBaO; Stacy Loeb[3,4], MD; Caitlin Bakker[5], MISt; Alastair D Lamb[6,7], MBChB; Christopher J Weight[1], MD

[1]Department of Urology, University of Minnesota, Minneapolis, MN, United States

[2]Department of Surgical Oncology, Peter MacCallum Cancer Centre, Melbourne, Australia

[3]Department of Urology, New York University, New York, NY, United States

[4]Department of Population Health, New York University, New York, NY, United States

[5]Health Science Libraries, University of Minnnesota, Minneapolis, MN, United States

[6]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom

[7]Department of Urology, Churchill Hospital Cancer Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

Corresponding Author:
Niranjan Jude Sathianathen, MBBS
Department of Urology
University of Minnesota
420 Delaware St SE
Minneapolis, MN, 55455
United States
Phone: 1 6603340774
Email: niranjan19@gmail.com

Abstract

Background: Social media coverage is increasingly used to spread the message of scientific publications. Traditionally, the scientific impact of an article is measured by the number of citations. At a journal level, this conventionally matures over a 2-year period, and it is challenging to gauge impact around the time of publication.

Objective: We, therefore, aimed to assess whether Web-based attention is associated with citations and to develop a predictive model that assigns relative importance to different elements of social media coverage: the #SoME_Impact score.

Methods: We included all original articles published in 2015 in a selection of the highest impact journals: The New England Journal of Medicine, The Lancet, the Journal of the American Medical Association, Nature, Cell, and Science. We first characterized the change in Altmetric score over time by taking a single month's sample of recently published articles from the same journals and gathered Altmetric data daily from the time of publication to create a mixed effects spline model. We then obtained the overall weighted Altmetric score for all articles from 2015, the unweighted data for each Altmetric component, and the 2-year citation count from Scopus for each of these articles from 2016 to 2017. We created a stepwise multivariable linear regression model to develop a #SoME_Score that was predictive of 2-year citations. The score was validated using a dataset of articles from the same journals published in 2016.

Results: In our unselected sample of 145 recently published articles, social media coverage appeared to plateau approximately 14 days after publication. A total of 3150 articles with a median citation count of 16 (IQR 5-33) and Altmetric score of 72 (IQR 28-169) were included for analysis. On multivariable regression, compared with articles in the lowest quantile of #SoME_Score, articles in the second, third, and upper quantiles had 0.81, 15.20, and 87.67 more citations, respectively. On the validation dataset, #SoME_Score model outperformed the Altmetric score (adjusted R^2 0.19 vs 0.09; P<.001). Articles in the upper quantile of #SoME_Score were more than 5 times more likely to be among the upper quantile of those cites (odds ratio 5.61, 95% CI 4.70-6.73).

Conclusions: Social media attention predicts citations and could be used as an early surrogate measure of scientific impact. Owing to the cross-sectional study design, we cannot determine whether correlation relates to causation.

(J Med Internet Res 2020;22(4):e12288) doi:10.2196/12288

KEYWORDS

bibliometrics; online social networking; online systems; online intervention; social media

Introduction

A direct reflection of the digital age, scientific work is primarily disseminated Web-based rather than in the conventional hard copy format [1]. The academic community has embraced the internet as a medium for discussion and debate with the increasing emergence of social media–facilitated journal clubs [2]. Although these facets of academia have evolved with the times and technology, measures of scientific impact still generally rely on the traditional citation counts and lag behind. Furthermore, there are important limitations to using citation count in this manner [3], primarily the time required for citations to mature, and therefore, impactful articles can only be classified as so retrospectively when they may already be obsolete. It is increasingly clear that the digital footprint harbored by each item of Web-based information contains a wealth of data that can be used to make inferences about it. Altmetric is a platform that captures the Web-based attention received by academic articles from several sources, including news, blogs, Twitter, Facebook, Sina Weibo, Wikipedia, policy documents, Q&A, F1000/Publons/PubPeer, YouTube, Reddit/Pinterest, LinkedIn, Open Syllabus, Google+, and Patents. It includes a wealth of information that could be used to circumvent the time delay to formal citations and thereby provide an earlier measure of scientific impact. This has substantial implications on a communal and individual level. These Web-based attention metrics can be used to potentially identify scientific works that are *game changers* prospectively, which would increase awareness of these works and potentially lead to earlier implementation of the recommendations arising from their results. Similarly, individuals will be able to receive credit for their work years earlier, which may form the foundation for further work and success because they could conceivably leverage this early indicator of scientific merit to endorse their applications for sponsorship or support [4].

Initial work in this area has demonstrated that Web-based metrics are associated with scholarly impact. Eysenbach [5] reported that the top-tweeted articles from the Journal of Medical Internet Research (highly tweeted after 7 days compared to other articles in the same issue) are also more likely to be the most highly cited articles from that journal. This relationship has also been confirmed in ecological journals [6]. Thelwall et al [7] also demonstrated a moderate correlation between Twitter, Facebook, research highlights, blogs, mainstream media, forums, Q&A, and citations in a heterogeneous sample of articles. However, there was no overall weighted score at the time of research because the Altmetric system was still undergoing development and has evolved considerably since then, with different components now being included. In addition, these works have mostly focused on articles from a single journal, and therefore, the applicability of these findings to other journals is unclear. It should also be noted that Web-based academic presence has grown tremendously since 2013, and thus, these results may not be applicable today. Nonetheless, there is a growing body of evidence that Web-based attention metrics are associated with scholarly impact and urge the scientific community to recognize this and attribute credit accordingly. This is further reflected by the increasing Web-based presence of journals that further potentiate the dissemination of research [8].

We aimed to characterize the dynamics in Altmetric scores over time immediately following publication and assess whether there is an association between Altmetric score or any of its components and citation count for articles published in high-impact scientific and clinical journals. We further intended to develop a new index that can be used to predict scientific impact soon after publication.

Methods

Sample

All articles published in Cell, Nature, Science, the Journal of the American Medical Association, the New England Journal of Medicine, and The Lancet during the 2015 calendar year were included for analysis. We used articles from the 2016 calendar year from the same journals as a test set. The latter 3 journals were considered to be clinically oriented journals, whereas the former 3 were classified as being scientifically oriented. We selected these journals because they represent high-impact publications in the clinical and scientific fields. Information on each of these publications was obtained from Scopus by matching digital object identifiers (DOIs), including the type of document and yearly cumulative citation counts. Only publications classified as *Articles* or *Articles in Press* in Scopus were included in the analysis to only assess original articles; that is, publications classified as *Editorial*, *Conference Paper*, *Letter*, *News*, *Note*, *Review*, and *Short Survey* were excluded because they were not believed to be peer-reviewed, novel scientific contributions. We also excluded articles with missing classifications. We obtained Altmetric data for each of the included publications by matching DOIs. This included the automatically calculated, weighted Altmetric score generated by Altmetric, which is an approximation of the attention a particular research output has received based on the raw number of news, blogs, Twitter, Facebook, Sina Weibo, Wikipedia, policy documents, Q&A, F1000/Publons/PubPeer, YouTube, Reddit/Pinterest, LinkedIn, Open Syllabus, Google+, and Patents. We also obtained the individual weightings for each of the elements in calculating the Altmetric score that are available Web-based. We included the impact factor of each journal as a continuous covariable in our model.

Altmetric counts are cumulative and cannot be obtained for retrospective periods, and therefore, the data represent the total Altmetric score at the time of search (October 2017). Therefore, to characterize the change in Altmetric score and establish that the total Altmetric score at the time of search represents the

meaningful Altmetric score early after article publication, we included articles from the same journal that was published within 2 days of the Altmetric search and tracked their Altmetric score daily for the first fortnight and then every 3 days for the next fortnight and then weekly for the last 2 weeks.

Statistical Analysis

Before analysis, normal distribution and homogeneity of variances were assessed. Data that were not normally distributed were described using median and IQR and compared using the Kruskal-Wallis test. The correlation between 2-year citation counts and Altmetric scores—weighted and individual components—was estimated by calculating the Spearman rank correlation coefficient, as the dependent variables were not normally distributed. To develop a #SoME_Score that could be used to predict citations, we performed a multivariable linear model using the forward stepwise regression method based on Akaike information criteria to determine the best fitting model [9]. We adjusted for each component of the Altmetric score in addition to time since publication. We defined outliers as those having a Cook distance 4 times greater than the mean.

To assess the change in Altmetric score from the time of publication, we fit mixed effects spline models. Cubic spline models were chosen because the data were nonlinear, and therefore, the data were allowed to fit separate curves for each section of time. Cubic spline knots were placed at 5-day intervals to characterize the change in score over this time frame.

All *P* values were 2 sided, and the statistical significance set at the .05 level. Data analysis was performed in R (R Foundation for Statistical Computing) version 3.4.

Results

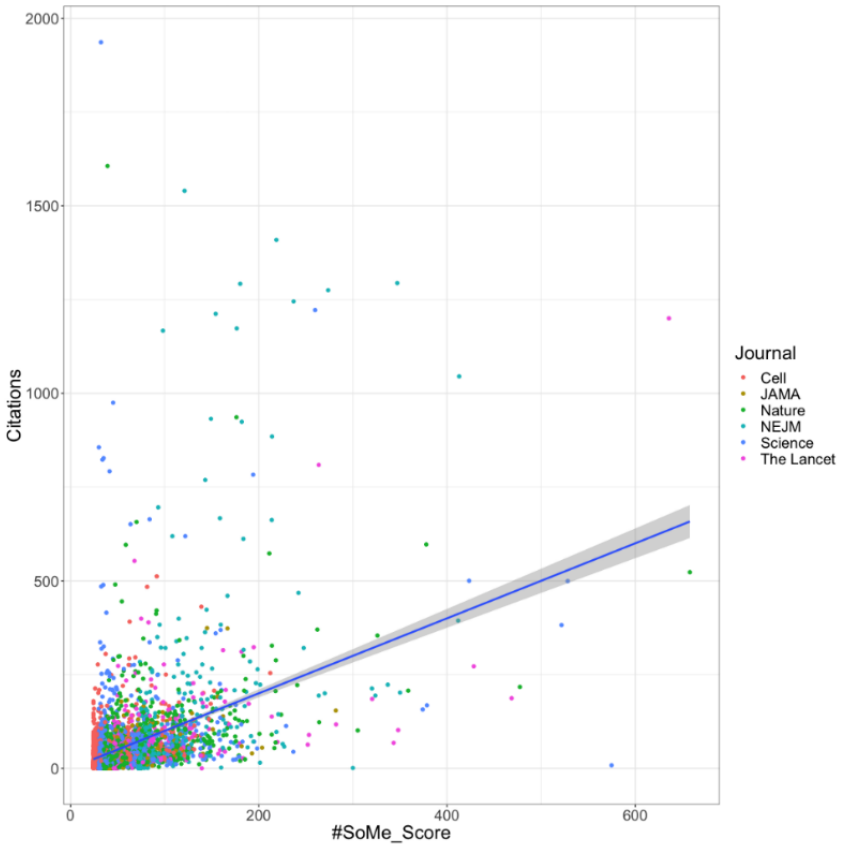
Inclusion Criteria

A total of 3510 articles met the inclusion criteria and were included for analysis. Of these, 15.52% (545/3510), 28.34% (995/3510), 25.25% (883/3510), 8.43% (296/3510), 11.22% (394/3510), and 11.31% (397/3510) were published in Cell, Nature, Science, the Journal of the American Medical Association, the New England Journal of Medicine, and The Lancet, respectively. Included articles had a median citation count of 37 (IQR 12-74). There was a difference in median citations between articles in clinical and scientific journals (32 vs 39; *P*<.001).

The median Altmetric score for all the articles was 72 (IQR 28-169). There was no difference in the Altmetric score between clinical and scientific journals (72 vs 72; *P*=.10). The adjusted R^2 of the Altmetric model in the training set was 0.045. With stepwise regression, only news, blog, policy, Peer Review, Wiki, F1000, and the journal impact factor were included in the model to generate the #SoME_Score (Multimedia Appendix 1). The median #SoME_Score for all articles was 47.86 (IQR 36.73-76.58). On multivariable regression, the #SoME_Score had a significant impact on citation count (Figure 1; *P*<.001). The adjusted R^2 of the #SoME_Score model in the training set was 0.168, which was significantly better than the Altmetric model (*P*<.001). Compared with articles in the lowest quantile of #SoME_Score, articles in the second, third, and upper quantiles had 0.81 (SE 5.5), 15.20 (SE 5.5), and 87.67 (SE 5.5) more citations, respectively. Articles in the upper quantile of #SoME_Score were more than 5 times more likely to be among the upper quantile of those cites (odds ratio [OR] 5.61, 95% CI 4.70-6.73). Conversely, publications in the top quantile of the Altmetric score did not act as a predictor of high citation count (OR 0.98, 95% CI 0.34-2.88).

On the validation dataset, #SoME_Score model outperformed the Altmetric score (adjusted R^2 0.19 vs 0.09; *P*<.001).

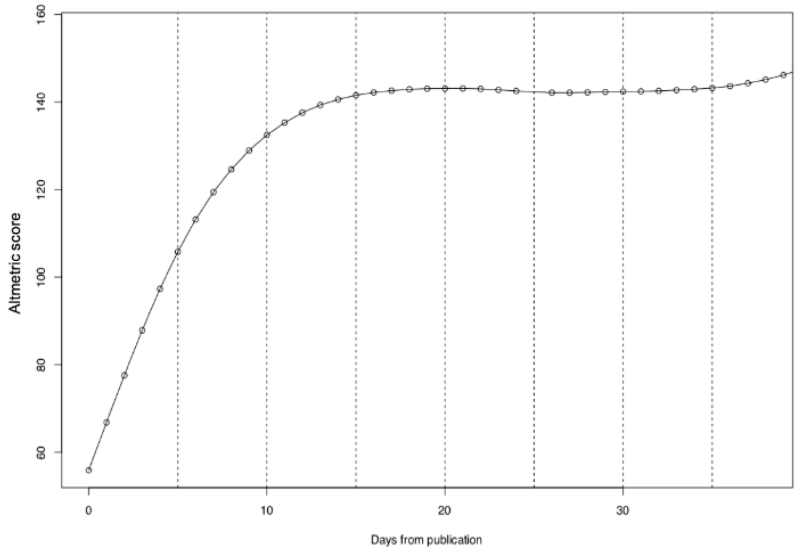
Figure 1. Association between #SoME_Score and citation count. JAMA: Journal of the American Medical Association; NEJM: New England Journal of Medicine.



Altmetric Score Trend

A total of 145 articles published in the same 6 journals between September 18 and September 22, 2017, were included in the analysis for the Altmetric score trend. Data were collected on the cumulative Altmetric score of these articles for a mean 40 days after publication. The increase in the Altmetric score appears to plateau approximately 14 days after publication (Figure 2).

Figure 2. Altmetric score trend.



Journal of Medical Internet Research Validation Results

We evaluated the use of the #SoME_Score in articles published in the Journal of Medical Internet Research in the 2016 calendar year. Of the 420 articles included in this analysis, the median Altmetric score was 8 (IQR 4-19), and the median 2-year citation count was 6 (IQR 3-11). #SoME_Score acted as a significant predictor of citations in this external validation dataset (P<.01). Articles that were among the highest quartile of #SoME_Scores

were nearly 3 times more likely to be among the highest quartile of those cited (OR 2.88, 95% CI 1.71-4.83). In contrast to the primary dataset of articles, articles in the Journal of Medical Internet Research in the top 25% of Altmetric scores were also more likely to be those with a high citation count (OR 3.28, 95% CI 2.02-5.30).

Discussion

Principal Findings

We observed an association between social media attention of scholarly articles and academic impact measured by citation count. Furthermore, we were able to identify the most important components of the Altmetric score that are associated with citations and then developed a new index, the #SoME_Score, which can be calculated in the immediate postpublication period (2-3 weeks) and be used to predict scholarly impact years down the road. We suggest that this score may be best used to predict high-impact papers as defined as those with citations in the highest quartile. This is in line with the findings from the study by Eysenbach [5], who reported that highly tweeted articles (measured by the Twimpact Factor, which is the number of tweets after 7 days) were 11 times more likely to be highly cited than less tweeted articles. However, the Twimpact factor was only generated from articles published in the Journal of Medical Internet Research, which inherently are likely to generate more Web-based attention because of the nature of its subject. Our findings were shown to be robust across other medical and scientific journals, which often report on topics that are not easily discerned by the general public. It should be noted that these findings represent an association only and not causation, that is, high Web-based presence may lead to increased citations or the impactful nature of certain studies may be underpinning the early Web-based attention it receives [5].

It is important to note that not all components of the Altmetric score were correlated with citations, which, therefore, served as the impetus for the development of the #SoME_Score. Only the news, blog, policy, Peer Review, and F1000 components showed a meaningful correlation to citation count. The overall Altmetric score is derived by applying weights to each component, which correlates to its *attention*, that is, its exposure and engagement in the Web-based sphere. This is a clear shortcoming of using the overall Altmetric score to predict academic impact because the weights applied are not a reflection of the contribution of each component to citation count. Aside from the limited sample, this could partly explain previous studies, which did not find a relationship between Altmetric and citations [10]. Furthermore, the overall Altmetric score is not specific for the scientific community as the measure of *attention* used to determine the weights will be largely driven by the general population of which academics only form a small proportion. This is why articles on populous subjects, such as sex and politics, have phenomenally high Altmetric scores but relatively low citation counts (Multimedia Appendices 2 and 3). In contrast, the weights applied to each component in the #SoME_Score is specific because citations are driven by academics. These fundamental benefits of the #SoME_Score suggest that it could be adopted as a predictive index of scholarly impact.

To the best of our knowledge, this is the first study to evaluate the trend in Altmetric scores for scientific articles and show an early plateau within 2 weeks of publication. This highlights both the speed of information dissemination in the internet age and the short window of spotlight a publication receives. Furthermore, the fleeting nature of Web-based attention is partly a reflection of the sheer volume of scientific work entering the literature [11]. The number of articles entering PubMed annually has increased by 62.5% between 2003 and 2013 [12]. We are, therefore, faced with the challenge of filtering signal from the noise [13]. #SoME_Impact Score could help amplify the signal.

Limitations

In terms of limitations, we only included articles from high-impact clinical and scientific journals to improve our power to detect a relationship, given the increased average citations of these publications. Given that these journals are highly selective for impact stories, our findings may not be generalizable to works in other journals, and the #SoME_Impact Score would need further validation to confirm similar predictive ability. However, we have reported on the generalizability of this score to articles published in the Journal of Medical Internet Research. Furthermore, the Web-based presence of biomedical academics is continually evolving with, for example, increasing numbers engaging in Twitter, and therefore, it is possible that components that were shown not to have a correlation with citation count will become significant in the future [14]. Moreover, we only evaluated total and individual components included in the Altmetric score, and further research could explore whether there are alternative metrics or components that can be used to improve predictive ability.

Conclusions

Social media attention can be used as an early surrogate measure of academic impact. This could lead to academics being recognized and receiving early credit for their work instead of having to wait for citation counts to mature. Further work is required to validate these findings in the wider biomedical literature and nonbiomedical fields.

Acknowledgments

SL is supported by The Edward Blank and Sharon Cosloy-Blank Family Foundation and the Prostate Cancer Foundation. CW is funded by the Climb 4 Kidney Cancer Foundation and the National Institutes of Health. CW declares travel and speaker fees from Abbott Molecular.

Conflicts of Interest

SL declares reimbursed travel from Sanofi, consulting fees from Bayer and Lumenis, and equity in Gilead.

Multimedia Appendix 1
Beta coefficients for #SoME_Score.
[DOCX File , 13 KB - jmir_v22i4e12288_app1.docx]

Multimedia Appendix 2
Outliers.
[DOCX File , 17 KB - jmir_v22i4e12288_app2.docx]

Multimedia Appendix 3
Association between #SoMe_Score and citation count with outliers.
[DOCX File , 483 KB - jmir_v22i4e12288_app3.docx]

References

1. Solomon DJ. Digital distribution of academic journals and its impact on scholarly communication: looking back after 20 years. J Acad Libr 2013 Jan;39(1):23-28 [FREE Full text] [doi: 10.1016/j.acalib.2012.10.001]
2. Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D. Globalization of continuing professional development by journal clubs via microblogging: a systematic review. J Med Internet Res 2015 May 23;17(4):e103 [FREE Full text] [doi: 10.2196/jmir.4194] [Medline: 25908092]
3. Vucovich LA, Baker JB, Smith JT. Analyzing the impact of an author's publications. J Med Libr Assoc 2008 Jan;96(1):63-66 [FREE Full text] [doi: 10.3163/1536-5050.96.1.63] [Medline: 18219384]
4. Dinsmore A, Allen L, Dolby K. Alternative perspectives on impact: the potential of ALMs and altmetrics to inform funders about research impact. PLoS Biol 2014 Dec;12(11):e1002003 [FREE Full text] [doi: 10.1371/journal.pbio.1002003] [Medline: 25423184]
5. 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]
6. Peoples BK, Midway SR, Sackett D, Lynch A, Cooney PB. Twitter predicts citation rates of ecological research. PLoS One 2016;11(11):e0166570 [FREE Full text] [doi: 10.1371/journal.pone.0166570] [Medline: 27835703]
7. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do altmetrics work? Twitter and ten other social web services. PLoS One 2013;8(5):e64841 [FREE Full text] [doi: 10.1371/journal.pone.0064841] [Medline: 23724101]
8. Nason GJ, O'Kelly F, Kelly ME, Phelan N, Manecksha RP, Lawrentschuk N, et al. The emerging use of Twitter by urological journals. BJU Int 2015 Mar;115(3):486-490 [FREE Full text] [doi: 10.1111/bju.12840] [Medline: 24925047]
9. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol 2004 Oct;53(5):793-808. [doi: 10.1080/10635150490522304] [Medline: 15545256]
10. Ruan QZ, Chen AD, Cohen JB, Singhal D, Lin SJ, Lee BT. Alternative metrics of scholarly output: the relationship among Altmetric score, Mendeley reader score, citations, and downloads in plastic and reconstructive surgery. Plast Reconstr Surg 2018 Mar;141(3):801-809. [doi: 10.1097/PRS.0000000000004128] [Medline: 29481413]
11. Larsen PO, von Ins M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. Scientometrics 2010 Oct;84(3):575-603 [FREE Full text] [doi: 10.1007/s11192-010-0202-z] [Medline: 20700371]
12. Vardakas KZ, Tsopanakis G, Poulopoulou A, Falagas ME. An analysis of factors contributing to PubMed's growth. J Inform 2015 Jul;9(3):592-617 [FREE Full text] [doi: 10.1016/j.joi.2015.06.001]
13. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Oct 21;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]
14. Mohammadi E, Thelwall M, Kwasny M, Holmes KL. Academic information on Twitter: a user survey. PLoS One 2018;13(5):e0197265 [FREE Full text] [doi: 10.1371/journal.pone.0197265] [Medline: 29771947]

Abbreviations

DOI: digital object identifier
OR: odds ratio

Edited by G Eysenbach; submitted 21.09.18; peer-reviewed by X Guo, J Triemstra, M Roberts; comments to author 03.02.19; revised version received 07.06.19; accepted 12.06.19; published 17.04.20.

Please cite as:
Sathianathen NJ, Lane III R, Murphy DG, Loeb S, Bakker C, Lamb AD, Weight CJ
Social Media Coverage of Scientific Articles Immediately After Publication Predicts Subsequent Citations - #SoME_Impact Score: Observational Analysis
J Med Internet Res 2020;22(4):e12288
URL: https://www.jmir.org/2020/4/e12288
doi:10.2196/12288
PMID:32301733

Original Paper

HoloLens-Based Vascular Localization System: Precision Evaluation Study With a Three-Dimensional Printed Model

Taoran Jiang[1*], PhD, MD; Dewang Yu[2*], MD; Yuqi Wang[2], MD; Tao Zan[1], PhD, MD; Shuyi Wang[2], PhD; Qingfeng Li[1], PhD, MD

[1]Department of Plastic and Reconstructive Surgery, Shanghai 9th People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Qingfeng Li, PhD, MD
Department of Plastic and Reconstructive Surgery
Shanghai 9th People's Hospital, Shanghai Jiao Tong University School of Medicine
No. 639 Zhizaoju Road
Shanghai, 200011
China
Phone: 86 21 64397277
Fax: 86 21 64397277
Email: dr.liqingfeng@yahoo.com

Abstract

Background: Vascular localization is crucial for perforator flap transfer. Augmented reality offers a novel method to seamlessly combine real information with virtual objects created by computed tomographic angiography to help the surgeon "see through" the skin and precisely localize the perforator. The head-mounted display augmented reality system HoloLens (Microsoft) could facilitate augmented reality–based perforator localization for a more convenient and safe procedure.

Objective: The aim of this study was to evaluate the precision of the HoloLens-based vascular localization system, as the most important performance indicator of a new localization system.

Methods: The precision of the HoloLens-based vascular localization system was tested in a simulated operating room under different conditions with a three-dimensional (3D) printed model. The coordinates of five pairs of points on the vascular map that could be easily identified on the 3D printed model and virtual model were detected by a probe, and the distance between the corresponding points was calculated as the navigation error.

Results: The mean errors were determined under different conditions, with a minimum error of 1.35 mm (SD 0.43) and maximum error of 3.18 mm (SD 1.32), which were within the clinically acceptable range. There were no significant differences in the errors obtained under different visual angles, different light intensities, or different states (static or motion). However, the error was larger when tested with light compared with that tested without light.

Conclusions: This precision evaluation demonstrated that the HoloLens system can precisely localize the perforator and potentially help the surgeon accomplish the operation. The authors recommend using HoloLens-based surgical navigation without light.

(J Med Internet Res 2020;22(4):e16852) doi:10.2196/16852

KEYWORDS

augmented reality; HoloLens; perforator flap; vascular localization; reconstructive surgery; 3D Printing

Introduction

High variability in perforator size and course, along with distorted anatomical landmarks after injury are the two key challenges in precisely predicting the location of the perforator in a vascular flap transfer procedure. Precisely locating the perforator can be difficult for new microsurgeons even without these key challenges. Computed tomographic angiography (CTA) has been used as a noninvasive and effective tool for vascular mapping to plan the positioning of perforator flaps

[1,2] by revealing the characteristics of perforators, such as the caliber, length, and course [3]. Using the three-dimensional (3D) objects reconstructed by CTA data, the relation between the perforator and the underlying skeleton and nearby muscle and skin can be immediately identified. This enables the surgeon to identify the specific perforator anatomy of the patient. Accordingly, CTA is increasingly applied as a standard preoperative procedure in planning flap design, and its superiority over Doppler ultrasound has been well-established [4].

However, the information and 3D objects supplied by CTA are typically evaluated on a two-dimensional screen, which limits the ability to encompass the true depth and the possibilities for transferring detailed information onto the patient's body during the operation. Augmented reality (AR) has emerged as a novel method to overcome these barriers. In AR, virtual objects are seamlessly combined with real information to generate a "see-through" image. In particular, AR allows for 3D objects, including the vasculature, bone, muscle, and skin, to be overlaid onto the patient's body during the operation, serving as a guide to precisely and rapidly localize the perforator.

Without a head-mounted display (HMD), AR images can only be displayed on a monitor, which requires the surgeon to repeatedly switch his/her view between the monitor and operating field. This action can reduce the surgical precision for some delicate operations such as perforator localization. HoloLens, developed by Microsoft Corporation, is considered a very suitable AR HMD for surgical practice [5]. When wearing HoloLens, surgeons are able to see the AR images in a single view. Another key advantage of HoloLens is that it is not only a display system but is also a self-contained computer equipped with a camera and sensors, thereby avoiding the need to install any other hardware to realize the AR effect, which could further help to reduce the risk of contamination in the operating room.

HoloLens has been tentatively used in several surgical fields to date, including heart surgery [6], visceral surgery [7], and orthopedic surgery [8]. Most of these studies focused on the feasibility and performance of HoloLens; however, few studies have examined the precision of HoloLens in surgical practice. We previously verified the feasibility of AR-based perforator localization [9]. To improve the accuracy of this localization, HoloLens was applied to reduce the error caused by switching the view. In this study, the precision of this HoloLens-based perforator localization system was evaluated in a simulated operating room using a 3D printed model under different conditions, including with or without light, different intensities of light, and different visual angles.

Methods

Model Design

This study was conducted using a 3D printed model of a volunteer's deep femoral artery and its branches. The volunteer underwent a thin-cut (0.1 mm) CTA scan (Siemens, Germany) with administration of a nonionic contrast agent (Omnipaque, China) and the DICOM files were imported into Mimics 19.0 software (Materialise, Belgium) to create a virtual model, including a vascular map and surrounding soft tissues. A holder was designed by CAD software 3D Max (Autodesk, Mill Valley, CA, USA) to fix the model so that it could be placed stably on the operating table. The holder could also be used to place a marker, which is very important during registration. The 3D virtual model with the vascular map, surrounding soft tissues, marker, and holder was then printed for use in precision analysis. Data of the virtual model of the vascular map and marker were then input into HoloLens.

Augmented Reality Effect Workflow

The localization app was written within the Unity framework (Unity Technologies, San Francisco, CA, USA), which is a custom-developed HoloLens C# Universal Windows Platform app. The initial registration of the virtual model for real-world application was realized using a marker. Once launched, the app started the HoloLens built-in camera to acquire the view of the operative field. When the app finds the marker in the view, it registers the real marker with the virtual marker. The relative position between the virtual marker and the virtual model was the same as that between the real marker and the 3D printed model; consequently, the 3D printed model was registered with the virtual model at the time of registration of the real marker and virtual marker. After registration, the app generated AR images that were projected into the optical see-through lens incorporated in HoloLens. By tracking the position of the real marker though the camera, the position and angle of the virtual model automatically change according to the wearer's perspective to obtain registered AR images in real time. The complete workflow was as follows. First, the surgeon wore the HoloLens and launched the app. Second, the app registered the virtual model and the real world, and generated the appropriate AR images. Finally, the wearer localized the vasculature under guidance of the AR images.

Localization Precision Analysis

Five pairs of points on the vascular map that could be easily identified on the 3D printed model and virtual model were selected for precision analysis (Figure 1). The coordinates of these pairs of points were recorded by a Micron Tracker tracking device (Claron Technology, Toronto, ON, Canada), which can be used to detect and calculate the space location through a probe (distance error<0.2 mm). To eliminate the error caused by hand tremor, a mechanical arm was used to hold the probe (Figure 2).

Figure 1. Three-dimensional printed model of the vascular map (A) and the augmented reality image with the virtual model of the vasculature overlapped onto the printed model (B). P1-P5, 5 points on the three-dimensional printed vascular map selected for precision analysis; P1'-P5', 5 corresponding points on the virtual model.

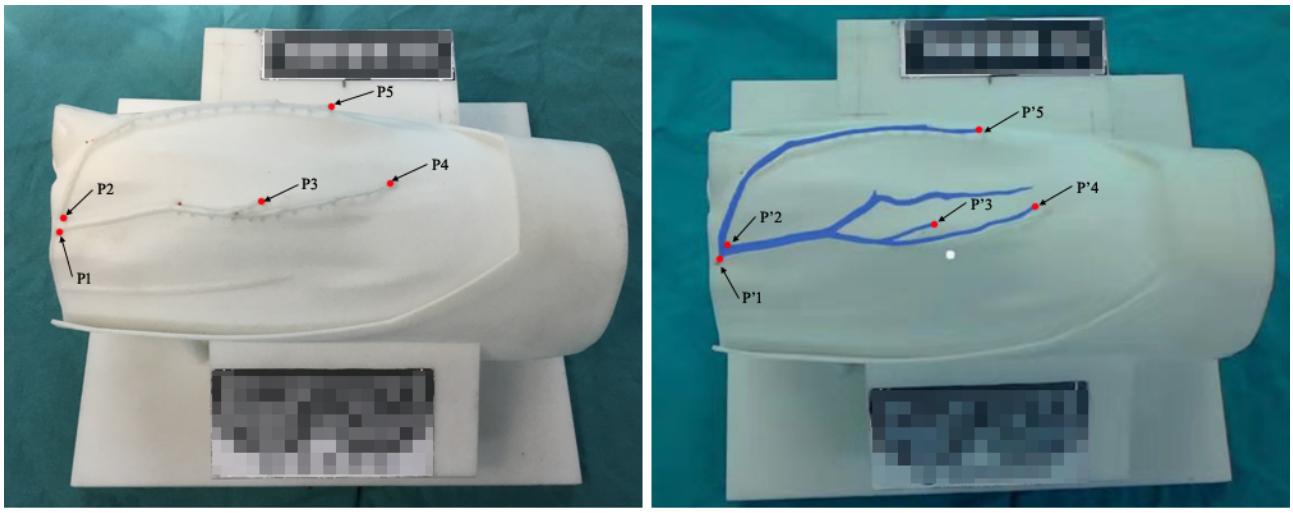
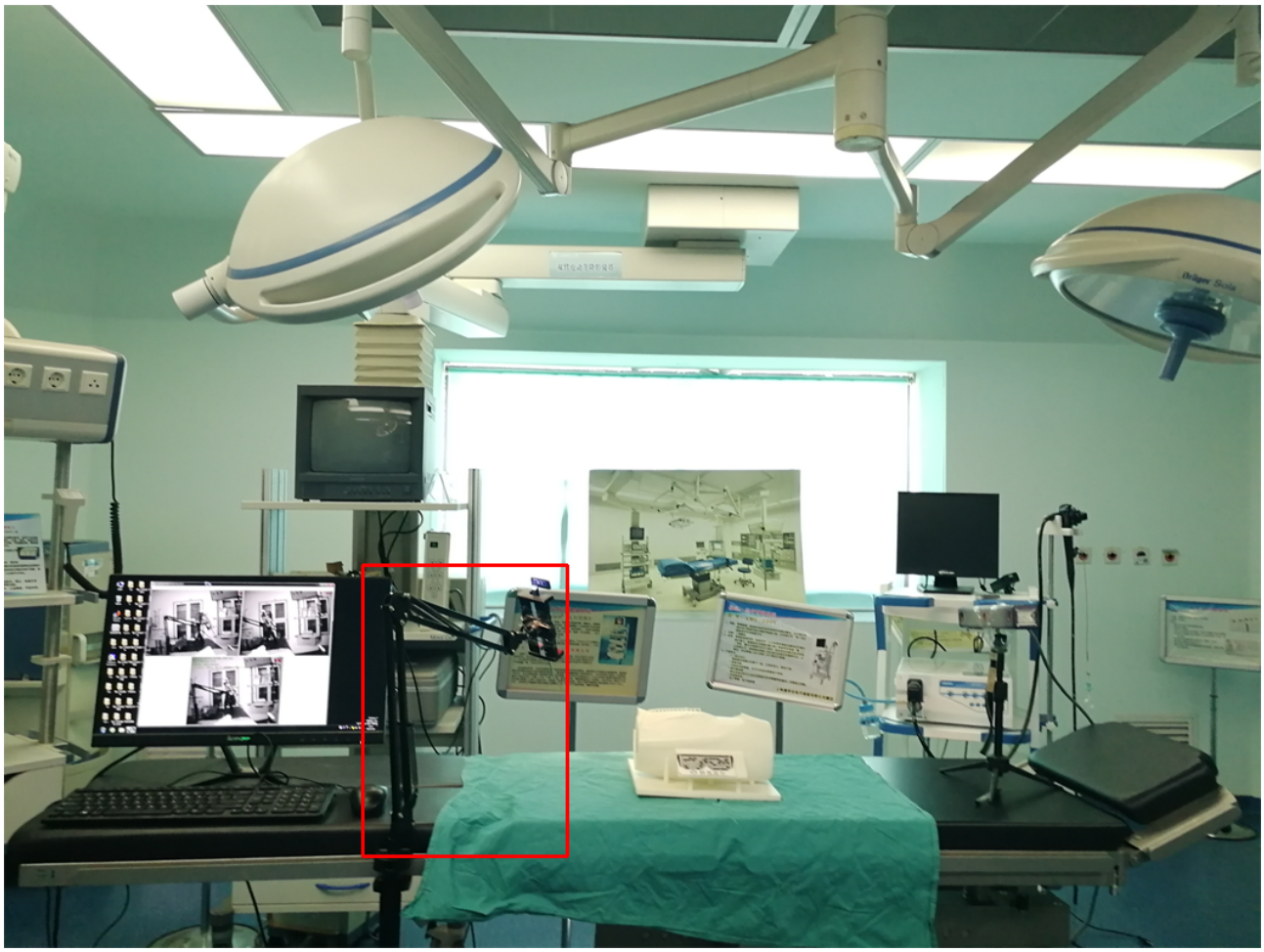


Figure 2. The simulated operating room used for the test. The probe was held by a mechanical arm.



The tests were carried out by 7 operators in a simulated operating room under different conditions, including with or without light, different intensities of light, and different visual angles (Figure 3). First, the coordinates of tested points in the 3D printed model were recorded without superimposing the virtual model. Then, the AR image was generated, and the coordinates of tested points in the virtual model were recorded. The distance between the corresponding points was then calculated with the following formula, and the distances were averaged to determine the overall error:

Figure 3. The coordinates of points were recorded by a tracking device (Micron Tracker) under different conditions. (A) The coordinates of points on the three-dimensional (3D) printed model were detected under the horizontal visual angle without light. (B) The coordinates of points on the 3D printed model were detected under the oblique visual angle without light. (C) The coordinates of points on the 3D printed model were detected under the vertical visual angle without light. (D) The coordinates of points on the virtual model were detected under the oblique visual angle without light. (E) The coordinates of points on the 3D printed model were detected under the oblique visual angle with light. (F) The coordinates of points on the virtual model were detected under the oblique visual angle with light.

Distance = square root $([X1 – X2]^2 + [Y1 – Y2]^2 + [Z1 – Z2]^2)$

Where X1, Y1, and Z1 are the coordinates of the point in the 3D printed model, and X2, Y2, and Z2 are the coordinates of the corresponding point in the virtual model.

Two types of errors were considered in this study: static error and motion error. The latter included the error caused by head rotation (ie, error caused by rotating the head to look around and then returning the head to the initial state without body motion) and error caused by walking (ie, walking around the operation room and then returning to the initial position). Figure 4 shows the workflow of the test. The detailed content of the test is shown in Table 1.

Figure 4. Workflow of the test. AR: augmented reality; 3D: three-dimensional.

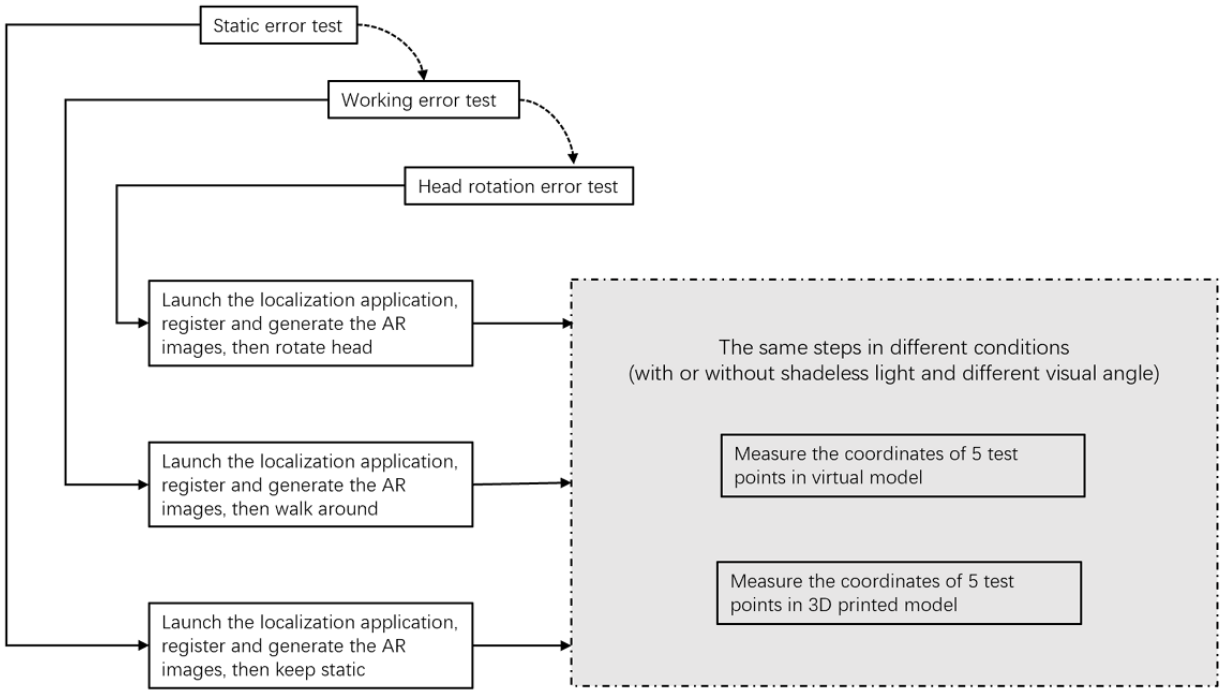


Table 1. Detailed content of the performance test.

Error value	Concrete content
SE-V	Average static error of 5 test points without light under vertical observation
SE-O	Average static error of 5 test points without light under oblique observation (45°)
SE-H	Average static error of 5 test points without light under horizontal observation
SE-(V/O/H)-L	Average static error of 5 test points under low light intensity and the best observation angle
SE-(V/O/H)-M	Average static error of 5 test points under medium light intensity and the best observation angle
SE-(V/O/H)-H	Average static error of 5 test points under high light intensity and the best observation angle
SE-(V/O/H)-SH	Average static error of 5 test points under superhigh light intensity and the best observation angle
HE-(V/O/H)	Average head rotation error of 5 test points without light under the best observation angle
WE-(V/O/H)	Average walking error of 5 test points without light under the best observation angle

Statistical Analysis

Continuous variables are presented as the mean and SD and were analyzed by paired Student *t* tests. A *P* value less than .05 was considered to represent a statistically significant difference.

Results

The results are summarized in Tables 2-5. The static error was first recorded under different visual angles without light, demonstrating no significant differences (horizontal vs oblique *P*=.25; horizontal vs vertical *P*=.25; oblique vs vertical *P*=.10; Table 2).

Since the oblique visual angle was similar to the surgeon's visual angle, further tests were performed under the oblique visual angle, and the static errors were recorded with different light intensities. There was no significant difference among the errors tested under these light intensities (low vs medium *P*=.25; low vs high *P*=.32; low vs superhigh *P*=.08; medium vs high *P*=.32; medium vs superhigh *P*=.43; high vs superhigh *P*=.09; Table 3).

However, the difference between the errors tested with and without light was statistically significant (*P*=.02; Table 4).

Finally, the motion errors under an oblique angle without light were recorded. Compared with the static error, there was no significant difference in either the head rotation (*P*=.25) or walking (*P*=.30) motion errors (Table 5).

Table 2. Static error tested without light under different visual angles (mm).

Operator	Visual angle, degrees		
	0 (Horizontal)	45 (Oblique)	90 (Vertical)
1	1.28	0.63	1.63
2	1.80	1.47	1.28
3	0.74	1.01	1.36
4	1.83	1.3	1.73
5	1.94	1.63	1.43
6	1.26	1.45	2.31
7	1.44	1.96	1.67
Mean (SD)	1.47 (0.42)	1.35 (0.43)	1.63 (0.34)

Table 3. Static error under different light intensities (mm).

Operator	Light Intensity			
	Low	Medium	High	Superhigh
1	2.70	1.33	1.91	1.93
2	2.82	4.00	2.55	3.25
3	3.93	4.23	5.03	4.60
4	1.74	1.99	1.74	2.27
5	2.97	2.26	3.05	2.99
6	4.69	6.39	4.89	5.31
7	1.11	1.77	1.54	1.94
Mean (SD)	2.85 (1.21)	3.14 (1.81)	2.96 (1.46)	3.18 (1.32)

Table 4. Static error with or without light (mm).

Operator	With light (low intensity)	Without light
1	2.70	0.63
2	2.82	1.47
3	3.93	1.01
4	1.74	1.3
5	2.97	1.63
6	4.69	1.45
7	1.11	1.96
Mean (SD)	2.85 (1.21)	1.35 (0.43)

Table 5. Motion errors under an oblique angle without light (mm).

Operator	Walking	Head rotation	Static
1	1.06	1.34	0.63
2	2.28	0.84	1.47
3	1.59	3.35	1.01
4	1.02	1.50	1.3
5	1.35	1.44	1.63
6	1.38	1.28	1.45
7	1.5	1.74	1.96
Mean (SD)	1.45 (0.42)	1.50 (0.80)	1.35 (0.43)

Discussion

Principal Findings

We previously verified the feasibility of AR-based perforator localization. Under the guidance of localization, a surgeon successfully dissected a beagle's thoracodorsal artery perforator without any prior anatomical knowledge of the dog [9]. The mean error of the localization system was 3.5 mm. However, the HMD (nVisor ST60, NVIS Company, USA) used in this previous system was merely an AR display without any other function, and thus this system had to be equipped with a computer workstation and a 3D camera, which increases the complexity of the system as well as the risk of contamination in the operating room. In addition, the nVisor ST60 HMD is very heavy (1.3 kg), which can be uncomfortable for a surgeon if required to wear the device for a long time. Therefore, we subsequently used HoloLens as the HMD. HoloLens is not only an HMD but is also a self-contained computer that has a built-in camera and sensors, and weighs only 579 g. In addition, HoloLens can be operated with simple hand gestures and voice commands instead of touch [10]. Therefore, the virtual model could be operated (including transparency adjustment and access to anatomical information) in real time while remaining sterile. However, it remained to be determined whether HoloLens-based vascular localization meets clinical requirements (error of localization less than 5 mm). The present study demonstrates that the HoloLens-based vascular localization system could precisely localize vessels with a minimum mean error of 1.35 mm (SD 0.43) and maximum mean error of 3.18 (SD 1.32), which were within the clinically acceptable range of 5 mm [11].

To maximally simulate the state of operation, the static errors were tested under different operating conditions, including different visual angles and different light intensities, and motion errors were also tested to simulate the typical movements of surgeons. None of the differences in the errors tested under different visual angles was statistically significant, and there was also no difference between the static error and motion error. Thus, the system has good robustness, and can remain stable and precise in different states. Since an opaque virtual model could block out real information, the virtual model should be displayed in a semitransparent state. Since light used in the operation could affect observation of the semitransparent virtual model, we also tested the static errors under different light intensities. No differences in the errors tested under different light intensities were found, whereas the difference in the errors tested with or without light was significant. Based on this finding, we recommend using HoloLens-based surgical navigation without light (ie, turn off the light when localizing the vasculature).

Prospects

Modern medicine is developing toward personalized and precision treatment. Image-assisted surgical navigation systems could enable individual surgical planning and precise surgical procedures, which largely involve the use of virtual and augmented/mixed reality techniques. In virtual reality, the user is completely separated from the real world and is highly immersed in a virtual world, whereas in augmented/mixed reality, the virtual elements are overlaid onto the user's reality and thus appear to be part of the real world. With a virtual reality–based surgical navigation system, the surgical instruments must be registered for projection into the virtual world, and the surgeon must switch the view from the virtual world to the real world to perform surgery. By contrast, with an AR-based surgical navigation system, the surgeon can see the virtual anatomical model and the real world simultaneously, so that the surgery can be performed directly without switching the view and surgical instrument registration. In recent decades, image-assisted surgical navigation systems have largely depended on virtual reality, but have gradually transferred to AR, which we believe will become the mainstream. Microsoft HoloLens could simplify and popularize the use of AR-based image-assisted surgical navigation systems, with potential benefits of low weight, less hardware, and gesture control.

Perforator flaps have been widely used in reconstructive surgery owing to their multiple advantages such as low morbidity at the donor site, good reconstruction and appearance of the recipient site, flexible design, and short postoperative recovery time [12]. Perforator dissection is a standard and important procedure for perforator flap transplantation, and also the most difficult process due to the uncertainty in predicting the anatomical location of the perforator and high variability in perforator size and course [13]. In this context, HoloLens-based perforator localization could play a pivotal role by offering individual 3D perforator information during the operation that is projected directly onto the patient with precise registration.

Conclusion

The results of this study demonstrated that the HoloLens-based vascular localization system could precisely localize the perforator and might help a surgeon accomplish the operation. Although the precision of this system reached the clinical requirement, reduction of the error is still an important issue that deservers further study. The potential sources of error could be produced from the very beginning of the process from computed tomography/magnetic resonance imaging data acquisition through to the end of the surgical procedure, including imaging error, registration error, tracking error, and human error. Limitations of the hardware could also be a source of error, such as the perceptual limit of HMD [14]. Further study is ongoing and upcoming to confirm and seek solutions to eliminate or diminish these error sources. Clinical research will also be carried out to verify the feasibility of the system. Additionally, clinical application may encounter some problems such as the anatomical deformation between the CTA scan and operation, and the muscle relaxation caused by anesthesia. Thus, further research on precision is still needed, and a body fixation device may need to be integrated in the system to reduce the influence of anatomical deformation.

Acknowledgments

This work was supported by Shanghai Municipal Commission of Health and Family Planning Program to TJ (No. 20194Y0033).

Conflicts of Interest

None declared.

References

1. Rozen WM, Paddle AM, Chubb D, Wilson J, Grinsell D, Ashton MW. Guiding local perforator flaps with preoperative imaging: revealing perforator anatomy to improve flap design. Plast Reconstr Surg 2012 Jul;130(1):130-134. [doi: 10.1097/PRS.0b013e31824ecaff] [Medline: 22743880]
2. Ngaage LM, Hamed R, Oni G, Di Pace B, Ghorra DT, Koo BMedSci BC, et al. The Role of CT Angiography in Assessing Deep Inferior Epigastric Perforator Flap Patency in Patients With Pre-existing Abdominal Scars. J Surg Res 2019 Mar;235:58-65. [doi: 10.1016/j.jss.2018.09.059] [Medline: 30691846]
3. Davis CR, Jones L, Tillett RL, Richards H, Wilson SM. Predicting venous congestion before DIEP breast reconstruction by identifying atypical venous connections on preoperative CTA imaging. Microsurgery 2019 Jan;39(1):24-31 [FREE Full text] [doi: 10.1002/micr.30367] [Medline: 30182489]
4. Rozen WM, Phillips TJ, Ashton MW, Stella DL, Gibson RN, Taylor GI. Preoperative imaging for DIEA perforator flaps: a comparative study of computed tomographic angiography and Doppler ultrasound. Plast Reconstr Surg 2008 Jan;121(1):9-16. [doi: 10.1097/01.prs.0000293874.71269.c9] [Medline: 18176200]
5. Moosburner S, Remde C, Tang P, Queisner M, Haep N, Pratschke J, et al. Real world usability analysis of two augmented reality headsets in visceral surgery. Artif Organs 2019 Jul;43(7):694-698. [doi: 10.1111/aor.13396] [Medline: 30485464]
6. Brun H, Bugge R, Suther L, Birkeland S, Kumar R, Pelanis E. Mixed reality holograms for heart surgery planning: first user experience in congenital heart disease. Eur Heart J Cardiovasc Imaging 2019 Aug 01;20(8):883-888 [FREE Full text] [doi: 10.1093/ehjci/jey184] [Medline: 30534951]
7. Shi L, Luo T, Zhang L, Kang Z, Chen J, Wu F, et al. [Preliminary use of HoloLens glasses in surgery of liver cancer]. Zhong Nan Da Xue Xue Bao Yi Xue Ban 2018 May 28;43(5):500-504 [FREE Full text] [doi: 10.11817/j.issn.1672-7347.2018.05.007] [Medline: 29886465]
8. Condino S, Turini G, Parchi PD, Viglialoro RM, Piolanti N, Gesi M, et al. How to Build a Patient-Specific Hybrid Simulator for Orthopaedic Open Surgery: Benefits and Limits of Mixed-Reality Using the Microsoft HoloLens. J Healthc Eng 2018 Nov 1;2018:5435097. [doi: 10.1155/2018/5435097] [Medline: 30515284]
9. Jiang T, Zhu M, Zan T, Gu B, Li Q. A Novel Augmented Reality-Based Navigation System in Perforator Flap Transplantation - A Feasibility Study. Ann Plast Surg 2017 Aug;79(2):192-196. [doi: 10.1097/SAP.0000000000001078] [Medline: 28509695]
10. Hong K, Sakamoto Y, Irani P. The Use of Head-Worn Augmented Reality Displays in Health Communications. Stud Health Technol Inform 2019;257:163-169. [Medline: 30741190]
11. Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, et al. Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. Eur Radiol Exp 2018 Jan;2(1):2 [FREE Full text] [doi: 10.1186/s41747-017-0033-2] [Medline: 29708204]
12. Koshima I, Soeda S. Inferior epigastric artery skin flaps without rectus abdominis muscle. Br J Plast Surg 1989 Nov;42(6):645-648 [FREE Full text] [doi: 10.1016/0007-1226(89)90075-1] [Medline: 2605399]
13. Rozen WM, Ashton MW, Pan W, Kiil BJ, McClure VK, Grinsell D, et al. Anatomical variations in the harvest of anterolateral thigh flap perforators: a cadaveric and clinical study. Microsurgery 2009;29(1):16-23. [doi: 10.1002/micr.20550] [Medline: 18942652]

14. Condino S, Carbone M, Piazza R, Ferrari M, Ferrari V. Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks. IEEE Trans Biomed Eng 2020 Feb;67(2):411-419. [doi: 10.1109/TBME.2019.2914517] [Medline: 31059421]

Abbreviations

3D: three-dimensional
AR: augmented reality
CTA: computed tomographic angiography
HMD: head-mounted display

Edited by G Eysenbach; submitted 31.10.19; peer-reviewed by D Amiras, M Eckert, K Kanno; comments to author 23.12.19; revised version received 22.01.20; accepted 10.02.20; published 17.04.20.

Please cite as:
Jiang T, Yu D, Wang Y, Zan T, Wang S, Li Q
HoloLens-Based Vascular Localization System: Precision Evaluation Study With a Three-Dimensional Printed Model
J Med Internet Res 2020;22(4):e16852
URL: http://www.jmir.org/2020/4/e16852/
doi:10.2196/16852
PMID:32301738

Original Paper

Age-Related Differences in the Evaluation of a Virtual Health Agent's Appearance and Embodiment in a Health-Related Interaction: Experimental Lab Study

Carolin Straßmann[1], Dr; Nicole C Krämer[2], Prof Dr; Hendrik Buschmeier[3], Dr; Stefan Kopp[3], Prof Dr

[1]Computer Science Institute, University of Applied Sciences Ruhr West, Bottrop, Germany

[2]Social Psychology: Media and Communication, University Duisburg-Essen, Duisburg, Germany

[3]Social Cognitive Systems Group, CITEC, Bielefeld University, Bielefeld, Germany

Corresponding Author:
Carolin Straßmann, Dr
Computer Science Institute
University of Applied Sciences Ruhr West
Lützowstraße 5
Bottrop, 46236
Germany
Phone: 49 20888254794
Email: carolin.strassmann@hs-ruhrwest.de

Abstract

Background: Assistive technologies have become more important owing to the aging population, especially when they foster healthy behaviors. Because of their natural interface, virtual agents are promising assistants for people in need of support. To engage people during an interaction with these technologies, such assistants need to match the users´ needs and preferences, especially with regard to social outcomes.

Objective: Prior research has already determined the importance of an agent's appearance in a human-agent interaction. As seniors can particularly benefit from the use of virtual agents to maintain their autonomy, it is important to investigate their special needs. However, there are almost no studies focusing on age-related differences with regard to appearance effects.

Methods: A 2×4 between-subjects design was used to investigate the age-related differences of appearance effects in a human-agent interaction. In this study, 46 seniors and 84 students interacted in a health scenario with a virtual agent, whose appearance varied (cartoon-stylized humanoid agent, cartoon-stylized machine-like agent, more realistic humanoid agent, and nonembodied agent [voice only]). After the interaction, participants reported on the evaluation of the agent, usage intention, perceived presence of the agent, bonding toward the agent, and overall evaluation of the interaction.

Results: The findings suggested that seniors evaluated the agent more positively (liked the agent more and evaluated it as more realistic, attractive, and sociable) and showed more bonding toward the agent regardless of the appearance than did students. In addition, interaction effects were found. Seniors reported the highest usage intention for the cartoon-stylized humanoid agent, whereas students reported the lowest usage intention for this agent. The same pattern was found for participant bonding with the agent. Seniors showed more bonding when interacting with the cartoon-stylized humanoid agent or voice only agent, whereas students showed the least bonding when interacting with the cartoon-stylized humanoid agent.

Conclusions: In health-related interactions, target group–related differences exist with regard to a virtual assistant's appearance. When elderly individuals are the target group, a humanoid virtual assistant might trigger specific social responses and be evaluated more positively at least in short-term interactions.

(J Med Internet Res 2020;22(4):e13726) doi:10.2196/13726

KEYWORDS

virtual health advisor; age-related differences; human-agent interaction; appearance; embodiment

Introduction

As care persons are lacking and, at the same time, most current industrial societies have an aging population, assistive technologies are of great interest [1]. Owing to their natural interface and their ability to communicate in a human-like way [2], virtual agents are ubiquitously applicable. Consequently, several authors have described the use of virtual agents as a promising approach in terms of assistive elderly care [1,3,4], especially in the area of health applications [5]. There are several groups of people in need of support, for whom cognitive limitations raise problems for mastering daily life activities. These people might have problems with regular activities, such as having meals, drinking sufficiently, taking medications, and meeting social contacts [4]. More precisely, assistive agents can provide reminders to drink or take medications, foster physical activity or social gatherings, and guide users regarding household activities [1]. This application was found to be acceptable and well usable in these kinds of target groups [4]. Within this application, the agent needs to not only process tasks correctly, but also demonstrate and use social skills [5], as it is integrated in daily life and is often used in vulnerable target groups. As a virtual agent was found to trigger social responses similar to humans [6], its appearance was found to affect the human-agent interaction regarding multiple variables [7-9]. In this regard, the appearance of the agent should match the needs of special users to achieve highly efficient assistive technology. However, regarding different experiences and expectations, the preferences and needs of target groups might broadly differ. Therefore, this study aimed to investigate target group–related differences in the perception and evaluation of a virtual assistant and its appearance in an assisted-living health-related scenario. As seniors are the main target group for which virtual assistants are highly beneficial [3,4], this study investigated age-related differences by comparing students to seniors.

Although seniors in general are more skeptical about the use of technologies, such as virtual agents [10], they seem to be less critical when asked after an actual interaction with these technologies. Prior research showed that seniors evaluated these agents more positively as compared with students in general [11]. This might be caused by fewer experiences with such technologies and by the resulting lower expectations. Thus, the following can be hypothesized: seniors will evaluate the agent and the interaction therewith more positively regarding its person perception (a), liking (b), usage intention (c), usefulness (d), and enjoyment (e) as compared with students (hypothesis 1 [H1]).

Nevertheless, as mentioned earlier, younger people are more familiar with virtual agents and their use. This familiarity with the usage of these technologies most likely enhances the perceived ease of use. Therefore, the following hypothesis can be assumed: students will rate the agent as easier to use as compared with seniors (hypothesis 2 [H2]).

To date, to the authors' knowledge, no research has investigated the effect of appearance on the preferences of different age groups in an actual human-agent interaction. However, some research has indicated that age-related differences exist in the evaluation of an agent's appearance. Findings with regard to e-commerce [12] report that seniors prefer an abstract appearance, as it is less distracting than a human or even a human who uses movements. The authors further showed that participants preferred an animal appearance over a human appearance, a humanoid agent was too distracting, and participants did not like technical entities to simulate a human [12]. However, the research did not systematically distinguish between different facets of appearance variables. Prior research [13] highlighted that it is important to distinguish systematically between variables and that the appearance variables species, realism, and embodiment are of specific interest. With regard to the results of Chattaraman et al [12], it is still unclear what is meant by an abstract agent. The level of abstractness can depend on the species and degree of realism. Beside these methodological inaccuracies, the findings contradict prior research [13]. Here, seniors clearly preferred humanoid and realistic agents to other species, as they were more familiar with human interaction. By contrast, they mentioned feeling stultified by nonhumanoid characters and that cartoon-stylized agents are for kids. Based on these results, it was concluded that seniors can only take humanoid agents seriously and therefore will evaluate them more positively. Additionally, Tsiouri et al [5] found in qualitative focus groups that elderly individuals prefer a realistic agent, as they want to be able to look into its eyes. This supports the findings from the study by Straßmann and Krämer [13]. The contradictions between those findings can most likely be explained through an applied context [8]. In e-commerce, the focus is on the presented product instead of personal communication and assistance, and therefore, users might prefer differently designed agents. As virtual assistance is applied in health-related domains in this research, the results of Straßmann and Krämer [13] might be better applicable. Further research [14] complements the assumptions and demonstrates that the species is more important in the evaluation process of a virtual agent's appearance for seniors and that seniors evaluate machine-like agents less positively than humanoid ones. Based on these features, for seniors, no differences between different degrees of realism are assumed, whereas they are expected to evaluate a humanoid agent more positively. Accordingly, the following hypothesis is assumed: seniors will evaluate a humanoid agent's appearance (regardless of its degree of realism) more positively regarding its person perception (a), liking (b), and usage intention (c) as compared with an agent having a machine-like appearance (hypothesis 3 [H3]).

Furthermore, a prior qualitative study [13] stressed that seniors prefer an embodied agent to a nonembodied agent, as they like to have something to address during an interaction. By contrast, students highlighted the advantages of a nonembodied agent, as nonembodied agents are not restricted to one device or a certain screen. Therefore, students are expected to evaluate nonembodied agents more positively than embodied agents. Thus, the following hypotheses can be formulated: seniors will evaluate an embodied agent more positively regarding its person perception (a), liking (b), and usage intention (c) as compared with a nonembodied agent (hypothesis 4 [H4]) and students will evaluate an agent represented through a voice only more

positively regarding its person perception (a), liking (b), and usage intention (c) as compared with an embodied agent (hypothesis 5 [H5]).

In focused application, social processes are of high interest. Prior studies demonstrated that users show bonding with virtual agents [15,16]. In interpersonal relationships and attraction, physical attractiveness and similarity are key variables. The appearance of an agent, especially its species and realism, can affect the perceived similarity and, of course, the perceived attractiveness. Therefore, how these appearance variables influence participants' bonding and trust needs to be investigated. Findings of prior research [13] demonstrated that seniors felt more trust toward embodied agents and bonded more with humanoid agents. In contrast, students did not mention these perceptions and seemed to try to avoid such social processes [13]. Nevertheless, according to the media equation theory [17], such processes are seen to occur automatically for all human beings. In summary, whether the appearance of a virtual agent affects social processes, such as bonding and trust, and whether these are influenced by user age should be investigated. Therefore, the following research question is posed: How are bonding and trust affected by the agent's appearance and the users' age group (research question 1 [RQ1])?

Methods

Study Design

This study aimed to investigate the effects of appearance and embodiment regarding a human-agent interaction and further examine the moderating effect of age, as a possible target group is elderly individuals in need of support and health advice. Therefore, a Wizard-of-Oz study with a 2 (age group) × 4 (appearance) between-subjects design was conducted.

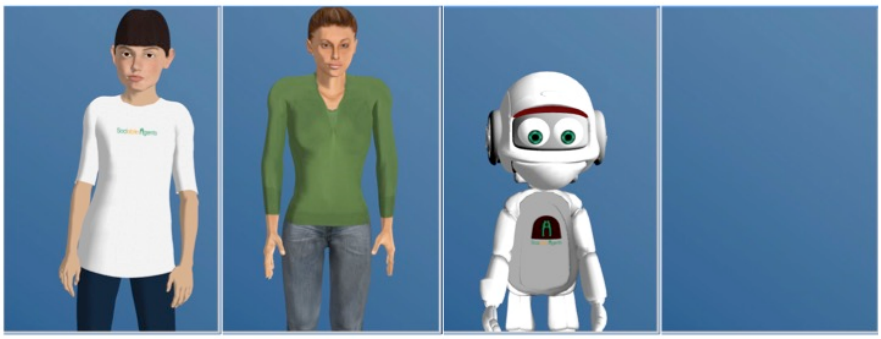
Sample

Overall, 130 people participated in this study. To investigate age-related differences, two different age groups were invited to participate in this study. In total, 84 students (mean age 23.65, SD 3.84; range 18-38 years) and 46 seniors (mean age 70.93, SD 9.05; range 51-89 years) interacted with a virtual agent and evaluated it thereafter. Both groups differed significantly with regard to age ($F_{1,129}$=1732.19, P<.001, η_{part}^2=0.931). Unfortunately, sex was not balanced, with more women (81/130, 62.3%) than men (49/130, 37.7%) participating. However, there were no differences in sex distribution between the two age groups (χ^2_1=1.59, P=.21). Students and seniors were further equally distributed in all four experimental conditions (χ^2_3=0.32, P=.96). Nevertheless, both groups differed in their prior experiences with virtual agents (χ^2_1=31.40, P<.001). Among the students, many (54/84, 64%) had interacted with a virtual agent in the past; however, among seniors, few (6/46, 13%) had interacted with a virtual agent in the past, and thus, the majority of seniors had no prior experience with this technique.

Stimulus Material

In this study, a real human-agent interaction was tested in a Wizard-of-Oz setting. Participants interacted with an agent that was manipulated with regard to its appearance. The behavior and interaction contents were kept constant among all conditions, and participants were randomized to one of four conditions. Owing to the between-subjects design, participants interacted with one of the following four agents differing in appearance: two humanoid agents, one machine-like agent, and one agent without embodiment (Figure 1). To investigate the effect of species in an actual interaction study, humanoid appearances (Billie and Character) and a machine-like appearance (Vince) were used. Further, the influence of realism was tested by comparing a cartoon-stylized human (Billie) and a more realistic character (Character). To investigate the effect of different degrees of realism, both appearances were chosen according to realism. Billie's overall degree of realism was rather low, and it had a cartoon-stylized shade. Although proportions were rather natural and not stylized, the resolution was more unrealistic than realistic owing to the material and texture. In contrast, Character was characterized by a higher degree of realism, as many details were obtainable, no stylized shade or proportions were used, and the resolution was more realistic (although not completely photorealistic). Additionally, this study aimed to investigate the influence of embodiment by comparing embodied characters with a voice-only version of the agent (Figure 1).

Figure 1. Overview of the used appearances. From left to right: Billie, Character, Vince, and Voice Only.



As participants interacted with the agent, an actual agent with an underlying skeleton, which could be animated to move and talk, was needed. Owing to these technical restrictions, the possible design decisions and usable appearances were limited. Thus, the manipulation used was not as controlled as desired. Nevertheless, it is a great advantage to investigate the effect of appearance in an interaction situation, where participants are able to communicate with a virtual agent. As described earlier,

appearances can be used to explore the effects of species, realism, and embodiment in an actual human-agent interaction. During an interaction, the agent used the same voice under all conditions. Its nonverbal behavior was also kept constant as much as possible.

The participants' task was to fill in a health diary and to schedule appointments in a calendar. This is a realistic and possible scenario in the application of health-related daily life assistance, as a virtual agent might be able to help with health choices, provide reminders in this regard, and help plan and structure the day. The diary entries were presented in tables and were supplemented by matching icons to adapt to people in need of support, who might have difficulties reading and understanding the textual inputs. Figure 2 presents an example of the appearance of the entries.

Figure 2. Presented tables during the interaction with the agent.

Overall, the interaction with the agent lasted for about 15 to 20 minutes. To guarantee a controlled and stable dialogue under all conditions, the interaction was prescripted. However, the wizard had a chance to respond to participants' answers by including specific attributes in the prescripted dialogue (eg, the agent asked about the participants' favorite sport and the wizard typed the answer, so that it was included in the prescripted response of the agent). Moreover, the wizard had the option to type free responses that had not been prescripted. However, this option was only used when participants digressed from the interaction topic and the wizard had to lead them back on the topic.

Measurements

In order to examine the effect of appearance on the evaluation of the virtual agent, the person perception of the agent was assessed. Five different concepts were measured with 28 items overall on a 5-point semantic differential as follows: perceived realism (seven items, eg, "fake-natural"), likability (seven items, eg, "unfriendly-friendly"), trustworthiness (five items, eg, "not trustworthy-trustworthy"), competence (five items, eg, "incompetent-competent"), and attractiveness (four items, eg, "unattractive-attractive"). The items were adapted from prior person perception measurements [7,18]. All scales showed good reliability (Cronbach α <.808).

Supplementing this measure, participants' liking of the agent was measured with an ad-hoc scale comprising five items (eg, "I think I would like this agent") rated on a 5-point Likert scale (ranging from 1 [totally disagree] to 5 [totally agree]). The internal consistency of the scale was good (Cronbach α=.869).

Additionally, the usage intention of the participants was assessed. The perceived usefulness was measured with three items (eg, "I think the agent is useful to me"), and participants' intention to use the virtual agent was measured with three items (eg, "I think I will use the agent during the next few days"). All items were rated on a 5-point Likert scale (ranging from 1 [totally disagree] to 5 [totally agree]). Both scales showed excellent internal consistency (Cronbach α <.934).

Furthermore, trust in the virtual agent was queried with two items (eg, "I would trust the robot if it gave me advice") rated on a 5-point Likert scale (ranging from 1 [totally disagree] to 5 [totally agree]). The reliability of this scale was excellent (Cronbach α=.930).

As the relationship between user and agent is the focus of this work, participants' bonding with the agent was measured using the bonding subscale of the Working Alliance Inventory [19]. The 12 included items (eg, "I feel uncomfortable with the agent") were rated on a 5-point Likert scale (ranging from 1

[totally disagree] to 5 [totally agree]), and there was good reliability (Cronbach α=.842).

To evaluate the interaction, enjoyment (five items, eg, "I enjoyed the agent talking to me"), ease of use (five items, eg, "I think I will know quickly how to use the agent"), and sociability (four items, eg, "I consider the robot a pleasant conversational partner") were measured. Participants evaluated all items on a 5-point Likert scale (ranging from 1 [totally disagree] to 5 [totally agree]). Cronbach α values demonstrated acceptable internal consistency for enjoyment and sociability (Cronbach α <.799), whereas the consistency for ease of use was not acceptable (Cronbach α=.544). Nevertheless, as this concept was of high relevance in this study, it was included in further analyses, but the results need to be discussed cautiously.

In addition, participants' intended health behavior was measured to determine whether there was any influence by the virtual agent. This was measured with five items (eg, "Eat a well-balanced diet" and "Eat fresh fruits and vegetables") [20]. As this scale mostly involves behaviors regarding diet, three items concerning physical exercises from Cunningham and Kwon [21] were added (eg, "I am planning to be physically active on a regular basis next week"). Participants rated all items on a 5-point Likert scale (ranging from 1 [totally disagree] to 5 [totally agree]). The reliability of this scale was good (Cronbach α=.831).

Several control variables were measured in the first part of the questionnaire. They included tendency to anthropomorphize [22] (10 items, eg, "I sometimes wonder if my computer deliberately runs more slowly after I have shouted at it"; Cronbach α=.796), prior experiences, participants' attitude (3 items, eg, "I think it is a good idea to use the virtual agent"; Cronbach α=.752), anxiety (4 items, eg, "If I should use the virtual agent, I would be afraid to make mistakes with it"; Cronbach α=.725) [18], and negative attitudes toward virtual agents [23] (14 items, eg, "I would feel uneasy if agents really had emotions"; Cronbach α >.618). At the end, sociodemographic variables, such as age and sex, were measured.

Procedure

When participants came into the laboratory, the experimenter welcomed them. To obtain informed consent, they were informed about the background and procedure of the study. At first, all participants filled in questionnaires about personality traits, prior experiences, and other control variables. When participants finished the first part of the questionnaire, the experimental part of the study began. Participants were asked to interact with the virtual agent in the context of a health diary. During the interaction, the experimenter left the room and participants were alone with the virtual agent. They were told that the agent interacts autonomously and can understand and react to their speech and behavior, but in fact, a Wizard-of-Oz setting was used, where a confederate ("wizard") controlled the agent from an adjacent room. Participants were asked to start the interaction with "Hello Billie." The wizard replied to this, and the interaction began. After the interaction with the agent was finished, the second questionnaire part was started, where dependent variables were assessed. At the end, the experimenter debriefed the participants and offered an incentive (either money or course credits).

Results

Assessment

This study aimed to investigate the effects of species, realism, and embodiment more closely in a real human-agent interaction. Therefore, planned contrasts were used to analyze the data with regard to specific comparisons and assumed hypotheses as follows: (1) embodiment (contrast 1), Billie, Vince, and Character versus Voice Only; (2) species (contrast 2), Billie and Character versus Vince; and (3) realism (contrast 3), Billie versus Character.

Furthermore, when significant interactions were found, according to Field [24], those effects were further investigate using simple effects, in which the effects of age groups at individual levels of the different appearances were assessed.

Person Perception

In order to test the influence of the different appearances and age groups on the person perception of the virtual assistant, a two-way multivariate analysis of variance (MANOVA) was conducted with appearance and age group as independent variables and perceived realism, likability, attractiveness, trustworthiness, and competence as dependent variables.

Using Pillai trace, there were significant effects of age group (V=0.21, $F_{5,118}$=6.17, P<.001) and appearance (V=0.21, $F_{15,360}$=1.84, P<.001) on person perception.

To test the hypothesis H1a, a univariate test was performed. The results indicated that seniors and students differed significantly in their evaluation of perceived realism ($F_{1,129}$=9.07, P=.003, η_p^2=0.069) and attractiveness ($F_{1,129}$=21.16, P<.001, η_p^2=0.148) of the agent. Seniors evaluated the agent in general as more realistic and more attractive than did students. No significant differences between the age groups were found for likability ($F_{1,129}$=1.12, P=.29, η_p^2=0.009), trustworthiness ($F_{1,129}$=2.34, P=.13, η_p^2=0.019), and competence ($F_{1,129}$=3.75, P=.055, η_p^2=0.030) (Table 1). Overall, the hypothesis H1a was only partly supported regarding attractiveness and had to be rejected for likability, trustworthiness, and competence.

Table 1. Person perception evaluation of the different age groups in general.

Variable	Score, mean (SD)		
	Seniors	Students	Overall
Realism	3.22 (1.01)	2.70 (0.81)	2.89 (0.91)
Likability	4.11 (0.76)	3.95 (0.71)	4.01 (0.73)
Attractiveness	3.79 (0.91)	3.01 (0.87)	3.29 (0.96)
Trustworthiness	3.45 (0.80)	3.19 (0.86)	3.28 (0.85)
Competence	4.13 (0.61)	3.86 (0.76)	3.95 (0.71)

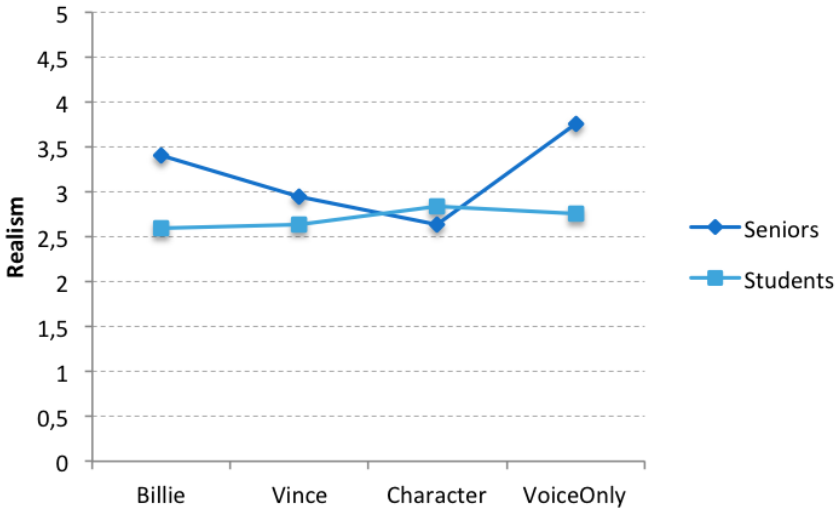
Univariate tests further revealed a significant difference between appearances regarding their perceived competence ($F_{3,129}$=2.98, P=.03, η_p^2=0.068). Running the planned contrast, results of the analyses showed a significant effect of embodiment (t_{126}=−2.21, P=.03, r=0.19), where an embodied character was evaluated as less competent than the Voice Only condition (Table 2). Furthermore, a marginally significant effect of appearance on likability was found ($F_{3,129}$=2.63, P=.05, η_p^2=0.061). Again, planned contrasts were used to explore this effect more deeply. Here, a significant difference in species was found (t_{126}=−2.52, P=.01, r=0.22), where Vince was evaluated as more likable than Billie and Character (Table 2). The tests yielded no significance for realism ($F_{3,129}$=2.20, P=.09, η_p^2=0.051), attractiveness ($F_{3,129}$=1.99, P=.12, η_p^2=0.047), and trustworthiness ($F_{3,129}$=0.60, P=.62, η_p^2=0.015).

Table 2. Person perception evaluation for the different appearances regardless of the age group.

Variable	Score, mean (SD)				
	Billie	Vince	Character	Voice Only	Overall
Realism	2.89 (0.90)	2.76 (0.78)	2.77 (0.84)	3.12 (1.09)	2.89 (0.91)
Likability	3.82 (0.77)	4.20 (0.58)	3.82 (0.72)	4.15 (0.78)	4.01 (0.73)
Attractiveness	3.14 (0.95)	3.49 (0.98)	3.03 (0.99)	3.47 (0.88)	3.29 (0.96)
Trustworthiness	3.16 (1.00)	3.46 (0.78)	3.21 (0.77)	3.28 (0.83)	3.28 (0.85)
Competence	3.78 (0.76)	4.04 (0.66)	3.81 (0.71)	4.18 (0.67)	3.95 (0.71)

According to Pillai trace, there was no significant interaction effect of both dependent variables for person perception (V=0.09, $F_{15,360}$=0.74, P=.75). In contrast, the univariate tests showed a significant interaction effect of appearance and age group with regard to perceived realism ($F_{3,129}$=2.71, P=.048, η_p^2=0.063) but not likability ($F_{3,129}$=1.18, P=.32, η_p^2=.028), attractiveness ($F_{3,129}$=0.47, P=.71, η_p^2=0.011), trustworthiness ($F_{3,129}$=0.73, P=.54, η_p^2=0.018), and competence ($F_{3,129}$=1.33, P=.27, η_p^2=0.032). To analyze the interaction effect for perceived realism in more detail, simple effects were assessed, and it was found that seniors evaluated Billie ($F_{1,122}$=6.76, P=.01) and Voice Only ($F_{1,122}$=7.74, P=.002) as more realistic than did students (Billie: mean score 3.40, SD 0.83 vs 2.59, SD 0.86; Voice Only: mean score 3.76, SD 1.15 vs 2.76, SD 0.88) (Figure 3). As only an interaction for realism was found, hypotheses H3a, H4a, and H5a were not supported by the current data.

Figure 3. Interaction effect of appearance and age group for perceived realism. The realism scale ranged from 1 (totally disagree) to 5 (totally agree).



Liking of the Agent

As effects of age and appearance on the users' liking of the agent were assumed, a two-way analysis of variance (ANOVA) was performed with appearance and age group as independent variables and liking of the agent as a dependent variable. A significant difference between the age groups was found ($F_{1,129}$=10.71, P=.001, η_p^2=0.081), where seniors liked the agents more in general than did students (mean score 3.01, SD 0.98 vs 2.39, SD 1.03). Thus, hypothesis H1b was supported. However, only a marginally significant effect of appearance ($F_{3,129}$=2.41, P=.07, η_p^2=0.056) and no significant interaction effect ($F_{3,129}$=0.99, P=.40, η_p^2=0.024) emerged. Planned contrasts for the appearance effect revealed an influence of species (t_{126}=−2.71, P=.008, r=0.23). Users reported liking the agent more in the condition where Vince (mean score 2.93, SD 1.12) was presented than in both humanoid conditions (Billie: mean score 2.39, SD 0.92; Character: mean score 2.28, SD 0.87). In addition, interaction effects between age and appearance variables (H3–5b) were hypothesized, but the present findings did not support these hypotheses.

Ease of Use, Perceived Usefulness, and Usage Intention

Using two-way MANOVA with appearance (four factors: Billie, Character, Vince, and Voice Only) and age group (two factors: students and seniors) as independent variables and ease of use, perceived usefulness, and usage intention as dependent variables, it was tested how the age groups differed from each other and how these variables were influenced by the appearances of different agents. Based on Pillai trace, the results indicated significant main effects for age group (V=0.98, $F_{3,120}$=3.86, P=.01) and appearance (V=0.14, $F_{9,366}$=1.98, P=.04) and also a significant interaction effect for both (V=0.16, $F_{9,366}$=2.25, P=.02).

With regard to the univariate tests, both age groups differed in their evaluation of the agent's ease of use ($F_{1,129}$=4.07, P=.002, η_p^2=0.073), with students rating the agent as easier to use as compared with the finding for seniors (Table 3). This result is in line with the assumed hypothesis H2, where students were expected to state higher ease of use values as compared with that for seniors. Usage intention ($F_{1,129}$=0.40, P=.59, η_p^2=0.002) and perceived usefulness ($F_{1,129}$=0.76, P=.39, η_p^2=0.006) did not yield statistically significant effects for age differences. Thus, hypotheses H1c and H1d were not supported by the data, as students and seniors reported the same levels of perceived usefulness and usage intention.

Regarding the univariate tests, no significant differences among appearances in participants' usage intention ($F_{3,129}$=1.20, P=.31, η_p^2=0.029), ease of use ($F_{3,129}$=1.50, P=.22, η_p^2=0.035), and usefulness ($F_{3,129}$=1.49, P=.22, η_p^2=0.035) occurred (Table 4).

Table 3. Usage intention, ease of use, and perceived usefulness evaluations of the different age groups in general.

Variable	Score, mean (SD)		
	Seniors	Students	Overall
Usage intention	2.78 (1.27)	2.54 (1.27)	2.63 (1.27)
Ease of use	3.77 (0.63)	4.17 (0.69)	4.03 (0.69)
Usefulness	2.96 (1.27)	2.80 (1.25)	2.86 (1.25)

Table 4. Usage intention, ease of use, and perceived usefulness evaluation for the different appearances regardless of the age group.

Variable	Score, mean (SD)				
	Billie	Vince	Character	Voice Only	Overall
Usage intention	2.70 (1.25)	2.72 (1.45)	2.33 (1.26)	2.72 (1.10)	2.63 (1.27)
Ease of use	3.81 (0.65)	4.02 (0.72)	4.09 (0.64)	4.19 (0.71)	4.03 (0.69)
Usefulness	2.65 (1.32)	2.98 (1.19)	2.56 (1.27)	3.20 (1.20)	2.86 (1.25)

However, a significant interaction effect of age group and appearance was found for perceived usefulness ($F_{3,129}$=5.06, P=.002, η_p^2=0.111) and usage intention ($F_{3,129}$=3.38, P=.02, η_p^2=0.077) but not for ease of use ($F_{3,129}$=1.96, P=.12, η_p^2=0.046) (Figure 4).

Again, simple effects were used to investigate the interaction effects. For perceived usefulness, seniors evaluated the usefulness of Billie (the cartoon-stylized humanoid agent) higher than did students (mean score 3.61, SD 1.32 vs 2.10, SD 0.97) ($F_{1,122}$=12.42, P=.001).

In line with this finding, seniors reported higher usage intention as compared with students (mean score 3.58, SD 1.20 vs 2.19, SD 0.99) after interaction with Billie ($F_{1,122}$=9.63, P=.002) (Figure 5).

Based on these results, the hypothesis H3b was partly supported, as it was assumed that seniors show greater usage intention for a humanoid agent. However, this was only true for a cartoon-stylized human. No such differences between both target groups were found in the evaluation of embodied and nonembodied agents; therefore, the hypotheses H4b and H5b were rejected.

Figure 4. Interaction effect of appearance and age group for perceived usefulness. The usefulness scale ranged from 1 (totally disagree) to 5 (totally agree).

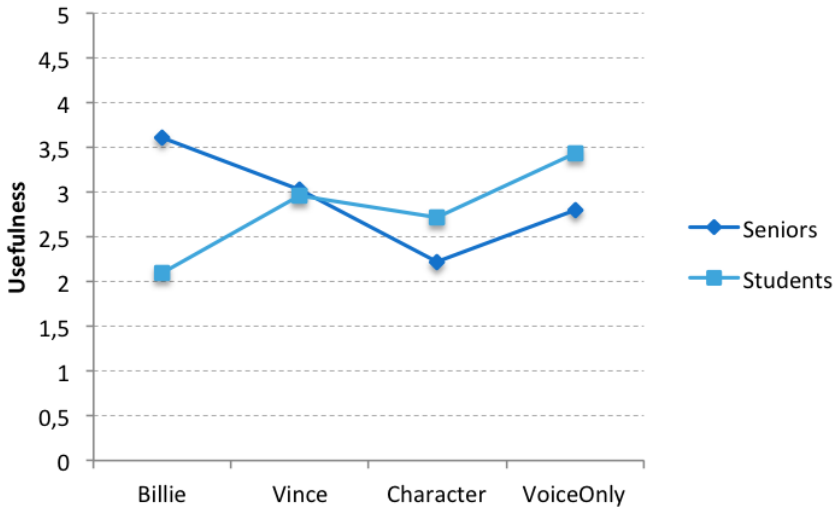
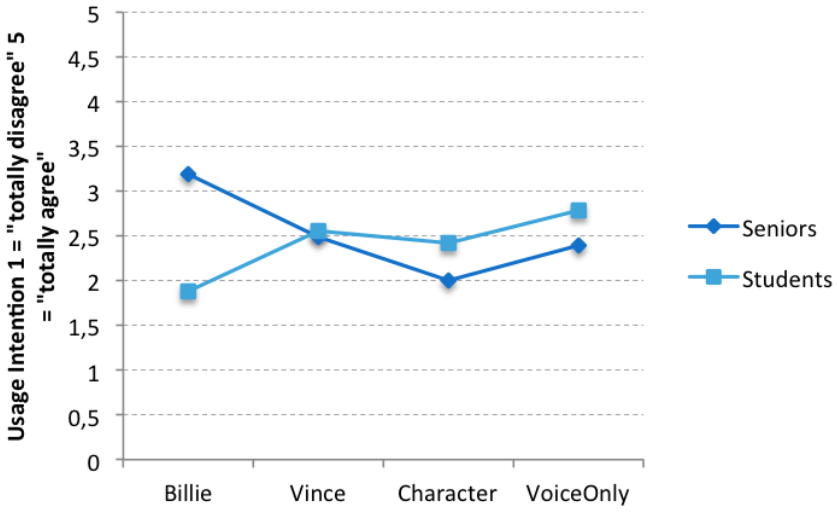


Figure 5. Interaction effect of appearance and age group for usage intention. The usage intention scale ranged from 1 (totally disagree) to 5 (totally agree).



Bonding, Trust, and Sociability

As research question RQ1 aimed to investigate the effects of appearance and age group on bonding and trust, multiple two-way ANOVAs were performed to answer this research question.

The influences of age group and appearance on participants' bonding were analyzed using two-way ANOVA with appearance (four factors: Billie, Character, Vince, and Voice Only) and age group (two factors: students and seniors) as independent variables and bonding as a dependent variable. Analyses revealed a significant main effect for age group ($F_{1,126}$=11.46, P=.001, η_p^2=0.088) and a significant interaction effect for both independent variables ($F_{3,129}$=3.67, P=.01, η_p^2=0.085), whereas the different appearances did not differ in participants' bonding ($F_{3,129}$=1.09, P=.36, η_p^2=0.027). Referring to descriptive values, seniors reported higher feelings of bonding toward the agent than did students (mean score 3.40, SD 0.79 vs 2.94, SD 0.67).

To explore the interaction effect between age group and appearance further, simple effects were used. The results indicated that students and seniors showed different bonding behavior under the Billie ($F_{1,119}$=13.89, P<.001) and Voice Only ($F_{1,122}$=9.18, P=.003) conditions, with seniors showing higher bonding under both conditions as compared with the findings for students (Figure 6 and Table 5).

Figure 6. Interaction effect of appearance and age group for participant bonding. The bonding scale ranged from 1 (totally disagree) to 5 (totally agree).

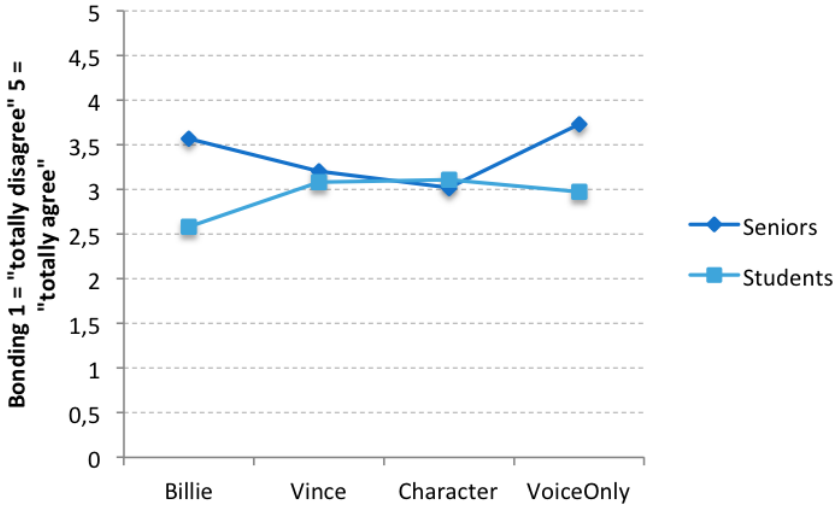
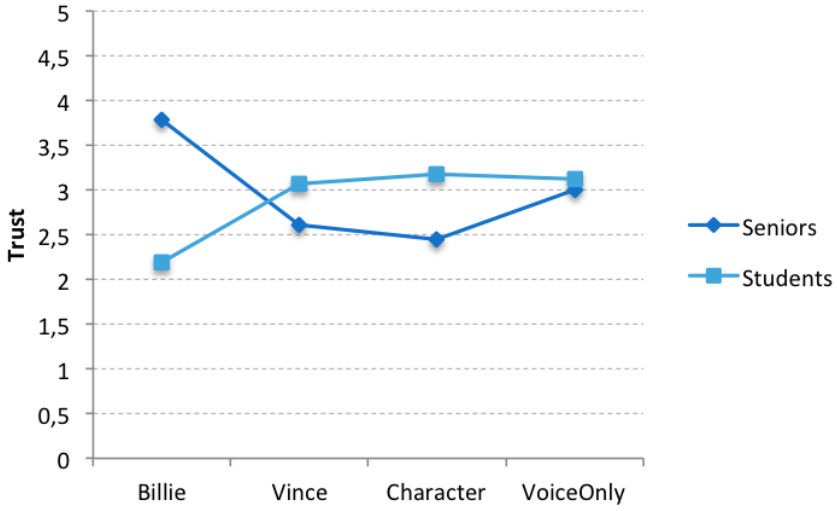


Table 5. Scores for all appearances among the age groups for bonding, trust, and sociability.

Variable	Score, mean (SD)							
	Billie		Vince		Character		Voice Only	
	Seniors	Students	Seniors	Students	Seniors	Students	Seniors	Students
Bonding	3.57 (0.94)	2.58 (0.45)	3.20 (0.67)	3.08 (0.75)	3.02 (0.78)	3.11 (0.76)	3.73 (0.67)	2.98 (0.58)
Trust	3.79 (0.94)	2.19 (1.04)	2.62 (1.14)	3.07 (1.28)	2.44 (1.10)	3.18 (1.16)	3.00 (1.28)	3.12 (1.13)
Sociability	4.13 (0.81)	3.00 (0.74)	4.04 (0.72)	3.81 (0.88)	3.36 (0.99)	3.44 (0.84)	4.38 (0.85)	4.05 (0.83)

Another two-way ANOVA was performed with the same independent variables (age group and appearance) and with trust as a dependent variable. Although no significant main effects of age group ($F_{1,129}$=0.13, P=.72, η_p^2=0.001) and appearance ($F_{3,129}$=0.32, P=.81, η_p^2=0.008) were found for trust, the analyses yielded a significant interaction effect for both variables ($F_{3,129}$=6.19, P=.001, η_p^2=0.132). Simple effects further revealed that seniors showed more trust for Billie than did students ($F_{1,122}$=14.93, P<.001), whereas both age groups did not differ regarding trust for the other appearances (Figure 7 and Table 5).

Figure 7. Interaction effect of appearance and age group for participants' trust toward the agent. The trust scale ranged from 1 (totally disagree) to 5 (totally agree).

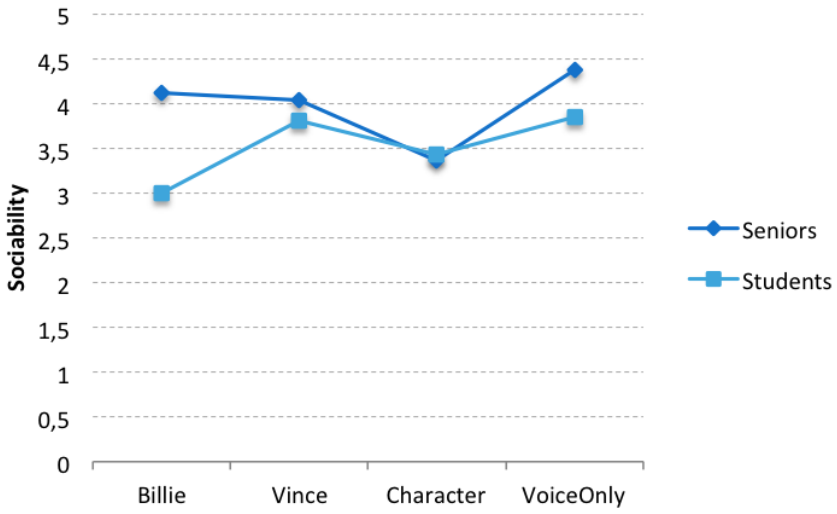


In addition, two-way ANOVA with both independent variables and perceived sociability of the agent was performed to analyze whether the different appearances were perceived as sociable interaction partners and whether the age group had any effect. The analysis revealed a significant difference between age groups ($F_{1,129}$=8.88, P=.003, η_p^2=0.068), where seniors in general evaluated the agent as more sociable than did students (mean score 4.02, SD 0.88 vs 3.53, SD 0.87).

Furthermore, a significant influence of appearance was found ($F_{3,129}$=4.57, P=.005, η_p^2=0.101). The planned contrasts reveal a significant difference in perceived sociability between embodied and nonembodied agents (t_{126}=−2.72, P=.007, r=0.24) and a significant difference between machine-like and humanoid agents (t_{126}=−2.64, P=.009, r=0.23). Participants rated the Voice Only agent as more sociable than the three embodied agents (Voice Only: mean score 4.05, SD 0.83 vs Billie: 3.41, SD 0.93; Vince: 3.89, SD 0.82; and Character: 3.41, SD 0.87). Additionally, Vince (the machine-like agent) was found to evoke more sociability than both humanoid agents (Billie and Character).

In addition, a significant interaction effect for both variables was noted ($F_{3,129}$=2.80, P=.04, η_p^2=0.064). The same pattern that was found before was also obtainable for sociability, as on referring to simple effects, seniors and students differed in their evaluation of Billie ($F_{1,122}$=14.48, P<.001) but not in their evaluation of the other appearances. Again, seniors rated Billie (the cartoon-stylized humanoid agent) as more sociable than did students (Figure 8 and Table 5), whereas no differences between both groups regarding the other appearances were noted.

Figure 8. Interaction effect of appearance and age group for perceived sociability. The sociability scale ranged from 1 (totally disagree) to 5 (totally agree).



Enjoyment

As seniors are generally assumed to perceive the agent and the interaction more positively (H1), hypothesis H1e claims that seniors will enjoy the interaction with the agent more than students. The results of two-way ANOVA with age group and appearance as independent variables and enjoyment of the interaction as a dependent variable revealed a significant age effect ($F_{1,126}$=7.26, P=.008, η_p^2=0.057). In line with prior findings, seniors reported the interaction as more enjoyable than did students (mean score 4.20, SD 0.82 vs 3.72, SD 0.89). Thus, the hypothesis H1e was supported by the present findings. However, the analyses yielded no significant effect for appearance ($F_{3,126}$=1.61, P=.19, η_p^2=0.039) or a significant interaction effect ($F_{3,126}$=2.40, P=.07, η_p^2=0.057).

Discussion

Results Summary and Interpretation

This study aimed to investigate age-related differences in the effects of species, realism, and embodiment on the perception and evaluation of agents in a health-related human–agent interaction, as no previous study has tested the impact of age on the perception and evaluation of appearance variables. To close this research gap, a laboratory study was conducted, in which four different appearances (cartoon-stylized human: Billie, realistic human: Character, cartoon-stylized robot: Vince, and nonembodied voice only condition: Voice Only) were tested in a between-subjects design (N=130) with two different age groups (students and seniors).

According to the findings of Rosenthal-von der Pütten et al [11], hypothesis H1 assumed that seniors will evaluate the agent and the interaction therewith more positively as compared with students. This was partly supported for person perception (H1a) and supported for liking (H1b) and enjoyment (H1e) but not for seniors' usage intention (H1c) and perceived usefulness (H1d). In line with prior findings [11], seniors liked the agent more and enjoyed the interaction more than did students. As students probably have more experiences of interactions with virtual agents and therefore have higher and more specific expectations, it is more difficult to match their expectations and impress them. By contrast, seniors can be assumed to have mostly not interacted with a virtual agent before, and they probably have less frequent points of contact with such technologies. Therefore, they seem to have more appreciation for the technology and its functions. Additionally, the application itself is more beneficial for seniors, who need more support in their daily life. Prior research has already emphasized the influence of experience on the perceived usefulness of the technology and the intention to use it [25]. Although the interaction was designed to be suitable for both groups, the general application of daily life assistance is more adapted to people in need of support. It was therefore assumed that seniors perceive the agent as more useful and show higher usage intentions. However, these assumptions were not confirmed by the data, as no differences between the two groups were found. Potentially, social desirability influenced the seniors' ratings. Qualitative research with elderly participants demonstrated that some elderly individuals are afraid of asking and accepting help and that elderly women especially consider the reception of help as a loss of independence [26]. Seniors (especially women, who mainly participated in this study) might want to hide their potential need for support and maintain the illusion that they do not need any help [26]. As a consequence, they might provide lower usage intention and usefulness ratings than the actual ratings. In line with this, Yaghoubzadeh et al [4] found a third-person effect [27] and reported that elderly individuals perceive virtual assistants as useful for a third person but not themselves. In summary, seniors might be afraid to be perceived as vulnerable, and therefore, they might state lower usefulness and usage intention. Another reason might lie in the fact that the seniors who participated in the study were required to visit the laboratory autonomously, which means that they needed to be mobile and healthy. Therefore, these seniors might not perfectly match the target group in need of support.

Nevertheless, with regard to the identified interaction effects, seniors rated the agent's usefulness and their usage intention

higher than did students when a specific appearance was presented. Seniors appear to perceive a cartoon-stylized humanoid agent as more useful and prefer to use an agent with this appearance.

Furthermore, in line with the hypothesis H2, students perceived the agent as easier to use than did seniors. This might be explained by differences in technical skills between the two groups. Students are mainly described as digital natives, who are highly familiar with the use of technology, whereas seniors are not described as digital natives. In addition, seniors' self-efficacy with regard to these entities might be lower [28]. Users of the agent only need to talk to the agent, and therefore, the actual use of the agent should be equally easy for both groups as no technical skills are needed. Nevertheless, it has been demonstrated that seniors are more skeptical [10] and are assumed to have lower self-efficacy [28], and thus, seniors perceived the agent as less easy to use than did students.

Concerning the interaction effects of appearance variables and user age, several target group-specific differences in participants' evaluations were assumed. Because there is scant prior research testing the preferences of senior users with regard to an agent's appearance, according to the evidence of prior research [13,14], it was assumed that seniors prefer a humanoid appearance over a machine-like appearance (H3). Moreover, with reference to the statements in qualitative interviews [13], where seniors stated that they preferred an embodied agent because an interaction with it is more familiar and that they would like to have an interlocutor who can be addressed (eg, looking in the eyes) during the interaction, it was assumed that seniors evaluate an embodied agent more positively than a nonembodied agent (H4). In parallel, in line with students' interview statements, the opposite was hypothesized for students, as they instead reported preferring a solely speech-based system that is more ubiquitous and not restricted to a specific screen (H5). Overall, the results showed the same pattern for multiple dependent variables. Seniors evaluated Billie (the cartoon-stylized humanoid agent) more positively than did students, whereas no differences between the target groups with regard to the other appearances were noted.

Seniors perceived Billie and Voice Only as more realistic than did students. These findings contradict the intended manipulation of realism, as it was assumed that Character would be perceived as more realistic than Billie. Prior studies have already demonstrated the importance of behavioral realism [29,30]. Accordingly, it might have been that the presented behavior of the agent matched the appearance of Billie the most and this concurrency evoked higher overall realism. However, the agent had the same behavior under all conditions, so this cannot explain the differences between the two target groups. It was shown that seniors rely more on the species of the agent and that realism is not a crucial variable for them [14]. Therefore, the differences in perceived realism might have been positive side effects of the generally more positive perception of Billie. Overall, the manipulation of both humanoid appearances might have been too subtle, so the intended manipulation was not successful. Future studies should address realism more closely.

Nevertheless, it needs to be questioned why no difference in likability or overall liking was found, although several other positive effects of a carton-stylized humanoid agent for seniors were noted. It might further be possible that participants relied more on the interaction and the agent's behavior for their realism evaluation and that, in some way, the interactions involving Billie and Voice Only were perceived as more realistic. However, the interaction itself was designed in a maximally controlled way, where only specific prescripted answers based on a decision tree were chosen by the wizard. Nevertheless, small differences may have necessarily occurred as the agent was asked to respond to the specific answers of the participants. This limitation cannot be ruled out during an actual interaction study, where the agent should be perceived as responsive and relational. The findings cannot support the hypotheses H3a and H3b, as only differences in perceived realism and not in any other variables of the agent's person perception were noted and no interaction effects of the target group and appearance for participants' liking of the agent were found.

In addition, Billie was perceived as more useful and more sociable by seniors and evoked higher usage intentions, trust, and bonding in seniors as compared with students. Hence, the hypothesis H3c was supported, as seniors showed higher usage intention after the interaction with the humanoid agent Billie as compared with students. This assumption can also be extended to effects on social outcomes (RQ1). Seniors rated sociability higher for the cartoon-like humanoid agent Billie and showed more trust toward and bonding with agents with this appearance. These findings might be explained by the user groups' expectations. Seniors might expect a virtual agent not to be photorealistic, whereas students are more used to different forms of virtual characters. Thus, it is possible that the appearance of Billie matched the seniors' expectations the most. As predicted, there were more positive outcomes for a humanoid agent than for a cartoon-stylized one. These results indicate that regarding the application of a virtual agent in the context of daily life assistance for people in need of support, a carton-stylized humanoid appearance might be beneficial, as seniors showed, among other things, higher usage intentions, more trust, and more bonding. In the aforementioned application field, these outcomes might be very helpful, as steady intense usage of the agent is the aim. When the agent's appearance is designed in a way that usage intention increases and when higher feelings of trust and bonding occur, regular usage behavior can be fostered. However, it needs to be acknowledged that the aforementioned results are only applicable under specific circumstances. As the findings resulted from a single short-term interaction in a Wizard-of-Oz setting, only limited deductions can be made for real-world interactions over a longer period. Because of the Wizard-of-Oz setting, the agent responded with higher accuracy than current state-of-the-art agents, which might affect the evaluation and responses toward the agent. Although the quick responses of the wizard might match the expectations participants had of the humanoid agent, when these expectations are not fulfilled in real application, people might lose interest to interact with the agent [14]. Thus, further long-term and field studies are needed to support the current findings and to derive better insights for possible design guidelines. These limitations and other shortcomings will be discussed in the next section.

Limitations and Future Work

Although this study offered valuable insights into appearance effects within a human-agent interaction and particularly highlighted the impact of the target group, some shortcomings need to be discussed.

First, the small sample size and unbalanced distribution of age groups need to be noted. Very few seniors participated in this study, and the age cutoff was rather low for seniors (at least 50 years). Additionally, no age cutoff for students was used (studying at the university was used as a criterion for this group). During the recruitment of participants, it was aimed to maintain a balanced distribution of both groups. However, it was challenging to find elderly volunteers who could participate in a laboratory study, as they needed to be mobile and able to visit the university on their own. Nevertheless, as this study already showed the influence of the target group, studies that investigate this effect more closely with larger and more balanced samples are needed. Furthermore, other target groups that match the application field should be considered.

Moreover, the used appearances and agents had to match certain technical criteria to perform an interaction study. Therefore, the appearances could not be manipulated as systematically as desired, and they have more variance than the presented factors (species, realism, and embodiment). Although stimuli involving similar hair styles and clothing styles were chosen for the humanoid agents, they differed in not only their degree of realism but also other factors (eg, perceived age). In addition, no significant difference with regard to perceived realism was found for the realism contrast. Thus, participants appeared to perceive both humanoid agents as equally realistic. This means that the manipulation targeted needs to be evaluated with caution. Future studies should design appearances that are systematically varied and where other confounding variables are mostly eliminated. Additionally, only a limited number of appearance variables could be addressed in this experiment, and future studies should investigate the effect of other appearance variables and forms. In particular, when it comes to the realism of an agent, the uncanny valley theory [31] needs to be discussed. This research did not aim to investigate the existence of an uncanny valley, and only two stimuli that differed in realism were used. Nevertheless, as age-related differences in the evaluation and responses to different appearance variables were found, future studies should also investigate whether there are age-related differences in the perception of an uncanny valley.

In this study, no static material but an actual interaction with an agent was used. Within this interaction, the agent had behavior related to the participants' responses. Although the interaction was scripted and the agent's answers were limited to a certain set of possible reactions, this behavior might have influenced the interaction and the participants' evaluation of the agent afterwards. As the interaction was aimed to be enjoyable and relatable, the agent's answers had to rely on the prior statements of the participants. Thus, there were minor differences in the interaction and the agent's responses. However, these minor differences cannot be prevented if an actual relational interaction is required. Behavioral realism was not the focus of this study, and therefore, it should be investigated further in the future.

Like most prior studies, this study's results are only based on a single interaction and one point of measurement. However, the envisioned application of autonomous living and health assistance aims for a steady longitudinal usage of a virtual assistant [3]. Initial studies showed that the perception of an agent might change with time [32]. Accordingly, it still needs to be asked how the effects of appearance develop over time. Furthermore, concepts like trust and bonding do evolve over time and are more relevant in long-term interactions. Thus, long-term studies are needed to investigate the development of the presented findings in multiple interactions.

Conclusion

This study investigated the effects of species, realism, and embodiment with regard to age-related differences on a health-related human–agent interaction. Therefore, a between-subjects laboratory study with a Wizard-of-Oz setting was used, where four different appearances and two different age groups were examined. The interaction was embedded in a virtual health assistance scenario, and participants filled in a virtual health diary together with the speech-based agent. The results emphasize the importance of the target group, as age-related differences were found in the general evaluation and in the evaluation of appearances. Seniors showed higher usage intention, trust, and bonding with a humanoid agent having cartoon stylization than did students. The realism of the agent was not found to affect the evaluation. Thus, when a virtual assistant is designed, the target group needs to be determined first. For seniors, a carton-stylized humanoid agent might be more appreciated, as it enhances usage intention and social processes. This is at least true in a short-term interaction and when a flawless interaction is provided. Overall, they appear to have a stronger need for social presence represented by a virtual human, as they are used to it from human-human communication. On the other hand, students appear not to rely on these social cues represented by appearance. This might be caused by their higher experiences with technologies and virtual agents. In summary, a health advisor for seniors should be designed with a humanoid appearance, as this fosters the interpersonal relationship and usage intention of the technology.

Conflicts of Interest

None declared.

References

1. Tsiourti C, Maher BM, Quintas J, Loke B, Jochem I, Albuquerque Lopes J, et al. A Virtual Assistive Companion for Older Adults: Design Implications for a Real-World Application. 2016 Presented at: SAI Intelligent Systems Conference; 2016; London, UK p. 556. [doi: 10.1007/978-3-319-56994-9_69]
2. Cassell J. Embodied conversational interface agents. Commun. ACM 2000;43(4):70-78 [FREE Full text] [doi: 10.1145/332051.332075]
3. Kopp S, Brandt M, Buschmeier H, Cyra K, Freigang F, Krämer NC, et al. Conversational Assistants for Elderly Users – The Importance of Socially Cooperative Dialogue. 2018 Presented at: AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications co-located with the Federated AI Meeting; 2018; Stockholm, Schweden.
4. Yaghoubzadeh R, Kramer M, Pitsch K, Kopp S. Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People. 2013 Presented at: International Conference on Intelligent Virtual Agents; August 29-31, 2013; Edinburgh, UK p. 79-91 URL: https://doi.org/10.1007/978-3-642-40415-3_7
5. Tsiourti C, Joly E, Wings C, Moussa MB, Wac K. Virtual Assistive Companions for Older Adults: Qualitative Field Study and Design Implications. 2014 Presented at: International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2014; Oldenburg, Germany URL: https://doi.org/10.4108/icst.pervasivehealth.2014.254943 [doi: 10.4108/icst.pervasivehealth.2014.254943]
6. Hoffmann L, Krämer NC, Lam-chi A, Kopp S. Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In: Ruttkay Z, Kipp M, Nijholt A, Vilhjálmsson HH, editors. Intelligent Virtual Agents. IVA 2009. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2009:159-165.
7. McDonnell R, Breidt M, Bülthoff HH. Render me real? ACM Trans Graph 2012 Jul 01;31(4):1-11 [FREE Full text] [doi: 10.1145/2185520.2185587]
8. Ring L, Utami D, Bickmore TW. The Right Agent for the Job?? The Effects of Agent Visual Appearance on Task Domain. 2014 Presented at: International Conference on Intelligent Virtual Agents; August 27-29 2014; Boston, MA, USA p. 374-384 URL: https://doi.org/10.1007/978-3-319-09767-1_49
9. van Wissen A, Vinkers C, van Halteren A. Developing a Virtual Coach for Chronic Patients: A User Study on the Impact of Similarity, Familiarity and Realism. 2016 Presented at: International Conference on Persuasive Technology; April 5-7, 2016; Salzburg, Austria p. 263-275 URL: https://doi.org/10.1007/978-3-319-31510-2_23
10. Scopelliti M, Giuliani MV, Fornara F. Robots in a domestic setting: a psychological approach. Univ Access Inf Soc 2005 Jul 19;4(2):146-155 [FREE Full text] [doi: 10.1007/s10209-005-0118-1]
11. Rosenthal-von der Pütten AM, Straßmann C, Yaghoubzadeh R, Kopp S, Krämer NC. Dominant and submissive nonverbal behavior of virtual agents and its effects on evaluation and negotiation outcome in different age groups. Computers in Human Behavior 2019 Jan;90:397-409 [FREE Full text] [doi: 10.1016/j.chb.2018.08.047]
12. Chattaraman V, Kwon W, Gilbert JE, In Shim S. Virtual agents in e‐commerce: representational characteristics for seniors. Jnl of Res in Interact Mrkting 2011 Oct 18;5(4):276-297 [FREE Full text] [doi: 10.1108/17505931111191492]
13. Straßmann C, Krämer NC. A Categorization of Virtual Agent Appearances and a Qualitative Study on Age-Related User Preferences. 2017 Presented at: International Conference on Intelligent Virtual Agents; 2017; Stockholm, Schweden p. 413-422 URL: https://doi.org/10.1007/978-3-319-67401-8_51
14. Straßmann C, Krämer NC. A Two-Study Approach to Explore the Effect of User Characteristics on Users' Perception and Evaluation of a Virtual Assistant's Appearance. MTI 2018 Oct 02;2(4):66 [FREE Full text] [doi: 10.3390/mti2040066]
15. Bickmore TW, Caruso L, Clough-Gorr K, Heeren T. 'It's just like you talk to a friend' relational agents for older adults. Interacting with Computers 2005 Dec;17(6):711-735 [FREE Full text] [doi: 10.1016/j.intcom.2005.09.002]
16. Ren J, Schulman D, Jack B, Bickmore TW. Bickmore. In: Supporting longitudinal change in many health behaviors. 2014 Presented at: ACM conference on Human factors in computing systems (CHI); 2014; Toronto, Canada p. 1657-1662 URL: https://doi.org/10.1145/2559206.2581201
17. Reeves B, Nass C. The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places [Book Review]. Cambridge: CSLI Publications and Cambridge University Press; 1996.
18. Heerink M, Kröse B, Evers V, Wielinga B. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. Int J of Soc Robotics 2010 Sep 4;2(4):361-375 [FREE Full text] [doi: 10.1007/s12369-010-0068-5]
19. Horvath AO, Greenberg LS. The development of the Working Alliance Inventory. In: Greenberg LS, Pinsoff W, editors. The psychotherapeutic process: A research handbook. New York, NY: Guildford Press; 1986:529-556.
20. Moorman C, Matulich E. A Model of Consumers' Preventive Health Behaviors: The Role of Health Motivation and Health Ability. Journal of Consumer Research 1993;20(2):208-228 [FREE Full text]
21. Cunningham GB, Kwon H. The Theory of Planned Behaviour and Intentions to Attend a Sport Event. Sport Management Review 2003 Nov;6(2):127-145. [doi: 10.1016/S1441-3523(03)70056-4]
22. Neave N, Jackson R, Saxton T, Hönekopp J. The influence of anthropomorphic tendencies on human hoarding behaviours. Personality and Individual Differences 2015 Jan;72:214-219 [FREE Full text] [doi: 10.1016/j.paid.2014.08.041]
23. Nomura T, Kanda T, Suzuki T. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. AI & Soc 2005 Aug 26;20(2):138-150 [FREE Full text] [doi: 10.1007/s00146-005-0012-7]
24. Andy F. GLM 3: Factorial Designs. In: Andy Field - Discovering Statistics Using SPSS. Thousand Oaks, CA: Sage Publications; 2018:607-648.

25. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204 [FREE Full text] [doi: 10.1287/mnsc.46.2.186.11926]
26. Roe B, Whattam M, Young H, Dimond M. Elders' needs and experiences of receiving formal and informal care for their activities of daily living. J Clin Nurs 2001 May;10(3):389-397 [FREE Full text] [doi: 10.1046/j.1365-2702.2001.00485.x] [Medline: 11820549]
27. Davison WP. The Third-Person Effect in Communication. Public Opinion Quarterly 1983;47(1):1-15. [doi: 10.1086/268763]
28. Rosenthal-von der Pütten A, Bock N, Brockmann K. Not Your Cup of Tea?? How Interacting With a Robot Can Increase Perceived Self-efficacy in HRI and Evaluation. 2017 Presented at: ACM/IEEE International Conference on Human-Robot Interaction - HRI; 6-9 March 2017; Vienna, Austria URL: https://doi.org/10.1145/2909824.3020251
29. Bailenson JN, Swinth K, Hoyt C, Persky S, Dimov A, Blascovich J. The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. Presence: Teleoperators and Virtual Environments 2005 Aug;14(4):379-393 [FREE Full text] [doi: 10.1162/105474605774785235]
30. von der Pütten AM, Krämer NC, Gratch J, Kang S. "It doesn't matter what you are!" Explaining social effects of agents and avatars. Computers in Human Behavior 2010 Nov;26(6):1641-1650 [FREE Full text] [doi: 10.1016/j.chb.2010.06.012]
31. Mori M, MacDorman K, Kageki N. The Uncanny Valley [From the Field]. IEEE Robot. Automat. Mag 2012 Jun;19(2):98-100 [FREE Full text] [doi: 10.1109/mra.2012.2192811]
32. Bergmann K, Eyssel F, Kopp S. A Second Chance to Make a First Impression? How Appearance and Nonverbal Behavior Affect Perceived Warmth and Competence of Virtual Agents over Time. In: Nakano Y, Neff M, Paiva A, Walker M, editors. Intelligent Virtual Agents. IVA 2012. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2012:126-138.

Abbreviations

ANOVA: analysis of variance
MANOVA: multivariate analysis of variance
RQ: research question

Edited by C Eichenberg; submitted 15.02.19; peer-reviewed by H Williamson, C Tsiourti; comments to author 26.03.19; revised version received 21.05.19; accepted 06.06.19; published 23.04.20.

Please cite as:
Straßmann C, Krämer NC, Buschmeier H, Kopp S
Age-Related Differences in the Evaluation of a Virtual Health Agent's Appearance and Embodiment in a Health-Related Interaction: Experimental Lab Study
J Med Internet Res 2020;22(4):e13726
URL: http://www.jmir.org/2020/4/e13726/
doi:10.2196/13726
PMID:32324146

Letter to the Editor

Comment on "Feasibility of a New Cuffless Device for Ambulatory Blood Pressure Measurement in Patients With Hypertension: Mixed Methods Study"

Noud van Helmond[1], MD; Wim J Jongejan, MD; Robert A Hirsh[1], MD

[1]Department of Anesthesiology, Cooper Medical School of Rowan University, Cooper University Health Care, Camden, NJ, United States

Corresponding Author:
Noud van Helmond, MD
Department of Anesthesiology
Cooper Medical School of Rowan University
Cooper University Health Care
One Cooper Plaza
2nd Floor Dorrance Bldg Suite# D206
Camden, NJ, 08103
United States
Phone: 1 8569687336
Email: vanhelmond-noud@cooperhealth.edu

Related Articles:

Comment on: http://www.jmir.org/2019/6/e11164/

Comment in: http://www.jmir.org/2020/4/e16205/

(J Med Internet Res 2020;22(4):e15952) doi:10.2196/15952

KEYWORDS

blood pressure measurement; cuffless; hypertension

We read the article from Ogink and colleagues [1] describing the use of cuffless blood pressure (BP) measurement for home BP measurement. We applaud the authors' intention to study the relevance of cuffless BP measurement in real-world conditions and have a few comments and questions that may improve our understanding of this study's results.

First, we previously commented [2] on a prior study performed by the same group [3]. We highlighted problems with the Checkme's accuracy and its marketing in the United States as a systolic BP monitor without the required Food and Drug Administration approval. We are surprised that the authors refer to the prior study performed by Schoot et al [3] as promising in terms of European Society of Hypertension accuracy standards and by their repeated erroneous claim that the Checkme has Food and Drug Administration approval for measurements of systolic BP [1], after acknowledging the Checkme's shortcomings related to accuracy and regulatory approval in a response to our letter [4]. It is important to note that the referenced CE (Conformitè Europëenne) certification constitutes conformity with electromagnetic safety standards but is not a certification for demonstrated accuracy and precision of systolic BP measurements.

Second, in the referenced study, the weak correlation (R=0.47) between paired home cuff systolic BP measurements and Checkme measurements, as well as a large absolute difference between these measurements (eg, 44% measurements differed by >10 mm Hg) confirms the inaccuracy of individual Checkme systolic BP measurements. Over the course of 3 weeks, the average of all twice-daily, duplicate, Checkme measurements (84 total measurements in each subject) correlated better (R=0.75) with the average of all once-weekly, duplicate home cuff BP measurements over 3 weeks (6 total measurements in each subject), but these means still varied by 5-15 mm Hg for 64% of measurements. The authors suggest that the average of a large number of Checkme measurements can be used by physicians to adjust medication. Considering that the mean treatment response of antihypertensive medications lies in the 5-15 mm Hg range for systolic BP [5], how do the authors envision Checkme measurements to be used in clinical care?

Finally, a major reported advantage of the Checkme is its user-friendliness as assessed by semistructured interviews. We note that individuals were excluded if they were not able to perform the Checkme measurement correctly after 20-40 minutes of instruction. Can the authors please share how many

patients were not enrolled after failure to perform the measurement after instruction? Further, we wonder if the required measurement frequency of Checkme as compared to conventional measurements (14 times as much) was assessed in the structured interviews?

Conflicts of Interest

NvH reports being an inventor on pending patent applications related to vital sign measurement. RAH is the inventor on an issued patent related to vital sign measurement. WJJ reports no disclosures or conflicts of interest.

References

1. Ogink PAM, de Jong JM, Koeneman M, Weenk M, Engelen LJLPG, van Goor H, et al. Feasibility of a New Cuffless Device for Ambulatory Blood Pressure Measurement in Patients With Hypertension: Mixed Methods Study. J Med Internet Res 2019 Jun 19;21(6):e11164 [FREE Full text] [doi: 10.2196/11164] [Medline: 31219050]
2. van Helmond N, Plante TB, Joseph JI. Blood Pressure Measurement Validation Off the Cuff? Comment on "A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study". J Med Internet Res 2018 Oct 22;20(10):e10089 [FREE Full text] [doi: 10.2196/10089] [Medline: 30348627]
3. Schoot TS, Weenk M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study. J Med Internet Res 2016 May 05;18(5):e85 [FREE Full text] [doi: 10.2196/jmir.5414] [Medline: 27150527]
4. Bredie S, van de Belt T, van Goor H. Authors' Reply: Comment on "A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study". J Med Internet Res 2018 Oct 22;20(10):e12477 [FREE Full text] [doi: 10.2196/12477] [Medline: 30348625]
5. Paz MA, de-La-Sierra A, Sáez M, Barceló MA, Rodríguez JJ, Castro S, et al. Treatment efficacy of anti-hypertensive drugs in monotherapy or combination: ATOM systematic review and meta-analysis of randomized clinical trials according to PRISMA statement. Medicine (Baltimore) 2016 Jul;95(30):e4071 [FREE Full text] [doi: 10.1097/MD.0000000000004071] [Medline: 27472680]

Abbreviations

BP: blood pressure
CE: Conformitè Europëenne

Edited by G Eysenbach; submitted 21.08.19; peer-reviewed by S Bredie; accepted 28.09.19; published 22.04.20.

Please cite as:
van Helmond N, Jongejan WJ, Hirsh RA
Comment on "Feasibility of a New Cuffless Device for Ambulatory Blood Pressure Measurement in Patients With Hypertension: Mixed Methods Study"
J Med Internet Res 2020;22(4):e15952
URL: https://www.jmir.org/2020/4/e15952
doi:10.2196/15952
PMID:32319964

Letter to the Editor

Authors' Reply to: Comment on "Feasibility of a New Cuffless Device for Ambulatory Blood Pressure Measurement in Patients With Hypertension: Mixed Methods Study"

Sebastian J H Bredie[1], MD, PhD; Jelske M de Jong[2], MSc; Tom H van de Belt[2], PhD; Harry van Goor[3], MD, PhD

[1]Department of Internal Medicine, Radboud University Medical Center, Nijmegen, Netherlands

[2]REshape Center for Innovation, Radboud University Medical Center, Nijmegen, Netherlands

[3]Department of Surgery, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Sebastian J H Bredie, MD, PhD
Department of Internal Medicine
Radboud University Medical Center
PO Box 9101
Nijmegen, 6500 HB
Netherlands
Phone: 31 243618819
Email: bas.bredie@radboudumc.nl

Related Articles:

Comment on: http://www.jmir.org/2019/6/e11164/

Comment on: http://www.jmir.org/2020/4/e15952/

(J Med Internet Res 2020;22(4):e16205) doi:10.2196/16205

KEYWORDS

cuffless blood pressure; pocket devices; ambulant blood pressure

We thank Noud van Helmond and colleagues [1] for the critical assessment of our study results.

In a number of studies, we have recently evaluated the possible value of a handy device (Viatom's Checkme) for measuring multiple vital parameters, including cuffless blood pressure (BP) determination.

The Checkme device entered the medical domain after it was originally designed for the consumer market. This makes it very interesting and necessary to scientifically investigate its use in patients.

After a comparison with a common BP monitor [2] and an evaluation of the self-assessment results by admitted patients [3], we recently reported the results of its use in ambulatory BP measurement [4]. In all these studies, both quantitative and qualitative aspects of the use of the Checkme were scientifically assessed.

We are aware of the questions about validity and certification raised by van Helmond et al [1], and we are pleased that through this platform, we can discuss the issues that we have already covered extensively in our manuscripts. Regarding the validity of Checkme's systolic BP results, we stated, as discussed extensively in our previous comment on their letter [5], that as long as there is no adequate validation protocol specifically for cuffless BP monitors, a formal validation study in accordance with leading protocols is impossible. Thus, in its current form, it is too early to implement a device such as Checkme in daily practice. We found that a real-life comparison currently gives the best insight into the potential value. In their study of both of a smartwatch and a portable health device (Checkme), van Helmond et al [6] concluded that the Bodimetrics device was more accurate, possibly due to calibration immediately prior to the study. However, the BP device still failed to meet the accuracy guidelines of the Association for the Advancement of Medical Instrumentation validation protocol, from which van Helmond et al derived their investigation. This protocol assumes that a device should actually be capable of measuring BP without an initial calibration reference measurement. This is peculiar, since it is precisely for the use of cuffless BP monitors that a validation measurement with a traditional BP monitor is required (for estimation of vascular compliance using pulse-oxymetry and electrocardiogram). Only then can an estimate of the BP with these two signals be made. The argument that accuracy improves after a validation measurement taken shortly before a cuffless measurement [6] is therefore not

valid. Precisely, the choice of reference BP monitors and conditions under which measurement is to be made are not included in the current validation protocols and are the reason that regulatory authorities such as the US Food and Drug Administration could not release Checkme for BP measurement.

With regard to the interpretation of the accuracy of the absolute BP values in our home measurement study, we disagree with van Helmond et al [1]. The clinical practice of treating hypertension is increasingly based on home measurement of BP. Here, measurement variation due to patient and environmental factors is taken for granted by the practitioners, since titration of the treatment based on these home measurements ultimately has a better clinical outcome than treatment based on office measurements. Although the Checkme is user-friendly, disruptive factors such as those found in home measurements cannot be excluded. A comparison of home measurement with different devices illustrates this phenomenon and will never have a strong agreement. That is not our message either. It is all about obtaining many measurement results in order to titrate medical treatment. With only a few reports available, including the study by Schoot et al [2], it is too early to promote cuffless measurements on a large scale. However, devices that make use of this technique do appear to be useful at present in, for example, outpatient BP measurement.

Finally, we agree with van Helmond et al [1] that a major advantage of the Checkme is its user-friendliness. All participants were able to take the measurement (12 participants in total). As mentioned in the article, one participant was excluded because the calibration of Checkme did not succeed. The cause for this is unknown, but the measurement itself was performed correctly. Calibration failure was, however, observed in about 10% of the participants in our previous studies [2-4], and this may be a reason for not using Checkme. The cause for this issue is mostly unclear. If a patient could use the Checkme device, it became clear that he or she started measuring BP much more often and that this self-assessment was a positive point raised in the interviews.

We hope that our joint effort to scientifically examine these new devices leads to optimization of self-monitoring technology and eventually to improved patient care.

Conflicts of Interest

None declared.

References

1. van Helmond N, Jongejan WJ, Hirsh RA. Comment on "Feasibility of a New Cuffless Device for Ambulatory Blood Pressure Measurement in Patients with Hypertension: Mixed Methods Study". J Med Internet Res 2020 Apr;22(4):e15952 [FREE Full text] [doi: 10.2196/15952]
2. Schoot TS, Weenk M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study. J Med Internet Res 2016 May 05;18(5):e85 [FREE Full text] [doi: 10.2196/jmir.5414] [Medline: 27150527]
3. Weenk M, van Goor H, van Acht M, Engelen LJLPG, van de Belt TH, Bredie SJH. A smart all-in-one device to measure vital signs in admitted patients. PLoS One 2018 Feb 12;13(2):e0190138 [FREE Full text] [doi: 10.1371/journal.pone.0190138] [Medline: 29432461]
4. Ogink PA, de Jong JM, Koeneman M, Weenk M, Engelen LJLPG, van Goor H, et al. Feasibility of a New Cuffless Device for Ambulatory Blood Pressure Measurement in Patients With Hypertension: Mixed Methods Study. J Med Internet Res 2019 Jun 19;21(6):e11164 [FREE Full text] [doi: 10.2196/11164] [Medline: 31219050]
5. Bredie SJH, van de Belt T, van Goor H. Authors' Reply: Comment on "A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study". J Med Internet Res 2018 Oct 22;20(10):e12477 [FREE Full text] [doi: 10.2196/12477] [Medline: 30348625]
6. van Helmond N, Freeman CG, Hahnen C, Haldar N, Hamati JN, Bard DM, et al. The accuracy of blood pressure measurement by a smartwatch and a portable health device. Hosp Pract (1995) 2019 Oct;47(4):211-215. [doi: 10.1080/21548331.2019.1656991] [Medline: 31423912]

Abbreviations

BP: blood pressure

Edited by G Eysenbach; submitted 10.09.19; this is a non–peer-reviewed article;accepted 28.09.19; published 22.04.20.

Please cite as:
Bredie SJH, de Jong JM, van de Belt TH, van Goor H
Authors' Reply to: Comment on "Feasibility of a New Cuffless Device for Ambulatory Blood Pressure Measurement in Patients With Hypertension: Mixed Methods Study"
J Med Internet Res 2020;22(4):e16205
URL: https://www.jmir.org/2020/4/e16205
doi:10.2196/16205
PMID:32319954

Corrigenda and Addenda

Table Correction: 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:

Correction of: https://www.jmir.org/2019/5/e13599/

(J Med Internet Res 2020;22(4):e18032) doi:10.2196/18032

In the original published paper "Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention" (J Med Internet Res 2019; 21(5):e13599), the authors noticed an error in Table 5. This was caused by a coding error on the authors' electronic survey. The original table listed the 10-item Center for Epidemiological Studies Depression Scale (CESD-10) score for the time point "Post session 4" lower than it should have been.

The score was initially listed as:

> *Mean 9.4, SD 6.8 (row 1, column 3), P=.84 (row 1, column 5).*

The correct score is:

> *Mean 11.1, SD 4.7 (row 1, column 3), P=.88 (row 1, column 5).*

The changes were not significant, and do not affect the overall findings of the paper.

The correction will appear in the online version of the paper on the JMIR website on April 28, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 29.01.20; this is a non–peer-reviewed article;accepted 17.02.20; published 28.04.20.

Please cite as:
Lindauer A, McKenzie G, LaFazia D, McNeill L, Mincks K, Spoden N, Myers M, Mattek N, Teri LL
Table Correction: Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention
J Med Internet Res 2020;22(4):e18032
URL: https://www.jmir.org/2020/4/e18032
doi:10.2196/18032
PMID:32343671

Corrigenda and Addenda

Correction: 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:

Correction of: https://www.jmir.org/2019/5/e13876/

(J Med Internet Res 2020;22(4):e18639) doi:10.2196/18639

In "OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes: (J Med Internet Res 2019;21(5):e13876), there were errors which were not identified during the proofing stage.

The original published Methods did not include the following sentence about the excluded respondents:

> *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.*

The new sentence has now been added to Methods of the paper under the subheading "Statistical Analysis", as follows:

> *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.*

Additionally, the original published Results incorrectly included the following sentence under the subheading "Accessing and Reading Notes":

> *After excluding those with unconfirmed note reading status, 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.*

This sentence has been corrected to the following:

> *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.

Finally, in the flowchart in Figure 1, the bottom left box included the following incorrect wording:

5,072 Discordant information about note reading status

4,443 Portal data conflicts with self-report of reading

629 Do not remember reading notes in the past 12 months, or missing

This wording has been corrected to the following:

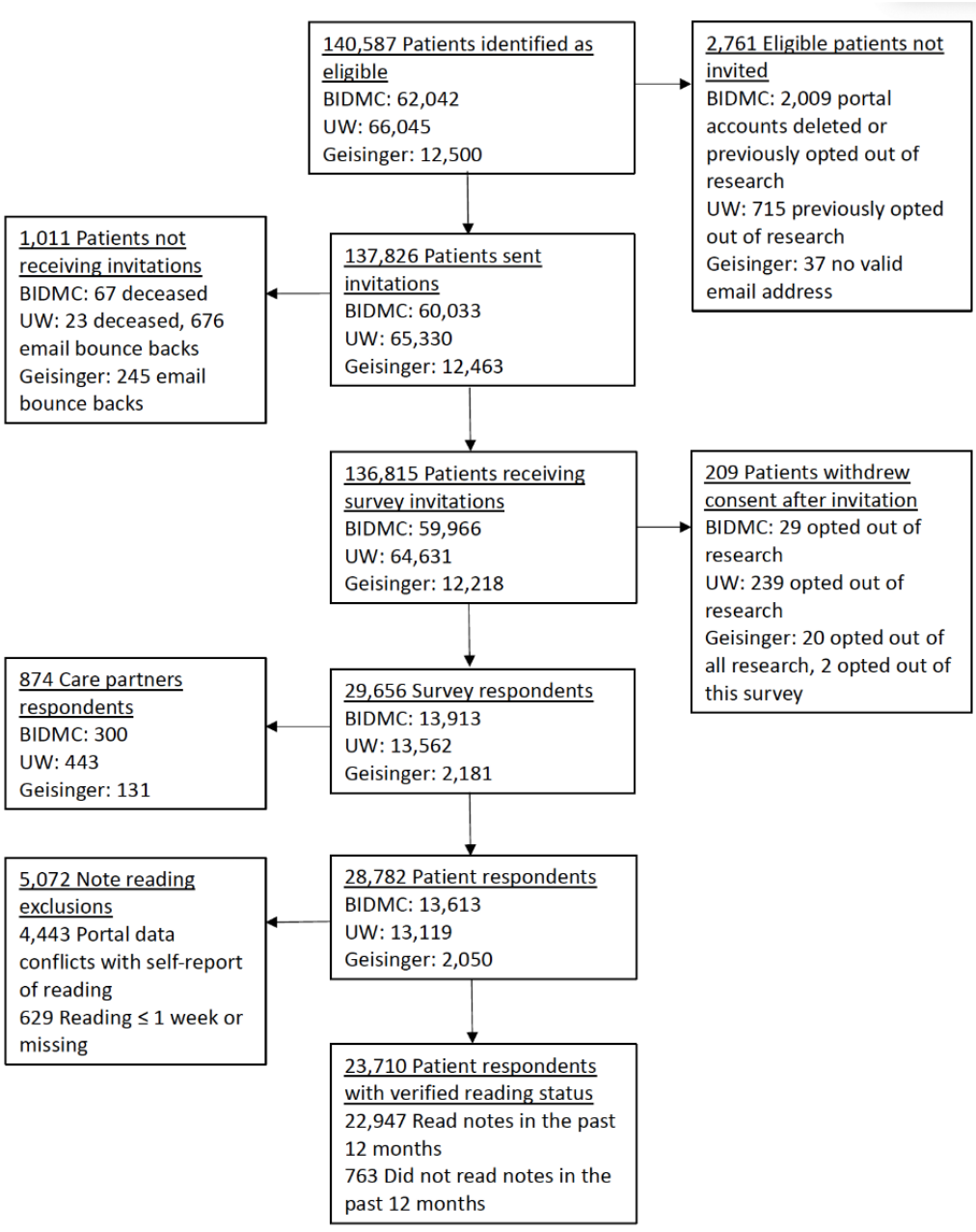
5,072 Note reading exclusions

4,443 Portal data conflicts with self-report of reading

629 Reading ≤1 week or missing

The corrected Figure 1 can be seen below.

Figure 1. Study flow diagram. BIDMC: Beth Israel Deaconess Medical Center; UW: University of Washington Medicine.



The correction will appear in the online version of the paper on the JMIR website on April 30, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 10.03.20; this is a non–peer-reviewed article;accepted 11.03.20; published 30.04.20.

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
Correction: OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes
J Med Internet Res 2020;22(4):e18639
URL: https://www.jmir.org/2020/4/e18639
doi:10.2196/18639
PMID:32352924

Original Paper

Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study

Kai Gong[1], MD; Zhong Xu[1], MS; Zhefeng Cai[2], BS; Yuxiu Chen[2], BS; Zhanxiang Wang[1], MD

[1]The Internet Hospital of the First Affiliated Hospital of Xia'men University, Xia'men City, China

[2]Zoenet Health Company Limited, Xia'men City, China

Corresponding Author:
Zhanxiang Wang, MD
The Internet Hospital of the First Affiliated Hospital of Xia'men University
10 Shanggu Road, Siming District
Xia'men City, 361003
China
Phone: 86 13063035888
Email: 493834266@qq.com

Abstract

Background: During the spread of the novel coronavirus disease (COVID-19), internet hospitals in China were engaged with epidemic prevention and control, offering epidemic-related online services and medical support to the public.

Objective: The aim of this study is to explore the role of internet hospitals during the prevention and control of the COVID-19 outbreak in China.

Methods: Online epidemic-related consultations from multicenter internet hospitals in China during the COVID-19 epidemic were collected. The counselees were described and classified into seven type groups. Symptoms were recorded and compared with reported patients with COVID-19. Hypochondriacal suspicion and offline visit motivation were detected within each counselees' group to evaluate the social panic of the epidemic along with the consequent medical-seeking behaviors. The counselees' motivation and the doctors' recommendation for an offline visit were compared. Risk factors affecting the counselees' tendency of hypochondriacal suspicion and offline visit motivation were explored by logistic regression models. The epidemic prevention and control measures based on internet hospitals were listed, and the corresponding effects were discussed.

Results: A total of 4913 consultations were enrolled for analysis with the median age of the counselees at 28 years (IQR 22-33 years). There were 104 (2.12%) healthy counselees, 147 (2.99%) hypochondriacal counselees, 34 (0.69%) exposed counselees, 853 (17.36%) mildly suspicious counselees, 42 (0.85%) moderately suspicious counselees, 3550 (72.26%) highly suspicious counselees, and 183 (3.72%) severely suspicious counselees. A total of 94.20% (n=4628) of counselees had epidemic-related symptoms with a distribution similar to those of COVID-19. The hypochondriacal suspicion (n=2167, 44.11%) was common. The counselees' motivation and the doctors' recommendation for offline visits were inconsistent (P<.001) with a Cohen kappa score of 0.039, indicating improper medical-seeking behaviors. Adult counselees (odds ratio [OR]=1.816, P<.001) with epidemiological exposure (OR 7.568, P<.001), shortness of breath (OR 1.440, P=.001), diarrhea (OR 1.272, P=.04), and unrelated symptoms (OR 1.509, P<.001) were more likely to have hypochondriacal suspicion. Counselees with severe illnesses (OR 2.303, P<.001), fever (OR 1.660, P<.001), epidemiological exposure history (OR 1.440, P=.01), and hypochondriacal suspicion (OR 4.826, P<.001) were more likely to attempt an offline visit. Reattending counselees (OR 0.545, P=.002) were less motivated to go to the offline clinic.

Conclusions: Internet hospitals can serve different types of epidemic counselees, offer essential medical supports to the public during the COVID-19 outbreak, reduce the social panic, promote social distancing, enhance the public's ability of self-protection, correct improper medical-seeking behaviors, reduce the chance of nosocomial cross-infection, and facilitate epidemiological screening, thus, playing an important role on preventing and controlling COVID-19.

(J Med Internet Res 2020;22(4):e18908) doi:10.2196/18908

KEYWORDS

internet hospital; telemedicine; novel coronavirus disease; pandemic; prevention; control; coronavirus; COVID-19; public health; infectious disease

Introduction

From late 2019 to early 2020, an outbreak of novel coronavirus disease (COVID-19) spread throughout China and soon became a global concern [1-3]. The Chinese government had adopted a series of administrative measures to stop the spread of the epidemic [4], including promulgating decrees that required the domestic internet hospitals to vigorously carry out remote medical services in response to the epidemic [5]. Under such circumstances, the internet hospitals in China, which is a new approach to outpatient health care that provides health services via internet technologies [6], were engaged in the epidemic prevention and control, opening up free COVID-19 consultation services as the main form of remote medical services for the public during the epidemic. This is the first time that internet hospitals have been involved in the response to an infectious public health incident. Its role during the epidemic prevention and control has yet to be explored. This study analyzed the details of the free epidemic consultations from multicenter internet hospitals in China during the COVID-19 outbreak. Through user profiling, we assessed the social panic and the public's medical needs during the COVID-19 outbreak, revealed the effects of internet hospitals on the epidemic prevention and control, and expounded and explored the managing strategies to make the internet hospitals play a greater role in the infectious public health emergency responses.

Methods

Data Sources

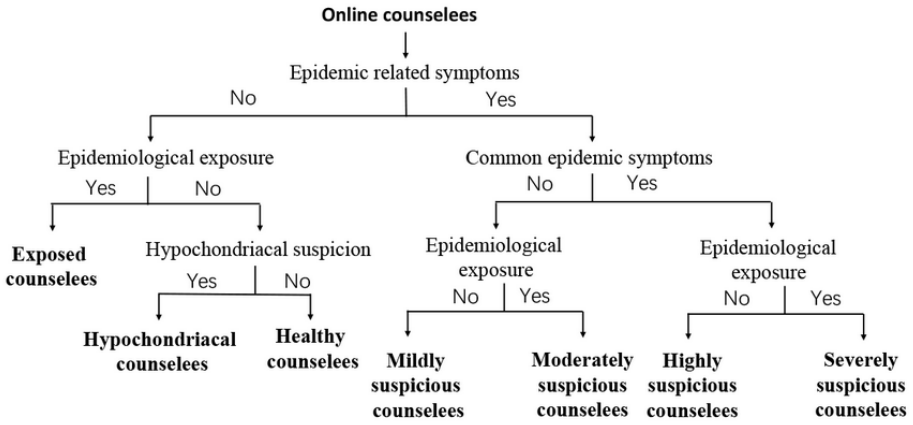
We collected 8913 consecutive deidentified free online consultations generated between January 25, 2020, and February 25, 2020, from 30 general public internet hospitals in 11 provinces of China outside of the Hubei area. The consultants were certified doctors from the general public hospitals, and the counselees were local residents who were not admitted to the offline hospitals and supposed to have epidemic-related questions. All of the data were extracted from the platform of Zoenet Health Company Limited [7], which cooperates with public hospitals to run the online medical services. A total of 405 invalid consultations containing duplicated or nonmedical contents were removed. Through semantic analysis, 563 repetitive and 3032 unrelated consultations, in which the counselees had no related symptoms and asked questions irrelevant to the epidemic, were also excluded. Each of the remaining consultations were analyzed artificially by one trained researcher and rechecked by another. Any divergence from this procedure was resolved by discussion with all researchers.

Study Definitions

Variables including age, sex, symptoms, reattendance, epidemiological exposure history, hypochondriacal suspicion, offline visit recommendation, and offline visit motivation were recorded. The epidemic-related symptoms were classified into common and uncommon epidemic symptoms. Fever (axillary temperature of 37.5°C or higher), cough, expectoration, myalgia, and fatigue were classified into common symptoms from which most of the patients with COVID-19 outside of Wuhan in China suffered from as described by Xiao-Wei Xu et al [8]. Symptoms including mild fever (axillary temperature between 37°C and 37.5°C), nasal congestion, headache, sore throat, shortness of breath, diarrhea, chills, nausea, and vomiting were recognized as uncommon epidemic symptoms that may potentially be caused by COVID-19 [9-11]. Other symptoms (eg, palpitation, dizziness, unexplained abdominal pain, eye discomfort) were also recorded and categorized as unrelated symptoms. Reattendance means the counselees had once consulted doctors either through online or offline approaches before the current consultation. Epidemiological exposure refers to the history of travel or residence in Wuhan and surrounding areas within the Hubei province or other communities with case reports, as well as the history of contact with patients with COVID-19 or contact with people with epidemic-related symptoms from Wuhan and surrounding areas or from communities with case reports. All of the above epidemiological exposure should happen within 14 days before the onset of illness. Hypochondriacal suspicion is whether the counselees had clearly expressed their concern of being potentially infected by the novel coronavirus. Offline visit recommendation means whether the online doctors had suggested an offline consultation. The counselees who had received offline visit recommendations were considered as being in severe condition. Offline visit motivation means whether the counselees had expressed their attempts to go to the offline clinics.

We classified the counselees into seven type groups, including healthy counselees, hypochondriacal counselees, exposed counselees, mildly suspicious counselees, moderately suspicious counselees, highly suspicious counselees, and severely suspicious counselees (Figure 1). The healthy counselees had neither epidemic-related symptoms nor any epidemiological exposure history, simply asking epidemic-related questions without hypochondriacal suspicion. The hypochondriacal counselees had hypochondriacal suspicion without any epidemic exposure nor any epidemic-related symptoms. The exposed counselees had epidemiological exposure history without epidemic-related symptoms. The mildly suspicious counselees had certain uncommon epidemic symptoms without any common epidemic symptoms nor epidemiological exposure history. The moderately suspicious counselees had both uncommon epidemic symptoms and epidemiological exposure history without any common epidemic symptoms. The highly suspicious counselees had common epidemic symptoms without epidemiological exposure history. The severely suspicious counselees had both common epidemic symptoms and epidemiological exposure history.

Figure 1. Classification of different counselees.



Statistical Analysis

The amount and the percentage of counselees with positive hypochondriacal suspicion, offline visit motivation, and an offline visit recommendation were counted and calculated within different counselee groups. All symptoms were extracted and compared with patients with reported COVID-19 [11]. Categorical variables were compared using the chi-square test. The Cohen kappa score was calculated to assess the consistency of the counselees' motivation and the doctors' recommendation for offline visits.

Univariate and multivariate logistical regression analyses were conducted to predict the risk factors for hypochondriacal suspicion and offline visit motivation. Predictors that were statistically significant in the univariate analyses (P<.05) were included in multivariate logistic regression models, with odds ratios and CIs calculated. Epidemiological exposure history, adulthood, sex, severe illness condition, reattendance, and symptoms were considered as the predictors of hypochondriacal suspicion; all of the previously mentioned predictors, with the addition of hypochondriacal suspicion, were considered as the predictors of offline visit motivation.

Statistical analysis was done by Python (version 3.6; Python Software Foundation). The forest figure showing the results of logistic regressions was drawn by R (version 3.6.2; R Foundation for Statistical Computing). P<.05 was considered statistically significant.

Ethics Statement

Ethic approval was obtained from the medical research ethics committee of the first affiliated hospital of Xiamen University, Xiamen, China (protocol number 3502Z2020YJ05) before the start of the study.

Results

Characteristics of the Counselees

A total of 4913 consultations were finally enrolled for analysis including 2031 (41.34%) males and 2882 (58.66%) females. The median age was 28 years (IQR 22-33 years). All children younger than 12 years had online consultations completed by their guardians. Epidemiological exposure history was reported for 259 (5.27%) counselees. Epidemic-related symptoms were reported for 4628 (94.20%) counselees, and 3733 (75.98%) had common-epidemic symptoms. Hypochondriacal suspicion was reported for 2165 (44.07%) counselees, and 869 (17.69%) were motivated to do an offline visit. A total of 190 (3.87%) were in severe condition with an affirmative offline visit recommendation. Only 2 severe cases had no epidemic-related symptoms, but both of them had hypochondriacal suspicion: a 39-year-old male who felt precordial discomfort and a 2-year-old girl who was found in a drowsy state by her parents. Most severe counselees were children less than 10 years old (102/190, 53.7%).

After classification of all the counselees, 104 (2.12%) were healthy counselees, 147 (2.99%) were hypochondriacal counselees, 34 (0.69%) were exposed counselees, 853 (17.36%) were mildly suspicious counselees, 42 (0.85%) were moderately suspicious counselees, 3550 (72.26%) were highly suspicious counselees, and 183 (3.72%) were severely suspicious counselees. Hypochondriacal suspicion and offline-visit motivation were common within different types of counselees. However, fewer counselees had received affirmative offline-visit recommendations (Table 1).

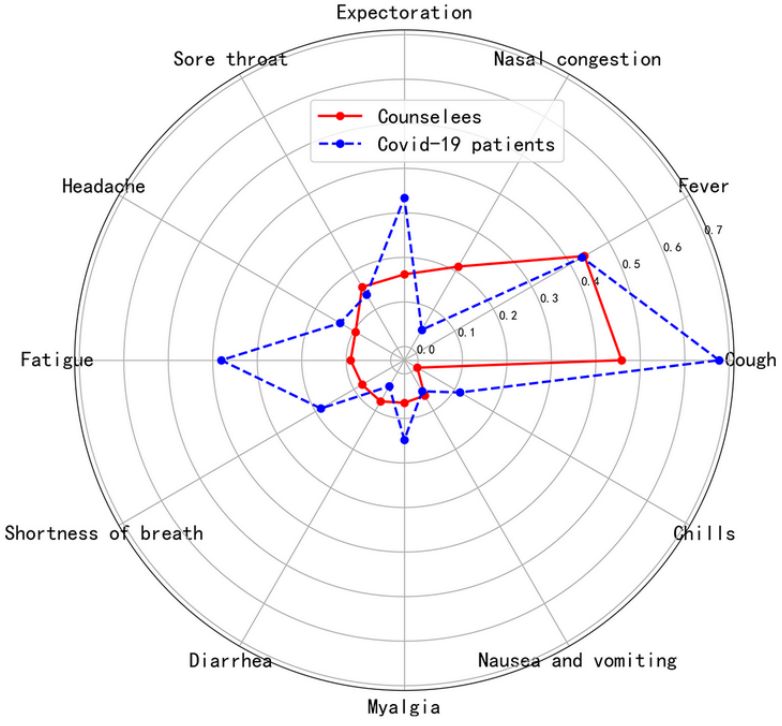
Table 1. Hypochondriacal suspicion, offline visit motivation, and offline visit recommendation in different counselees' type groups.

Counselee type	Hypochondriacal suspicion, n (%)	Offline visit motivation, n (%)	Offline visit recommendation, n (%)
Healthy counselees, n=104	0 (0.0)	3 (2.9)	0 (0.0)
Hypochondriacal counselees, n=147	147 (100.0)	18 (12.2)	2 (1.4)
Exposed counselees, n=34	32 (94.1)	12 (35.3)	0 (0.0)
Mildly suspicious counselees, n=853	419 (49.1)	132 (15.5)	1 (0.1)
Moderately suspicious counselees, n=42	38 (90.5)	18 (42.9)	0 (0.0)
Highly suspicious counselees, n=3550	1378 (38.8)	635 (17.9)	185 (5.2)
Severely suspicious counselees, n=183	151 (82.5)	51 (27.9)	2 (1.1)
Total number, N=4913	2165 (44.07)	869 (17.69)	190 (3.87)

Distribution of the Counselees' Symptoms

For the 4628 counselees with epidemic-related symptoms, cough (n=2118, 45.76%) was the most common symptom, followed by fever (n=2021, 43.67%), nasal congestion (n=981, 21.20%), expectoration (n=752, 16.25%), sore throat (n=735, 15.88%), headache (n=443, 9.57%), fatigue (n=415, 8.97%), shortness of breath (n=365, 7.89%), diarrhea (n=350, 7.56%), myalgia (n=301, 6.50%), nausea and vomiting (n=282, 6.09%), and chills (n=16, 0.35%) (see Figure 2).

Figure 2. The symptom distribution of online counselees and COVID-19 patients. COVID-19: coronavirus disease.



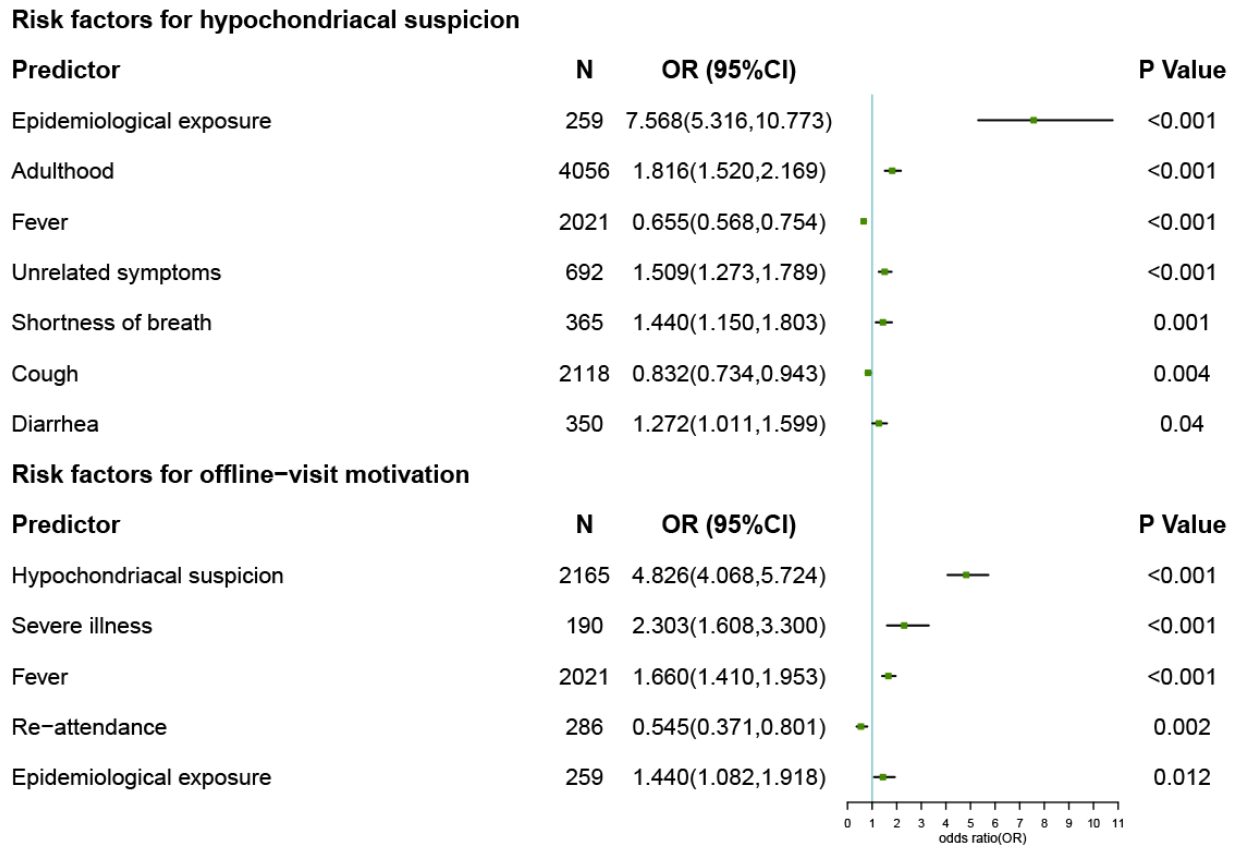
Improper Medical-Seeking Behaviors

The counselees' motivation and the doctors' recommendation for offline visits were significantly different (χ^2_1=13.4230, P<.001) with a Cohen kappa score of 0.039. Of the 190 severe conditions, there were 137 (72.1%) counselees who did not attempt an offline visit for further treatment. Of the 4728 nonsevere cases, 816 (17.26%) counselees with mild conditions were motivated for offline medical care without an affirmative recommendation.

Risk Factors for Hypochondriacal Suspicion and Offline Visit Motivation

The multivariate logistic regression models showed that epidemiological exposure, adulthood, shortness of breath, diarrhea, and unrelated symptoms can independently increase the probability of hypochondriacal suspicion; fever and cough can reduce the probability of hypochondriacal suspicion. Severe illness, fever, epidemiological exposure, and hypochondriacal suspicion can increase the probability of offline visit motivation; however, reattending counselees were less likely to have offline visit motivation (Figure 3).

Figure 3. The independent risk factors for hypochondriacal suspicion and offline visit motivation.



Discussion

Principal Findings

Our results showed that the epidemic of COVID-19 brought panic and hypochondria to the public, further inducing improper health-seeking behaviors and increased demand of medical care services. Along with the arrival of information times, the traditional management style could not adapt to the public's needs during the period of the COVID-19 outbreak. Internet hospitals may serve different types of epidemic counselees, helping prevent and control the epidemic of COVID-19 in China.

Medical Contradictions Brought by COVID-19

As the epidemic of COVID-19 in China overlapped with the high incidence for common cold and seasonal influenza [12], many residents got certain respiratory or other symptoms similar to COVID-19 (Figure 2). A recent large epidemiological study found that physical symptoms (eg, myalgia, dizziness, coryza) and poor self-rated health statuses were significantly associated with a greater psychological impact of the COVID-19 outbreak and higher levels of stress, anxiety, and depression among Chinese individuals [13]. Our results showed that nearly half of the counselees had hypochondriacal suspicion; besides epidemiological exposure, both epidemic-related and unrelated symptoms can exacerbate hypochondria, which can encourage counselees' unnecessary motivation for an offline visit. This is why a considerable proportion of counselees were motivated for an offline visit without the doctor's recommendation. On the contrary, most of the counselees in severe conditions were reluctant to go to an offline clinic for fear of cross-infection. The counselees' motivation was significantly different from the doctors' recommendation, leading to improper health-seeking behaviors, bringing further harm to the public health.

Under the stress of the epidemic, an increasing number of people need professional medical guidance, but the offline hospital visits should be strictly restricted to cut off nosocomial transmission routes. The demanding medical services and the inaccessibility of medical care became one of outstanding contradictions during the COVID-19 outbreak. The internet hospitals may solve this dilemma by offering equitable and inclusive online services to assist the epidemic control [14].

Internet Hospitals in China During the COVID-19 Epidemic

After recent years of development, the internet hospitals in China can now break through the limitations of time and space with excellent accessibility, providing a variety of medical services to all citizens [15]. Through internet connections, interdisciplinary and cross-regional collaborations can be accomplished to improve the capability of dealing with emerging diseases. The Chinese government had encouraged internet hospitals to join the epidemic prevention and control efforts at the beginning of the COVID-19 outbreak [5] and confirmed their role as one important part of the joint epidemic prevention and control system [16]. On March 15, 2020, the first professional standard, "Specification for online consultation service for infectious disease epidemic situation," was published

on the national group standard information platform of China [17], requiring that internet hospitals provide 24/7 online services in response to the epidemic, including prehospital services such as initial screening and medical education; intrahospital services such as offline service appointment and offline visit guidance; and posthospital services such as psychological counseling, post services for medical records, report interpretation for reattending patients, and drug delivery services for patients with chronic diseases. These services should cover all potential medical needs of the public during the epidemic. Meanwhile, the internet hospitals were required to establish intact traceable health files for each counselee and share the data with the supervising departments including the Centers for Disease Control and Prevention (CDC) and local health commissions.

Effects of Internet Hospitals on the Epidemic Prevention and Control

Through the above measures, internet hospitals may help prevent and control the COVID-19 epidemic with both primary and auxiliary functions. In terms of primary functions, first, internet hospitals may reduce the crowd gatherings in offline hospitals through multiple approaches. Through online education and propaganda [18] as well as psychological interventions, the online medical services not only teach the public essential epidemic-protective skills, but also alleviate the social panic and help release the public's hypochondriacal suspicions, thus, reducing the unnecessary offline hospital visits and enhancing psychological resilience [19]. This is consistent with our results that there was less offline visit motivation for the reattending counselees. Meanwhile, the various handy services provided by internet hospitals may reduce the patients' repeated visits. These measures together may effectively reduce people gathering in offline hospitals [20]. Furthermore, internet hospitals run by public hospitals can connect the online services with the offline medical procedures, offering online triage services and guiding the counselees to the corresponding offline departments according to their symptoms [21,22]. This would make the suspected cases walk through isolated channels to the specialized outpatient clinic, thus, reducing people's contact within hospitals and further reducing the chance of getting infected [8]. Furthermore, this protects our medical staff, who are the major force in confronting the epidemic [23], from a massive offline workload and unnecessary occupational exposure [24,25]. Second, the internet hospitals may play a greater role through the integration of online resources and offline epidemic control efforts. As most of the provinces in China have initiated a level-1 public health response to control COVID-19, a joint prevention and control system with online information and an offline screening network was established [26]. By sharing information, internet hospitals can help recognize individuals with a higher probability of being infected. Combined with the offline screening conducted by the epidemiologic investigation organizations and the community offices, online consultations and follow-ups may reduce the risk of missed screenings. Meanwhile, self-isolation [27,28] was required for all symptomatic counselees and epidemiologically close contacts [29]. By identifying these counselees, the internet hospitals can facilitate offline supervision for potentially undocumented cases and promote social distancing [30].

In the aspect of auxiliary functions, internet hospitals provide basic medical support to the public during the epidemic. Affected by the outbreak, numerous nonemergency outpatient departments were closed in Chinese hospitals, causing most offline clinics to be unavailable for the public. Through internet hospitals, the patients could keep in touch with their attending doctors. For those who had new mild symptoms, the doctors could give professional advice on self-management of care and treatment. For those in severe conditions, online doctors may guide them to visit offline hospitals as soon as possible in case of deterioration. Moreover, during the period of individual self-isolation, internet hospitals may join the social capital, improving quality of life by reducing anxiety and stress [31].

Limitations of Internet Hospitals

Fully understanding the positive role played by internet hospitals during the epidemic, we should also realize their limitations. First, online consultation is an indirect way of communication. Due to the lack of information such as physical and auxiliary examinations, online doctors may only give rough medical advice for primary care patients. Second, most of the internet hospitals in China currently offer only passive order-based services to the public. In terms of epidemic screening, it is necessary to cooperate with offline approaches to make better use of the internet hospitals' online advantages. Third, the audience of internet hospitals has not yet covered the whole population due to the difference in public acceptance, which can be reflected by the unbalanced distribution of the counselees' ages. Furthermore, the accessibility of the internet is another limitation of internet hospitals.

Directions for Future Efforts

To make better use of internet hospitals during the epidemic, more efforts are needed, such as recruiting more doctors, especially psychologists [32] and pediatricians, to join the online services while ensuring that each attending doctor has mastered the latest epidemic prevention and control knowledge; discovering the public's needs in a timely manner to adjust our response strategies; improving the usability of the internet hospital apps; strengthening the propaganda to expand the user base; and cooperating with the communities and the CDC to improve joint control and prevention mechanisms [33]. Meanwhile, a more standardized consultation service guideline is needed, enacting different response strategies according to different counselee types. In our study, online education and propaganda were needed for all counselees; self-isolation guidance was needed for all suspicious and exposed counselees; psychological intervention was badly needed for hypochondriacal counselees; and offline visit recommendations were essential when the counselees were in severe conditions. The data of the suspicious counselees and their corresponding risk levels should be highlighted and shared through the joint control and prevention system run by the official agencies for further interventions.

Limitations and Conclusions

This study has some limitations. First, our data was collected outside of the Hubei province. In relatively low-prevalence areas, the hypochondria and panic rate might be underestimated. Meanwhile, the characteristics of the counselees were collected according to their description rather than standardized questionnaires. Some symptoms might be neglected by the counselees and, thus, also underestimated. Second, there are other forms of online consultation services, including paid services held by general public hospitals or private companies [34], which need further estimation. Moreover, the follow-up information after the online consultations were unavailable due to privacy reasons. More longitudinal studies are required to evaluate the counselees' compliance to the online services. Further studies may focus on how to optimize the online services during the epidemic and expand internet hospitals' beneficial influences on the public [35]. As COVID-19 has quickly became a global threat, all countries should consider a combination of response measures with telemedicine platforms involved [36]. According to our results, remote medical services are badly needed for the panicked public. Internet hospitals can make targeted and tailored medical interventions for various types of counselees and help prevent and control the epidemic of COVID-19 in China.

Acknowledgments

This study was funded by the special project for COVID-19 prevention and control of Xiamen city (3502Z2020YJ05).

Authors' Contributions

ZW contributed to the study concept and design. ZC and YC contributed to the acquisition of data. KG and ZX contributed to the semantic and statistical analysis. KG drafted the manuscript. All authors contributed to the interpretation of the results and gave final approval of the manuscript.

Conflicts of Interest

ZC and YC are employees of Zoenet Health Company Limited, which cooperated with public hospitals to offer the free online consultation services.

Multimedia Appendix 1
The complete results of logistic regression models.
[DOC File , 78 KB - jmir_v22i4e18908_app1.doc]

References

1. Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. BMJ 2020 Mar 18;368:m1090. [doi: 10.1136/bmj.m1090] [Medline: 32188598]
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
3. Mahase E. Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. BMJ 2020 Mar 12;368:m1036. [doi: 10.1136/bmj.m1036] [Medline: 32165426]
4. Wu Z, McGoogan JM. haracteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:e. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
5. National Health Commission of the People's Republic of China. 2020 Feb 07. Notice on accomplishing online consultation services in the epidemic prevention and control (Medical Letter 2020 No. 112) URL: http://www.nhc.gov.cn/yzygj/s7652m/202002/32c3e98988894fa18280e4543d2710c7.shtml
6. 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 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
7. Zoenet Health Company. URL: https://www.zoenet.cn/
8. Xu X, Wu X, Jiang X, Xu K, Ying L, Ma C, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ 2020 Feb 19;368:m606 [FREE Full text] [doi: 10.1136/bmj.m606] [Medline: 32075786]
9. Zhang J, Dong X, Cao Y, Yuan Y, Yang Y, Yan Y, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 2020 Feb 19:e. [doi: 10.1111/all.14238] [Medline: 32077115]
10. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513. [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]

11. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Feb 28:e [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
12. Su W, Liu T, Geng X, Yang G. Seasonal pattern of influenza and the association with meteorological factors based on wavelet analysis in Jinan City, Eastern China, 2013-2016. PeerJ 2020;8:e8626. [doi: 10.7717/peerj.8626] [Medline: 32195046]
13. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):e [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
14. Berger ZD, Evans NG, Phelan AL, Silverman RD. Covid-19: control measures must be equitable and inclusive. BMJ 2020 Mar 20;368:m1141. [doi: 10.1136/bmj.m1141] [Medline: 32198146]
15. 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 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
16. The State Council's Comprehensive Team for Joint Prevention and Control of New Coronavirus Pneumonia. 2020 Feb 26. Notice on launching online services to further strengthen epidemic prevention and control in Hubei URL: http://www.gov.cn/xinwen/2020-02/27/content_5483977.htm
17. Zhejiang Digital Economy Association. 2020 Mar 15. Specification for online consultation service for infectious disease epidemic situation URL: http://www.ttbz.org.cn/StandardManage/Detail/33876/
18. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
19. Ho CS, Chee CY, Ho RC. Mental health strategies to combat the psychological impact of COVID-19 beyond paranoia and panic. Ann Acad Med Singapore 2020 Jan;49(1):1-3 [FREE Full text] [Medline: 32200399]
20. Little P, Stuart B, Andreou P, McDermott L, Joseph J, Mullee M, et al. Primary care randomised controlled trial of a tailored interactive website for the self-management of respiratory infections (Internet Doctor). BMJ Open 2016 Apr 20;6(4):e009769 [FREE Full text] [doi: 10.1136/bmjopen-2015-009769] [Medline: 27098821]
21. Gibson A, Randall D, Tran DT, Byrne M, Lawler A, Havard A, et al. Emergency department attendance after telephone triage: a population-based data linkage study. Health Serv Res 2018 Apr;53(2):1137-1162 [FREE Full text] [doi: 10.1111/1475-6773.12692] [Medline: 28369871]
22. Li L, Lake R, Raban MZ, Byrne M, Robinson M, Westbrook J, et al. Medication-related calls received by a national telenursing triage and advice service in Australia: a retrospective cohort study. BMC Health Serv Res 2017 Mar 14;17(1):197 [FREE Full text] [doi: 10.1186/s12913-017-2135-1] [Medline: 28288619]
23. Bowdle A, Munoz-Price LS. Preventing infection of patients and healthcare workers should be the new normal in the era of novel coronavirus epidemics. Anesthesiology 2020 Mar 25:e. [doi: 10.1097/ALN.0000000000003295] [Medline: 32195701]
24. Chang D, Xu H, Rebaza A, Sharma L, Dela Cruz CS. Protecting health-care workers from subclinical coronavirus infection. Lancet Respir Med 2020 Mar;8(3):e13. [doi: 10.1016/S2213-2600(20)30066-7] [Medline: 32061333]
25. Dewey C, Hingle S, Goelz E, Linzer M. Supporting clinicians during the COVID-19 Pandemic. Ann Intern Med 2020 Mar 20:e [FREE Full text] [doi: 10.7326/M20-1033] [Medline: 32196544]
26. Deng S, Peng H. Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. J Clin Med 2020 Feb 20;9(2):e [FREE Full text] [doi: 10.3390/jcm9020575] [Medline: 32093211]
27. Niu Y, Xu F. Deciphering the power of isolation in controlling COVID-19 outbreaks. Lancet Glob Health 2020 Apr;8(4):e452-e453 [FREE Full text] [doi: 10.1016/S2214-109X(20)30085-1] [Medline: 32199105]
28. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
29. Leung C. The difference in the incubation period of 2019 novel coronavirus (SARS-CoV-2) infection between travelers to Hubei and nontravelers: the need for a longer quarantine period. Infect Control Hosp Epidemiol 2020 Mar 18:1-3. [doi: 10.1017/ice.2020.81] [Medline: 32183920]
30. Desai AN, Patel P. Stopping the spread of COVID-19. JAMA 2020 Mar 20:e. [doi: 10.1001/jama.2020.4269] [Medline: 32196079]
31. Xiao H, Zhang Y, Kong D, Li S, Yang N. Social capital and sleep quality in individuals who self-isolated for 14 days during the coronavirus disease 2019 (COVID-19) outbreak in January 2020 in China. Med Sci Monit 2020 Mar 20;26:e923921 [FREE Full text] [doi: 10.12659/MSM.923921] [Medline: 32194290]
32. Lima CKT, Carvalho PMDM, Lima IDAAS, Nunes JVADO, Saraiva JS, de Souza RI, et al. The emotional impact of Coronavirus 2019-nCoV (new Coronavirus disease). Psychiatry Res 2020 Mar 12;287:112915. [doi: 10.1016/j.psychres.2020.112915] [Medline: 32199182]
33. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science 2020 Mar 16:e. [doi: 10.1126/science.abb3221] [Medline: 32179701]

34. Li Y, Yan X, Song X. Provision of paid web-based medical consultation in China: cross-sectional analysis of data from a medical consultation website. J Med Internet Res 2019 Jun 03;21(6):e12126 [FREE Full text] [doi: 10.2196/12126] [Medline: 31162129]
35. Ioannidis JPA. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. Eur J Clin Invest 2020 Mar 19:e13222. [doi: 10.1111/eci.13222] [Medline: 32191341]
36. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, WHO Strategic Technical Advisory Group for Infectious Hazards. COVID-19: towards controlling of a pandemic. Lancet 2020 Mar 28;395(10229):1015-1018. [doi: 10.1016/S0140-6736(20)30673-5] [Medline: 32197103]

Abbreviations

CDC: Centers for Disease Control and Prevention
COVID-19: coronavirus disease
OR: odds ratio

Edited by G Eysenbach; submitted 27.03.20; peer-reviewed by R Ho, MA Bahrami; comments to author 01.04.20; revised version received 02.04.20; accepted 03.04.20; published 14.04.20.

Please cite as:
Gong K, Xu Z, Cai Z, Chen Y, Wang Z
Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study
J Med Internet Res 2020;22(4):e18908
URL: http://www.jmir.org/2020/4/e18908/
doi:10.2196/18908
PMID:32250962

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org