Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 19 (2017), Issue 8 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Original Paper

Key Components in eHealth Interventions Combining Self-Tracking and Persuasive eCoaching to Promote a Healthier Lifestyle: A Scoping Review

Aniek J Lentferink[1,2,3], MSc; Hilbrand KE Oldenhuis[2], PhD; Martijn de Groot[3], PhD; Louis Polstra[2], PhD; Hugo Velthuijsen[2], PhD; Julia EWC van Gemert-Pijnen[1], PhD

[1]Centre for eHealth & Wellbeing Research, Departement of Psychology, Health, and Technology, University of Twente, Enschede, Netherlands

[2]Marian van Os Centre for Entrepreneurship, Hanze University of Applied Sciences, Groningen, Netherlands

[3]Quantified Self Institute, Hanze University of Applied Sciences, Groningen, Netherlands

Corresponding Author:
Aniek J Lentferink, MSc
Centre for eHealth & Wellbeing Research
Departement of Psychology, Health, and Technology
University of Twente
Cubicus Bldg
10 De Zul
Enschede, 7522 NJ
Netherlands
Phone: 31 505956217
Fax: 31 534892895
Email: a.j.lentferink@utwente.nl

Abstract

Background: The combination of self-tracking and persuasive eCoaching in automated interventions is a new and promising approach for healthy lifestyle management.

Objective: The aim of this study was to identify key components of self-tracking and persuasive eCoaching in automated healthy lifestyle interventions that contribute to their effectiveness on health outcomes, usability, and adherence. A secondary aim was to identify the way in which these key components should be designed to contribute to improved health outcomes, usability, and adherence.

Methods: The scoping review methodology proposed by Arskey and O'Malley was applied. Scopus, EMBASE, PsycINFO, and PubMed were searched for publications dated from January 1, 2013 to January 31, 2016 that included (1) self-tracking, (2) persuasive eCoaching, and (3) healthy lifestyle intervention.

Results: The search resulted in 32 publications, 17 of which provided results regarding the effect on health outcomes, 27 of which provided results regarding usability, and 13 of which provided results regarding adherence. Among the 32 publications, 27 described an intervention. The most commonly applied persuasive eCoaching components in the described interventions were *personalization* (n=24), *suggestion* (n=19), *goal-setting* (n=17), *simulation* (n=17), and *reminders* (n=15). As for self-tracking components, most interventions utilized an accelerometer to measure steps (n=11). Furthermore, the medium through which the user could access the intervention was usually a mobile phone (n=10). The following key components and their specific design seem to influence both health outcomes and usability in a positive way: *reduction* by setting short-term goals to eventually reach long-term goals, *personalization* of goals, *praise* messages, *reminders* to input self-tracking data into the technology, use of *validity-tested* devices, *integration of self-tracking and persuasive eCoaching*, and provision of face-to-face instructions during *implementation*. In addition, health outcomes or usability were not negatively affected when more *effort* was requested from participants to input data into the technology. The data extracted from the included publications provided limited ability to identify key components for adherence. However, one key component was identified for both usability and adherence, namely the provision of *personalized* content.

Conclusions: This scoping review provides a first overview of the key components in automated healthy lifestyle interventions combining self-tracking and persuasive eCoaching that can be utilized during the development of such interventions. Future

studies should focus on the identification of key components for effects on adherence, as adherence is a prerequisite for an intervention to be effective.

(J Med Internet Res 2017;19(8):e277) doi:10.2196/jmir.7288

KEYWORDS

telemedicine; review; health promotion; remote sensing technology

Introduction

Health Promotion and Technology

Improving healthy lifestyle behavior is an effective strategy to decrease mortality and increase health-related quality of life [1,2]. Current digital health technologies provide meaningful contributions to the design of healthy lifestyle interventions and the dissemination of such interventions [3]. A combination of self-tracking, goal-setting, and feedback in automated interventions has been indicated by many to be an effective approach for increasing healthy lifestyle behavior [3-5]. Self-tracking is "the practice of systematically recording information about one's diet, health, or activities, typically by means of a mobile phone, so as to discover behavioral patterns that may then be adjusted to help improve one's physical or mental well-being" [6]. Components that might be important for self-tracking are the self-tracking device, validity, the effort required of the participant to perform self-tracking, and the presentation of summary data to the user [7].

Persuasive eCoaching

Goal-setting and feedback are components that can be provided via so-called persuasive eCoaching. This new term is a contraction of the terms "persuasive technology" and "eCoaching." We refer to persuasive eCoaching as the use of technology during coaching to motivate and stimulate (groups of) people to change attitudes, behaviors, and rituals [8]. Oinas-Kukkonen and Harjumaa's persuasive system design (PSD) model [9] describes such persuasive technologies that are expected to positively influence health behavior change. This PSD model builds upon earlier research by Fogg [10] and divides the persuasive components into 4 main categories: primary task support, dialogue support, system credibility support, and social support. These categories contain additional components such as personalization and reminders. To make the PSD model more complete for persuasive eCoaching, some coaching components that can be provided via technology can be added, namely educational coaching, goal-setting, and feedback.

New Opportunities and Challenges

The integration of self-tracking and persuasive eCoaching in fully automated healthy lifestyle interventions creates new opportunities for healthy lifestyle management. First, self-tracking devices enable the objective tracking of lifestyle behavior such as physical activity, heart rate, or sleep. This objective measurement of one's lifestyle pattern can be more reliable than people's own estimations based on their memory and biological sensing of their lifestyle patterns [11-13]. More reliable measurements could become an essential component in lifestyle behavior change, enabling a greater awareness of people's current lifestyles [14]. Second, data from wearable devices can generate automated, personally relevant feedback 24/7. Previous research suggests that this just-in-time tailored feedback contributes to the sustainable use of the intervention [8,15,16]. Third, more and more people own devices that are suitable for eHealth interventions [3]. Even among ethnic minorities and the elderly, the use of mobile phones and computers is rising [17-19]. This suggests a certain scalability for such interventions and maybe even cost-effectiveness due to the fact that no human effort is required to carry them out.

Besides these opportunities, applying the combination of self-tracking and persuasive eCoaching in automatic eHealth interventions also gives rise to a few challenges. These challenges concern privacy issues, trust, and ethics due to personally sensitive data being obtained and stored [20-22]. Concerning ethics, suggestions based on self-tracking data that are invalidated or unsupervised might end up being incorrect or even harmful [21]. In addition, individuals need to be able to understand and interpret the self-tracking data [23].

Identifying Key Components

Despite the challenges, the combination of self-tracking and persuasive eCoaching to promote a healthier lifestyle is promising [3-5], and consequently, interventions employing this combination are becoming more common [5]. To our knowledge, no literature review has been conducted to identify the key components of such interventions. Knowledge about these key components can serve as input for future development of healthy lifestyle interventions that combine self-tracking and persuasive eCoaching, which in turn might increase the effect on health outcomes, usability, and adherence. Usability and adherence are important effect measures of eHealth interventions as they are prerequisites for the intervention to positively influence health or health behavior. In addition, it is worthwhile to identify the specific way a key component should be designed to create positive effects on health outcomes, usability, and adherence. "Effect on health outcomes" here means the effects of the lifestyle intervention on both changes in healthy lifestyle behavior (eg, an increase in physical activity) as well as changes in health status (eg, improved blood levels or weight loss). "Usability" here means the user's satisfaction with the technology and its ease of use [24]. "Adherence" here means the extent to which the technology is used as intended [15].

Key components of interest are self-tracking components (eg, type of device and presentation of summary data to the user), persuasive eCoaching components (eg, elements of the PSD model such as personalization and suggestion), and other intervention components (eg, the underlying behavior change theory and cocreation with end users). This review addresses the following research questions: (1) What are key components

for the effectiveness on health outcomes, usability, and/or adherence of automated healthy lifestyle interventions combining self-tracking and persuasive eCoaching? and (2) In which way should key components be designed to contribute to effectiveness on health outcomes, usability, and/or adherence?

Overarching Project

This review is part of an overarching project for the development of a workplace stress management intervention that combines self-tracking and persuasive eCoaching. To ensure systematic and holistic development and implementation of the eHealth intervention, the Center for eHealth Research (CeHRes) roadmap is adhered to throughout this project [8]. This evidence-based roadmap aims to improve the uptake and impact of eHealth technologies and is based on a participatory development approach, persuasive design techniques, and business modeling. The first step consists of contextual inquiry. This step aims to identify key components from the literature and from users and other stakeholders who will affect—or will be affected by—the intervention.

Methods

Scoping Review Methodology

As technology continues to evolve rapidly, this particular scoping review methodology was chosen for this review study because it allowed us to obtain a quick overview of the current literature on the topic. The fact that this field is rapidly evolving is illustrated by the development of Fitbit self-tracking devices. Ever since the first Fitbit tracker was released at the end of 2009, 13 more Fitbit trackers have been released [25].

Another reason to conduct a scoping review on this topic is that a scoping review is not limited to randomized controlled trials (RCTs) [26,27]. To properly identify the scope of this topic, studies evaluating the effect on health outcomes, usability, and adherence are required. Studies regarding the latter two will primarily be qualitative studies [24].

Arksey and O'Malley's scoping review methodology [26] was applied. This methodology comprises the following steps: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) consultation. A number of additional recommendations by Levac et al [27] were followed, namely: providing a clear purpose for the scoping review, review of full-text articles by 2 independent reviewers to decide on their inclusion, collectively developing the data-charting form with the research team, continually extracting data and updating the data-charting form, inclusion of the consultation step (an optional step according to Arksey and O'Malley [26]), and providing a clear consultation purpose.

Identifying Relevant Studies and Study Selection

A systematic literature search was performed in PubMed, EMBASE, PsycINFO, and Scopus covering the period from January 1, 2013 to January 31, 2016. PubMed and EMBASE were chosen for their wide coverage of scientific journals, whereas PsycINFO was chosen for its specific relevance to this review's subject. Scopus was searched because of its multidisciplinary scope, which allows for identification of articles outside the medical field, such as in engineering. We decided to include no publication from before 2013 as technologies described in publications before 2013 seem less comparable with technologies described in newer publications. To illustrate, publications containing the search terms "Fitbit" and "smartwatch" increased from a negligible number before 2013 to hundreds from 2013 onward, reflecting the rise of personal monitoring devices [28]. These personal monitoring devices represent newer self-tracking technologies that simplify the collection and combining of personal data and enable more personalized healthy lifestyle interventions [29]. Including older publications, in which technological advances are not displayed, might lead to less relevant findings [30].

This study's search strategy was created in collaboration with a University of Twente librarian, based on 3 main components: (1) self-tracking, (2) persuasive eCoaching, and (3) healthy lifestyle interventions. Related search keywords were identified using MeSH and EMTREE terms, PubReMiner, synonyms, keywords from relevant articles, and self-determined search terms (see Multimedia Appendix 1).

Our aim was to include articles that described fully automated interventions. However, we found that many articles involved a fully automated intervention in addition to human coaching, which we call blended coaching. As the scoping review methodology allows for post hoc decisions [27], we then decided to also include blended coaching interventions because we expected to find relevant results in these studies. Other inclusion criteria were that the articles had to be written in English or Dutch and had to be journal articles. Excluded publications included reviews, study protocols, study populations outside the age range of 18-66 years, publications lacking empirical data, and paper-based or personally reported tracking. This age range is in line with the target group of our overarching research project that focuses on the working population. In a lot of European countries, the retirement age is gradually increasing toward 67 years [31].

The results of the search query were uploaded into the EndNote X7 reference manager (Thomson Reuters, Philadelphia, PA, USA) and independently assessed by two reviewers to decide on their inclusion based on title, abstract, and full-text (the review team was comprised of AL and HO for selection based on title and abstract, and AL, HO, LP, and MG for selection based on full-text, with AL reviewing all full-text articles). Differences were fully discussed until consensus was reached.

In addition to the electronic database search, manual searching was performed in *JMIR mHealth and uHealth* for issues dated from January 1, 2013 to January 31, 2016. In addition, a check was performed on the bibliographic reference lists of publications that remained after full-text selection of the search query or manual searching and did not describe interventions involving blended coaching, to identify any additional eligible publications.

The electronic database search and manual searching resulted in 394 publications and 59 publications, respectively, 98 of which were duplicates. After the final full-text selection, 27 publications remained [32-58]. The check of the reference lists

resulted in 5 additional publications [59-63] (see the flowchart in Figure 1).

Charting the Data

A data-charting form was created by the research team that included the following: study characteristics (eg, title, participants, outcomes of interest, and effectiveness), intervention characteristics (eg, short description of the intervention, self-tracking components, and persuasive eCoaching components), and advantages and limitations of the intervention and research according to the authors or reviewers (see Table 1). Next, the data-charting form was improved by several iterations between researchers and 2 consensus meetings of the whole research team.

Figure 1. Flowchart of the selection process. Note: ST=self-tracking, PeC=Persuasive eCoaching.

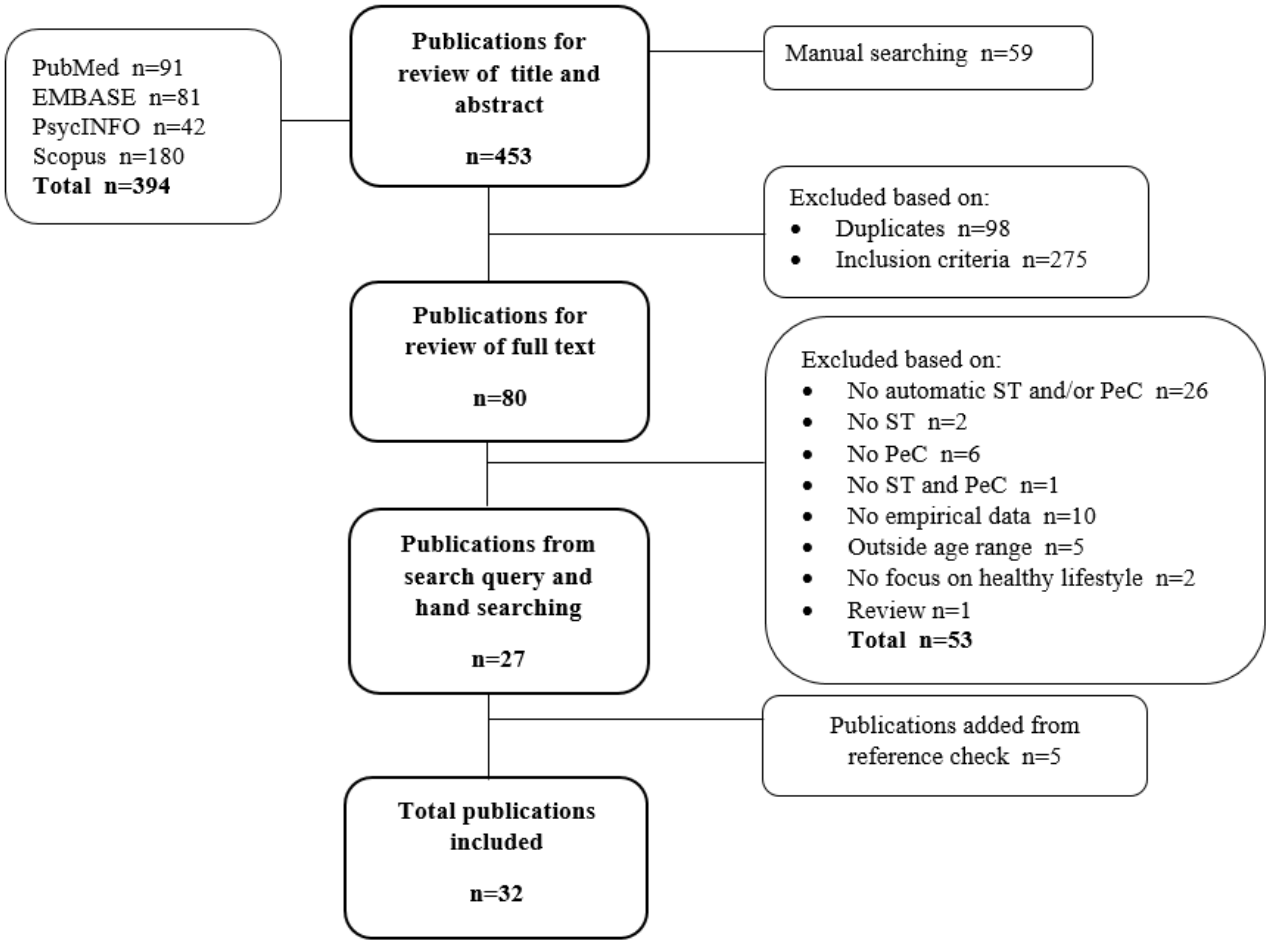


Table 1. Components of the data-charting form.

Category	Component
Study characteristics	Title
	Author (Year)
	Study set-up
	comparison intervention
	study quality
	Objective study
	Participants
	Country of the study
	Duration
	Outcomes of interest
	Secondary outcomes
	Measuring instruments
	Validity of measuring instruments
	Effect on health outcomes (high effective, low effective, and ineffective)
	Usability
	Adherence
Intervention characteristics	Intervention setting (lifestyle, chronic disease, or mental health)
	Country of the intervention
	Persuasive eCoaching components
	components from the PSD model
	social support in general
	educational coaching
	goal-setting
	feedback
	Self-tracking components
	in general
	device
	measurement variable
	the participant's effort
	presentation of summary data
	duration of device usage
	device placement
	validity
	Other intervention components
	short description of the technology
	the intervention's aim
	theory applied
	results from other research applied
	cocreation
Advantages and limitations according to author or reviewer	Advantages and limitations of intervention
	Advantages and limitations of research

Table 2. Categorization of interventions by effect on health outcomes by Morrison et al [64].

Effectiveness	Criteria
High effective	The intervention led to statistically significant improvement on the majority of outcome measures.
	The intervention was more effective or as effective as comparison groups.
	The intervention was more effective than control groups without an intervention or waiting lists.
Low effective	The intervention led to statistically significant improvement on the minority of outcome measures.
	The intervention was as effective or less effective than comparison groups.
	The intervention was more effective than control groups without an intervention or waiting lists.
Ineffective	The intervention led to no statistically significant improvements on any of the outcome measures.
	The intervention was no more effective than control groups without an intervention or waiting lists.

Using the framework by Morrison et al [64], the interventions were divided into 3 categories in terms of their effect on health outcomes: high effective, low effective, and ineffective (see Table 2). For example, if the intervention group showed statistically significant improvements for steps per day and body mass index but not for blood pressure as a result of within-group analyses and was more effective than the comparison intervention group as a result of between-group analysis, the intervention was categorized as "high effective." No distinction has been made between the different outcomes, as long as they were related to health (eg, healthy lifestyle behavior or health status). Study quality was assessed by evaluating the rigor of the study designs based on the established hierarchy of study designs [65].

Data on usability was extracted from included publications if participants in the study expressed a preference for a component or a specific component's design that increased their level of satisfaction regarding the technology or its ease of use.

Data on adherence was extracted if the studies described the way in which participants should use the intervention and presented results on the participants' adherence to that intended use. Originally, we intended to divide the studies into incremental categories of adherence. Unfortunately, the data extracted from the studies did not allow us to do so. Additionally, data on adherence was extracted if participants expressed an expectation that a specific component could increase their adherence in using the technology.

Persuasive eCoaching components were extracted from included publications using the PSD model [9]. Solely, persuasive components were coded when they were executed by the technology and not by human effort, which is in line with the use of the PSD model as described in the review by Kelders et al [15]. As data extraction progressed, we decided to include 3 persuasive eCoaching components in the data-charting form: *educational coaching*, *goal-setting*, and *feedback*. The reason for this decision was that these components were often described in the intervention's design and comprised coaching strategies that could be delivered via technology. In addition, the specific reasons for allowing social support in the design were often omitted. Consequently, we could not link design elements to specific social support components in the PSD model and therefore created the component *social support in general* (see Multimedia Appendix 2 for an overview).

The self-tracking components and other intervention components were identified using qualitative analysis of the data from publications. Components were added or changed continuously as the qualitative data analysis progressed.

The consistency of the data-charting form was discussed by two reviewers (AL and HO), who focused on data extraction performed by one reviewer (AL) on 4 articles with various study designs (3 studies evaluating the effect on health outcomes and 1 study on usability) [32,35,50,53]. After their discussion, it appeared that persuasive eCoaching components and the advantages and limitations of the research and interventions were more prone to reviewer subjectivism than other components. Therefore, we decided to extract the data from 3 articles [42,46,49] gathered by two reviewers (AL and HO) independently of these components to increase consensus with regards to the data interpretation.

Collating, Summarizing, and Reporting the Results

All relevant data was coded using the data-charting form in ATLAS.ti version 7.5 (Scientific Software Development GmbH, Berlin), a qualitative software package. In addition, short summaries were obtained from the data-charting forms to provide quick overviews. Qualitative analysis was used due to our interest in how and why components were applied and to observe patterns in the application of the components and their contribution to the effect on health outcomes and usability [66]. Following Arksey and O'Malley [26] and Levac et al [27], descriptive numerical summaries and thematic analyses were used for data analysis, resulting in an approach that is akin to a "narrative review" [26]. First, a descriptive numerical summary was used to create a numerical overview of specific self-tracking components, persuasive eCoaching components, and other components in the interventions categorized by their effect on health outcomes. Components were identified as key components if at least 50% of the interventions that showed effectiveness on health outcomes (high and low effective interventions) included the component. This 50% rule was applied to all persuasive eCoaching components, with exception of feedback, the self-tracking component *validity*, and the other intervention components, *theory applied*, *results from other research applied*, *cocreation*, *design testing*, *integration of self-tracking and persuasive eCoaching*, and *blended coaching*. Other components included in the data extraction were of a descriptive nature and could therefore not be treated as

dichotomous components for which percentages could be calculated.

Second, thematic analysis was applied to obtain more insight into the various components' specific designs and if these specific designs relate to the effectiveness of the interventions on health outcomes, usability, and/or adherence. When patterns were observed linking components and effectiveness, these components were then identified as key components. Additionally, thematic analysis directed the process of creating the data-charting form.

Consultation

The aim of this consultation was to give meaning to and assess the applicability of the results by obtaining insight from other perspectives, beyond the research team's own perspectives [26,27]. The consultation was carried out during the 11th International Conference on Persuasive Technology. The preliminary results of this scoping review [67] were presented and input was requested from experts in several fields during the workshop on Behavior Change Support Systems (BCSS 2016): Epic for Change, the Pillars for Persuasive Technology for Smart Societies. This consultation adjusted the scope to the way in which components are designed, to get a clear idea of how and why specific components do or do not contribute to effects on health outcomes, usability, and/or adherence.

Results

Characteristics of Included Studies

General Characteristics

Most studies were carried out in the United States [32,33,35,37,40,46,49-52,57,60,63], followed by the Netherlands [34,45,53,55,56,62].

Of the 32 included publications, 27 in total described an intervention [32,33,35,37-45,47-49,51-58,60-63], 17 of which [32,35,37,39,41,43-45,48,49,51-53,57,60,61,63] evaluated the effects of that intervention on health outcomes. Of these 17 studies, 16 were RCTs (highest level in the hierarchy of study designs [65]) and 1 was a quasi-RCT study (second highest level) [43]. In addition, 10 were categorized as high effective [32,39,41,45,48,49,51,60,61,63], 4 as low effective [32,37,44,57], and 3 as ineffective [35,43,52]. Additionally, 25 publications [32-36,38-43,46-51,54-59,62,63] included results on usability, 18 of which [32,33,35,36,38-43,48,49,51,54,56,57,62,63] were based on people's experiences after having used the technology and 6 of which were based on expectations [34,46,50,55,58,59]. Only 1 study addressed usability results based on experiences as well as those based on expectations [47]. As for adherence, 8 publications included information about the intended use of the intervention [40,41,45,47,51,53,60,63], and 5 publications included information about expectations regarding components that could increase adherence [36,46,47,59,62] (see Multimedia Appendix 3 for a summary of the included publications).

Intervention Characteristics

General Characteristics

Out of those publications that described the design of an intervention, 17 were developed in a healthy lifestyle setting [32,33,35,37,39,42-45,49,51,52,57,58,61-63] and 10 in a chronic disease setting [38,40,41,47,48,53-56,60]. Furthermore, about half of the interventions described included the application of a certain theory in the design. The most frequently applied theories were social cognitive theory [33,35,38,49,52], transtheoretical theory [32,35,39,52,55], and self-regulation theory [38,47,63]. Additionally, 6 studies included descriptions of cocreation with end users [42,47,52,55,56,58]. The medium most often used to execute the intervention was a mobile phone app [37,38,40,48,49,52,55,60-62]. Other mediums were a computer [35,39,43,45,54,63], a combination of computer and mobile phone for text messages (short message service, SMS) [32,41,42,44,51,57,58], a combination of computer and a mobile phone app [33,53,56], or just a mobile phone for text messages [56]. Finally, 10 interventions involved blended coaching [38,40,44,45,47,48,54-56,60].

Persuasive eCoaching Characteristics

The persuasive technology category from the PSD model applied most often was primary task support, followed by dialogue support. System credibility support was applied sparingly. For the most part, the identified persuasive eCoaching components were *personalization* (85%, or 23/27), *goal-setting* (74%, or 20/27), *suggestion* (70%, or 19/27), *simulation* (56%, or 15/27), and *reminders* (52%, or 14/27).

Self-Tracking Characteristics

Most interventions used an accelerometer for self-tracking [33,37,38,40,41,45,49,53,55-57]. Other devices used in multiple interventions were pedometers [32,35,39,44,47,48,51,54,58,60] and smart scales [48,52,63]. Five of these self-tracking devices [32,39,42,45,56] were described as tested for validity. In addition, the effort required of the participant to input data into the technology was either none, that is, automatic transfer of data [33,38,43,48,52,55,56,62,63], manually entering data [32,37,39,47-49,51,54,58,60], or uploading data [35,41,42,45,57]. Four studies made no mention of the transmission of data to the technology [40,44,53,61]. The type of electronic data collected was usually the number of steps taken [32,33,35,38-41,44,47,48,51,53,54,56-58,60,61]. Furthermore, data regarding weight [48,52,63], heart rate [42], and other types of physical activity outcomes was collected, such as distance [33,38,43], intensity [38,41,45,55-57], time [38,42,43,45,55,57,60], and/or energy expenditure [43,45,49,61]. The electronic data was either presented to the participant as summary data via visual presentation in a graph, chart, or bar [33,35,37,38,40,42,48,49,51,53,55,56,58,61,63], as summary data via a message [47], or in a life log with a list of activities [49]. Eleven interventions drew a comparison between the current behavior and the goal [33,37,38,48,51,53,55,56,58,61].

Key Components

An overview of the key components, categorized by effect measures (health outcomes, usability, and adherence), can be found in Multimedia Appendix 4. This table also provides an

overview of which studies the key components are based on. Results regarding components have not been presented if too little data was present (results regarding that component from only one study) or no clear pattern could be observed between the component and the effectiveness (on health outcomes, usability, and/or adherence). Moreover, key components were not separately described for interventions utilizing blended coaching or automatic coaching, as too little data was present or no differences were observed for the key components between the 2 types of coaching in terms of their effects. The same holds for key components from studies describing results regarding usability based on expectations, and studies based on experiences. The key components presented below are divided according to the 3 effect measures: health outcomes, usability, and adherence.

Key Components for a Positive Effect on Health Outcomes

Persuasive eCoaching Key Components

In the category of primary task support, *reduction* [32,39,41,49,51,60,63], *personalization* [32,37,39,41,45,48,49,51,53,57,60,61], and *simulation* [37,39,48,51,53,57,61,63] were identified as key components, as they were included in at least 50% of the interventions that were effective in terms of health outcomes. As for dialogue support, *reminders* [32,37,44,51,53,57,60] and *suggestion* [32,37,39,45,48,49,51,53,57,60,63] were identified as key components. No key components were identified in the categories of system credibility support and social support. *Goal-setting* was determined to be another key component for persuasive eCoaching [32,37,39,41,49,51,53,60,61,63] (see Table 3 for an overview).

Design of the Persuasive eCoaching Key Components

In studies evaluating the effect on health outcomes, the *reduction* component was designed in 1 of 3 ways: (1) setting short term goals to eventually reach long term goals [32,39,41,49,51,60,63], (2) providing low effort behavior suggestions [35,49,63], or (3) helping the user solve a problem [35]. *Personalization* was most often implemented to adjust goals or feedback [32,35,37,39,41,43,45,48,51-53,60,61,63] but not so much for the user's ability to set technical features, such as their ability to control prompts and layout [32,57]. The personalization of feedback was mostly based on self-tracking data or reaching goals [32,39,41,43,45,48,52,53,60,61]. The *simulation* component consisted of an overview of the collected data over time in a graph [35,37,48,51,53,57,61,63] or in a message [39]. *Reminders* were usually sent daily [32,35,37,51,60] and were either task reminders regarding self-tracking [32,51,52,60] or reminders to perform health behavior [35,37,44,53,57]. As for *suggestion*, these messages were often personalized [35,37,45,49,51,63] and contained suggestions on how to perform the intended behavior [32,37,39,45,49,51-53,57,60,63] or suggestions for behavior change [45,48,63]. Some suggestion messages were of a motivational nature [32,53], such as "you have taken more rest, please go for a walk" [53]. Apart from one study [53], the *goal-setting* component was usually personalized [32,39,49,51,52,60]. Finally, goals were either assigned to the user [32,37,41,51,53,60] or the user could choose personal goals [35,39,49].

Design of the Persuasive eCoaching Components and Effectiveness on Health Outcomes

When comparing the use of *reduction* among the 3 categories of effectiveness, 6 out of 7 [32,39,41,51,52,60] high effective interventions used reduction by setting short-term goals to eventually reach the ultimate long-term goal. This was not done in the other, low effective or ineffective, interventions.

When comparing the effectiveness and the application of *personalization*, it became apparent that 4 out of 9 high effective interventions [32,39,51,60] used personalization to *set goals*, whereas none of the low effective and only one ineffective intervention [52] applied this strategy. Moreover, high effective studies were the only ones to personalize goals by means of self-tracking data [32,51,60]. In addition, differences were observed in the number of personalized components in the interventions, with 5 out of 9 high effective studies [32,39,41,51,60] personalizing 2 or more components in comparison with 1 out of 3 low effective [57] and 1 out of 3 ineffective interventions [52].

It was observed that 2 out of 3 high effective interventions that applied *reminders* used those reminders to ask the participant to input behavioral data into the technology [51,60], whereas ineffective and low effective interventions only used reminders on changing health behavior [35,37,44,52,53,57].

No clear pattern was observed between the 3 categories of effectiveness on health outcomes and the specific design of the simulation and suggestion components.

Other persuasive eCoaching key components for which patterns were observed regarding their effectiveness on health outcomes were the inclusion of *praise* messages [32,39,51,60] and *tunneling* by providing advice based on how well the participant changed the desired behavior [41,45,63]. These components were only ever applied in high effective interventions.

Self-tracking Key Components

The *validity* component was applied in 21% (3/14) of the interventions [32,39,45] that showed effectiveness on health outcomes. Based on the 50% rule, the validity of the self-tracking device is thus not considered to be a key component.

Table 3. Interventions' persuasive eCoaching elements, ordered by their effect in terms of health outcomes.

Persuasive eCoaching category	Effective interventions (n=14) n (%)	Total number of studies evaluating effect on health outcomes (n=17) n (%)
Primary task support		
Reduction	7 (50)	8 (47)
Tunneling	4 (29)	5 (29)
Tailoring	3 (21)	4 (24)
Personalization	12 (86)	15 (88)
Simulation	8 (57)	10 (59)
Rehearsal	2 (14)	2 (12)
Dialogue support		
Praise	4 (29)	4 (24)
Rewards	2 (14)	3 (18)
Reminders	7 (50)	9 (53)
Suggestion	11 (79)	13 (76)
Similarity	2 (14)	2 (12)
Liking	2 (14)	2 (12)
Social role	0 (0)	1 (6)
System credibility support		
Trustworthiness	1 (7)	1 (6)
Expertise	1 (7)	1 (6)
Surface credibility	0 (0)	1 (6)
Real-world feel	1 (7)	1 (6)
Authority	0 (0)	0 (0)
Third-party endorsement	0 (0)	0 (0)
Verifiability	0 (0)	0 (0)
Social support		
Social support in general	2 (14)	3 (18)
Social learning	0 (0)	0 (0)
Social comparison	0 (0)	0 (0)
Normative influence	1 (7)	1 (6)
Social facilitation	0 (0)	1 (6)
Cooperation	0 (0)	0 (0)
Competition	0 (0)	0 (0)
Recognition	0 (0)	0 (0)
Other		
Educational coaching	6 (43)	7 (41)
Goal-setting	10 (71)	13 (76)

Design of the Self-Tracking Components and Effectiveness on Health Outcomes

When comparing the self-tracking *device* applied by effect on health outcomes, it was observed that accelerometers were only applied in the high effective interventions [41,45,49] and the low effective interventions [37,53,57], whereas not at all in the ineffective interventions. In addition, the most intensive *effort* was asked from participants in high effective interventions to input data into the technology. To illustrate, the low effective and ineffective interventions mostly applied uploading data [35,57] or automatic transfer of the data to the technology [43,52,53]. In the high effective interventions, participants were

asked for a more intensive approach than uploading data or doing nothing, such as sending a daily message with steps to the technology [32,48,51,60]. The latter was also applied in one low effective intervention [48]. Although the validity component was not identified as a key component based on the 50% rule, only in high effective studies good *validity* and reliability of the device were described [32,39,45].

Other Intervention Key Components

With respectively 71% (10/14) and 50% of the effective interventions applying *integration of self-tracking and persuasive eCoaching* [32,37,39,41,45,48,49,51,53,61] and *results from other research applied* [32,39,49,51,60,61,63], these components were identified as key components. The percentages of the other components were 29% (4/14) for *theory applied* [32,39,49,63], 14% (2/14) for *design testing* [39,51], and 0% for *cocreation*.

Design of the Other Intervention Key Components

The design of *self-tracking and persuasive eCoaching integration* usually involved the use of self-tracking data to provide feedback [32,37,41,45,48,49,51,53,60,63]. Some studies also used self-tracking data to set goals [32,51,60]. The following *results from other research* were used in intervention design: the application of a known protocol [51,60,63], methods that were evaluated as effective [32,49,61], and components from healthy lifestyle interventions that were evaluated as effective [39,43].

Design of the Other Intervention Components and Effectiveness on Health Outcomes

Ineffective interventions applied less intensive *implementation* strategies such as brief tutorial [35], instructions on paper [43], or nothing [52] in comparison with the high and low effective interventions, which used mostly face-to-face instructions [32,37,39,49,51,57,61,63].

Key Components for Usability

Persuasive eCoaching Key Components

An overview of all key components for usability can be found in Multimedia Appendix 4. The most apparent key components for usability are described below. In line with the key components for a positive effect on health outcomes, a pattern was observed between the following key components and a positive effect on usability:

- *Reduction* to simplify the performance of behavior [36,50,62]. In addition, participants found it useful to be able to set short-term goals [46,58]. They believed that it could contribute to their motivation [46,58]. In addition to the similarities with key components for health outcomes, users also appreciated the provision of means to simplify their performance of the behavior [58,62].
- *Personalization* of goals [50,56,58,59]. For the most part, users appreciated the ability to set personal goals because it fosters the observation of progress [50,58,59].
- *Praise* messages [42,47,55,59,62]. However, praise might require a different design for men and women, since gender differences were observed, with women appreciating praise more than men [59].
- *Reminders* were perceived useful by most [35,42,47,50]. However, the timing and frequency of the reminders are of importance to avoid annoyance, feelings of being checked up on, or guilt for not reaching the goal [42,47,50,57,59,62]. One study's participants expressed a preference for reminders to upload or enter data into the technology [47].
- *Simulation* to observe progress [33-35,47,50,55,59,62]. Users particularly appreciated visualization of self-tracking data to observe progress toward their goals [33-35,47,50,55,59,62]. A clear overview with only a few important features displayed was preferred overall [55,62]. However, people following physical activity guidelines preferred more detailed information [62].

In contrast to studies evaluating the effect on health outcomes, where results on *personalization* were mostly observed for the personalization of content, the participants' concerns regarding usability were mostly about the ability to set technical features such as the timing of the message, password protection, and layout. For example, not all participants were concerned about the safety of their self-tracking data [46,50,56,59], and some found that password protection interferes with the technology's usability [50,59]. In addition, participants would like to be able to decide whom to share data with [46,50]. These aspects also relate to the *trustworthiness* component. As for personalization of content, users acknowledged personalization as a practical solution [50] to account for the differences that existed among the various groups of users and even within groups of users [34,36,50,62]. Participants themselves also expressed a desire for the personalization of content [34,47,57,58]. Some participants felt that it would be meaningful to take personalization to the next level by using data mining to enable context-sensing and observe trends and patterns in personal data [46,47,50,59], which is also a form of *reduction*. However, others felt such extensive personalization would be unreliable, artificial, or unnecessary [47,50,59].

The *social support* component was rated negatively by most participants [34,50,59,62]. However, it appears that acceptability of social support was higher when receiving support via the technology from close friends, family, or peers [50,59,62]. However, a few participants did not like the idea of receiving support from family members, as they had previous negative experiences with support from family during behavior change [58]. In contrast, acceptability of social support was lower when the intervention used social media platforms open to everyone, such as Facebook [59,62].

In terms of users' perspective on *educational coaching*, the fact that most users had already been trying to change their behavior for quite some time and were already familiar with much of the information on the subject should be taken into account [47,58].

Self-Tracking Key Components

Overall, it was apparent from the studies on usability that users had a positive attitude regarding the self-tracking of behavioral outcomes [34,46,47,51,59,63]. One positive aspect mentioned by participants was that performing self-tracking increased their awareness [36,46,47,49,62].

In line with key components for effect on health outcomes, the *validity* of the device was perceived as important among users [46,56,59]. In addition, one publication reported on users' willingness to put in more *effort* if they felt doing so was justified by its added value [50]. Overall, most participants had a favorable attitude toward the automatic tracking of behavioral outcomes [46,59,62], although basic data entry was also perceived as acceptable [46,50,62].

As for the *measurement* component, self-tracking was found to have a potential demotivating effect when users were unable to capture all personally relevant data using self-tracking devices (eg, the use of an accelerometer when walking or running was not in fact their most common physical activity) [33,55,62].

Other Intervention Key Components

In line with key components for effect on health outcomes, participants acknowledged the advantages of *the integration of self-tracking and persuasive eCoaching* [32,34,35,46,51,56,59,62] and believed that an intervention incorporating this combination could successfully motivate or change behavior [36,40,42,47,54]. In addition, participants considered it useful to receive instructions on how to use the intervention [42,56].

Furthermore, participants preferred the use of mobile phones for intervention delivery to delivery via the computer [46,50,55,58,59,62]. One advantage of mobile phone apps the participants named was the ability to use the intervention whenever they wanted [36,47].

Most participants reported that it was preferable to have access to a health care professional on top of using the automated intervention [46,47,54,55,58,59]. Even though health care professionals were negative about the provision of feedback [55], they did see the advantage of such interventions to supplement in-person sessions, as it might increase their ability to anticipate and better understand the process of behavior change among clients [40,55].

Key Components for Adherence

No key components for adherence could be identified based on information about intended usage, as these results were only sparingly presented [40,41,45,47,51,53,60,63]. Out of the 8 studies that did present results on intended usage, only 6 presented data on the intended usage of the self-tracking component and not the intervention as a whole [41,45,47,51,60,63].

Based on participants' opinions, the following key components were identified: the *personalization* component, as users believed that personally relevant advice could increase adherence [46,59,62]; and the *design testing* component, as users said that adherence declined when a problem occurred while using the intervention [36,62].

Discussion

Findings

This scoping review aimed to identify key components of self-tracking and persuasive eCoaching in automated healthy lifestyle interventions that contribute to the effectiveness on health outcomes, usability, and adherence.

Key Components for Effect on Health Outcomes

A pattern was observed between the following key components and a positive effect on health outcomes: *reduction*, *personalization*, *simulation*, *suggestion*, *goal-setting*, *praise*, use of *valid* wearables and specifically accelerometers, *integration of self-tracking and persuasive eCoaching*, use of *results from other research* to inform design, and provision of face-to-face instructions during *implementation* of the intervention. A pattern was also observed between more *effort* by the participant to input the self-tracking data in the intervention and more effect on health outcomes. For the following key components, it appears that a specific design is required for the component to have a positive influence on health outcomes: *reduction* by setting short term goals to eventually reach long-term goals, *personalization* of goals using self-tracking data, personalization of multiple components, *tunneling* by provision of feedback based on how well the user changed their behavior, and *reminders* to input data into the technology.

Similar to this scoping review's results, other recent reviews on eHealth also observed the contribution to effectiveness of reminders [68,69], personalization [70], and integration of self-tracking and persuasive eCoaching [68,71]. In addition, one review found that less persuasive technologies were extracted from the system credibility support and social support categories [14]. This could either indicate that designers do not pay enough attention to these categories of persuasive technologies or that these components are often omitted in the description of the technology in publications. If the first is true, this might have consequences for the effectiveness of the intervention. For the system credibility support category, Harris et al [72] found that users engaged less with the technology when credibility was lacking. Neglect of the social support category will be addressed in further detail below.

Support for the importance of reminders can be found in the reviews by Neff and Fry [69] and Bardus et al [68]. However, more knowledge is needed about the reminder component's specific design and effectiveness [69,73,74]. This scoping review diminishes this research gap to some extent by indicating that sending reminders to signal self-tracking could increase effectiveness, which is consistent with the findings of one RCT study [75]. Reminders regarding behavior change appeared to be less effective. One possible explanation for this is that reminders regarding behavior change remind users of their failure to change behavior [76]. Knowledge about other aspects of the reminder component is also of importance, such as the proper frequency, timing, and the way in which users should be notified by reminders (eg, visual or audible cues). One review mentioned that a frequency of one reminder per day should be considered [71]. Another study found that sending event-based reminders (such as after breakfast) were more effective for health behavior change than time-based reminders (such as at a specified time) [74].

Requiring more effort from the participant to input data into the intervention appears to have a positive influence on

interventions' effect on health outcomes. One explanation for this, from the review by Kelders et al [15], is that if more action is required from the participant, it might make the participant more engaged with the intervention. In addition, one study on usability [50] mentioned that participants are willing to devote a higher level of effort (eg, manually entering data), as long as the effort is balanced by its added value (eg, more personally relevant feedback).

Key Components for Usability

Several key components were identified for a positive effect on usability. The most apparent key components for a positive effect on usability are described below. Similar to key components for effect on health outcomes, key components for usability were inclusion of *reduction*, *personalization*, *reminders*, *praise*, *simulation* to observe progress, use of *valid* wearables, *integration of self-tracking and persuasive eCoaching*, provision of face-to-face instructions during *implementation* of the intervention, and requesting more *effort* from the participant for input of self-tracking data into the technology. As for key components for effect on health outcomes, participants deemed the following specific key component designs to be preferable: *reduction* by setting short term goals to eventually reach long-term goals, *personalization* of goals, and *reminders* to input data into the technology. Participants also considered the frequency and timing of reminders to be important to avoid annoyance and acknowledged the advantages of personalizing several aspects of the design. Furthermore, a negative attitude toward *social support* was observed. To increase the acceptability of social support, designs should include the provision of social support via peers, close friends or family, and eliminate the use of social media platforms open to everyone. It was apparent that participants appreciated the delivery of the intervention via a mobile phone. On top of that, participants and health care professionals liked the idea of using the automated intervention as a supplement to in-person sessions.

A recent qualitative review on engagement with digital health interventions obtained mostly similar results [76]. Similarities include the importance of reminders, personalization, the ability to use the intervention 24/7, a suitable supplement to in-person sessions, and provision of reduction to observe trends and patterns. A preference for automated self-tracking was also observed in line with our results. However, our scoping review also uncovered that users may also be willing to accept having to put in some level of effort for self-tracking. It is also worth noting that other studies on usability of eHealth have indicated that a positive attitude exists concerning self-tracking [77-79], inclusion of praise [80], personalization of goals [80,81], the ability to observe progress [80], use of validity-tested devices [78], and that not everyone is concerned with privacy issues [82].

The studies presenting results on usability are a way for us to learn what is most valued or noticed by users about healthy lifestyle interventions combining self-tracking and persuasive eCoaching. A few observations can be made about this topic. First, we observed that the studies evaluating an intervention for effect on health outcomes were mostly focused on the personalization of intervention content, whereas studies on usability were mostly focused on the personalization of technical features. This could indicate that the importance of technical feature personalization for users is not given sufficient attention during intervention development. On the other hand, it is also possible that results on technical feature personalization were simply described less in studies evaluating the effect on health outcomes in comparison with studies on usability. One element of usability is the ease of use of a technology and studies on usability might, therefore, focus more on technical features.

Second, social influence and support are often components of traditional health behavior change models recognized in earlier research as effective for changing behavior [83-85]. But even though social support might be effective in changing behavior, we also observed an unfavorable attitude in many participants toward social support. In addition, to our knowledge, no study found strong evidence for the contribution that social support might make in automated interventions towards improving health behavior [86]. This could be explained by the low usage of the social support component observed in publications included in this review [38,58], as well as in other research [20,87]. As advocated by Riley et al [88], applying traditional health behavior change models might not be the best fit for healthy lifestyle interventions via technology due to their interactive and adaptive character. The social support component probably requires different strategies via technology than via face-to-face provision of social support.

Third, it was observed that not everyone appreciated a high level of *personalization* of feedback messages via data mining in order to discover patterns. The observation of patterns helps users become aware of their way of living and the consequences thereof. Although awareness is a first important step in behavior change [14], some people might prefer not to discover patterns they were not aware of.

Finally, this review's results show that participants would appreciate the ability to consult a health care professional during the intervention. However, health care professionals seemed less open to this. Including consultation by a health care professional would also be a costly way of increasing usability. In terms of their effect on health outcomes, earlier research found automated interventions to be as effective as interventions that include human coaching [15]. Furthermore, this scoping review observed positive effects on health behavior change not only in blended coaching interventions but also in fully automated interventions. Therefore, including human coaching is probably not an essential component. This viewpoint is also supported by previous research [89].

Key Components for Adherence

Few studies included in this review described adherence that concerned the intended usage of the intervention. When intended usage was described, most of the information dealt with the self-tracking part and not the intervention as a whole. In addition, most studies that presented data on the usage of specific components did not state the intended usage in advance. This was also observed in another review [15]. Key components for a positive effect on adherence could therefore only be identified based on participants' expectations. According to

users, adherence could be increased by the *personalization* of content and the performance of *design testing* to eliminate problems during usage of the intervention. Personalization of content has been recognized in another review as a facilitator for adherence [70].

Recommendations for Future Design and Research

First, the key components identified in this scoping review are identified as separate components. However, components might interact with each other and could lead to different outcomes than those anticipated based on knowledge of single components. One review on e-mental health interventions indicated that some combinations of persuasive technology components do indeed differ in terms of their synergy [90]. Further research is recommended to identify the most effective combination and dosage of the key components in healthy lifestyle interventions combining self-tracking and persuasive eCoaching. To date, most studies evaluating eHealth designs apply the traditional RCT method [91]. However, RCTs are often too time-consuming to keep up with the speed of technological developments and can explain little about separate elements and their contribution to effectiveness [91-93]. Riley and Rivera [94], Hekler et al [93], and Pham et al [91] advocated for new strategies to identify and design effective intervention components—"opening the black box." Technology can provide a meaningful contribution to such discoveries, for example, by means of the "Model Predictive Control" [94]. This strategy changes the intensity and combination of intervention components on a daily basis by using the monitoring data provided by participants' responses and other contextual factors.

Another strategy is suggested by Sieverink et al [94]. This strategy does not only attempt to open the black box, it also contributes to more insight into adherence. Sieverink et al provided preliminary results for the development of a log data protocol for eHealth technologies to identify their adherence level and effect on health outcomes [94]. They suggest collecting log data on the usage and intensity of usage for specific intervention components to be able to draw conclusions regarding adherence and linking such log data with effects on health outcomes to be able to draw conclusions regarding adherence to specific components and their effects on health outcomes.

Future research and design should focus specifically on the reminder design, social support, and the observation of patterns through data mining, as different designs seemed to influence effect on health outcomes and/or usability. It would be interesting for future research to test variations on components' designs and their effects on health outcomes and usability.

Another recommendation is the application of personalization to account for the variation in preferences between groups of participants and even within groups of participants. Besides the fact that it is a practical way to account for the existing differences between users, participants also considered personalization to be useful. In addition to the application of personalization in the design, the observed differences both among and within groups also suggest that a needs assessment is required before and during the design phase of an intervention using self-tracking and persuasive eCoaching. Although cocreation is often mentioned as an important aspect in design models [95,96], only a few publications cited in this review described anything regarding cocreation between intervention developers and the target group. The importance of cocreation has been indicated as important to increase satisfaction with the design by other reviews [67,69]. Similar to the component of cocreation, few publications described anything about the theory underpinning the design of the intervention. The main message of other reviews in the field of eHealth is that the design of current eHealth interventions is often not based on existing theory [67,69]. The use of theory has been recognized as resulting in higher effectiveness [97].

Strengths and Limitations

One of the strengths of this research is that by applying qualitative research methods, an attempt has been made to not only describe which components might contribute to effectiveness but also which specific component design is most effective in terms of health outcomes, usability, and adherence. This is important because applying components from one theory in different interventions can result in various designs, whereas applying components from different theories can result in interventions with quite similar designs [93]. Another strength is the fact that we used both data from RCT studies and studies in real-life settings, providing a more realistic overview of the opportunities and challenges for interventions that combine self-tracking and persuasive eCoaching in practice than we would have been able to by relying only on results from RCTs.

One limitation of this scoping review is that potential biases might have influenced the results. First, publication bias could be present, indicated by the absence of negative effects reported and the higher number of high effective interventions included in this scoping review versus low effective or ineffective interventions. With results from only 3 ineffective studies, we could not come to conclusions about any one key component being more often applied in effective interventions compared to ineffective interventions. Due to this limitation, we introduced the 50% rule to identify key components. It should be mentioned that the most commonly applied components in interventions were, therefore, more likely to be identified as key components.

Second, we observed that interventions described in publications from 2013 already differ to some extent from interventions described in publications from 2016. The importance of certain identified key components for effectiveness might increase or decrease due to new technological developments. The main differences were the more frequent use of accelerometers and mobile sensors for self-tracking in newer publications and the delivery of the intervention via mobile phone in newer publications in comparison with computer in older publications. This trend is likely to continue [73,98]. An example of a key component that may become more important is *the ability to enable or disable observation of trends and patterns*. The use of mobile phone sensors enables collection of a wide spectrum of personal data. As indicated by this scoping review's results, not every user is open to intensive data mining. As another example, the importance of applying the proper *frequency and timing* of reminders may increase . Mobile phone interventions can use a broader set of tools to send reminders than

computer-based interventions. In addition, reminders cannot be ignored as easily due to visual or audible alerts [73].

Third, we did not make a distinction between health outcomes. It could be the case that interventions targeting a more intermediate health outcome (eg, an effect on physical activity instead of an effect on blood pressure) were more easily identified as high effective studies. Fourth, the extraction of data concerning persuasive eCoaching components is somewhat subjective, which was observed by the two researchers during data extraction comparison. Finally, we did not code intervention components from the actual interventions because technology is a rapidly evolving field of research and this would have taken a significant amount of time. Choices such as this one are characteristic of the scoping review methodology [26]. These limitations limited us in making a definite list of key components. However, we attempted to provide a first impression of key components in this relatively new field of research.

Conclusions

To our knowledge, this scoping review provides a first overview of key components and effects on health outcomes, usability, and adherence. The following key components and their specific design both seem to influence health outcomes and usability in a positive way: *reduction* by setting short term goals to eventually reach long-term goals, *personalization* of goals, *praise*, *reminders* to input self-tracking data into the technology, use of *validity-tested* devices, *integration of self-tracking and persuasive eCoaching*, and provision of face-to-face instruction during *implementation*. In addition, health outcomes or usability were not affected when more *effort* was requested from participants to input data into the technology. Unfortunately, we were limited in our ability to identify key components for adherence. Still, one key component identified for both usability and adherence is the provision of *personalized* content. Identification of key components for adherence is highly important because adherence is a prerequisite for interventions to be effective. This scoping review provides a first overview, and future research is needed to confirm the key components identified for effect on health outcomes and usability, identify key components for adherence, and study whether the key components represent an effective combination of components.

Acknowledgments

This study is partly funded by Menzis. This funding source had no involvement in the study's design, execution, or reporting. We would like to thank Olga Kulyk, PhD, for her cooperation for our presentation of the preliminary results during the conference workshop and Peter Noort, MSc, for his contribution to developing the search strategy.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full search query per database.

[PDF File (Adobe PDF File), 24KB - jmir_v19i8e277_app1.pdf]

Multimedia Appendix 2

Principles and examples of persuasive eCoaching components.

[PDF File (Adobe PDF File), 31KB - jmir_v19i8e277_app2.pdf]

Multimedia Appendix 3

Overview and characteristics of included publications.

[PDF File (Adobe PDF File), 60KB - jmir_v19i8e277_app3.pdf]

Multimedia Appendix 4

Overview of identified key components and their specific design for effects on health outcomes, usability, and/or adherence.

[PDF File (Adobe PDF File), 31KB - jmir_v19i8e277_app4.pdf]

References

1. Block G, Sternfeld B, Block CH, Block TJ, Norris J, Hopkins D, et al. Development of Alive! (A Lifestyle Intervention Via Email), and its effect on health-related quality of life, presenteeism, and other behavioral outcomes: randomized controlled trial. J Med Internet Res 2008 Nov 19;10(4):e43 [FREE Full text] [doi: 10.2196/jmir.1112] [Medline: 19019818]

2. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med 2012 Sep;55(3):163-170. [doi: 10.1016/j.ypmed.2012.06.017] [Medline: 22735042]
3. Dallery J, Kurti A, Erb P. A new frontier: integrating behavioral and digital technology to promote health behavior. Behav Anal 2015 Aug 23;38(1):19-49 [FREE Full text] [doi: 10.1007/s40614-014-0017-y] [Medline: 27347477]
4. Fawcett T. Mining the quantified self: personal knowledge discovery as a challenge for data science. Big Data 2015 Dec;3(4):249-266. [doi: 10.1089/big.2015.0049] [Medline: 27441406]
5. Grossglauser M, Saner H. Data-driven healthcare: from patterns to actions. Eur J Prev Cardiol 2014 Nov;21(2 Suppl):14-17. [doi: 10.1177/2047487314552755] [Medline: 25354949]
6. Oxford Dictionaries. Self-tracking URL: https://en.oxforddictionaries.com/definition/us/self-tracking [accessed 2017-01-05] [WebCite Cache ID 6nI3I4Zf0]
7. Kim J. A qualitative analysis of user experiences with a self-tracker for activity, sleep, and diet. Interact J Med Res 2014 Mar 04;3(1):e8 [FREE Full text] [doi: 10.2196/ijmr.2878] [Medline: 24594898]
8. 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 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
9. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Communications of the Association for Information Systems 2009 Mar;24(1):485-500 [FREE Full text]
10. Fogg BJ. Persuasive technology: using computers to change what we think and do. San Francisco: Morgan Kaufmann Publishers; 2003.
11. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical activity questionnaire comprehension: lessons from cognitive interviews. Med Sci Sports Exerc 2009 Feb;41(2):336-343 [FREE Full text] [doi: 10.1249/MSS.0b013e318186b1b1] [Medline: 19127192]
12. Archer E, Pavela G, Lavie CJ. The inadmissibility of what we eat in America and NHANES dietary data in nutrition and obesity research and the scientific formulation of national dietary guidelines. Mayo Clin Proc 2015 Jul;90(7):911-926 [FREE Full text] [doi: 10.1016/j.mayocp.2015.04.009] [Medline: 26071068]
13. Baker FC, Maloney S, Driver HS. A comparison of subjective estimates of sleep with objective polysomnographic data in healthy men and women. J Psychosom Res 1999 Oct;47(4):335-341. [Medline: 10616227]
14. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
15. 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]
16. Oinas-Kukkonen H, Harjumaa M. Towards deeper understanding of persuasion in software and information systems. In: Proceedings of the First International Conference on Advances in Computer-Human Interaction. Washington: IEEE Computer Society; 2008 Presented at: ACHI '08; 2008 Feb 10-15; USA. [doi: 10.1109/ACHI.2008.31]
17. Smith A. Pew Internet. 2014. Older Adults and Technology Use URL: http://www.pewinternet.org/files/2014/04/PIP_Seniors-and-Tech-Use_040314.pdf [accessed 2017-01-09] [WebCite Cache ID 6nNvMVvPM]
18. Smith A. Pew Internet. 2015. U.S. Smartphone Use in 2015 URL: http://www.pewinternet.org/files/2015/03/PI_Smartphones_0401151.pdf [accessed 2017-01-09] [WebCite Cache ID 6nNvQJ1g5]
19. Brown A, López G, Lopez M. Pew Hispanic. 2016. Hispanics and mobile access to the internet URL: http://www.pewhispanic.org/2016/07/20/3-hispanics-and-mobile-access-to-the-internet/ [accessed 2017-01-09] [WebCite Cache ID 6nNvVewli]
20. Brouwer ER. Legality and data protection law: The forgotten purpose of purpose limitation. In: Besselink L, Pennings F, Prechal S, editors. The Eclipse of the Legality Principle in the European Union. The Netherlands: Kluwer Law International; 2011:273-294.
21. Schermer BW. The limits of privacy in automated profiling and data mining. CLSR 2011 Feb;27(1):45-52. [doi: 10.1016/j.clsr.2010.11.009]
22. Tene O, Polonetsky J. Privacy in the age of big data: a time for big decisions. 2012 Feb. URL: https://www.stanfordlawreview.org/online/privacy-paradox-privacy-and-big-data/ [WebCite Cache ID 6nI9nyTpn]
23. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
24. Rubin J, Chisnell D. Handbook of usability testing: how to plan, design and conduct effective tests. 2nd ed. Indianapolis, Indiana: John Wiley & Sons; 2008.
25. Marshall G. Wareable. 2016. The story of Fitbit: How a wooden box became a $4 billion company URL: http://www.wareable.com/fitbit/youre-fitbit-and-you-know-it-how-a-wooden-box-became-a-dollar-4-billion-company [accessed 2016-11-28] [WebCite Cache ID 6mMLn8guP]
26. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

27. 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]
28. Hoy MB. Personal activity trackers and the quantified self. Med Ref Serv Q 2016;35(1):94-100. [doi: 10.1080/02763869.2016.1117300] [Medline: 26794199]
29. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: A randomized pilot study. Obesity (Silver Spring) 2016 Aug;24(8):1653-1659. [doi: 10.1002/oby.21536] [Medline: 27367614]
30. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
31. ETK. 2017. Retirement Ages in Member States URL: http://www.etk.fi/en/the-pension-system-2/the-pension-system/international-comparison/retirement-ages/ [accessed 2017-04-03] [WebCite Cache ID 6pRSoZrgk]
32. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013 Dec 09;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
33. Al Ayubi SU, Parmanto B, Branch R, Ding D. A persuasive and social mHealth application for physical activity: a usability and feasibility study. JMIR Mhealth Uhealth 2014 May 22;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
34. Belmon LS, Middelweerd A, Te Velde SJ, Brug J. Dutch young adults ratings of behavior change techniques applied in mobile phone apps to promote physical activity: a cross-sectional survey. JMIR Mhealth Uhealth 2015 Nov 12;3(4):e103 [FREE Full text] [doi: 10.2196/mhealth.4383] [Medline: 26563744]
35. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
36. Casey M, Hayes PS, Glynn F, OLaighin G, Heaney D, Murphy AW, et al. Patients' experiences of using a smartphone application to increase physical activity: the SMART MOVE qualitative study in primary care. Br J Gen Pract 2014 Aug;64(625):e500-e508 [FREE Full text] [doi: 10.3399/bjgp14X680989] [Medline: 25071063]
37. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May;20(5):1091-1101. [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
38. Choo S, Kim JY, Jung SY, Kim S, Kim JE, Han JS, et al. Development of a weight loss mobile app linked with an accelerometer for use in the clinic: usability, acceptability, and early testing of its impact on the patient-doctor relationship. JMIR Mhealth Uhealth 2016 Mar 31;4(1):e24 [FREE Full text] [doi: 10.2196/mhealth.4546] [Medline: 27032541]
39. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015 Feb 09;17(2):e38 [FREE Full text] [doi: 10.2196/jmir.3402] [Medline: 25665498]
40. Forman DE, LaFond K, Panch T, Allsup K, Manning K, Sattelmair J. Utility and efficacy of a smartphone application to enhance the learning and behavior goals of traditional cardiac rehabilitation: a feasibility study. J Cardiopulm Rehabil Prev 2014;34(5):327-334. [doi: 10.1097/HCR.0000000000000058] [Medline: 24866355]
41. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-term effectiveness of a comprehensive internet-based and patient-specific telerehabilitation program with text messaging support for cardiac patients: randomized controlled trial. J Med Internet Res 2015 Jul 23;17(7):e185 [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]
42. González C, Herrero P, Cubero JM, Iniesta JM, Hernando ME, García-Sáez G, et al. PREDIRCAM eHealth platform for individualized telemedical assistance for lifestyle modification in the treatment of obesity, diabetes, and cardiometabolic risk prevention: a pilot study (PREDIRCAM 1). J Diabetes Sci Technol 2013 Jul 01;7(4):888-897 [FREE Full text] [Medline: 23911170]
43. Hunter RF, Tully MA, Davis M, Stevenson M, Kee F. Physical activity loyalty cards for behavior change: a quasi-experimental study. Am J Prev Med 2013 Jul;45(1):56-63. [doi: 10.1016/j.amepre.2013.02.022] [Medline: 23790989]
44. Long JE, Ring C, Bosch JA, Eves F, Drayson MT, Calver R, et al. A life-style physical activity intervention and the antibody response to pneumococcal vaccination in women. Psychosom Med 2013 Oct;75(8):774-782. [doi: 10.1097/PSY.0b013e3182a0b664] [Medline: 23922400]
45. Marsaux CF, Celis-Morales C, Fallaize R, Macready AL, Kolossa S, Woolhead C, et al. Effects of a web-based personalized intervention on physical activity in European adults: a randomized controlled trial. J Med Internet Res 2015 Oct 14;17(10):e231 [FREE Full text] [doi: 10.2196/jmir.4660] [Medline: 26467573]
46. Miyamoto SW, Henderson S, Young HM, Pande A, Han JJ. Tracking health data is not enough: a qualitative exploration of the role of healthcare partnerships and mHealth technology to promote physical activity and to sustain behavior change. JMIR Mhealth Uhealth 2016 Jan 20;4(1):e5 [FREE Full text] [doi: 10.2196/mhealth.4814] [Medline: 26792225]
47. Morton K, Sutton S, Hardeman W, Troughton J, Yates T, Griffin S, et al. A text-messaging and pedometer program to promote physical activity in people at high risk of type 2 diabetes: the development of the PROPELS follow-on support program. JMIR Mhealth Uhealth 2015 Dec 15;3(4):e105 [FREE Full text] [doi: 10.2196/mhealth.5026] [Medline: 26678750]

48. Oh B, Cho B, Han MK, Choi H, Lee MN, Kang HC, et al. The effectiveness of mobile phone-based care for weight control in metabolic syndrome patients: randomized controlled trial. JMIR Mhealth Uhealth 2015 Aug 20;3(3):e83 [FREE Full text] [doi: 10.2196/mhealth.4222] [Medline: 26293568]
49. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
50. Ramanathan N, Swendeman D, Comulada WS, Estrin D, Rotheram-Borus MJ. Identifying preferences for mobile health applications for self-monitoring and self-management: focus group findings from HIV-positive persons and young mothers. Int J Med Inform 2013 Apr;82(4):e38-e46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.05.009] [Medline: 22704234]
51. Steinberg DM, Levine EL, Askew S, Foley P, Bennett GG. Daily text messaging for weight control among racial and ethnic minority women: randomized controlled pilot study. J Med Internet Res 2013 Nov 18;15(11):e244 [FREE Full text] [doi: 10.2196/jmir.2844] [Medline: 24246427]
52. Svetkey LP, Batch BC, Lin PH, Intille SS, Corsino L, Tyson CC, et al. Cell phone intervention for you (CITY): a randomized, controlled trial of behavioral weight loss intervention for young adults using mobile technology. Obesity (Silver Spring) 2015 Nov;23(11):2133-2141. [doi: 10.1002/oby.21226] [Medline: 26530929]
53. Tabak M, Vollenbroek-Hutten MM, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]
54. Tomkins-Lane CC, Lafave LM, Parnell JA, Rempel J, Moriartey S, Andreas Y, et al. The spinal stenosis pedometer and nutrition lifestyle intervention (SSPANLI): development and pilot. Spine J 2015 Apr 01;15(4):577-586. [doi: 10.1016/j.spinee.2014.10.015] [Medline: 25452012]
55. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 02;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
56. Vorrink SN, Kort HS, Troosters T, Lammers JW. A mobile phone app to stimulate daily physical activity in patients with chronic obstructive pulmonary disease: development, feasibility, and pilot studies. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e11 [FREE Full text] [doi: 10.2196/mhealth.4741] [Medline: 26813682]
57. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792. [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
58. Waterlander W, Whittaker R, McRobbie H, Dorey E, Ball K, Maddison R, et al. Development of an evidence-based mHealth weight management program using a formative research process. JMIR Mhealth Uhealth 2014 Jul 11;2(3):e18 [FREE Full text] [doi: 10.2196/mhealth.2850] [Medline: 25098337]
59. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
60. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
61. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391. [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
62. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 01;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
63. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]
64. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
65. Centre for Reviews and Dissemination. Core principles and methods for conducting a systematic review of health interventions. In: Systematic Reviews: CRD's guidance for undertaking reviews in health care. UK: CRD, University of York; Jan 2009.
66. Flick U. An introduction to qualitative research, 5th edition. London: SAGE Publications; 2014.
67. Lentferink A, Oldenhuis H, Kulyk O, De Groot M, Polstra L, Velthuijsen H, et al. Self-tracking and Persuasive eCoaching in Healthy Lifestyle Interventions: Work-in-progress Scoping Review of Key Components. Cham: Springer International Publisher AG; 2016 Presented at: Fourth International Workshop on Behavior Change Support Systems (BCSS'16); 2016 April 5-7; Salzburg, Austria p. 15-35 URL: http://ceur-ws.org/Vol-1573/

68. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
69. 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]
70. 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]
71. Park LG, Beatty A, Stafford Z, Whooley MA. Mobile phone interventions for the secondary prevention of cardiovascular disease. Prog Cardiovasc Dis 2016;58(6):639-650. [doi: 10.1016/j.pcad.2016.03.002] [Medline: 27001245]
72. Harris PR, Sillence E, Briggs P. The effect of credibility-related design cues on responses to a web-based message about the breast cancer risks from alcohol: randomized controlled trial. J Med Internet Res 2009 Aug 25;11(3):e37 [FREE Full text] [doi: 10.2196/jmir.1097] [Medline: 19709989]
73. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 1;2(1):91-101. [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
74. Stawarz K, Cox AL, Blandford A. Beyond self-tracking and reminders: designing smartphone apps that support habit formation. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York: ACM; 2015 Presented at: CHI '15; 2015 April 18-23; Seoul, Republic of Korea. [doi: 10.1145/2702123.2702230]
75. Whitton AE, Proudfoot J, Clarke J, Birch MR, Parker G, Manicavasagar V, et al. Breaking open the Black Box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015 Mar 04;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]
76. 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]
77. Cheng Q, Church J, Haas M, Goodall S, Sangster J, Furber S. Cost-effectiveness of a population-based lifestyle intervention to promote healthy weight and physical activity in non-attenders of cardiac rehabilitation. Heart Lung Circ 2016 Mar;25(3):265-274. [doi: 10.1016/j.hlc.2015.07.002] [Medline: 26669813]
78. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: a qualitative inquiry. J Telemed Telecare 2016 Apr;22(3):172-178. [doi: 10.1177/1357633X15595178] [Medline: 26199275]
79. Shaw RJ, Steinberg DM, Bonnet J, Modarai F, George A, Cunningham T, et al. Mobile health devices: will patients actually use them? J Am Med Inform Assoc 2016 May;23(3):462-466. [doi: 10.1093/jamia/ocv186] [Medline: 26911820]
80. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
81. Henry BL, Moore DJ. Preliminary findings describing participant experience with iSTEP, an mHealth intervention to increase physical activity and improve neurocognitive function in people living with HIV. J Assoc Nurses AIDS Care 2016;27(4):495-511. [doi: 10.1016/j.jana.2016.01.001] [Medline: 26847379]
82. Greenfield R, Busink E, Wong CP, Riboli-Sasco E, Greenfield G, Majeed A, et al. Truck drivers' perceptions on wearable devices and health promotion: a qualitative study. BMC Public Health 2016 Jul 30;16:677 [FREE Full text] [doi: 10.1186/s12889-016-3323-3] [Medline: 27475984]
83. Glanz K, Rimer BK, Viswanath K, editors. Health behavior and health education: theory, research, and practice. 4th ed. San Francisco: John Wiley & Sons; 2008.
84. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
85. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychol 2008 Oct;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
86. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011 Jul 22;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]
87. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]
88. 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]
89. Vorderstrasse A, Lewinski A, Melkus GD, Johnson C. Social support for diabetes self-management via eHealth interventions. Curr Diab Rep 2016 Jul;16(7):56. [doi: 10.1007/s11892-016-0756-0] [Medline: 27155606]

90. Wildeboer G, Kelders SM, van Gemert-Pijnen JE. The relationship between persuasive technology principles, adherence and effect of web-Based interventions for mental health: A meta-analysis. Int J Med Inform 2016 Dec;96:71-85. [doi: 10.1016/j.ijmedinf.2016.04.005] [Medline: 27117057]
91. 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 09;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
92. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016 Jun;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
93. Riley WT, Rivera DE. Methodologies for optimizing behavioral interventions: introduction to special section. Transl Behav Med 2014 Sep;4(3):234-237 [FREE Full text] [doi: 10.1007/s13142-014-0281-0] [Medline: 25264463]
94. Sieverink F, Kelders SM, Akkersdijk S, Poel M, Siemons L, van Gemert-Pijnen J. Work in Progress: A Protocol for the Collection, Analysis, and Interpretation of Log Data from eHealth Technology. Cham: Springer International Publisher AG; 2016 Presented at: Fourth International Workshop on Behavior Change Support Systems co-located with the 11th International Conference on Persuasive Technology; 2016 April 5-7; Salzburg, Austria.
95. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. International Journal of CoCreation in Design and the Arts 2008;4(1):5-18. [doi: 10.1080/15710880701875068]
96. Van Gemert-Pijnen J, Peters O, Ossebaard H, editors. Improving eHealth. The Netherlands: Eleven International Publishing; 2013.
97. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
98. Spanakis EG, Santana S, Tsiknakis M, Marias K, Sakkalis V, Teixeira A, et al. Technology-based innovations to foster personalized healthy lifestyles and well-being: a targeted review. J Med Internet Res 2016 Jun 24;18(6):e128 [FREE Full text] [doi: 10.2196/jmir.4863] [Medline: 27342137]

Abbreviations

CeHRes roadmap: Center for eHealth Research roadmap
PSD model: persuasive system design model
RCT: randomized controlled trial
SMS: Short Message Service

Edited by G Eysenbach; submitted 09.01.17; peer-reviewed by JR Bautista, K Ng, J Hensel, J Apolinário-Hagen; comments to author 23.02.17; revised version received 03.04.17; accepted 05.06.17; published 01.08.17.

Please cite as:
Lentferink AJ, Oldenhuis HKE, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JEWC
Key Components in eHealth Interventions Combining Self-Tracking and Persuasive eCoaching to Promote a Healthier Lifestyle: A Scoping Review
J Med Internet Res 2017;19(8):e277
URL: http://www.jmir.org/2017/8/e277/
doi:10.2196/jmir.7288
PMID:28765103

Original Paper

Behavior Change Techniques in Physical Activity eHealth Interventions for People With Cardiovascular Disease: Systematic Review

Orlaith Mairead Duff[1], BSc (Hons); Deirdre MJ Walsh[1], MSc, PhD; Bróna A Furlong[1], PhD; Noel E O'Connor[2], PhD; Kieran A Moran[3], PhD; Catherine B Woods[4], PhD

[1]MedEx Wellness, School of Health and Human Performance and Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland

[2]Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland

[3]School of Health and Human Performance and Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland

[4]Department of Physical Education and Sport Sciences, Faculty of Education and Health Sciences, University of Limerick, Limerick, Ireland

Corresponding Author:
Kieran A Moran, PhD
School of Health and Human Performance and Insight Centre for Data Analytics
Dublin City University
Room A246 (Albert College)
School of Health and Human Performance, Dublin City University
Dublin, D9
Ireland
Phone: 353 1 700 ext 8011
Fax: 353 7008888
Email: kieran.moran@dcu.ie

Abstract

Background: Cardiovascular disease (CVD) is the leading cause of premature death and disability in Europe, accounting for 4 million deaths per year and costing the European Union economy almost €196 billion annually. There is strong evidence to suggest that exercise-based secondary rehabilitation programs can decrease the mortality risk and improve health among patients with CVD. Theory-informed use of behavior change techniques (BCTs) is important in the design of cardiac rehabilitation programs aimed at changing cardiovascular risk factors. Electronic health (eHealth) is the use of information and communication technologies (ICTs) for health. This emerging area of health care has the ability to enhance self-management of chronic disease by making health care more accessible, affordable, and available to the public. However, evidence-based information on the use of BCTs in eHealth interventions is limited, and particularly so, for individuals living with CVD.

Objective: The aim of this systematic review was to assess the application of BCTs in eHealth interventions designed to increase physical activity (PA) in CVD populations.

Methods: A total of 7 electronic databases, including EBSCOhost (MEDLINE, PsycINFO, Academic Search Complete, SPORTDiscus with Full Text, and CINAHL Complete), Scopus, and Web of Science (Core Collection) were searched. Two authors independently reviewed references using the software package Covidence (Veritas Health Innovation). The reviewers met to resolve any discrepancies, with a third independent reviewer acting as an arbitrator when required. Following this, data were extracted from the papers that met the inclusion criteria. Bias assessment of the studies was carried out using the Cochrane Collaboration's tool for assessing the risk of bias within Covidence; this was followed by a narrative synthesis.

Results: Out of the 987 studies that were identified, 14 were included in the review. An additional 9 studies were added following a hand search of review paper references. The average number of BCTs used across the 23 studies was 7.2 (range 1-19). The top three most frequently used BCTs included information about health consequences (78%, 18/23), goal setting (behavior; 74%, 17/23), and joint third, self-monitoring of behavior and social support (practical) were included in 11 studies (48%, 11/23) each.

Conclusions: This systematic review is the first to investigate the use of BCTs in PA eHealth interventions specifically designed for people with CVD. This research will have clear implications for health care policy and research by outlining the BCTs used in eHealth interventions for chronic illnesses, in particular CVD, thereby providing clear foundations for further research and developments in the area.

(J Med Internet Res 2017;19(8):e281) doi:10.2196/jmir.7782

KEYWORDS

systematic review; exercise; behavior; telemedicine; cardiovascular disease

Introduction

Cardiovascular disease (CVD) is the leading cause of mortality worldwide, accounting for 30% of global deaths and 48% of deaths in Europe [1]. Cardiac rehabilitation (CR), which is used to reduce the impact of CVD and to promote healthy behaviors and active lifestyles for those with CVD [2], has been shown to improve physical health and decrease subsequent morbidity and mortality rates in CVD populations [3]. The main modality of CR is exercise. Two systematic reviews of exercise-based CR, which included 48 randomized controlled trials (RCTs), showed a 20% reduction in all-cause mortality and a 27% reduction in cardiac mortality at 2 to 5 years [4,5].

The efficacy of standard CR has been extensively reviewed. In terms of mortality rates, a systematic review and meta-analysis of 34 RCTs (n=6111 myocardial infarction patients) showed that those who attended CR had a lower risk of all-cause mortality than non-attendees (odds ratio=0.74, 95% CI 0.58-0.95) [6]. With respect to hospital admissions, a Cochrane review of 33 RCTs (n=4740 patients with heart failure) showed that CR reduced the risk of overall hospitalization (relative risk, RR=0.75, 95% CI 0.62-0.92; absolute risk reduction, ARR=7.1%; number needed to treat, NNT=15) and hospitalization for heart failure (RR=0.61, 95% CI 0.46-0.80; ARR=5.8%; NNT=18) [7]. A US observational study (n=635 coronary heart disease [CHD] patients) reported improvements in depression, anxiety, and hospital scores after CR [8]. CR has also been found to improve psychological well-being and to facilitate an improvement in quality of life. One of the most significant benefits of CR exercise training to participants is the improvement in aerobic capacity and cardiorespiratory fitness [9].

Even though CR has been shown to be effective, adherence to these programs is generally suboptimal. Participation rates in CR are documented at less than 50% worldwide [10]. Results from a Cochrane systematic review revealed that common barriers to adherence to CR programs included accessibility and parking at local hospitals, a dislike of group environments and work or domestic commitments [3]. In 2012, a *Heart* journal editorial concluded that CR should not only focus on content such as CHD risk factor modification and medication adherence but should also focus on the delivery mechanisms, thereby offering a range of different delivery methods for people according to their preferences and needs and potentially addressing the issue of low levels of participation [11]. The delivery of CR to date has largely been center-based, either in hospitals or community centers. However, in more recent times, there has been a shift toward a more home-based model of care. A systematic review by Dalal and colleagues [3] found that both home- and center-based forms of CR are equally effective in improving clinical and health-related quality of life outcomes in patients with CVD, suggesting the further provision of additional evidence-based home CR programs. A Cochrane review found that home-based interventions may be superior in terms of adherence to exercise, especially in the long term [12]. This would ensure that patients are given the choice of participating in a more traditional supervised center-based program or a home-based program, based on their personal preference.

The emerging area of electronic health (eHealth), defined as the use of information and communication technologies (ICTs) for health [13] may provide this alternative home-based delivery method. Interventions that encompass ICT (eg, Internet- and mobile-based communications and wearable monitors) enable the efficient delivery of educational resources, individually tailored health and wellness programs, as well as time-unlimited feedback, coaching and support [14]. Technology solutions for physical activity (PA) uptake and monitoring are being undertaken as a new mode of facilitating behavior change and may impact the current delivery of CR [15]. Telerehabilitation solutions refer to the use of ICT to provide rehabilitation services to people. Literature in this area for cardiac patients indicates that such interventions are feasible and effective when compared with conventional center-based CR [16].

Furthermore, eHealth interventions have been showing promising results in CR, supporting behavior change, clinical improvement, and improved social functioning. In 2013, Beatty and colleagues [17] conducted a review of mobile interventions for CR, identifying only three studies for inclusion. More recently, the interest in eHealth and mobile health (mHealth) has risen dramatically, indicating the increased focus in this field over recent years. Buys and colleagues [15] investigated the interest among cardiac patients in technology-enabled cardiovascular rehabilitation. Of the 298 patient (77% male; mean age 61.7 years [SD 14.5]) questionnaires included in the analysis, 97% had a mobile phone and 91% used the Internet. PA monitoring was reported by 12% of the respondents. Overall, cardiac patients showed a high interest in CR support through the Internet (77%) and mobile phones (68%). These findings suggest that patients with CVD show an interest in technology-enabled home-based CR, potentially allowing exercise-based rehabilitation programs to be more effective by making them more accessible, personalized, and more interactive with patients.

Behavior change techniques (BCTs) are integral to the design of complex health service interventions such as CR. A BCT is defined as “an observable, replicable, and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior, that is, a technique is proposed to be an ‘active ingredient’” [18]. The Medical Research Council (MRC) guidelines recommend the application of behavior change theory within complex health service interventions to allow for a theoretical understanding of behavior change [19]. The National Institute for Health and Care Excellence (NICE; [20]) guidelines on individual-level behavior change interventions aimed at changing health-damaging behaviors such as unhealthy diet,

physical inactivity, excessive alcohol consumption, unsafe sex, and smoking, recommend the use of evidence-based BCTs, which have been proven to be effective at changing behavior such as goals and planning, feedback and monitoring, and social support. Despite this guidance, few interventions pay close attention to the behavior change theory and techniques used to design their interventions. In particular, the poor description of interventions in research protocols and published reports presents a barrier for future design of complex interventions [21], as it is difficult to identify the active and effective components of the intervention [18]. The proliferation of eHealth interventions requires the coding of such interventions to facilitate future research to compare accurately across interventions. With that in mind, this systematic review aims to identify the key BCTs applied in eHealth PA interventions for adults with CVD.

Methods

This systematic review is reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance. The inclusion criteria for studies were as follows: human- and quasi-RCTs, published and unpublished, of PA eHealth interventions for adults (aged ≥18 years) clinically diagnosed with CVD. Studies were included if the main intervention component was delivered via a computer, mobile phone, tablet, or phone (eg, mobile phone app, emails, text messages, and phone calls) with the primary or secondary aim of increasing the PA level of the user. The interventions could be delivered to groups or individuals. The inclusion criteria was kept quite broad to identify as many studies as possible with PA as a primary or secondary outcome, as well as studies that had PA as a component of the intervention.

The behavior change taxonomy version 1 was used to identify the specific BCTs used within the included studies [18]. Two researchers coded for the BCTs using the taxonomy.

Outcome Measures

A description of the BCTs and their frequency of use in the 23 eHealth interventions reviewed were classified using a BCT taxonomy by Michie and colleagues. Due to the heterogeneous nature of the studies differing in PA outcome measures and time points, we were unable to carry out a meta-analysis examining the effectiveness of the BCTs in relation to the PA outcomes.

Search Methods for the Identification of Studies

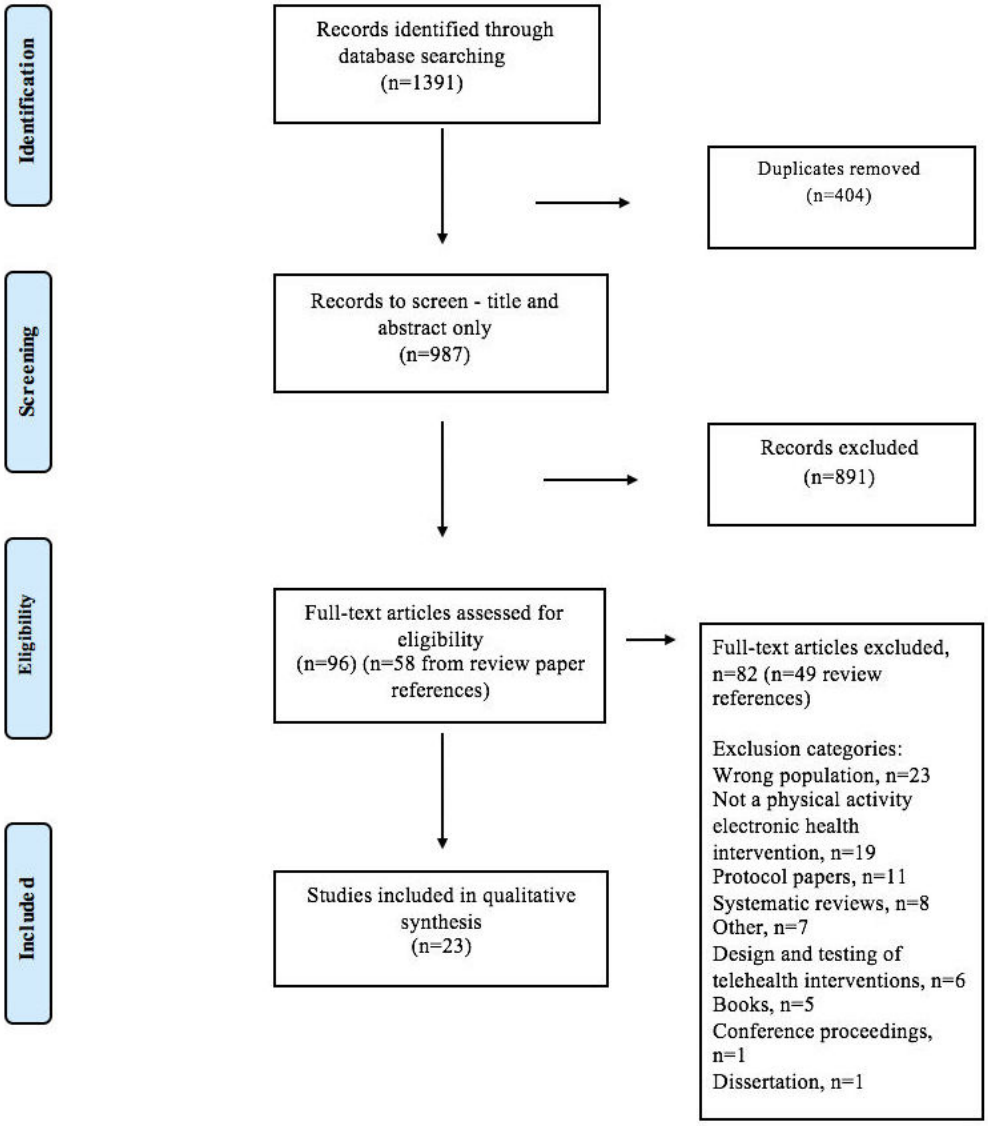
Seven electronic databases were searched, including MEDLINE (via EBSCOhost, 2000-2016), PsycINFO (via EBSCOhost, 2000-2016), Academic Search Complete (via EBSCOhost, 2000-2016), SPORTDiscus (via EBSCOhost, 2000-2016), CINAHL Complete (via EBSCOhost, 2000-2016), Scopus (2000-2016) and Web of Science (Core Collection; 2000-2016).

The search was restricted to studies published in English between 2000 and 2016. The search strategy used keywords relating to PA, eHealth interventions, CVD and adults, as well as appropriate synonyms. Boolean operators were used to expand, exclude, or join keywords in the search, using the terms “AND” and “OR.” In all the databases, the searches were limited to the fields of abstract and title only. The search strategy for all databases is illustrated in the Multimedia Appendix 1.

Selection of Studies

Figure 1 shows the PRISMA flow diagram of reviewed and included studies. One researcher conducted the database search. All the studies identified following the database search were then uploaded to the Web-based systematic review software package “Covidence” (Veritas Health Innovation). First, a title and abstract review of all studies was completed independently by 2 authors. Any disagreements were discussed until a consensus was reached, or a third reviewer helped to resolve the discrepancy. A record was kept of all the studies excluded and the reason for exclusion via Covidence. Second, all the studies that met the inclusion criteria went through a full-text screening process by the 2 authors independently. Again, any disagreements between the authors on the eligibility of the studies were reviewed by a third author. Additional studies were also identified for inclusion by reviewing the reference lists of review papers through a hand search.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of reviewed and included studies.



Data Extraction

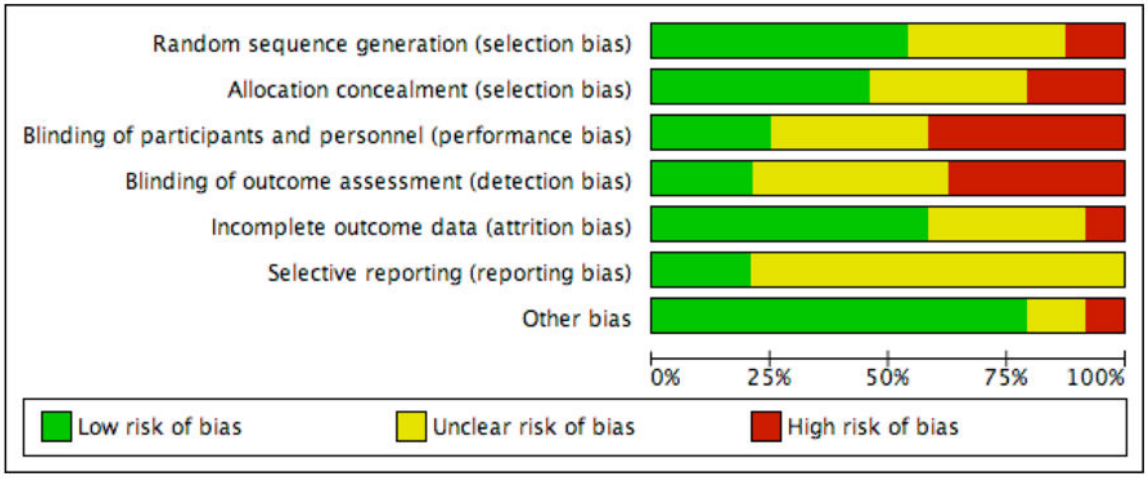
Data from the studies were extracted independently by 2 review authors using a data extraction template. Data extracted from the studies included study title, authors, country, year, patient group (sample size), inclusion criteria, study design, technology involvement, assessment, intervention details, outcomes, theory involved, BCTs identified, and results. No blinding to study author, institution, or journal occurred during the screening process for the study.

If multiple publications of the same study were identified, the team would try to extract and combine all the available data; where there was doubt, the original publication would be given priority. If data seemed to be missing from a study, we tried to obtain the missing data through correspondence with the study authors. The review team resolved any disagreements regarding study eligibility through group discussion.

Assessment of Risk Bias

Two reviewers assessed each study for risk of bias (high, low, or unclear) using the Cochrane Collaboration's risk of bias tool [22]. A third review author acted as arbitrator, if necessary. The results of the risk of bias assessment were then exported to RevMan to create a visual representation of the publication bias (see Figure 2).

Figure 2. Risk of bias graph: review authors' judgments about each risk of bias item presented as percentages across all included studies.



Assessing for Heterogeneity

Diversity across the studies was assessed qualitatively in terms of eHealth intervention, patient characteristics, and outcome measures.

Data Synthesis

Following the extraction of data from the studies, careful consideration was given to the appropriateness of conducting a meta-analysis. As the studies were too heterogeneous to combine statistically, the data were synthesized qualitatively.

Behavior Change Techniques

To gain an understanding of the types of BCTs used in PA eHealth interventions in this patient population, 2 authors screened the included studies and coded the BCTs used in each study using Michie and colleagues BCT taxonomy [18].

Results

The search criteria returned 1391 studies through the database search. A total of 404 duplicates were removed, leaving 987 studies to screen. The title and abstracts of the studies were then screened by 2 reviewers, resulting in 891 records excluded for not meeting the inclusion criteria. The authors reviewed the full text of 96 studies, identifying 14 studies for inclusion in this review. From a hand search of review paper references, an additional 58 studies were identified as potentially eligible. Following a full-text review of these papers, nine studies were included in the review. Therefore, a total of 23 studies were included in the qualitative synthesis.

Study Characteristics

For an overview of the included studies and the PA results, see Multimedia Appendix 2. Of the 23 studies included, 14 studies comprised an Internet or Web-based program and/or mobile phone intervention [23-36], three studies were telephone interventions [37-39], two studies used a telehealth device [40,41], and the remaining two studies comprised a form of telemonitoring [42,43]. Single studies comprising videoconferencing [44] and virtual reality wraparound screens [45] were also found. Of the 20 studies with a control group, 17 involved "usual care" as the control. Usual care predominately pertained to receiving standard CR services [25-29,31,33-37,39-42,44,45]. Eight studies were carried out in Europe [23,24,27,28,30,32,39,43], whereas seven of the studies were conducted in North America/South America [29,32,34,36,37,40,41]. Three studies apiece were conducted in Australia [25,35,38] and New Zealand [26,31,44], and two studies were conducted in Asia [25,42].

The majority of participants were recruited from hospitals/medical centers [23-27,29,30,32-41,43,45]. One study recruited participants from a general practitioner CHD registry [30], whereas another recruited from a CR referral list [38]. Tomita and colleagues [34] recruited participants from 3 hospitals and 2 health insurance companies. One study recruited participants from primary and community health services [35]. Outcomes were assessed from 3 weeks [41] to 16 months [39], with the average end point across the 23 studies at 4.5 months.

Behavior Change Techniques

Only two out of the 23 studies explicitly mentioned the BCTs applied [26,27]. From the other studies, 2 reviewers coded the BCTs from the program description. Multimedia Appendix 3 outlines the number of BCTs used in each study as well a comprehensive list of the techniques used. The average number of BCTs employed in the included studies was 7.2 (range: 1-14). The top three most frequently used BCTs were information about health consequences (78%; 18/23), goal setting (behavior; 74%, 17/23) and joint third, self-monitoring of behavior and social support (practical) (48%; 11/23 each) (see Multimedia Appendix 4). The Text4Heart study conducted by Dale and colleagues [26] employed 14 BCTs—using the maximum number of BCTs out of all the studies. These were goal setting (behavior), problem solving, review outcome goals, feedback on behavior, self-monitoring of behavior, social support (unspecified), instruction on how to perform the behavior, information about health consequences, demonstration of the behavior, social comparison, prompts/cues, graded tasks, credible source, and reduce negative emotions. A study by Barnason and colleagues [41] used the lowest number of BCTs

of the 23 studies included in the review, employing just one BCT - graded tasks.

The most common BCT group used in the 23 included studies was feedback and monitoring, whereas the second most common group was goals and planning. This was followed by social support. Four groups that did not appear in any of the 23 included studies were identity, scheduled consequences, self-belief, and covert learning.

Multimedia Appendix 4 outlines the frequency of use of the BCTs across the 23 studies, the BCT taxonomy group, and an example of how a BCT was incorporated into a study. Only two BCTs were used in over 70% of the studies; these were 5.1 information about health consequences (78%) and 1.1 goal setting (behavior; 74%). Additionally, four BCTs were used in over 40% of the studies; these included 2.2 feedback on behavior (43%), 2.3 self-monitoring of behavior (48%), 3.2 social support (practical; 48%), and 4.1 instruction on how to perform the behavior (43%). Several BCTs, including 10.3 nonspecific reward, 12.1 restructuring the physical environment, 12.5 adding objects to the environment, 11.1 pharmacological support, 6.1 demonstration of the behavior, 6.2 social comparison, 1.7 review outcome goals, 10.4 social reward, and 1.8 behavioral contract were only used in one study (see Multimedia Appendix 4 for more details).

Table 1. Frequency of behavior change techniques (BCTs) used in the studies with improved physical activity (PA) outcome.

Behavior change technique label	Total number of studies N=8 n (%)
1.1 Goal setting (behavior)	6 (75)
5.1 Information about health consequences	6 (75)
2.2 Feedback on behavior	5 (63)
4.1 Instruction on how to perform the behavior	5 (63)
2.3 Self-monitoring of behavior	4 (50)
3.2 Social support (practical)	4 (50)
3.1 Social support (unspecified)	3 (38)
9.1 Credible source	3 (38)
1.2 Problem solving	2 (25)
1.5 Review behavior goals	2 (25)
3.3 Social support (emotional)	2 (25)
7.1 Prompts/cues	2 (25)
8.7 Graded tasks	2 (25)
11.2 Reduce negative emotions	2 (25)
1.4 Action planning	1 (13)
2.4 Self-monitoring of outcomes of behavior	1 (13)
2.6 Biofeedback	1 (13)
2.7 Feedback on outcomes of behavior	1 (13)
10.4 Social reward	1 (13)
11.1 Pharmacological support	1 (13)
1.3 Goal setting (outcome)	0 (0)
1.7 Review outcome goals	0 (0)
1.8 Behavioral contract	0 (0)
2.1 Monitoring of behavior by others without feedback	0 (0)
2.5 Monitoring of outcomes of behavior without feedback	0 (0)
6.1 Demonstration of the behavior	0 (0)
6.2 Social comparison	0 (0)
10.3 Nonspecific reward	0 (0)
12.1 Restructuring the physical environment	0 (0)
12.5 Adding objects to the environment	0 (0)

Behavior Change Techniques Linked to Improved Physical Activity Outcomes

Eight of the 15 interventions that had PA as an outcome measure reported statistically significant improvements in PA between the experimental and control groups. Goal setting (behavior) and information about health consequences were the most frequently used BCTs across the eight studies (n=6 each). This was followed by feedback on behavior and instruction on how to perform the behavior, which were incorporated in five studies each. The following BCTs were also included in the interventions that had an improved PA outcome at the final end point: self-monitoring of behavior, social support (practical), social support (unspecified), credible source, problem solving, review behavior goals, social support (emotional), prompts/cues, graded tasks, reduce negative emotions, action planning, self-monitoring of outcomes of behavior, biofeedback, feedback on outcomes of behavior, social reward, and pharmacological support (Table 1).

It is worth noting that the interventions that did not demonstrate a significant increase in PA (n=5) were at par with the level achieved in standard CR, as no significant differences between the control and experimental groups were found. This is an important finding, as it highlights the fact that the eHealth interventions were at par with or were significantly better at improving PA levels of cardiac patients when compared with standard cardiac services. This emphasizes the potential of eHealth interventions in a CR setting.

To further examine the efficacy of the individual BCTs, the interventions were grouped into four groups depending on whether PA was measured objectively or subjectively and whether there was a difference between experimental and control groups. Once the interventions were grouped, we sought to examine whether there were any common BCTs used across the studies (see Multimedia Appendix 5). This task allowed us to examine whether there were any similarities between the interventions in terms of the BCTs they employed. Objective and self-report studies with no difference between experimental and control groups were the only groups with similarities in the BCTs they employed. Social support (practical) and information about health consequences were employed in all self-report studies where there was no PA difference between the experimental and control groups. Goal setting (behavior) and feedback on behavior were employed in all PA objectively measured interventions where no significant difference was found between groups at the final end point. However, there were no similarities in the BCTs used across all the effective interventions, regardless of whether PA was measured objectively or subjectively. Furthermore, the average number of BCTs used across significant interventions did not differ, as the studies that increased PA versus those that did not increase PA employed an average of seven BCTs.

Discussion

Summary

This systematic review comprised 23 studies reviewing the use of BCTs in PA eHealth interventions for adults with CVD. To our knowledge, this is the first review that aimed to identify the use of Michie and colleagues behavior change taxonomy in PA eHealth intervention studies among this population. The findings of the review indicate that on an average, 7.2 BCTs were employed in the 23 studies. Information about health consequences was the most frequently used technique, with 78% of the studies incorporating this technique into their intervention. This was followed closely by goal setting (behavior), which was used in 74% of the studies, with self-monitoring of behavior and social support (practical) each employed in 48% of the studies.

Although Michie and colleagues BCT taxonomy is comprised of 93 different techniques, the maximum number of techniques used in a single intervention was 14 [26]. These were goal setting (behavior), problem solving, review outcome goals, feedback on behavior, self-monitoring of behavior, social support (unspecified), instruction on how to perform the behavior, information about health consequences, demonstration of the behavior, social comparison, prompts/cues, graded tasks, credible source, and reduce negative emotions. The minimum number of techniques used in a study was one - graded tasks [41]. A failing of the studies included in this review was the poor description of the intervention components. Only two studies in the review specifically mentioned the BCTs incorporated in their interventions [26,27]. However, even though the paper by Devi and colleagues [27] listed the BCTs used, it failed to link the BCTs used to the intervention functions or components. In the study by Dale [26], the researchers provided only examples of text messages linked to BCTs. None of the studies gave a full account of the BCTs used in their studies and how these were linked to the intervention components. This finding is in line with previous research, where reviews of nearly 1000 behavior change outcome studies found that interventions were fully and accurately described in only 5% to 30% of experimental studies [46-49]. Overall, this lack of robust and detailed information on the intervention functions provides a significant barrier to better understand the effects and mechanisms of behavior change interventions and to inform the development of more effective interventions in the future [16].

Another key issue relating to the poor description of behavior change interventions is the inconsistent use of terminology. This variation in terminology used makes the coding of the techniques used even more difficult when reviewing behavior change interventions. For example, social support (unspecified) was coded for in 39% of the studies included in the review by the reviewers. Terminology varied across the studies where social support was coded; for example, one study used a social reinforcement network [24], another incorporated mentors into their intervention [35], whereas another study involved tutorials in their intervention [33]. The reviewers coded these examples as social support (unspecified); however, this BCT was not specifically mentioned in any of the studies. Therefore, there is a need to have consistent terminology and sufficient information on intervention components to allow for the replication of interventions that have been found to be effective. The lack of such information appears to be particularly problematic in behavioral interventions rather than in pharmacological ones

[21]. In a workshop, 26 multidisciplinary researchers were presented with behavioral or pharmacological intervention protocols and were asked whether the protocol provided sufficient information so that the study could be replicated in a practice setting. The researchers were less confident that they could replicate the behavioral interventions compared with the pharmacological interventions (t=6.45, P<.001) and concluded they would need more information for the replication of behavioral interventions (U=35.5, P=.02; [50]).

This review provides new and important information regarding the use of BCTs in eHealth PA for adults with CVD, highlighting the frequent use of the following BCTs: information about health consequences, goal setting (behavior), self-monitoring of behavior and social support (practical). However, it is clear that more robust and comprehensive interventions are needed, which systematically and coherently detail the BCTs used in the interventions. Identifying the active ingredients of the interventions will enable researchers to examine the effectiveness of these key intervention components, ensuring that the most effective BCTs are used regarding eHealth PA interventions for adults with CVD.

Strengths and Limitations

A major strength of this review was the authors' attempt to identify all relevant studies by using a comprehensive search strategy and multiple databases. The authors' also hand-searched review paper references to identify any additional studies that may have been relevant to the review. All the studies identified following the database search were then uploaded to the Web-based systematic review software package "Covidence" (Veritas Health Innovation). This allowed for a systematic and comprehensive approach to screening the studies and coding the reasons for exclusion. This software also enabled the screening for risk of bias in a simple and efficient way. From this, a visual representation of the publication bias was produced using RevMan.

A limitation of this review was the wide variability among the studies included, with study designs ranging from RCTs, to feasibility studies and pilot trials. However, it was necessary to include all the studies and not just RCTs to identify as many PA eHealth interventions as possible. There was also a lack of consistency in the measurement of PA across the studies, from subjective to objective assessments. The follow-up duration also varied significantly from 3 weeks to 16 months. This meant that it was impossible to pool the results in a meta-analysis.

Many studies measured the physical fitness of their participants, as opposed to their PA levels. Although all the interventions had a PA/exercise component to their eHealth intervention, some studies did not directly measure the PA level of the participants. We can therefore only infer from the studies that by increasing PA behavior, the physical fitness outcome improved. This inference of a causal relationship between PA and physical fitness is a limitation to these studies. Another limitation is the variety of methods used to measure PA, meaning that the comparison between studies is challenging and therefore determining the impact of specific BCTs is impossible.

Implications for Research and Practice

This systematic review highlights the need for more robust and comprehensive eHealth PA interventions for adults with CVD. Although the most frequently used BCTs were identified, it is worth noting that the majority of studies did not specifically detail the active ingredients of their interventions. Further work is also needed to determine the most appropriate measurement of PA among this population so that interventions use the best subjective and/or objective measurements, thereby ensuring that comparisons can be easily drawn across studies. This review also highlights the importance of identifying the BCTs used within a study and their link to the intervention components to understand the ingredients that bring about the desired behavior change. It is only by identifying these mechanisms of change that we can understand why an intervention was found to be effective or not.

Acknowledgments

This research has been funded by the Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289 and our industry partner, Acquis-bi based in Los Gatos, California.

Authors' Contributions

OD ran the keyword search in the chosen databases, reviewed all the papers for inclusion and exclusion, and also drafted, in large part, the first version of the manuscript. DW was the second reviewer who reviewed the papers for inclusion and exclusion in the review. DW and OD independently extracted data from the final papers for inclusion. CW was the third reviewer, if any discrepancies occurred between OD and DW in the review and data extraction processes. CW, DW, and BF revised and provided feedback on the drafts on the manuscript. KW and NOC also provided feedback on the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Keyword searches.

[PDF File (Adobe PDF File), 29KB - jmir_v19i8e281_app1.pdf]

Multimedia Appendix 2

Information on included studies.

[PDF File (Adobe PDF File), 93KB - jmir_v19i8e281_app2.pdf]

Multimedia Appendix 3

Behavior change techniques used in the included studies.

[PDF File (Adobe PDF File), 93KB - jmir_v19i8e281_app3.pdf]

Multimedia Appendix 4

Frequency of behavior change techniques (BCTs) used in the included studies.

[PDF File (Adobe PDF File), 36KB - jmir_v19i8e281_app4.pdf]

Multimedia Appendix 5

Link between Interventions and behavior change techniques.

[PDF File (Adobe PDF File), 66KB - jmir_v19i8e281_app5.pdf]

References

1. Allender S, Scarborough P, Peto V, Rayner M, Leal J, Luengo-Fernandez R, et al. BHF. 2008. European cardiovascular disease statistics 2008 URL: https://www.bhf.org.uk/publications/statistics/european-cardiovascular-disease-statistics-2008 [accessed 2017-07-26] [WebCite Cache ID 6sEpXW3WO]
2. Balady G, Ades PA, Comoss P, Limacher M, Pina IL, Southard D, et al. Core components of cardiac rehabilitation/secondary prevention programs: a statement for healthcare professionals from the American heart association and the American association of cardiovascular and pulmonary rehabilitation writing group. Circulation 2000 Aug 29;102(9):1069-1073 [FREE Full text] [Medline: 10961975]
3. Dalal H, Zawada A, Jolly K, Moxham T, Taylor RS. Home based versus centre based cardiac rehabilitation: Cochrane systematic review and meta-analysis. BMJ 2010 Jan 19;340:b5631 [FREE Full text] [doi: 10.1136/bmj.b5631]
4. Jolliffe JA, Rees K, Taylor RS, Thompson D, Oldridge N, Ebrahim S. Exercise-based rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2001(1):CD001800. [doi: 10.1002/14651858.CD001800] [Medline: 11279730]
5. Taylor R, Brown A, Ebrahim S, Jolliffe J, Noorani H, Rees K, et al. Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. Am J Med 2004 May 15;116(10):682-692. [doi: 10.1016/j.amjmed.2004.01.009] [Medline: 15121495]
6. Lawler P, Filion K, Eisenberg M. Efficacy of exercise-based cardiac rehabilitation post-myocardial infarction: a systematic review and meta-analysis of randomized controlled trials. Am Heart J 2011 Oct;162(4):571-584.e2. [doi: 10.1016/j.ahj.2011.07.017] [Medline: 21982647]
7. Sagar VA, Davies EJ, Briscoe S, Coats AJS, Dalal HM, Lough F, et al. Exercise-based rehabilitation for heart failure: systematic review and meta-analysis. Open Heart 2015;2(1):e000163 [FREE Full text] [doi: 10.1136/openhrt-2014-000163] [Medline: 25685361]
8. Lavie CJ, Milani RV. Adverse psychological and coronary risk profiles in young patients with coronary artery disease and benefits of formal cardiac rehabilitation. Arch Intern Med 2006 Sep 25;166(17):1878-1883. [doi: 10.1001/archinte.166.17.1878] [Medline: 17000945]
9. Menezes AR, Lavie CJ, Milani RV, Arena RA, Church TS. Cardiac rehabilitation and exercise therapy in the elderly: should we invest in the aged? J Geriatr Cardiol 2012 Mar;9(1):68-75 [FREE Full text] [doi: 10.3724/SP.J.1263.2012.00068] [Medline: 22783325]
10. Bjarnason-Wehrens B, McGee H, Zwisler A, Piepoli MF, Benzer W, Schmid J, Cardiac Rehabilitation Section European Association of Cardiovascular PreventionRehabilitation. Cardiac rehabilitation in Europe: results from the European cardiac rehabilitation inventory survey. Eur J Cardiovasc Prev Rehabil 2010 Aug;17(4):410-418. [doi: 10.1097/HJR.0b013e328334f42d] [Medline: 20300001]
11. Jones R, Furze G, Buckley J, Speck L, Carver C, Connolly S, et al. RAMIT presents an outdated version of cardiac rehabilitation. Heart 2012;98(8):672 [FREE Full text] [doi: 10.1136/heartjnl-2012-301766] [Medline: 22505465]

12. Ashworth NL, Chad KE, Harrison EL, Reeder BA, Marshall SC. Home versus center based physical activity programs in older adults. Cochrane Database Syst Rev 2005 Jan 25(1):CD004017. [doi: 10.1002/14651858.CD004017.pub2] [Medline: 15674925]
13. World Health Organisation. WHO. 2012. Legal frameworks for eHealth URL: http://apps.who.int/iris/bitstream/10665/44807/1/9789241503143_eng.pdf [accessed 2017-07-26] [WebCite Cache ID 6sErv9fRO]
14. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
15. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. BMC Med Inform Decis Mak 2016 Jul 19;16:95 [FREE Full text] [doi: 10.1186/s12911-016-0329-9] [Medline: 27431419]
16. Frederix I, Vanhees L, Dendale P, Goetschalckx K. A review of telerehabilitation for cardiac patients. J Telemed Telecare 2015 Jan;21(1):45-53. [doi: 10.1177/1357633X14562732] [Medline: 25475219]
17. Beatty A, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013 Nov 01;2(6):e000568 [FREE Full text] [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
18. 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]
19. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
20. National Institute for Health and Care Excellence. NICE. 2014. Behaviour change: individual approaches URL: https://www.nice.org.uk/guidance/ph49/resources/behaviour-change-individual-approaches-pdf-1996366337989 [accessed 2017-07-26] [WebCite Cache ID 6sEt1x24G]
21. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behaviour change interventions: the need for a scientific method. Implement Sci 2009 Jul 16;4:40 [FREE Full text] [doi: 10.1186/1748-5908-4-40] [Medline: 19607700]
22. Higgins J, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [Medline: 22008217]
23. Ammenwerth E, Woess S, Baumgartner C, Fetz B, van der Heidt A, Kastner P, et al. Evaluation of an integrated telemonitoring surveillance system in patients with coronary heart disease. Methods Inf Med 2015;54(5):388-397 [FREE Full text] [doi: 10.3414/ME15-02-0002] [Medline: 26395147]
24. Antypas K, Wangberg SC. An internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. Journal of Medical Internet Research 2014;16(3):e77 [FREE Full text] [doi: 10.2196/jmir.3132] [Medline: 24618349]
25. Chow C, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
26. Pfaeffli DL, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the text4heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
27. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]
28. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-term effectiveness of a comprehensive internet-based and patient-specific telerehabilitation program with text messaging support for cardiac patients: randomized controlled trial. J Med Internet Res 2015 Jul 23;17(7):e185 [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]
29. Lear SA, Singer J, Banner-Lukaris D, Horvat D, Park JE, Bates J, et al. Randomized trial of a virtual cardiac rehabilitation program delivered at a distance via the Internet. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):952-959 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.114.001230] [Medline: 25271050]
30. Lindsay S, Smith S, Bellaby P, Baker R. The health impact of an online heart disease support group: a comparison of moderated versus unmoderated support. Health Educ Res 2009 Aug;24(4):646-654. [doi: 10.1093/her/cyp001] [Medline: 19251770]
31. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: Results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]

32. Piotrowicz E, Korzeniowska-Kubacka I, Chrapowicka A, Wolszakiewicz J, Dobraszkiewicz-Wasilewska B, Batogowski M, et al. Feasibility of home-based cardiac telerehabilitation: results of TeleInterMed study. Cardiol J 2014;21(5):539-546. [doi: 10.5603/CJ.a2014.0005] [Medline: 24526507]
33. Reid R, Morrin LI, Beaton LJ, Papadakis S, Kocourek J, McDonnell L, et al. Randomized trial of an internet-based computer-tailored expert system for physical activity in patients with heart disease. Eur J Prev Cardiol 2012 Dec;19(6):1357-1364. [doi: 10.1177/1741826711422988] [Medline: 21903744]
34. Tomita MR, Tsai BM, Fisher NM, Kumar NA, Wilding G, Stanton K, et al. Effects of multidisciplinary internet-based program on management of heart failure. J Multidiscip Healthc 2008 Dec 01;2009(2):13-21 [FREE Full text] [Medline: 20505786]
35. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
36. Widmer R, Allison TG, Lerman LO, Lerman A. Digital health intervention as an adjunct to cardiac rehabilitation reduces cardiovascular risk factors and rehospitalizations. J Cardiovasc Transl Res 2015 Jul;8(5):283-292 [FREE Full text] [doi: 10.1007/s12265-015-9629-1] [Medline: 25946990]
37. Ades PA, Pashkow FJ, Fletcher G, Pina IL, Zohman LR, Nestor JR. A controlled trial of cardiac rehabilitation in the home setting using electrocardiographic and voice transtelephonic monitoring. Am Heart J 2000 Mar;139(3):543-548. [doi: 10.1016/S0002-8703(00)90100-5]
38. Furber S, Butler L, Phongsavan P, Mark A, Bauman A. Randomised controlled trial of a pedometer-based telephone intervention to increase physical activity among cardiac patients not attending cardiac rehabilitation. Patient Educ Couns 2010 Aug;80(2):212-218. [doi: 10.1016/j.pec.2009.11.012] [Medline: 20022201]
39. Hanssen T, Nordrehaug JE, Eide GE, Hanestad BR. Improving outcomes after myocardial infarction: a randomized controlled trial evaluating effects of a telephone follow-up intervention. Eur J Cardiovasc Prev Rehabil 2007 Jun;14(3):429-437. [doi: 10.1097/HJR.0b013e32801da123] [Medline: 17568244]
40. Artinian N, Harden JK, Kronenberg MW, Vander Wal JS, Daher E, Stephens Q, et al. Pilot study of a Web-based compliance monitoring device for patients with congestive heart failure. Heart Lung 2003 Jul;32(4):226-233 [FREE Full text] [doi: 10.1016/S0147-9563(03)00026-8] [Medline: 12891162]
41. Barnason S, Zimmerman L, Nieveen J, Schulz P, Miller C, Hertzog M, et al. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. Heart Lung 2009;38(5):364-376 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.005] [Medline: 19755186]
42. Lee Y, Hur S, Sohn J, Lee H, Park N, Cho Y, et al. Impact of home-based exercise training with wireless monitoring on patients with acute coronary syndrome undergoing percutaneous coronary intervention. J Korean Med Sci 2013 Apr;28(4):564-568 [FREE Full text] [doi: 10.3346/jkms.2013.28.4.564] [Medline: 23580444]
43. Scalvini S, Zanelli E, Comini L, Tomba MD, Troise G, Giordano A. Home-based exercise rehabilitation with telemedicine following cardiac surgery. J Telemed Telecare 2009;15(6):297-301. [doi: 10.1258/jtt.2009.090208] [Medline: 19720767]
44. Dalleck L, Schmidt LK, Lueker R. Cardiac rehabilitation outcomes in a conventional versus telemedicine-based programme. J Telemed Telecare 2011;17(5):217-221. [doi: 10.1258/jtt.2010.100407] [Medline: 21508077]
45. Chuang T, Sung W, Chang H, Wang R. Effect of a virtual reality-enhanced exercise protocol after coronary artery bypass grafting. Phys Ther 2006;86(10):1369-1377 [FREE Full text] [doi: 10.2522/ptj.20050335] [Medline: 17012641]
46. Dane A, Schneider B. Program integrity in primary and early secondary prevention: are implementation effects out of control? Clin Psychol Rev 1998 Jan;18(1):23-45. [Medline: 9455622]
47. Gresham F, Gansle KA, Noell GH. Treatment integrity in applied behavior analysis with children. J Appl Behav Anal 1993;26(2):257-263 [FREE Full text] [doi: 10.1901/jaba.1993.26-257] [Medline: 8331022]
48. Moncher F, Prinz R. Treatment fidelity in outcome studies. Clin Psychol Rev 1991 Jan;11(3):247-266 [FREE Full text] [doi: 10.1016/0272-7358(91)90103-2]
49. Odom S, Brown WH, Frey T, Karasu N, Lee Smith-Canter L, Strain PS. Evidence-based practices for young children with autism: contributions for single-subject design research. Focus Autism Other Dev Disabl 2016 Sep 14;18(3):166-175 [FREE Full text] [doi: 10.1177/10883576030180030401]
50. Michie S, Johnston M, Francis J, Hardeman W. Behaviour change interventions: developing a classification system. In: Workshop presented at the 1st annual conference of the UK society for behavioural medicine. 2005 Presented at: 1st annual conference of the UK society for behavioural medicine; 2005; London.

Abbreviations

ARR: absolute risk reduction
BCTs: behavior change techniques
CHD: coronary heart disease
CR: cardiac rehabilitation
CVD: cardiovascular disease

eHealth: electronic health
ICT: information and communication technology
mHealth: mobile health
MRC: Medical Research Council
NICE: National Institute for Health and Care Excellence
NNT: number needed to treat
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: relative risk
SFI: Science Foundation Ireland

Edited by G Eysenbach; submitted 30.03.17; peer-reviewed by A Fisher, R Sigal; comments to author 20.04.17; revised version received 14.06.17; accepted 01.07.17; published 02.08.17.

Please cite as:
Duff OM, Walsh DMJ, Furlong BA, O'Connor NE, Moran KA, Woods CB
Behavior Change Techniques in Physical Activity eHealth Interventions for People With Cardiovascular Disease: Systematic Review
J Med Internet Res 2017;19(8):e281
URL: http://www.jmir.org/2017/8/e281/
doi:10.2196/jmir.7782
PMID:28768610

Original Paper

How do eHealth Programs for Adolescents With Depression Work? A Realist Review of Persuasive System Design Components in Internet-Based Psychological Therapies

Lori Wozney[1*], PhD; Anna Huguet[1*], PhD; Kathryn Bennett[2*], PhD; Ashley D Radomski[3*], MSc; Lisa Hartling[3*], PhD; Michele Dyson[3*], PhD; Patrick J McGrath[1], PhD; Amanda S Newton[3*], PhD

[1]Centre for Research in Family Health, IWK Health Centre, Halifax, NS, Canada

[2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada

[3]Department of Pediatrics, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada

[*]these authors contributed equally

Corresponding Author:
Amanda S Newton, PhD
Department of Pediatrics
Faculty of Medicine & Dentistry
University of Alberta
3-526 Edmonton Clinic Health Academy
11405 87 Avenue
Edmonton, AB, T6G 1C9
Canada
Phone: 1 780 248 5581
Fax: 1 780 248 5625
Email: mandi.newton@ualberta.ca

Abstract

Background: Major depressive disorders are common among adolescents and can impact all aspects of their daily life. Traditional therapies, cognitive behavioral therapy (CBT), and interpersonal psychotherapy (IPT) have been delivered face-to-face. However, Internet-based (online) delivery of these therapies is emerging as an option for adolescents. Internet-based CBT and IPT involve therapeutic content, interaction between the user and the system, and different technological features embedded into the online program (eg, multimedia). Studies of Internet-based CBT and IPT for adolescent depression differ on all three aspects, and variable, positive therapy effects have been reported. A better understanding of the treatment conditions that influence therapy outcomes is important to designing and evaluating these novel therapies.

Objective: Our aim was to examine the technological and program delivery features of Internet-based CBT and IPT for adolescent depression and to document their potential relation to treatment outcomes and program use.

Methods: We performed a realist synthesis. We started with an extensive search of published and gray literature. We included intervention studies that evaluated Internet-based CBT or IPT for adolescent depression. We included mixed-methods and qualitative studies, theoretical papers, and policy/implementation documents if they included a focus on how Internet-based psychological therapy is proposed to work for adolescents with depression/depressive symptoms. We used the Mixed-Methods Appraisal Tool to assess the methodological quality of studies. We used the Persuasive System Design (PSD) model as a framework for data extraction and analysis to examine how Internet-based CBT and IPT, as technology-based systems, influence the attitudes and behaviors of system users. PSD components described for the therapies were linked to reported outcomes using a cross-case comparison method and thematic synthesis.

Results: We identified 19 Internet-based CBT programs in 59 documents. Of those, 71% (42/59) were of moderate to high quality. The PSD features surface credibility (competent “look and feel”), dialogue support (online program + in-person support), liking and similarity (esthetics and content appeal to adolescent users), the reduction and tunneling of therapeutic content (reducing online content into simple tasks, guiding users), and use of self-monitoring were present in therapies that resulted in improved therapy engagement, satisfaction, and adherence, as well as symptom and functional impairments.

Conclusions: When incorporated into Internet-based CBT for adolescent depression, PSD features may improve adolescent adherence, satisfaction, and depression-related outcomes. Testing of these features using hypothesis-driven dismantling approaches is recommended to advance our understanding of how these features contribute to therapy effectiveness.

(J Med Internet Res 2017;19(8):e266) doi:10.2196/jmir.7573

KEYWORDS

persuasive systems; mental health; Internet-based intervention; review; psychological therapy

Introduction

In their lifetimes, as many as 1 in every 9 adolescents will meet criteria for major depressive disorder (MDD) [1]. A comparable percentage of children and adolescents will also experience subthreshold depressive symptoms. MDD interferes with academic performance and school attendance, can have persistent impacts on daily functioning, and disrupts family and peer relationships [2-5]. For many adolescents, it is a recurrent and lifelong illness making access to timely assessment and treatment essential.

Cognitive behavioral therapy (CBT) and interpersonal psychotherapy (IPT) are recommended psychological therapies for adolescents with MDD [6]; yet, a considerable proportion of adolescents with MDD do not receive such services [7,8]. The reasons may include a lack of trained deliverers, inconvenient service times and locations, the social stigma associated with mental illness, discomfort discussing mental health problems, or a preference for self-help options [9,10].

While CBT and IPT have been traditionally delivered as face-to-face therapies, Internet-based (online) delivery of these therapies is proposed as a solution to access, availability, and uptake barriers. As technology-based treatments, Internet-based psychological therapies consist of (1) therapeutic content, (2) interaction between the user, their computer, and treatment material on the webpage, and (3) technological features embedded into the program (eg, multimedia, interactive treatment components). Recent systematic reviews of studies of Internet-based CBT and IPT for MDD have shown that Internet-based CBT and IPT therapies for clinical (adolescents diagnosed with MDD) and general (adolescents with subthreshold symptoms, adolescents considered at high risk for MDD) populations result in improvements in depressive symptoms and moderate to high satisfaction with the therapies [11-13]. Reviews have also highlighted variability across studies in terms of therapy attrition rates and the design features and functionalities of online delivery. This variability emphasizes the "black box" that remains for understanding how online CBT and IPT engage adolescent users, deliver therapeutic content, and lead to symptom improvements.

We report on a realist review that we conducted to examine the technological and program delivery features of Internet-based CBT and IPT for adolescent depression and to document their potential relation to treatment outcomes and program use. The realist approach provided a lens to explore two main questions: (1) Under what conditions are Internet-based CBT and IPT for adolescent depression being delivered? and (2) Within these conditions, what are the technological features of Internet-based CBT and IPT programs that may explain outcomes reported in studies?

Methods

Study Design

Our review used realist synthesis philosophy and principles as recommended by Pawson and Tilley [14] and is reported using Realist And Meta-narrative Evidence Synthesis: Evolving Standards (RAMESES) criteria [15]. Realist synthesis is an approach concerned with theory development and refinement of "how interventions work, for whom, and under what circumstances" [16]. From a realist lens, treatment effects are understood to be influenced by implementation context, and thus, Context-Mechanism-Outcome (C-M-O) configurations provide potential explanations about causal processes for treatment outcomes. Realist synthesis is also particularly useful when reviewing a complex and heterogeneous body of research (ie, diverse study designs, interventions used, outcomes measured) [14]. In this realist review, we examined the relationships between contextual factors for Internet-based CBT and IPT delivery (eg, conditions of use and the type of user, therapeutic content) and adolescent outcomes (eg, symptom reduction, satisfaction, therapy adherence), and the underlying mechanisms (eg, user behaviors, technological features/system design) that connect them.

Theory Identification

Many realist reviews begin by identifying a theory or theories to develop a preliminary list of C-M-O configurations. The evidence identified in the review is then used to determine which C-M-O configurations are upheld when reviewing the evidence. In contrast, we identified potential theories for our review using an iterative process during project development—brainstorming within the review team and reviewing literature on human-technology interaction and studies of Internet-based psychological therapies for adolescent depression. Persuasive System Design (PSD) emerged from this process as a key framework for our C-M-O configurations. Being derived from both behavior change models and information technology systems models, PSD provided a comprehensive framework for exploring how systems influence the attitudes and behaviors of system users [17,18]. The framework is composed of four system features that relate to the persuasiveness of an Internet-based therapy—primary task support, dialogue support, system credibility support, and social support (Table 1). An Internet-based therapy can take on a persuasive role and use persuasive mechanisms in online delivery (persuasive system design) by (1) conveying symbolic (eg, text, data graphs, icons) and sensory content (eg, real-time video, virtual worlds,

simulation) and (2) facilitating a social experience through the adoption of animate characteristics (ie, physical features, emotions, voice communication) and roles (eg, coach, pet, assistant, opponent), and using social dynamics (eg, greetings, apologies, taking turns) [17].

Using the PSD model, we then theorized what persuasive system features were likely to be linked to each outcome (ie, which features were mechanisms) reported for Internet-based CBT and IPT, and what delivery contexts were the most relevant in allowing that to happen. The end result of this discussion was a list of C-M-O configurations that would guide data analysis.

Search for Relevant Literature Realist methods call for a purposive and iterative approach. The search process began with a research librarian conducting a systematic search of academic databases in the psychology and health fields: Medline, CINAHL, Embase, PsycINFO for the period 2000-2016. Other search engines—Google and Beacon—and gray literature repositories (ACM Digital Library, OpenGrey, Canadian Agency for Drugs and Technologies in Health, IEEE Digital Library) were manually searched to identify gray literature such as government reports, community program evaluations, and conference proceedings for the same time period. Search terms were related to clinical area (ie, depression, mental health), modality (ie, online, Internet-based), and therapeutic approach (ie, CBT and IPT). The strategies for two searches are provided in Multimedia Appendix 1.

We also manually searched the table of contents in medical informatics journals (ie, Journal of Medical Internet Research, Internet Interventions, Journal of Cybertherapy and Rehabilitation, Journal of Telemedicine and Telecare). Snowball searching was also conducted (ie, reviewing eligible article reference lists) to identify relevant documents that may have been missed in the search process.

Table 1. The Persuasive Systems Design model.

Category	Persuasive feature	Definition
Primary task support	Reduction	Reduces complex behavior into simple tasks
	Tunneling	Guides a user through a process or experience
	Tailoring	Tailors the experience to the potential needs, interests, personality, or use context
	Personalization	Personalizes content (eg, allows you to customize the interface or populates your name)
	Self-monitoring	Keeps track of the user's performance or status towards goal achievement
	Simulation	Provides simulations to enable the user to observe link between cause and effect
	Rehearsal	Provides a way for user to rehearse a skill or task
Dialogue support	Praise	Offers praise as a form of feedback
	Rewards	Rewards target behaviors
	Reminders	Reminds the user of their target behavior
	Suggestion	Offers fitting suggestions
	Similarity	Reminds the user of themselves in some meaningful way
	Liking	Is visually attractive for the user
	Social role	Adopts a social role
System credibility support	Trustworthiness	Provide information that is truthful, fair, and unbiased
	Expertise	Provides information showing knowledge, experience, and competence
	Surface credibility	Has a competent look and feel
	Real-world feel	Provides information of the actual people behind its content and services
	Authority	Refers to people in the role of authority
	Third-party endorsement	Provides endorsements from other sources
	Verifiability	Provides means to verify the accuracy of program via outside sources
Social support	Social learning	Can use the system to observe others performing tasks or behaviors
	Social comparison	Can use the system to compare their performance with the performance of others
	Normative influence	Leverages normative influence or peer pressure
	Social facilitation	User is able to discern via the system that others are performing the behavior along with them
	Cooperation	Leverages drive to cooperate to complete tasks or behaviors
	Competition	Leverages drive to compete against others in completing a task or action
	Recognition	Offers public recognition for an individual or group

Literature Selection

Search results from academic databases were downloaded into Endnote (Thomson Reuters, Version 7.2) and then screened for eligibility by 2 trained raters (authors LW, AR). Inclusion criteria were as follows: (1) intervention studies (eg, clinical trials) were eligible for inclusion if they evaluated Internet-based CBT or IPT with adolescents with depression, (2) theoretical papers, mixed-methods and qualitative studies, and policy/implementation documents were eligible if they included a focus on how Internet-based CBT/IPT is proposed to work for adolescents with depression/depressive symptoms, and (3) documents that met Criteria 1 and 2 were eligible if they were published in English language. Any reviews (systematic, meta-analysis, etc) identified during our screening process were appraised to identify intervention studies and other potentially relevant documents.

During Stage 1 screening, the eligibility of a random subset (10 citations) was assessed independently by 2 team members (authors LW, AR), and interrater agreement was assessed within the "substantial" Kappa range (Cohen kappa=.74).

Literature Appraisal

The evidence for each Internet-based therapy was assessed for relevance and rigor by consensus of 2 reviewers (authors LW, AR). These two reviewers also conducted co-coding and debriefing activities periodically during analysis.

Relevance was defined as the level of contribution to the review, and rigor was defined by the methodological quality of a study conducted on the Internet-based therapy (intervention, mixed-methods, and qualitative studies). Relevance was assessed by reviewing the details provided for an Internet-based therapy's (1) context (eg, user, program features/design components), (2) mechanism(s): hypotheses as to how specific elements of the therapy worked, was proposed to work, or did not work, and (3) outcomes: reasons for therapy effect or lack of effect on specific adolescent outcomes. These details were obtained by reviewing documentation of usability evaluation, therapy/study protocols, and publications related to evaluations (eg, clinical intervention studies evaluating efficacy or effectiveness). Relevance was rated as low/none (no or little information), medium (some information), and high (well-described information).

The methodological quality of evidence (rigor) around each therapy was assessed, where possible, using the Mixed Methods Appraisal Tool (MMAT), an effective and practical quality assessment tool [19]. The tool includes sections for qualitative, mixed-methods, and quantitative studies. MMAT scores range from 0 (no criteria met) to 4 (all criteria met). Pawson describes the process of determining rigor as whether a "particular inference drawn by the authors has sufficient evidence to make a methodologically credible contribution to the test of a theory" [20]. Thus, we used the MMAT to assess the credibility of the reported findings based on the methodology described. Following MMAT guidelines, if publications were companion papers on the same data set, they were assessed as a set of publications.

Data Extraction

NVivo software (QSR International; Version 11) was used to extract data. We extracted data for several aspects of the context of the Internet-based therapy: (1) user context (eg, urban/rural, age group, sociocultural composition, clinical severity), (2) usage context (eg, therapy objectives, adjunct versus stand-alone therapy), and (3) technology context (eg, synchronicity, use of multimedia, software and bandwidth requirements, mobile phone versus desktop). We also extracted information on therapy design (ie, conditions under which an adolescent completed the therapy: sequence, structure, timing; CBT/IPT features). In instances where the description of the therapeutic content was not available in the article, we included information from the article's citations that described a therapy. We also extracted available information on PSD system features (primary task support, dialogue support, system credibility support, social support), therapy usage (eg, attrition, engagement), as well as clinical (eg, symptom reduction) and therapeutic (eg, therapeutic alliance) outcomes. Where available, information on full or partial C-M-O configurations was also extracted for individual therapies (ie, we sought data on particular therapies that could explain what Mechanism led to an Outcome, under which Contexts).

To promote consistency during data extraction, a coding guide with operational definitions for each code was used. PSD principles were used to guide the coding of mechanisms [14,18]. Two review team members (authors LW, AR) cross-referenced decisions regarding coding and extraction on a random subset of 10 articles and the remaining documents were coded by 1 reviewer (author LW).

Analysis and Synthesis Process

We used a multistep approach to identify and organize information about what contexts of Internet-based CBT/IPT and persuasive system design attributes may contribute to adolescent outcomes. Our initial list of C-M-O configurations was revised based on consensus between team members. Drawing from qualitative synthesis methods, we selected an "index" therapy (CATCH-IT [21]) that was conceptually rich (ie, high methodological rigor, and most complete and robust descriptions) and could be used as a starting point for lines of argument synthesis [22] (ie, a form of grounded theorizing). Using a cross-case comparison method, we initially compared each therapy to the index therapy using the data matrix produced during data extraction. Similar and recurring concepts and themes were pulled together in order to corroborate or refute the proposed C-M-O configurations. This process involved identifying concepts from one therapy and recognizing the same concepts in another therapy, though they may not be expressed using identical words [23]. Proposed C-M-O configurations were analyzed at different levels of abstraction (ie, within and across therapies) to determine the most robust and plausible explanations of how in a context, with a mechanism, outcomes could be generated.

As a final step we compiled a framework matrix for each C-M-O configuration in NVivo in order to map "demi-regularities", semi-predictable patterns of therapy outcomes [16]. This allowed the review team to examine which therapies provided what

evidence to support each C-M-O configuration. Conference calls with team members were used to discuss, amend, or confirm identified patterns. In instances where a C-M-O configuration was judged to not fully explain the demi-regularities, we returned to the data to refine the configuration.

Results

Figure 1 presents a flow diagram outlining the document search and appraisal process. A total of 15,760 unique and potentially eligible documents were reviewed for inclusion in this review. Of these, 59 documents were deemed eligible for inclusion: published studies (n=45), gray literature documents (n=8), and clinical trial protocols (n=6). These documents were published between 2006 and 2016, and they detailed 19 unique Internet-based psychological therapies.

Characteristics of Internet-based Therapies for Adolescent Depression

Structure and Delivery Features

An overview of the structure and delivery features of the 19 Internet-based therapies is provided in Table 2. The majority of the therapies (14/19, 79%) were adaptations from manualized (paper-based) programs, Internet-based therapies designed for adults, or other online programs. Six therapies were designed to concurrently treat depressive and anxiety disorders (Cope2Thrive, Chilled Plus, MoodGym, Mood Mechanic Course, ThisWayUp, Feeling Better), and two therapies were designed to also address problems with alcohol use (DEAL, iTread). Two therapies had significantly larger bodies of associated research than the others—of the reviewed documents, 19% (11/59) were related to the MoodGym program and 15% (9/59) were related to the CATCH-IT program.

Most therapies (14/19, 74%) were designed to support contact with a health care or teaching professional. However, the nature of this contact varied significantly across therapies ranging from a 10-minute, initial face-to-face motivational interview (CATCH-IT) to a fully synchronous, online chatroom moderated by a trained coach over multiple weeks (Master Your Mood). Three therapies included the option of synchronous computer-mediated communication (ie, chat, text messaging); most therapies relied on email-based communication. ChilledPlus, CATCH-IT, and CURB involved the adolescent's primary health care provider and adjunct education for parents as part of the therapy. For brief content descriptions of each Internet-based therapy, see Multimedia Appendix 2.

Table 2. Reported structure and delivery characteristics of Internet-based psychological therapies for adolescent MDD.

Program (country)	Participants		Program details				
	Target age (years)	Testing context[a]	Parent involvement	Time commitment	Contact		Adapted from
					Before program	During program	
Blues Blaster (USA) [24]	11-15	P	No	Total: 60-90 minutes over 1 wk 6 modules (10-15 minutes per module); 1 module/day	None	None	Face-to-face Coping With Depression
CATCH-IT (USA) [21,25-32]	14-21	P	Yes	Total: 660-840 minutes over 7-8 wks 11-14 modules (60 minutes per module); 1-2 modules/wk	In-person[b]	Phone	
Chilled Plus (AUS) [33,34]	12-17	T	Yes	Total: 600 minutes over 8 weeks 8 modules (60 minute per module + 30 minute phone calls); 1 module/wk	In-person	Email	Face-to-face Chilled
Cope2thrive (USA) [35,36]	13-18	P, T	No	Total: 350 minutes over 7 weeks 7 sessions (50 minutes per session); 1 session/wk	None	Email/ Phone	Face-to-face COPE group
CURB (USA) [37]	13-17	P	Yes	Total: 660-840 minutes over 7-8 weeks 11-14 modules (60 minutes per module); 1-2 modules/wk	In-person	Phone	CATCH-IT
DEAL (AUS) [38-41]	18-25	T	No	Total: 240 minutes over 4 wks 4 sessions (60 minutes per session)	None	Email	Computerized SHADE
DWD (CAN) [42,43]	13-18	P, T	No	Total: unspecified 8 sections	None	None	Manualized DWD
Feeling Better (CAN) [44,45]	16-30	T	No	Total: 120-200 minutes over 6-10 wks	None	Email/ Phone	Telehealth Family Help
iRFCBT (UK) [46]	15-22	P	No	Total: 360 minutes over 6-12 wks 6 modules (60 minutes per module); 1 module/1-2 wks	None	Email	Internet MindReSolve
iTreAD (AUS) [47]	18-30	T	No	Total: minimum 240 minutes + social networking over 12 months 4 sessions (60 minutes per module)	None	Email/ Online chat	Includes DEAL as component
Master Your Mood (NZ) [48-52]	16-25	T	No	Total: 720 minutes over 8 wks 8 modules (90 minutes per modules); 1 module/wk	None	Online chat	Face-to face group Grip op je dip
MAYA (Chile) [53-55]	12-18	T	No	Total: <20 minutes over <1 wk 1 session	None	None	
Mood Mechanic Course (AUS) [56,57]	18-25	T	No	Total: 1924 minutes over 8 wks 5 lessons	None	Email/ phone/ text	Internet UniWellbeing

Program (country)	Participants		Program details				
	Target age (years)	Testing context[a]	Parent involvement	Time commitment	Contact		Adapted from
					Before program	During program	
MoodGym (AUS) [58-68]	12-17	P, T	No	Total: 150-300 minutes over 2-3 wks 5 modules (30-60 minutes per module); 1-2 modules/wk	None[c]	None	
MoodHelper (USA) [69]	18-24	T	No	Total: unspecified 4 sessions	None	None	Internet for adults ODIN
OIPE (USA) [70]	12-17	T	No	Total: unspecified number of minutes over 12 wks 8 modules	In-person	Text	
Rebound (AUS) [71,72]	15-24	RP	No	Total: unspecified number of minutes over 12 wks User can select from 56 sessions (20 minutes per session)	None	Social network moderation	Internet Horysons for youth psychosis
SPARX (AUS) [73-79]	13-18	P, T	No	Total: 210 minutes 7 modules (30 minutes per module)	None[d]	Phone	CD-ROM SPARX
Thiswayup (AUS) [80,81]	12-16	P	No	Total: 228-263 minutes over 7 wks 7 modules (per module: 15-20 minutes online + 17.5 minute discussion); 1 module/wk	In-person	In-person	Face-to-face CLIMATE Schools

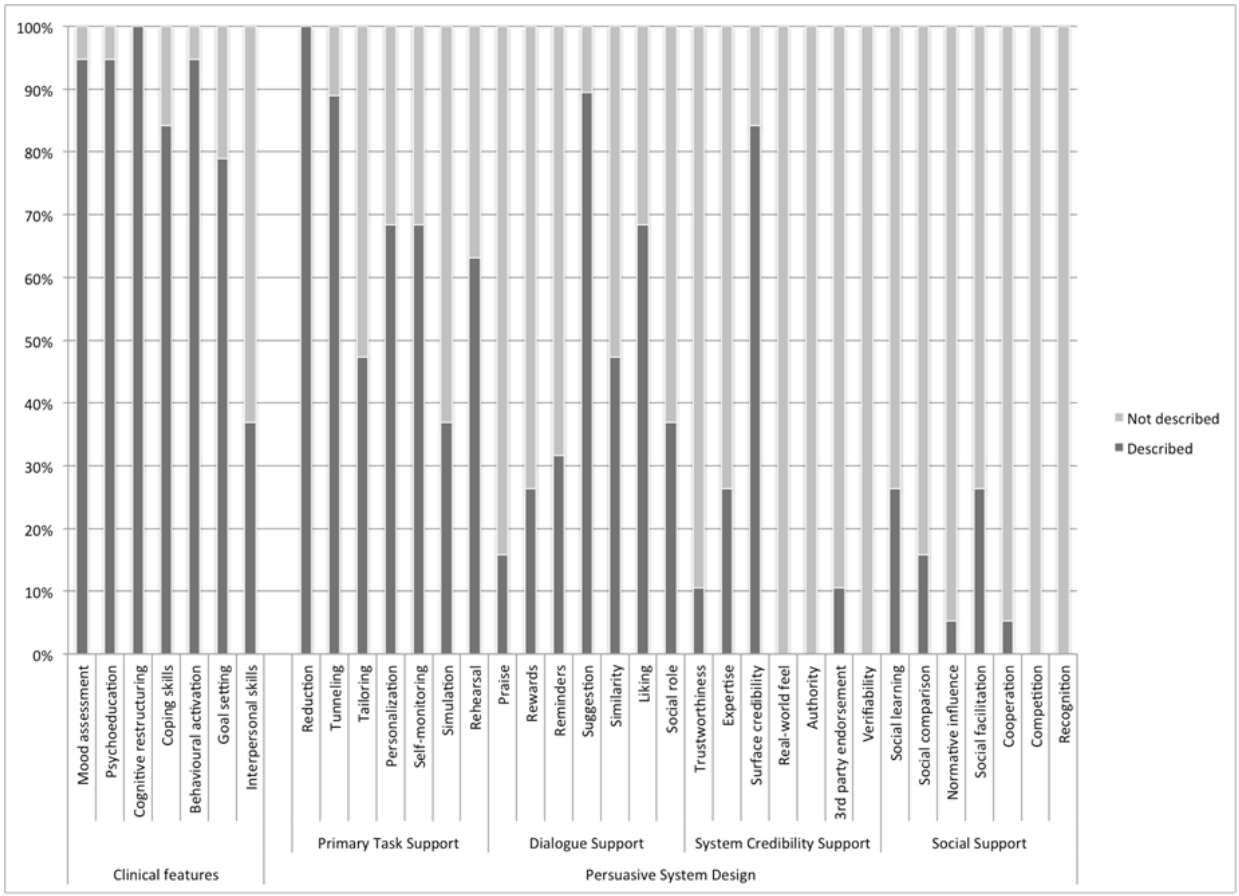
[a]Testing context refers to the type of population who received the therapy: P=prevention (ie, recruited participants with subthreshold depression), T=treatment (ie, inclusion criteria stipulated that participant meet threshold for depressive symptomology, risk, or diagnosis), RP=relapse prevention (ie, required participant have had a previous depressive episode)

[b]Contact with primary care provider was either motivational interview or brief advice.

[c]MoodGym was tested in different implementation contexts; some included no in-person contact and some with in-person contact.

[d]SPARX was tested in different implementation contexts; some included no in-person contact and some with in-person contact.

Figure 2. Reported therapeutic and persuasive system design features in evaluated Internet-based psychological therapies for adolescent MDD.



Psychological Therapy Approaches

All 19 therapies were based on principles of CBT and used essential "ingredients" described in clinical practice guidelines (eg, [82,83]). Fidelity to CBT was high for 14 therapies that described 6/6 core ingredients (Figure 2). Fidelity to CBT was modest for 4 therapies that described 4 or 5 out of 6 core ingredients, and low for 1 therapy (MAYA) that included 3/6 core ingredients. Seven therapies (CATCH-IT, CURB, DEAL, iRFCBT, Rebound, MAYA, Chilled Plus) also included an interpersonal therapeutic orientation (eg, social support, interpersonal conflict resolution) consistent with principles of IPT [84].

Persuasive System Design Features

The PSD features that were present in the therapies are also presented in Figure 2. Primary task support and dialogue support were found to be the most widely represented. All of the therapies included some reference to *reduction*, and a majority referenced some form of *tunneling* and *suggestions* (17/19 therapies). Aspects of system credibility support were not well represented in the documentation we reviewed for the therapies with the exception of *surface credibility* (16/19 therapies). Persuasive design features that leverage social support to motivate users were rarely, if ever, reported features.

Level of Contribution and Methodological Quality

Details of the level of contribution and quality assessments are provided in Table 3. Based on the level of contribution assessment, three therapies—MoodGym, SPARX, and CATCH-IT—were rated as having high contributions to all three dimensions relevant to the review questions: contexts, mechanisms, and outcomes. Twelve other therapies provided medium to high contributions to at least one of the three dimensions. Contributions from four therapies were low. Across the therapies, descriptions of mechanisms (how technology components were expected to function in relation to outcomes) and context were less developed than descriptions of outcomes. The methodological quality of studies associated with each therapy varied considerably. Half (28/57) of the documents were research studies eligible for MMAT ratings—15 randomized controlled trials, 3 nonrandomized trials, 3 quantitative descriptive studies, 4 mixed-method studies, and 3 qualitative studies. All studies met at least 2 of four MMAT criteria. Studies with lower ratings did not provide a clear description of the randomization process, had higher than 20% study dropout rates, or used nonvalidated measurement tools. Follow-up beyond 6 months was conducted only for the CATCH-IT program suggesting limited knowledge about post-program experiences of adolescents.

Table 3. The availability and contribution of evidence related to Internet-based psychological therapies.

Therapy	Level of contribution[a]			Documentation available for review			Associated MMAT scores[a]
	Context	Mechanism	Outcome	Usability	Protocol	Efficacy/Effectiveness	
CATCH-IT	High	High	High	[27,29,30]	[26]	[21,25,28]	4 [21], 4 [25], 3 [29], 3 [30]
MoodGym	High	High	High	None	None	[58-68]	3 [58], 3 [59], 2 [60], 3 [61], 2 [62], 4 [63-64], 4 [65], 2 [66], 3 [67] [68]
SPARX	High	High	High	[74,75,78]	[79]	[73,77]	4 [73], 3 [74], 3 [75], 4 [77], 3 [78]
Blues Blaster	High	High	Medium	None	None	[24]	3 [24]
DEAL	High	Medium	High	[38]	[40]	[39,41]	3 [38], 3 [39,41]
Master Your Mood	Medium	High	High	None	[51]	[48,49]	2 [48], 2 [49]
MoodHelper	Low	High	Medium	None	None	[69]	3 [69]
Feeling Better	Medium	Medium	Low	[44,45]	None	None	2 [44], 2 [45]
Thiswayup	Medium	Low	Medium	None	None	[80,81]	2 [80], 4 [81]
Maya	Low	Medium	Medium	[53,54]	None	[55]	2 [53-55]
OIPE	Low	Low	Medium	None	None	[70]	2 [70]
Mood Mechanic Course	Low	Low	Medium	None	None	[56]	3 [56]
Rebound	Medium	Medium	Low	[71]	None	[72]	3 [71], 4 [72]
Cope2thrive	Medium	Low	Medium	None	None	[35]	3 [35]
CURB	Medium	Low	None	[37]	None	None	N/A
iRFCBT	Low	Low	None	None	[46]	Ongoing trial	N/A
iTreAD	Low	Low	None	None	[47]	Ongoing trial	N/A
Chilled Plus	Low	Low	None	None	None	Ongoing trial	N/A
DWD	Low	Low	None	None	None	Open Access	N/A

[a]Following published guidelines for MMAT scoring, in instances where multiple documents reported on the same data set, a single MMAT score was calculated.

Context-Mechanism-Outcome Configurations

Of the candidate C-M-O configurations initially put forward using the PSD model, five configurations were substantively supported by available evidence (Table 4). We present the configurations with key examples of the contexts, mechanisms, and outcomes from the documents reviewed.

C-M-O Configuration 1

In this review, we found that *dialogue support* provided as a PSD feature by a therapy required real-time, in-person guidance to optimize therapy adherence. Five therapies contributed to this C-M-O configuration, which were connected to studies with a moderate mean MMAT score of 2.94 (SD 0.77).

Real-time guidance involved adolescents completing the Internet-based therapy with an individual (doctor, teacher, therapist) in a setting available to support them (or supervise them) while they completed the activities. Results from studies suggested that completion rates increased if the therapy was delivered with real-time guidance in contexts such as schools or connected with primary/secondary care versus having the adolescent complete the therapy on their own (self-guided) [21,58,80]. Even therapies that included more computer-generated or automated dialogue opportunities (ie, reminders, praise, automated suggestions) were able to optimize adherence only with real-time, in-person contact [27,73,80].

Studies of MoodGym have compared in-person guidance to self-guidance in a setting of their choice (eg, home) with synchronous support from the virtual guide (eg, avatar). One study found that there was a 10-fold difference between the approaches (favoring the in-person guidance) in terms of the number of online exercises adolescents completed [58]. The authors of the study stated that the comparison highlighted "the success of the monitored setting in increasing compliance" [58] Sethi et al have suggested that, "treating youth depression with a combination of face-to-face and online therapy is ideal...if the technology incorporates ways to interactively learn and practice at their own pace" [68]. Reasons for nonuse reported by adolescents in one MoodGym study showed that almost a third of adolescents indicated they stopped using the therapy because

they, "felt the need to talk to someone, rather than doing this program." [65].

Table 4. Summary of the C-M-O configurations substantively supported by evidence.

C-M-O configuration	C, M, O	Supporting programs
1. Computer-mediated dialogue required real-time support and monitoring to optimize therapy adherence.	C: real-time support M: dialogue support features O: adherence	CATCH-IT, MAYA, MoodGym, SPARX, ThisWayUp
2. Therapies with surface credibility led to engagement and satisfaction with the therapy.	C: user interface M: credibility support O: engagement and satisfaction	Blues Blaster, CATCH-IT, Feeling Better, Master Your Mood, MAYA, SPARX
3. Therapies that included liking and similarity features led to engagement and satisfaction with the therapy.	C: user interface M: liking and similarity O: engagement and satisfaction	Blues Blaster, CATCH-IT, CURB, DEAL, Feeling Better, MoodGym, MAYA, SPARX
4. Reduction and tunneling of therapy content were necessary for adolescents to complete more of the therapy.	C: user interface M: reduction and tunneling O: adherence	Blues Blaster, CATCH-IT, CURB, DEAL, Feeling Better, Master Your Mood, MAYA, MoodGym, SPARX
5. Self-monitoring was a key PSD component for facilitating symptom improvements among adolescent users with a MDD diagnosis or functional impairments.	C: users with a MDD diagnosis and/or functional impairments M: self-monitoring O: clinical outcomes	Blues Blaster, CATCH-IT, DEAL, Feeling Better, Master Your Mood, Mood Helper

C-M-O Configuration 2

In this review, we found that adolescents assess the *surface credibility* of an Internet-based therapy, and that this assessment impacts engagement and satisfaction with the therapy. Therapies with surface credibility have resulted in higher engagement and satisfaction. Others have argued that surface credibility is critical "as negative perceptions of the interface usability could influence the program's eventual effectiveness...acceptability... [and] long-term clinical value" [45]. Six therapies contributed to this C-M-O configuration, which were connected to studies with a high mean MMAT score of 3.10 (SD 0.79).

Potential therapy users quickly assess credibility [45] and look for programs that have good face validity and are appealing [45]. Adolescents have been found to be highly satisfied with programs that presented therapeutic key concepts in engaging ways (eg, video, animation, interactive exercises) [24] and have reported that multimedia resources make programs "easier to use" and "helped in learning the material" [24]. Features perceived as unacceptable to adolescents include games that are not engaging or interactive, sections that are difficult to navigate, and video streaming that lags [53,55]. Adolescents are significantly more likely to complete a therapy designed in response to adolescent consumer preferences [27]. In a study of Master Your Mood, 50% of participants quit before completing half of the program and attributed this in part, to the fact that they "did not feel motivated by the materials provided" [49].

C-M-O Configuration 3

Therapies that incorporated the PSD elements of *liking* (ie, visual appeal) and *similarity* (ie, a way for adolescents to recognize themselves in the program) were found to lead to increased adolescent engagement and satisfaction. Eight therapies contributed to this C-M-O configuration; supporting primary studies had a high mean MMAT score of 3.16 (SD 0.76).

Across programs, liking related to the program's appearance. The issue of color palette was consistently identified during usability testing [38,44]. For example, the culturally adapted version of CATCH-IT for African-American and Latino adolescents (CURB) was designed to "avoid the appearance of a 'school-like' experience" [36] and later redesigned after adolescents reported that the "initial color design appeared too 'boring'" [37].

Feedback on the experience of the online video game MAYA, which had low satisfaction and engagement, suggested that, "it would be desirable that [the game] portrayed a social context more similar to [the participant's] reality" [55]. A central theme for several programs has been similarity. Programs have found that cultural relevance, the use of personal characters and language for adolescents to relate to, as well as tailoring content based on user choices, has led to increased adolescent engagement and satisfaction [38,44,75]. As one participant in the Feeling Better program indicated, "it [tailored content] made me more interested in it, like rather than just kind of skipping through it" [44]. In exploring nonsignificant or mixed-results in several studies of MoodGym among adolescent populations, Sethi et al speculated that because the program was originally developed for adults, it "may need to be tailored to specific adolescents themes and examples to encourage engagement" [62].

C-M-O Configuration 4

The use of *reduction* and *tunneling* of therapeutic content has been found as necessary for adolescents to achieve therapy adherence. In this way, therapies that initiate young people into doing a series of therapeutic actions through incremental steps (small doses) rather than a one-time consolidated effort may

provide a more persuasive experience that leads to improved outcomes. This C-M-O configuration is supported by evidence from 9 therapies and studies with an average MMAT score of 2.81 (SD 0.66) suggesting overall moderate quality evidence.

Reducing the amount of text, improving navigational instructions, and reducing the length of therapy modules to improve ease of use were associated with higher adherence (eg, [30,32,39]). Heterogeneity of therapies made it challenging to disentangle under what contexts tunneling and reduction worked more/less persuasively in improving adherence. Large variations in program design: (1) time commitment per module (range 20-1924 minutes), (2) number of modules (range 4-14), (3) attrition/dropout out measures used, and (4) differences in self-paced versus locked content ("unlocked" until the completion of certain therapy tasks/activities) were problematic when attempting to compare these persuasive design features across therapies. Moreover, many of the therapies targeted a wide range of ages without providing discussions on how developmental level or cognitive ability were considered in terms of reduction and tunneling strategies. Less than half (8/19, 42%) of the therapies targeted young people with a spread of 7 years or more (eg, Feeling Better was designed for young adults 16-30 [14-year range] and Master Your Mood was designed for 16-25 year olds [9-year spread]).

C-M-O Configuration 5

A defining feature of the therapies that were successful in treating adolescents with severe symptomology at baseline was the use of *self-monitoring* as a PSD feature. This C-M-O configuration was supported by evidence from 6 therapies with high-quality evidence (mean MMAT score, mean 3.09, SD 0.54).

Increasing self-awareness of emotions is an important clinical program feature for treating depression as it prepares individuals for changing their cognitions, beliefs, and schemas [82,83]. Studies contributing to this C-M-O configuration were rated as having modest to high fidelity to CBT (5 or 6 out of 6 clinical features; see Figure 2). Thus, opportunities for self-monitoring likely played a role both as an active therapeutic ingredient but also as a PSD element. Two studies provide useful case examples of the impact of self-monitoring.

Mood Helper was one of the only pure stand-alone Internet-based therapies in our review [69]. Mood Helper comprises multiple opportunities for self-monitoring. Users (1) complete a series of brief auto-scored depression scales, (2) review their graphically displayed depression scores over time, (3) receive automated reminders to return to the website every few days, (4) are given access to a personal diary space, and (5) are guided in creating self-contracts to define goals and remind the user of progress towards them. A small but significant effect size favoring Mood Helper compared to the treatment as usual control condition was found, and researchers noted that participants using the website more intensely had higher average baseline depression scores [69]. This was especially significant since researchers found an unexpected dose-effect response where fewer minutes of the website usage were associated with greater depression symptom reduction. A post-hoc analysis showed that participants who improved more rapidly found the website less necessary and they discontinued use, "whereas those with more persistent depression may have continued for longer periods in the hopes of obtaining relief" [69]. This finding suggests that tailoring self-monitoring around level of depressive symptoms maybe be warranted as it might facilitate and likely trigger increased engagement in practicing skills and in turn, decreasing symptoms.

The second program, Blues Blaster, was designed primarily to teach adolescents how to monitor their mood and to do so using engaging methods [24]. As adolescents progress through each of the six modules they are "encouraged to track their mood ratings and fun activities which are depicted together on a graph to help illustrate the relationship between them" [24]. Adolescents also track personal "mood triggers" to help them plan ahead. Mean responses on the 6-point satisfaction items, "I liked seeing the graph of my mood and activities" (mean 4.52, SD 1.27) and "the mood and activities tracking form was easy to fill out" (mean 4.58, SD 1.15) showed strong support for the self-monitoring components of the program. Results of a randomized controlled trial demonstrated greater improvement for the Blues Blaster condition in depression levels, negative thoughts, behavioral activation, knowledge, self-efficacy, and school functioning compared to the information-only control condition [24]. A significant correlation was also found between total modules completed and the depression measure posttest change scores.

Discussion

Principal Considerations

Advances in technology have allowed for health care programs to connect users to treatments in dynamic ways. The dramatic growth of technologies designed to persuade and motivate represents a significant shift in focus toward end-user computing in health behavior change therapies [18]. Internet-based therapies for adolescent depression are one group of therapies that have incorporated persuasive design strategies to influence therapy outcomes. However, to date, such programs have not been reviewed and evaluated with this perspective. Not accounting for human-technology interaction in therapy design and evaluation limits our understanding of how Internet-based therapies work. In our view, this review of Internet-based CBT and IPT for adolescent MDD highlights the need to purposefully consider PSD features early in the program design process and to further develop PSD theory to help explain therapy outcomes. This review is also an important step in understanding how PSD features of Internet-based therapies may work based on existing literature. The C-M-O configurations generated by this review represent important hypotheses. The testing of specific program mechanisms (ie, PSD features, provider involvement) through hypothesis-driven, dismantling approaches is now necessary to advance the understanding of the effect of Internet-based therapies on users.

A key argument in favor of developing and offering Internet-based psychological therapies to young people is increased access [57] to effective mental health care. Therapies in our review were either stand-alone therapies meant to increase options for young people with depression [69] or used in

combination with existing health care [25] or school-based [58] services. Therapies that relied on existing services reported better engagement with the program and completion rates, although across all studies, attrition was a reported issue. Does availability matter if young people do not complete the program as intended? At this point in time, this question is hard to answer. It is necessary for future studies to examine whether young people stop using stand-alone therapies because of symptom improvement, changes in motivation, or because of the program features. Qualitative exploration of the type of support provided and required during adolescent use of programs used in combination with existing services would also provide valuable information on necessary "program ingredients." While adherence to Internet-based programs has traditionally been low [85], the use of PSD principles, or in-person support from existing services at critical program timepoints (eg, when dropout has been shown to occur, when therapeutic content is strenuous for the user) may address program engagement and adherence.

The findings from our review, in terms of surface credibility and liking, suggest that adolescents' visual experiences lead to esthetic and credibility judgments [86-88] and should also be taken into account. Bennett and Glasgow have argued that the most critical design gap in the current generation of Internet-based programs is the underuse of Web 2.0 features (eg, social media, user-generated content, collaborative consumption) [89]. In our review, many therapies were adapted from face-to-face or manualized programs. This may have resulted in a "replication" of components that were originally designed for face-to-face interactions rather than fully optimizing Web 2.0 features available for delivery.

Fogg has argued that persuasive systems will work only if the user has sufficient motivation and the user's ability is being adequately triggered to perform the new behavior [17]. High rates of noncompletion of Internet-based psychological therapies for adolescent MDD suggest the need for therapies to address motivation in the context of depression/depressive symptoms (eg, if a depressed adolescent is not motivated to log into a stand-alone program, what program supports are needed for behavioral activation?). That the use of specific PSD features—reduction, tunneling, self-monitoring—led to increased adolescent engagement and satisfaction, and improvement in depressive symptoms and functioning, suggests the cognitive load of therapies may also need to be addressed. The advantage of primary task supports—tunneling, reduction, and self-monitoring—is that these PSD features can increase positive affect (ie, "I can do this") [90,91]. Pairing primary task supports with other persuasive strategies such as self-monitoring, reminders, simulation-rehearsal, and suggestion-reward can be especially effective together [92,93]. Of the studies in this review, despite high attrition, most reported a positive clinical outcome associated with Internet-based program use, suggesting that we do not yet have a clear understanding of therapy "dosage" for adolescents and how "persuasive" a therapy needs to be (eg, the persuasive experience) to ensure optimal outcomes for adolescents. Future research is needed to strengthen our understanding of the relationship between treatment adherence and symptom change. Persuasive system design may be one testable model to facilitate knowledge generation. The studies in this review highlight that the relationship between engagement in therapy activities, adherence through to treatment completion, and treatment satisfaction are complexly related to clinical outcomes and require theoretical and empirical investigation.

Future studies should operationalize each PSD feature, hypothesize its intended effect, and measure its use and effect. As technology and methods of human-technology interaction evolve, this documentation and evidence will provide a valuable roadmap for the depression treatment field. Studies that use a factorial design or fractional factorial design would move the field forward by providing an opportunity to compare intervention groups that include multiple, and different combinations of, persuasive design functionality. For example, there could be conditions within both study arms in which some groups receive tailored feedback and others do not, some include a social networking forum and others do not, and some provide reminders and others do not. In this way, the impact of PSD features can be isolated. It might also be useful to develop hybrid designs that include both standard "randomization" as well as "preference" arms, in order to determine which groups of adolescents might be more attracted to certain PSD features.

The exploration of target population characteristics is also needed to determine how motivational (eg, readiness for change, self-regulatory skills), developmental/age-related, sociocultural, technical competency and modality features (synchronous, ambient, etc), and depression severity differentially interact with program mechanisms and impact adolescent outcomes. For example, research has shown gender differences in the perceived persuasiveness of numerous health intervention components, with females being more receptive to most persuasive behavior change strategies [94]. Under this approach, studies can identify mediators of therapy effect and examine both mediators and moderators as a means to identify program mechanisms (how) and for whom the therapy works best (who), respectively.

Review Strengths and Challenges

This review is the first to use a realist framework for studying Internet-based psychological therapies for adolescent depression. This framework allowed us to consider studies and theories together to understand how the therapies worked. We included numerous therapies identified in the gray literature, which allowed for a comprehensive appraisal of the current evidence base and reduced the risk of publication and selection bias. Previous reviews of Internet-based therapies for this population have focused only on empirical literature, and therefore, have provided limited insight into the complex causal pathways that may underpin therapy effects. Including multiple research designs, while challenging from a data integration standpoint, enabled the analysis to benefit from the strengths of each approach and corroborate findings across divergent contexts and theoretical orientations. From a realist perspective, this diversity has huge explanatory value and can help uncover contexts and conditions not typically captured in meta-analytic or traditional systematic reviews. In addition to offering a more thorough assessment of Internet-based therapies, this review

supersedes existing reviews by including substantially more therapies and documenting the body of work around each one. A further strength of the review is the use of gold standard review methods (notably, duplication of screening, quality assessment, and consensus-building with research team members).

The greatest challenge in applying the PSD model is that no explicit heuristics have been defined for it yet, and so nuances between different PSD features are still being mapped [95]. The model would gain more strength from explicitly defined scales and instructions for evaluating the implementation of each PSD principle. A further challenge in exploring these therapies from a sociotechnological frame is that studies largely employed traditional clinical reporting methods with very little attention paid to describing the informatics architecture and the *expected* role of technology functionalities. In this regard, usability studies included in our review provided much needed information. Because we reviewed the literature and evidence base around these therapies, as opposed to conducting a workflow analysis of each therapy, our appraisal was limited to what authors reported, which may or may not be a full reflection of the therapy's capabilities and design.

Another challenge in our review was the lack of information on the nature of adolescent's interaction with PSD features. For example, current descriptions of dialogue support provided in Internet-based therapies are lacking. While authors described using email, reminders, and options for peer engagement, there was little detail about actual engagement with these features (eg, How often were reminder emails triggered?, How many adolescents elected to publish journal entries to their peers?, How many adolescents spontaneously emailed their assigned coach and how often?). In terms of peer-based dialogue support, some have argued that there is not yet very strong evidence for what type of peer-based social support therapies ought to provide [96].

Conclusions

Results from our review suggest there is room for improvement in both designing and implementing Internet-based therapies for adolescent depression and in elucidating how persuasive mechanisms are designed and ultimately function. We offer that many of the assumptions that implicitly shape Internet-based therapy development and delivery—adolescents are highly competent technology users, adolescents want to complete programs on their own, the more persuasive design components the better, or that compliance will result in improved outcomes—are vastly under-acknowledged and are based on pervasive assumptions about adolescents, what they prefer, and what they need. Improved engagement of adolescents with MDD in the design and development of future therapies is crucial if we hope to provide effective Internet-based therapies for this population.

Acknowledgments

This work was supported by a knowledge synthesis grant from the Canadian Institutes of Health Research (CIHR: KRS2014). AN and LH are supported by CIHR New Investigator Salary Awards.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Review search strategies.

[PDF File (Adobe PDF File), 28KB - jmir_v19i8e266_app1.pdf]

Multimedia Appendix 2

Brief description of Internet-based psychological therapies for adolescent depression.

[PDF File (Adobe PDF File), 46KB - jmir_v19i8e266_app2.pdf]

References

1. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
2. Fergusson DM, Woodward LJ. Mental health, educational, and social role outcomes of adolescents with depression. Arch Gen Psychiatry 2002 Mar;59(3):225-231. [Medline: 11879160]
3. Keenan-Miller D, Hammen CL, Brennan PA. Health outcomes related to early adolescent depression. J Adolesc Health 2007 Sep;41(3):256-262 [FREE Full text] [doi: 10.1016/j.jadohealth.2007.03.015] [Medline: 17707295]
4. Jaycox LH, Stein BD, Paddock S, Miles JNV, Chandra A, Meredith LS, et al. Impact of teen depression on academic, social, and physical functioning. Pediatrics 2009 Oct;124(4):e596-e605. [doi: 10.1542/peds.2008-3348] [Medline: 19736259]

5. Puig-Antich J, Lukens E, Davies M, Goetz D, Brennan-Quattrock J, Todak G. Psychosocial functioning in prepubertal major depressive disorders. I. Interpersonal relationships during the depressive episode. Arch Gen Psychiatry 1985 May;42(5):500-507. [Medline: 3985760]
6. Birmaher B, Brent D, Bernet W, Bukstein O, Walter H, Benson RS, et al. Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. J Am Acad Child Adolesc Psychiatry 2007 Nov;46(11):1503-1526. [doi: 10.1097/chi.0b013e318145ae1c] [Medline: 18049300]
7. Cheung AH, Dewa CS. Mental health service use among adolescents and young adults with major depressive disorder and suicidality. Can J Psychiatry 2007 Apr;52(4):228-232. [doi: 10.1177/070674370705200404] [Medline: 17500303]
8. Merikangas KR, He J, Burstein M, Swendsen J, Avenevoli S, Case B, et al. Service utilization for lifetime mental disorders in U.S. adolescents: results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2011 Jan;50(1):32-45 [FREE Full text] [doi: 10.1016/j.jaac.2010.10.006] [Medline: 21156268]
9. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
10. Meredith LS, Stein BD, Paddock SM, Jaycox LH, Quinn VP, Chandra A, et al. Perceived barriers to treatment for adolescent depression. Med Care 2009 Jun;47(6):677-685. [doi: 10.1097/MLR.0b013e318190d46b] [Medline: 19434001]
11. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
12. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
13. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
14. Pawson R, Tilley N. Realistic evaluation. In: Realistic evaluation. London, UK: Sage; 1997.
15. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. J Adv Nurs 2013 May;69(5):1005-1022. [doi: 10.1111/jan.12095] [Medline: 23356726]
16. Pawson R. Evidence-based policy. In: Evidence-Based Policy: A Realist Perspective. London, UK: Sage Publications Ltd; 2006.
17. Fogg B. Motivating, influencing,persuading users. In: Sears A, Jacko JA, editors. The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. 3rd ed. Hillsdale, NJ: L. Erlbaum Associates Inc; 2002:358-370.
18. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Communications of the Association for Information Systems 2009;24(1):A-28.
19. Pluye P. Critical appraisal tools for assessing the methodological quality of qualitative, quantitative and mixed methods studies included in systematic mixed studies reviews. J Eval Clin Pract 2013 Aug;19(4):722. [doi: 10.1111/jep.12017] [Medline: 23279088]
20. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10 Suppl 1:21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]
21. Van Vorhees BW, Vanderplough-Booth K, Fogel J, Gladstone T, Bell C, Stuart S, et al. Integrative internet-based depression prevention for adolescents: a randomized clinical trial in primary care for vulnerability and protective factors. J Can Acad Child Adolesc Psychiatry 2008 Nov;17(4):184-196 [FREE Full text] [Medline: 19018321]
22. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Med Res Methodol 2009 Aug 11;9:59 [FREE Full text] [doi: 10.1186/1471-2288-9-59] [Medline: 19671152]
23. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
24. Makarushka M. Efficacy of an Internet-based intervention targeted to adolescents with subthreshold depression PhD. In: University of Oregon. Ann Arbor, MI: University of Oregon; 2011:1-119.
25. Hoek W, Marko M, Fogel J, Schuurmans J, Gladstone T, Bradford N, et al. Randomized controlled trial of primary care physician motivational interviewing versus brief advice to engage adolescents with an Internet-based depression prevention intervention: 6-month outcomes and predictors of improvement. Transl Res 2011 Dec;158(6):315-325. [doi: 10.1016/j.trsl.2011.07.006] [Medline: 22061038]
26. Gladstone TG, Marko-Holguin M, Rothberg P, Nidetz J, Diehl A, DeFrino DT, et al. An internet-based adolescent depression preventive intervention: study protocol for a randomized control trial. Trials 2015 May 01;16:203 [FREE Full text] [doi: 10.1186/s13063-015-0705-2] [Medline: 25927539]

27. Landback J, Prochaska M, Ellis J, Dmochowska K, Kuwabara SA, Gladstone T, et al. From prototype to product: development of a primary care/internet based depression prevention intervention for adolescents (CATCH-IT). Community Ment Health J 2009 Oct;45(5):349-354. [doi: 10.1007/s10597-009-9226-3] [Medline: 19641992]
28. Saulsberry A, Marko-Holguin M, Blomeke K, Hinkle C, Fogel J, Gladstone T, et al. Randomized Clinical Trial of a Primary Care Internet-based Intervention to Prevent Adolescent Depression: One-year Outcomes. J Can Acad Child Adolesc Psychiatry 2013 May;22(2):106-117 [FREE Full text] [Medline: 23667356]
29. Gladstone T, Marko-Holguin M, Henry J, Fogel J, Diehl A, Van Vorhees BW. Understanding adolescent response to a technology-based depression prevention program. J Clin Child Adolesc Psychol 2014;43(1):102-114. [doi: 10.1080/15374416.2013.850697] [Medline: 24245935]
30. Iloabachie C, Wells C, Goodwin B, Baldwin M, Vanderplough-Booth K, Gladstone T, et al. Adolescent and parent experiences with a primary care/Internet-based depression prevention intervention (CATCH-IT). Gen Hosp Psychiatry 2011;33(6):543-555. [doi: 10.1016/j.genhosppsych.2011.08.004] [Medline: 21958447]
31. Marko M, Fogel J, Mykerezi E, Van Vorhees BW. Adolescent internet depression prevention: Preferences for intervention and predictors of intentions and adherence. J Cyber Ther Rehab 2010;3(1):9-30.
32. Van Voorhees BW, Gladstone T, Cordel S, Marko-Holguin M, Beardslee W, Kuwabara S, et al. Development of a technology-based behavioral vaccine to prevent adolescent depression: A health system integration model. Internet Interventions 2015 Sep;2(3):303-313. [doi: 10.1016/j.invent.2015.07.004]
33. Macquarie U. Chilled Plus: Online anxiety and depression treatment for adolescents (12-17 yrs old). 2006. URL: https://www.mq.edu.au/__data/assets/pdf_file/0004/125077/ChilledPlus_FAQ_BC0472_V2_Approved_RD_181115.pdf [WebCite Cache ID 6o85Tdfgm]
34. Schniering C. The feasibility and outcome of Internet delivery of transdiagnostic cognitive-behaviour therapy for adolescent anxiety and depression. 2016. URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=369602 [WebCite Cache ID 6o87gw0Ls]
35. Melnyk BM, Amaya M, Szalacha LA, Hoying J, Taylor T, Bowersox K. Feasibility, Acceptability, and Preliminary Effects of the COPE Online Cognitive-Behavioral Skill-Building Program on Mental Health Outcomes and Academic Performance in Freshmen College Students: A Randomized Controlled Pilot Study. J Child Adolesc Psychiatr Nurs 2015 Aug;28(3):147-154. [doi: 10.1111/jcap.12119] [Medline: 26268362]
36. Melnyk M, Lusk P. Psychiatric Times. 2013. Opportunities and Challenges in Treating Adolescents and Young Adults With Major Depressive Disorder URL: http://www.psychiatrictimes.com/articles/opportunities-and-challenges-treating-adolescents-and-young-adults-major-depressive-disorder [accessed 2017-02-08] [WebCite Cache ID 6o86dRLIv]
37. Saulsberry A, Corden ME, Taylor-Crawford K, Crawford TJ, Johnson M, Froemel J, et al. Chicago Urban Resiliency Building (CURB): An Internet-Based Depression-Prevention Intervention for Urban African-American and Latino Adolescents. J Child Fam Stud 2012 Aug 14;22(1):150-160. [doi: 10.1007/s10826-012-9627-8]
38. Deady M, Kay-Lambkin F, Teesson M, Mills K. Developing an integrated, Internet-based self-help programme for young people with depression and alcohol use problems. Internet Interventions 2014 Jul;1(3):118-131. [doi: 10.1016/j.invent.2014.06.004]
39. Deady M, Mills KL, Teesson M, Kay-Lambkin F. An Online Intervention for Co-Occurring Depression and Problematic Alcohol Use in Young People: Primary Outcomes From a Randomized Controlled Trial. J Med Internet Res 2016 Mar 23;18(3):e71 [FREE Full text] [doi: 10.2196/jmir.5178] [Medline: 27009465]
40. Deady M, Teesson M, Kay-Lambkin F, Mills KL. Evaluating a brief, internet-based intervention for co-occurring depression and problematic alcohol use in young people: protocol for a randomized controlled trial. JMIR Res Protoc 2014 Feb 27;3(1):e6 [FREE Full text] [doi: 10.2196/resprot.3192] [Medline: 24583824]
41. Deady M. Comorbid depression and problematic alcohol use in young people: The development and evaluation of an online intervention. Australia: University of New South Wales; 2015.
42. Simon Fraser University. Dealing With Depression URL: http://dwdonline.ca/ [accessed 2017-02-08] [WebCite Cache ID 6o82fBDAE]
43. Blisker D, Gilbert M, Worling D, Garland J. Dealing with Depression: Antidepressant Skills for Teens. 2005. URL: http://www.comh.ca/publications/resources/dwd/DWD_PrintVersion.pdf [WebCite Cache ID 6o82MWeY3]
44. Bradley KL, Robinson LM, Brannen CL. Adolescent help-seeking for psychological distress, depression, and anxiety using an Internet program. International Journal of Mental Health Promotion 2012 Feb;14(1):23-34. [doi: 10.1080/14623730.2012.665337]
45. Currie SL, McGrath PJ, Day V. Development and usability of an online CBT program for symptoms of moderate depression, anxiety, and stress in post-secondary students. Computers in Human Behavior 2010 Nov;26(6):1419-1426. [doi: 10.1016/j.chb.2010.04.020]
46. Cook L, Watkins E. Guided, internet-based, rumination-focused cognitive behavioural therapy (i-RFCBT) versus a no-intervention control to prevent depression in high-ruminating young adults, along with an adjunct assessment of the feasibility of unguided i-RFCBT, in the REducing Stress and Preventing Depression trial (RESPOND): study protocol for

a phase III randomised controlled trial. Trials 2016 Jan 04;17:1 [FREE Full text] [doi: 10.1186/s13063-015-1128-9] [Medline: 26725476]
47. Kay-Lambkin FJ, Baker AL, Geddes J, Hunt SA, Woodcock KL, Teesson M, et al. The iTreAD project: a study protocol for a randomised controlled clinical trial of online treatment and social networking for binge drinking and depression in young people. BMC Public Health 2015 Oct 06;15:1025 [FREE Full text] [doi: 10.1186/s12889-015-2365-2] [Medline: 26444863]
48. van DZR, Kramer J, Gerrits R, Cuijpers P. Effectiveness of an online group course for depression in adolescents and young adults: a randomized trial. J Med Internet Res 2012 Jun 07;14(3):e86 [FREE Full text] [doi: 10.2196/jmir.2033] [Medline: 22677437]
49. Gerrits R, Van de Zanden R, Visscher R, Conijin B. Master your mood online: A preventive chat group intervention for adolescents. Australian E-Journal for the Advancement of Mental Health 2007;6(3):-.
50. van der Zanden R, Galindo-Garre F, Curie K, Kramer J, Cuijpers P. Online cognitive-based intervention for depression: exploring possible circularity in mechanisms of change. Psychol Med 2014 Apr;44(6):1159-1170. [doi: 10.1017/S003329171300175X] [Medline: 23866103]
51. van der Zanden R, Kramer JJ, Cuijpers P. Effectiveness of an online group course for adolescents and young adults with depressive symptoms: study protocol for a randomized controlled trial. Trials 2011 Aug 19;12:196 [FREE Full text] [doi: 10.1186/1745-6215-12-196] [Medline: 21854617]
52. Van der Zanden R, Curie K, Van Londen M, Kramer J, Steen G, Cuijpers P. Web-based depression treatment: associations of clients' word use with adherence and outcome. J Affect Disord 2014 May;160:10-13 [FREE Full text] [doi: 10.1016/j.jad.2014.01.005] [Medline: 24709016]
53. Carrasco AE. Acceptability of an adventure video game in the treatment of female adolescents with symptoms of depression. ResPsy 2016 Apr 18;19(1):-. [doi: 10.4081/ripppo.2016.182]
54. Carrasco AE. Psychoeducational design of an adventure video game as a tool in the treatment of depression. Nuevas Ideas en Informática Educativa TISE 2015;127:135.
55. Carrasco AE. Universidad de Chile. 2015. Design and pilot study of an adventure video game as a tool in psychotherapeutic processes of adolescent women with symptoms of depression URL: https://archiv.ub.uni-heidelberg.de/volltextserver/19010/1/Carrrasco_Dissertation_Heidelberg_Final.pdf [accessed 2017-07-12] [WebCite Cache ID 6ru5IsMbL]
56. Johnston L, Dear BF, Gandy M, Fogliati VJ, Kayrouz R, Sheehan J, et al. Exploring the efficacy and acceptability of Internet-delivered cognitive behavioural therapy for young adults with anxiety and depression: an open trial. Aust N Z J Psychiatry 2014 Sep;48(9):819-827. [doi: 10.1177/0004867414527524] [Medline: 24622977]
57. McLaughlin K. Virtual psychology clinic helps the hard to reach. J Physiotherapy 2016 Apr;62(2):120. [doi: 10.1016/j.jphys.2016.02.007]
58. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of Adherence by Adolescents to a Cognitive Behavior Therapy Website in School and Community-Based Settings. J Med Internet Res 2009;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
59. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]
60. O'Kearney R, Gibson M, Christensen H, Griffiths KM. Effects of a cognitive-behavioural internet program on depression, vulnerability to depression and stigma in adolescent males: a school-based controlled trial. Cogn Behav Ther 2006;35(1):43-54. [doi: 10.1080/16506070500303456] [Medline: 16500776]
61. O'Kearney R, Kang K, Christensen H, Griffiths K. A controlled trial of a school-based Internet program for reducing depressive symptoms in adolescent girls. Depress Anxiety 2009;26(1):65-72. [doi: 10.1002/da.20507] [Medline: 18828141]
62. Sethi S, Campbell AJ, Ellis LA. The Use of Computerized Self-Help Packages to Treat Adolescent Depression and Anxiety. Journal of Technology in Human Services 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
63. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]
64. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
65. Lillevoll KR, Vangberg HCB, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014 Jan 21;14:14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]
66. 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]
67. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]

68. Sethi S. Treating Youth Depression and Anxiety: A Randomised Controlled Trial Examining the Efficacy of Computerised versus Face-to-face Cognitive Behaviour Therapy. Australian Psychologist 2013 Feb 14;48(4):249-257. [doi: 10.1111/ap.12006]
69. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
70. Kobak KA, Mundt JC, Kennard B. Integrating technology into cognitive behavior therapy for adolescent depression: a pilot study. Ann Gen Psychiatry 2015;14:37 [FREE Full text] [doi: 10.1186/s12991-015-0077-8] [Medline: 26535048]
71. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy. ACM Trans Comput-Hum Interact 2014 Feb 01;21(1):1-26. [doi: 10.1145/2513179]
72. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2016 Jun 17:-. [doi: 10.1111/eip.12354] [Medline: 27311581]
73. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [Medline: 22517917]
74. Cheek C, Bridgman H, Fleming T, Cummings E, Ellis L, Lucassen MF, et al. Views of Young People in Rural Australia on SPARX, a Fantasy World Developed for New Zealand Youth With Depression. JMIR Serious Games 2014 Feb 18;2(1):e3 [FREE Full text] [doi: 10.2196/games.3183] [Medline: 25659116]
75. Shepherd M, Fleming T, Lucassen M, Stasiak K, Lambie I, Merry SN. The design and relevance of a computerized gamified depression therapy program for indigenous māori adolescents. JMIR Serious Games 2015 Mar 03;3(1):e1 [FREE Full text] [doi: 10.2196/games.3804] [Medline: 25736225]
76. The University of Auckland. SPARX Fact Sheet URL: https://www.fmhs.auckland.ac.nz/assets/fmhs/faculty/ABOUT/newsandevents/docs/SPARX%20Fact%20sheet.pdf [WebCite Cache ID 6o85D6g8P]
77. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
78. Cheek C, Fleming T, Lucassen MF, Bridgman H, Stasiak K, Shepherd M, et al. Integrating Health Behavior Theory and Design Elements in Serious Games. JMIR Ment Health 2015;2(2):e11 [FREE Full text] [doi: 10.2196/mental.4133] [Medline: 26543916]
79. Perry Y, Calear AL, Mackinnon A, Batterham PJ, Licinio J, King C, et al. Trial for the Prevention of Depression (TriPoD) in final-year secondary students: study protocol for a cluster randomised controlled trial. Trials 2015 Oct 12;16:451 [FREE Full text] [doi: 10.1186/s13063-015-0929-1] [Medline: 26458896]
80. Wong N, Kady L, Mewton L, Sunderland M, Andrews G. Preventing anxiety and depression in adolescents: A randomised controlled trial of two school based Internet-delivered cognitive behavioural therapy programmes. Internet Interventions 2014 Apr;1(2):90-94. [doi: 10.1016/j.invent.2014.05.004]
81. Silverstone PH, Bercov M, Suen VYM, Allen A, Cribben I, Goodrick J, et al. Initial Findings from a Novel School-Based Program, EMPATHY, Which May Help Reduce Depression and Suicidality in Youth. PLoS One 2015;10(5):e0125527 [FREE Full text] [doi: 10.1371/journal.pone.0125527] [Medline: 25974146]
82. McCarty CA, Weisz JR. Effects of psychotherapy for depression in children and adolescents: what we can (and can't) learn from meta-analysis and component profiling. J Am Acad Child Adolesc Psychiatry 2007 Jul;46(7):879-886 [FREE Full text] [doi: 10.1097/chi.0b013e31805467b3] [Medline: 17581452]
83. Weersing VR, Rozenman M, Gonzalez A. Core components of therapy in youth: do we know what to disseminate? Behav Modif 2009 Jan;33(1):24-47. [doi: 10.1177/0145445508322629] [Medline: 18955540]
84. beyondblue: the National Depression Initiative. 2011. Clinical practice guidelines: Depression in adolescents and young adults URL: http://docplayer.net/2638147-Clinical-practice-guidelines-depression-in-adolescents-and-young-adults-february-2011.html [WebCite Cache ID 6o81vBvW3]
85. 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]
86. van Vugt H, Konijn E, Hoorn J, Keur I, Eliëns A. Realism is not all! User engagement with task-related interface characters. Interacting with Computers 2007 Mar;19(2):267-280. [doi: 10.1016/j.intcom.2006.08.005]
87. Everard A, Galletta D. How presentation flaws affect perceived site quality, trust, and intention to purchase from an online store. J Manage Inform Syst 2005;22(3):56-95.
88. Robins D, Holmes J. Aesthetics and credibility in web site design. Inform Process Manag 2008 Jan;44(1):386-399. [doi: 10.1016/j.ipm.2007.02.003]
89. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]

90. Wildeboer G, Kelders SM, van Gemert-Pijnen JE. The relationship between persuasive technology principles, adherence and effect of web-Based interventions for mental health: A meta-analysis. Int J Med Inform 2016 Dec;96:71-85. [doi: 10.1016/j.ijmedinf.2016.04.005] [Medline: 27117057]
91. Drozd F, Lehto T, Oinas-Kukkonen H. In: Bang M, Ragnemalm EL, editors. Exploring perceived persuasiveness of a behavior change support system: a structural model. Berlin/Heidelberg: Springer; 2012:157-168.
92. Kelders S, van Gemert-Pijnen JE. Using log-data as a starting point to make eHealth more persuasive. In: Berkovsky S, Freyne J, editors. Persuasive Technology. Berlin/Heidelberg: Springer; 2013:99-109.
93. Räisänen T, Lehto T, Oinas-Kukkonen H. Practical findings from applying the PSD model for evaluating software design specifications. In: Ploug T, Hasle PH, Oinas-Kukkonen H, editors. Persuasive Technology. Berlin/Heidelberg: Springer; 2010:185-192.
94. Orji R. Exploring the Persuasiveness of Behavior Change Support Strategies and Possible Gender Differences. 2014 Presented at: Second International Workshop on Behavior Change Support Systems (BCSS); 2014; Canada p. 41-57.
95. Kelders SM, Oinas-Kukkonen H, Oörni A, van Gemert-Pijnen JE. Health Behavior Change Support Systems as a research discipline: A viewpoint. Int J Med Inform 2016 Dec;96:3-10. [doi: 10.1016/j.ijmedinf.2016.06.022] [Medline: 27432207]
96. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]

Abbreviations

CBT: cognitive behavioral therapy
C-M-O: context-mechanism-outcome
IPT: Interpersonal Psychotherapy
MDD: major depressive disorder
MMAT: mixed method appraisal tool
PSD: persuasive system design

Edited by J Torous; submitted 02.03.17; peer-reviewed by A Carcone, TR Soron, Y Byambasuren; comments to author 29.05.17; revised version received 05.06.17; accepted 30.06.17; published 09.08.17.

Please cite as:
Wozney L, Huguet A, Bennett K, Radomski AD, Hartling L, Dyson M, McGrath PJ, Newton AS
How do eHealth Programs for Adolescents With Depression Work? A Realist Review of Persuasive System Design Components in Internet-Based Psychological Therapies
J Med Internet Res 2017;19(8):e266
URL: http://www.jmir.org/2017/8/e266/
doi:10.2196/jmir.7573
PMID:28793983

Original Paper

The Effect of Integration of Self-Management Web Platforms on Health Status in Chronic Obstructive Pulmonary Disease Management in Primary Care (e-Vita Study): Interrupted Time Series Design

Esther PWA Talboom-Kamp[1,2], MBA, MD; Noortje A Verdijk[1,2], MSc, PhD; Marise J Kasteleyn[1], MSc, PhD; Lara M Harmans[2], MSc; Irvin JSH Talboom[3]; Ingrid Looijmans-van den Akker[4], MD, PhD; Nan van Geloven[5], MSc, PhD; Mattijs E Numans[1], MD, PhD; Niels H Chavannes[1], MD, PhD

[1]Leiden University Medical Center, Public Health and Primary Care Department, Leiden, Netherlands

[2]Saltro Diagnostic Center, Utrecht, Netherlands

[3]Zorgdraad Foundation, Oosterbeek, Netherlands

[4]Leidsche Rijn Julius Healthcare Centers, location Terwijde, Utrecht, Netherlands

[5]Leiden University Medical Center, Department of Medical Statistics and Bioinformatics, Leiden, Netherlands

Corresponding Author:
Esther PWA Talboom-Kamp, MBA, MD
Leiden University Medical Center
Public Health and Primary Care Department
Albinusdreef 2
2333ZA Leiden
Netherlands
Phone: 31 71 526 9111
Email: e.talboom@saltro.nl

Abstract

Background: Worldwide nearly 3 million people die from chronic obstructive pulmonary disease (COPD) every year. Integrated disease management (IDM) improves quality of life for COPD patients and can reduce hospitalization. Self-management of COPD through eHealth is an effective method to improve IDM and clinical outcomes.

Objectives: The objective of this implementation study was to investigate the effect of 3 chronic obstructive pulmonary disease eHealth programs applied in primary care on health status. The e-Vita COPD study compares different levels of integration of Web-based self-management platforms in IDM in 3 primary care settings. Patient health status is examined using the Clinical COPD Questionnaire (CCQ).

Methods: The parallel cohort design includes 3 levels of integration in IDM (groups 1, 2, 3) and randomization of 2 levels of personal assistance for patients (group A, high assistance, group B, low assistance). Interrupted time series (ITS) design was used to collect CCQ data at multiple time points before and after intervention, and multilevel linear regression modeling was used to analyze CCQ data.

Results: Of the 702 invited patients, 215 (30.6%) registered to a platform. Of these, 82 participated in group 1 (high integration IDM), 36 in group 1A (high assistance), and 46 in group 1B (low assistance); 96 participated in group 2 (medium integration IDM), 44 in group 2A (high assistance) and 52 in group 2B (low assistance); also, 37 participated in group 3 (no integration IDM). In the total group, no significant difference was found in change in CCQ trend (P=.334) before (–0.47% per month) and after the intervention (–0.084% per month). Also, no significant difference was found in CCQ changes before versus after the intervention between the groups with high versus low personal assistance. In all subgroups, there was no significant change in the CCQ trend before and after the intervention (group 1A, P=.237; 1B, P=.991; 2A, P=.120; 2B, P=.166; 3, P=.945).

Conclusions: The e-Vita eHealth-supported COPD programs had no beneficial impact on the health status of COPD patients. Also, no differences were found between the patient groups receiving different levels of personal assistance.

Trial Registration: Netherlands Trial Registry NTR4098; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4098 (Archived by WebCite at http://www.webcitation.org/6sbM5PayG)

(J Med Internet Res 2017;19(8):e291) doi:10.2196/jmir.8262

KEYWORDS

COPD; CCQ; health status; eHealth; self-management; integrated disease management; self-efficacy; Web-based platform; primary care; chronically ill; blended care

Introduction

Chronic obstructive pulmonary disease (COPD) is a slowly progressive lung disease and a main cause of morbidity and mortality in high-, middle-, and low-income countries [1]. Worldwide, nearly 3 million people die from COPD every year which, in 2012, was equal to 6% of all deaths globally [2,3].

According to current COPD guidelines, symptoms and airflow obstruction should be regularly monitored to modify treatment and identify complications at an early stage [4,5]. Monitoring should contribute to delaying disease progression and alleviate its manifestations; the most important primary care objective should be to improve quality of life (QoL) [6]. In primary care COPD studies, the mean score on the Clinical COPD Questionnaire (CCQ) reflects a mildly symptomatic COPD [7], and the health status of patients was found to decline over a longer period of time [8].

In the last decade, integrated disease management (IDM) was introduced to improve quality of care. An IDM program consists of different components of care in which different health care providers are cooperating and collaborating to provide efficient and good quality of care; IDM for COPD improves disease-specific QoL and exercise capacity and also reduces hospital admissions and hospital days per person [9].

To improve the quality and efficiency of IDM and reduce health care costs, self-management of COPD patients was introduced and has proven an effective method [10,11,12]. The core components of self-management include education, eliciting personalized goals, psychological coping strategies, improving compliance to treatment, behavioral change, and practical and social support [13,14]. Interventions to support self-management have shown reductions in hospital admissions and fewer sick days because of exacerbations [15,16]. Chronically ill patients who received person-centered care focusing on patient activation and goal setting are better self-managers [17], with self-efficacy as an important factor influencing self-management behaviors [18].

eHealth interventions are effective in stimulating self-management in chronic disease; patients are better able to cope with their illness and adapt their lifestyle, while eHealth support also reduces medical staff consultations [19]. The deployment of eHealth applications facilitates accessibility to health care, enhances patient understanding of their disease, sense of control, and willingness to engage in self-management [20,21]. Although patient attitudes and receptiveness toward eHealth applications are promising in certain groups of age and education [22,23,24], large-scale adoption of eHealth in daily practice still lags behind predictions in comparison with the explosive growth of other digital tools like Facebook and Twitter [25] (also, during recent years, online banking acceptance has increased rapidly worldwide [26]). A major challenge of eHealth in care coordination is to make it beneficial and easy to use for both health care providers and patients [27]. It is important that online self-management support is a fully integrated element of IDM; COPD and asthma patients tended to use an online application on a regular basis when the caregiver was involved, whereas patients on their own used the application only sporadically [28]. For clinicians, the eHealth evidence base needs strengthening, while for primary care practices, a learning process including staff training needs to be instituted [29]. Since advances in eHealth are not clear for patients who have never used it, it is necessary to provide patients with more and better information about the possibilities and potential benefits of eHealth to increase their self-efficacy and provide a feeling of more personal control in daily life [30]. Furthermore, poor user-friendliness of Web-based applications and the lack of push factors (frequent reminders or messages by caregivers) are a common cause of low usage or decline in usage [31].

Despite high expectations and numerous eHealth initiatives, implementation and use of eHealth applications is not yet common practice. Therefore, this e-Vita study investigated the effect of usage of eHealth platforms on the health status of COPD patients treated in primary care. In this paper, we describe 3 eHealth-supported care programs with different components that support the treatment of COPD patients through digital coaching. Two programs were applied as blended care (ie, implemented within usual care to explore the potential for daily health care practice), and one program was applied with the self-management platform as an independent component.

In our e-Vita study, use of the self-management platforms was higher when the platform was an integrated part of IDM and with adequate personal assistance on how to use the platform [32]. We hypothesize that use of the platforms will improve self-management skills and thereby stabilize the health status of COPD patients, with a greater effect with higher usage.

Methods

Study Design

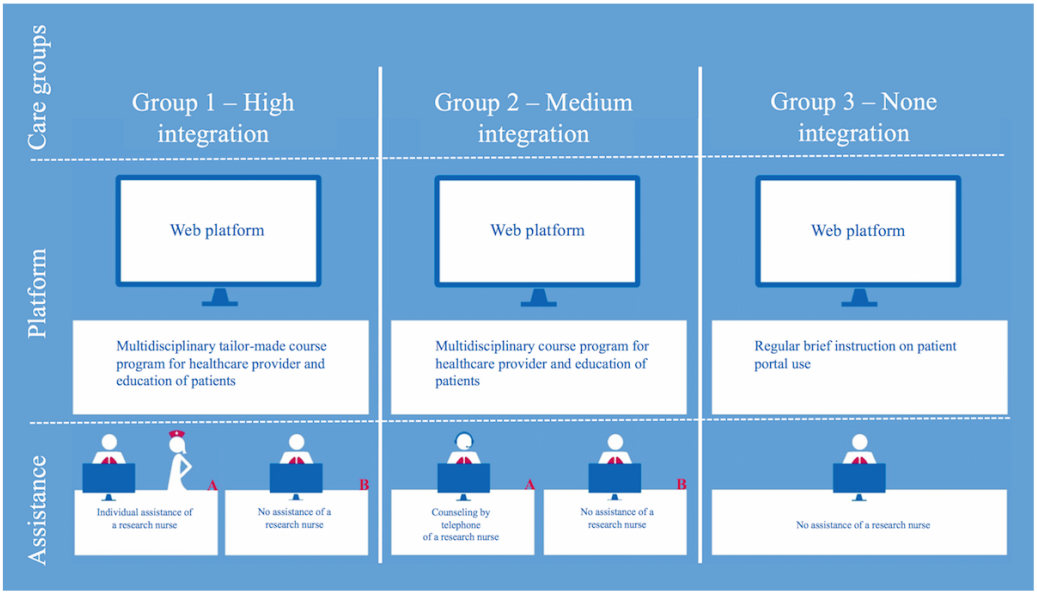
Full methodological details of this multilevel parallel cohort design have been reported previously [33]. For our implementation study in a real-life health care setting, we used an interrupted time series (ITS) design in a pragmatic trial in which data were collected at multiple time points before and after implementation to detect whether the intervention had a significantly greater effect than any underlying secular trend [34]. The ITS is performed according to the guidelines of the Effective Practice and Organization of Care Cochrane review group [35].

Figure 1 presents an overview of the study design. The study included 3 different care groups in primary care (groups 1 to 3); all patients started using the Web-based platforms (Multimedia Appendices 1 and 2).

In groups 1 and 2, we offered the patients blended care, and in group 3 the self-management platform was offered to the patients as an independent module. In group 1, the online platform was offered as a highly integrated part of the COPD IDM (High) with a tailor-made intensive course program on COPD for health care professionals that contained education on COPD, training about the possibilities of eHealth, and conversational techniques to approach patients in an equal and coaching way. All patients in group 1 started with a personal consultation with the primary care nurse with coaching on the necessity of self-management and explanations of their burden of disease and the eHealth program. Group 2 had a medium level of integration (Medium) with a basic course program for health care professionals on COPD that contained education on COPD and training about the possibilities of eHealth. All patients in group 2 started with a personal consultation with the primary care nurse with coaching and explanation of the self-management program. In groups 1 and 2, the first question the nurse asked was what patients would like to achieve in their daily life when they would have a lower burden of disease. With the platform, the patients could work with a coaching program on their personal goals, actions and health-related QoL [33]. In group 3, the online platform was offered without integration in COPD IDM (None); health care providers and patients received basic instructions on the platform.

Two different levels of assistance for patients were distinguished within groups 1 and 2: one with a high level of personal assistance (A) and the other with a low level of personal assistance (B). Patients were randomly subdivided into 2 groups. In group 1A, high-level support implied home visits to patients by a research nurse who coached and assisted in use of the platform. In group 2A, high-level support implied telephone consultation (3 times during the intervention period, scheduled after 3, 6, and 9 months) between the patient and a research nurse who explained use of the platform. In groups 1B and 2B, low-level support implied that the primary care nurse showed the patient only once how to use the platform. Patients in group 3 who used the online self-management platform had no support from the caregivers or research nurses. Both platforms were provided for the intervention period of 15 months.

Figure 1. Study design of the e-Vita online platform.



Participants

A total of 3 care groups participated. Patients were eligible when they were diagnosed with COPD according to the Global Initiative for Chronic Obstructive Lung Disease criteria (post-bronchodilator FEV_1/FVC <0.7) in accordance with the COPD Guidelines of the Dutch College of General Practitioners [36] and when they were being treated for COPD in primary care. Patients were excluded if they were unable to fill in questionnaires, had no access to the Internet, had a terminal illness, were immobile, or were severe substance abusers.

Recruitment of Patients

We started by recruiting the care groups: managing general practitioners (GPs) in groups 1, 2, and 3 decided to participate because they wanted to contribute to possible health care improvement. Members of the care groups (GPs) volunteered to participate.

Patients were invited to participate by letter via their own GP. When participants of the e-Vita study logged in and used the Web platform at least once, they were defined as users. Patients were defined as lost to follow-up if they did not log on to the platform for at least 12 months after signing informed consent or if they did not complete the digital questionnaires within the intervention period.

Ethics Approval and Consent to Participate

This study was conducted according to the principles of the Declaration of Helsinki (version 59, 2008) and in accordance with the Medical Research Involving Human Subjects Act. The study was approved by the Medical Ethics Committee of the Medical University Center of Leiden. All subjects provided written informed consent.

Intervention

The online self-management platforms were created by national experts on chronic disease management guided by interviews with COPD patients about their thoughts and feelings related to living with COPD; the experiences of professional COPD experts were also integrated. The main content of the platform consists of insight into personal health data, self-monitoring of health values (CCQ, Modified Medical Research Council Dyspnea scale [MRC]), education, and a coach for attaining personal goals. The educational and coaching programs were developed by the Lung Alliance of the Netherlands. The online self-management platform e-Vita is an initiative of the Dutch foundation Care Within Reach [37]. The patients received automated online reminders. Details on the online platforms are published elsewhere [33].

Outcome Measures

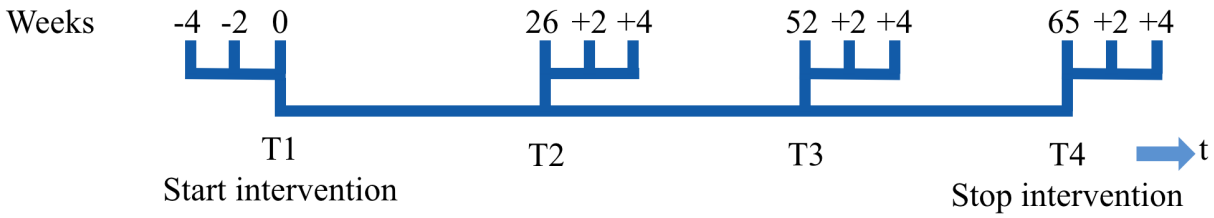
The primary outcome is a clinical one, expressed as health status (ie, the CCQ). This questionnaire was designed by Van der Molen [38] and consists of 10 items, each answered on a 7-point Likert scale. The CCQ comprises 3 domains: symptom state (4 items), functional state (4 items), and mental state (2 items). The CCQ total score is calculated as the mean of the sum of all items (minimum 0, maximum 6), with a higher value indicating lower health status. The CCQ is a reliable and valid questionnaire with a Cronbach alpha of 0.89-0.91.

Data collected at baseline included age, gender, education level, and total scores on the CCQ, MRC [39], Generalized Self-Efficacy Scale (GSES) [40], and EuroQol 5-dimension questionnaire (EQ5D) [41]. Education was self-reported using 8 response categories ranging from no formal education to graduate or professional level and converted into 3 levels (low, medium, high). In the main analyses, personal assistance for the participants (high assistance vs low assistance) and integration in IDM (integrated vs not integrated) were used as determinants.

Data Collection

Data were extracted from the log files of the self-management platforms. Figure 2 shows the measurement schedule of the CCQ. During the 15-month intervention period, there were 4 measurement periods with 3 CCQ questionnaires at each period (3 data points before intervention and 9 data points after intervention) in order to apply ITS analysis.

Figure 2. Interrupted time series measurements of the Clinical COPD Questionnaire per interval.



Sample Size Calculation

Generally, the health status of patients with COPD decreases over time. Studies on IDM in COPD in primary care show that a general increase in CCQ of 0.5 (SD 0.75) can be expected over a 1-year period [42,43]. In this study, we offered patients a Web platform in addition to their regular IDM program. Therefore, we hypothesized that the regular increase in CCQ (0.5 points per year) would change to stabilization of health status as compared to literature [44].

To measure differences in health status (>0.5 CCQ points) at 80% power (SD 0.75 and α=0.05), 37 patients needed to be included. Based on an estimated 20% drop-out during the study period, 45 (37/0.80) patients needed to be included. As we used 2 different implementation methods (with and without personal assistance) in 2 of the care groups, 2×45=90 patients needed to be included in those groups. In the third care group, because only 1 implementation method was used, a total of 45 patients were required to achieve sufficient statistical power.

Statistical Methods

Categorical baseline characteristics were reported as numbers and percentages, normally distributed continuous variables as means with standard deviations (SD), and nonnormally distributed variables as medians with interquartile ranges (IQRs). Characteristics between the 3 groups were examined using chi-square tests and Kruskal-Wallis tests.

ITS analyses were used to study time trends before and after intervention. Due to the correlation between the repeated measurements within a patient, we used multilevel linear regression modeling (mixed models) to analyze CCQ data in the total group as well as in the groups with high and low levels of assistance and in the 5 subgroups.

The analyses allowed us to quantify the effect of the intervention on CCQ versus the observed preintervention period. Estimates for regression coefficients corresponding to 2 standardized effect sizes were obtained: a direct change in the level of the CCQ (also called step change or jump) and a change in trend of the CCQ before and after the intervention [34].

Included in the model for the total group and 5 subgroups as fixed effects were time, treatment, and the interaction between time and treatment; the model comparing the groups with high and low assistance additionally contained the assistance group factor and the interaction of this factor. All models included a random intercept per patient. When there was a substantial improvement in the Akaike Information Criterion (used to assess the model fit score), an additional random slope (time) was used.

Because of a nonnormal distribution of the CCQ data, the log of the CCQ data was used as outcome in the multilevel linear regression models. The analysis model did not test or correct for seasonality; although seasonality influences exacerbations of COPD, it has no effect on the CCQ [45,46]. We visually assessed normality of the residuals to evaluate the validity of the assumptions of the mixed models analysis.

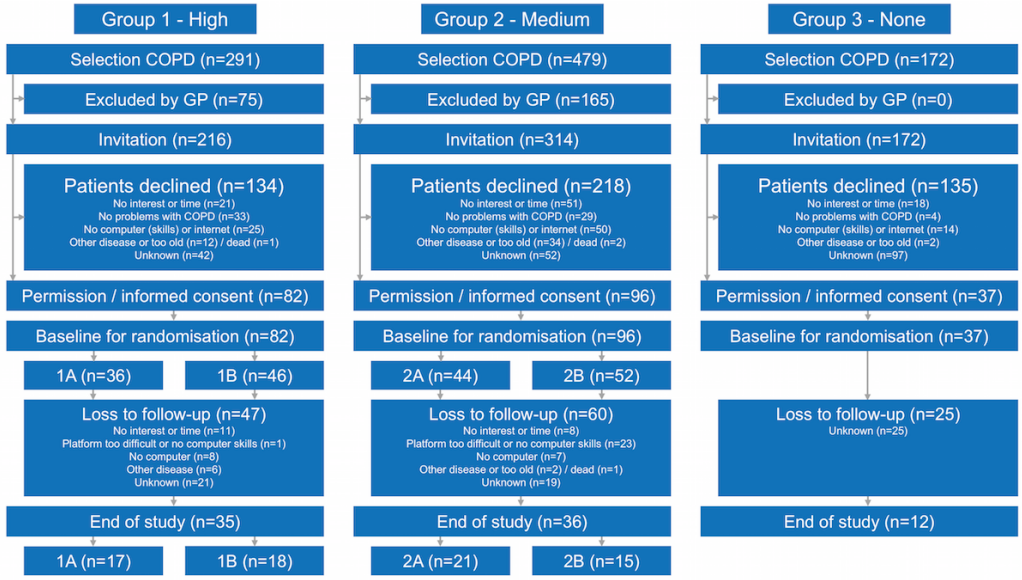
Results

Inclusion

A total of 942 COPD patients were selected from 3 care groups (Figure 3). The GPs of these care groups excluded 240 COPD patients from participation due to other diseases, treatment in hospital, or incompetency to participate in the program. Finally, 702 COPD patients were invited, of whom 215 (30.6%) agreed and provided informed consent.

The number of patients lost to follow-up (no log on to the platform after signing informed consent or not completing the entire intervention period) was 132; results of the nonparticipation analysis are reported elsewhere [33]. Figure 3 shows the reasons for loss to follow-up in groups 1 and 2; patients in group 3 were not asked for their reasons.

Figure 3. Flowchart of the e-Vita chronic obstructive pulmonary disease study.



Baseline Characteristics of Patients With Chronic Obstructive Pulmonary Disease

Table 1 presents the baseline demographic and clinical characteristics of the included COPD population (median age 66.6 years; 52.1% male). These patients had mildly symptomatic COPD which was reflected by a median MRC scale of 1.0 and a median CCQ of 1.2. Out of 215 participants, 89 (41.4%) filled in the online questionnaire for education level; most participants had a middle education level, reflected by a 4 or 5 on a scale of 1 to 7 (38/89, 42.7%). The median GSES was 3.3, and the median EQ-5D was 0.86. The characteristics age (χ^2_2=5.4, *P*=.07), education level (χ^2_4=2.2, *P*=.70), GSES (χ^2_2=1.7, *P*=.42), and EQ-5D (χ^2_2=2.4, *P*=.28) were similar in the 3 groups. There was a difference in the characteristic gender (χ^2_2=6.8, *P*=.03), with more male patients in group 1, and a difference in CCQ (χ^2_2=6.5, *P*=.04) and MRC scale (χ^2_2=11.3, *P*=.003), with a higher CCQ and MRC scale in group 2.

Table 1. Baseline characteristics of patients with chronic obstructive pulmonary disease in the e-Vita study.

Participants		Group 1 (high)		Group 2 (medium)		Group 3 (none)	Total
Assistance		High n=36	Low n=46	High n=44	Low n=52	n=37	n=215
Age, years, (IQR[a])		66.3 (61.0-79.2)	65.6 (61.3-73.4)	68.7 (64.0-78.3)	66.8 (60.3-75.1)	64.1 (61.5-69.2)	66.6 (61.4-74.7)
Males, n (%)		19 (52.8)	32 (69.6)	17 (38.6)	24 (46.2)	20 (54.1)	112 (52.1)
Education level, n (%)							
	Low	4 (28.6)	8 (38.1)	5 (22.7)	8 (42.1)	7 (53.8)	32 (36.0)
	Medium	7 (50.0)	8 (38.1)	11 (50.0)	8 (42.1)	4 (30.8)	38 (42.7)
	High	3 (21.4)	5 (23.8)	6 (27.3)	3 (15.8)	2 (15.8)	19 (21.3)
Questionnaire, median (IQR)							
	CCQ[b]	1.0 (0.6-1.9)	1.2 (0.8-1.6)	1.3 (0.9-2.1)	1.4 (1.1-2.1)	1.3 (0.6-1.8)	1.2 (0.8-1.9)
	MRC[c]	1.0 (1.0-3.0)	1.0 (1.0-2.0)	2.0 (1.0-3.0)	2.0 (1.0-2.0)	1.0 (1.0-1.0)	1.0 (1.0-2.0)
	GSES[d]	3.4 (3.1-3.7)	3.3 (3.0-3.8)	3.3 (2.8-3.5)	3.3 (3.1-3.7)	3.4 (3.3-3.7)	3.3 (3.0-3.7)
	EQ-5D[e]	0.85 (0.7-1.0)	0.89 (0.81-1.0)	0.85 (0.72-1.0)	0.84 (0.71-1.0)	0.9 (0.84-1.0)	0.86 (0.78-1.0)

[a]IQR: interquartile range.

[b]CCQ: Clinical COPD Questionnaire.

[c]MRC: Modified Medical Research Council Dyspnea score.

[d]GSES: Generalized Self-Efficacy Scale.

[e]EQ-5D: EuroQoL 5-Dimension Questionnaire.

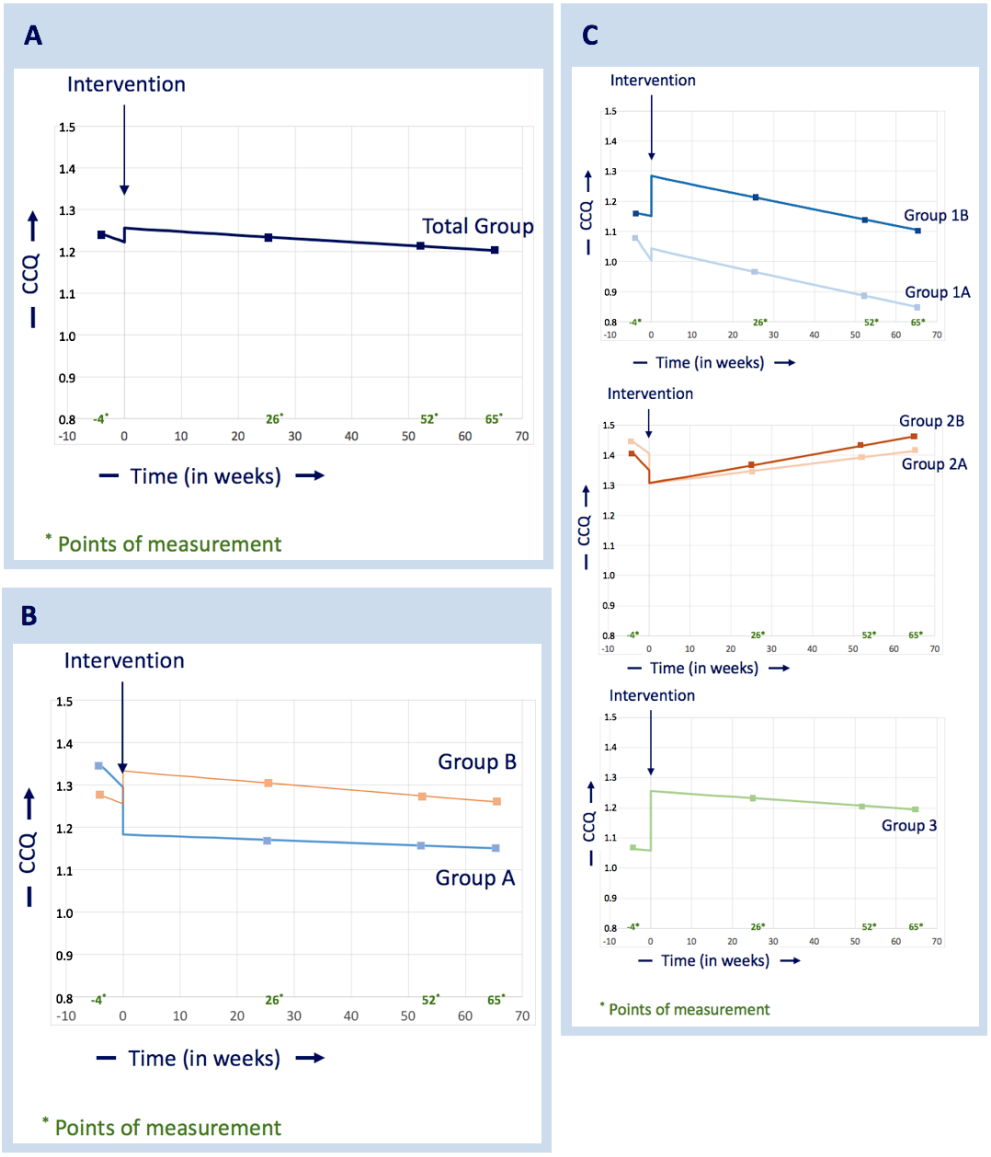
Health Status Changes

Figure 4 A shows the effect of the intervention on CCQ in the total patient group. The decrease before the intervention was 0.5% per month and after the intervention 0.08% per month; this difference was not significant (*P*=.334). The estimated direct change in the level of the CCQ slopes at the moment of the intervention (jump) was –0.015 (*P*=.421) implying that the CCQ trend was 1.5% lower before the intervention.

Figure 4 B shows the effect of the intervention on CCQ in the groups with a high level of personal assistance (A) and a low level of personal assistance (B). In group A, the preintervention decrease was 0.8% per month and the decrease after the intervention was 0.05% per month; in group B, the preintervention decrease was 1.8% per month and the decrease after the intervention was 0.1% per month. No significant difference was found in CCQ changes going from preintervention to postintervention between groups A and B (*P*=.429). The direct change in the level of the CCQ slopes at the moment of the intervention (jump) was 0.017 in group A and –0.033 in group B (implying that the CCQ trend was 1.7% higher before the intervention in group A and 3.3% lower in group B). There was no significant difference in the jumps (*P*=.207).

Figure 4 C shows the effect of the intervention on the CCQ in the 5 subgroups; no significant difference was found in the slope of the CCQ before and after the intervention (1A, *P*=.237; 1B, *P*=.991; 2A, *P*=.120; 2B, *P*=.166; 3, *P*=.945). The direct changes in the level of the CCQ slopes before and after the intervention (jump) were –0.0196 in group 1A, –0.0582 in group 1B, 0.0426 in group 2A, 0.0184 in group 2B, and –0.0874 in group 3 and were not significant.

Figure 4. (A) Total group analysis of health status, (B) analysis of health status in groups receiving high and low levels of assistance, and (C) subgroup analyses of health status.



Discussion

Principal Findings

This study investigated the effect of use of eHealth platforms on the health status of COPD patients in primary care. No changes in health status were found before and after introduction of the eHealth-supported COPD programs, and no differences were found between care groups with a high versus a low level of personal assistance.

It is essential to carefully review the design strategies for integrating eHealth applications within disease management [47] by means of a thorough evaluation and analysis of the results. A recent study analyzed the effect of wearable devices that monitored and provided feedback on physical activity among young adults with obesity; providing a group with this device resulted in less weight loss over 24 months [48]. In a trial of elderly persons with a high risk for prehospitalization, telemonitoring was offered for monitoring and reporting symptoms but did not result in lower hospitalizations or emergency room visits (although mortality was higher in this telemonitoring group) [49]. High expectations of eHealth should be preceded by evaluations in pragmatic studies on implementation [50].

In our study, changes in health status CCQ were not within the range of a minimal clinically important difference [51]. QoL and health status are determined by a significant number of factors [46]. We expect that eHealth interventions will be effective in stimulating self-management and stabilizing health status in COPD patients when these patients use the platforms for a longer period of time. In earlier studies, a worsening in health status was found for primary care patients over a longer period of time [8]; this finding was not confirmed in our e-Vita study. In Dutch primary care, the standards of IDM are high, with a wide variety in the implementation of interventions [7]; this might explain the absence of a worsening of health status before the intervention in our study. The introduction and

integration of eHealth within IDM will not make a significant difference in the short term due to the high standard of IDM.

In our study, the median baseline CCQ score of 1.2 was low compared with scores in other primary care COPD studies, which reflects mildly symptomatic COPD [7]. This limits the room for stabilization or improvement in our primary outcome (ie, ceiling effect).

In a patient population with more severe COPD (patient-data meta-analysis from 2016), self-management interventions improved health-related QoL at 12 months but not at 6 months [52]; this confirms our observation that long-term use of platforms is necessary for an effect on health status. In our research, the platforms are probably not sufficiently customized to the wishes or needs of COPD patients to provide sufficient motivation to use the platform on a regular basis for a longer period; in our e-Vita study, a significant number of users stopped using the platforms (attrition) [32].

The change in level of CCQ (positive/negative) at the start of the intervention might be explained by the participant rise in consciousness regarding their health status, thereby completing the questionnaire more critically after explanation from a health care professional. Similar to our study, in a randomized controlled trial (RCT) with asthma patients, the QoL was enhanced over the first 3 months after starting to use a self-management portal [53].

The effect of eHealth cannot easily be evaluated in a classical RCT; integrating eHealth in IDM is a complex intervention in a multidisciplinary care process. Pragmatic trials frequently include complex interventions and often involve the skills and experience of health care professionals [54] and are, therefore, more suitable for eHealth studies.

EHealth technologies for chronic conditions can be used to enhance self-management and revise the chronic care model; patients who actively participate in their care achieve valuable and sustained improvement in well-being [55,56]. Findings in many eHealth studies suggest that the use of a personal health record or self-management platform can promote an informed or activated patient and augment the chronic care model for self-management support and productive interactions even though a direct dosage-effect relationship (usually analyzed in a classical RCT) is not common in eHealth studies [57]. Also in our e-Vita study, interpretation of the results on use and health status cannot be made in a direct dosage-effect relationship. Use of the self-management platform was higher when the platform was an integrated part of IDM, with trained caregivers encouraging patients to use the platform and with personal assistance about how to use the platform, but without a significant change in health status. Based on current literature and the e-Vita COPD study, we conclude that eHealth-supported self-management integrated into usual care can help patients with COPD manage their disease better.

Further studies based on this study and current literature are needed to establish the mechanisms most likely to ensure the successful development and implementation of Web-based self-management interventions, including considerations about how the intervention is integrated in IDM and how it enhances the patient's self-management behavior to stimulate long-term use of self-management platforms with a stabilizing effect on health status [11,58].

Strengths and Limitations

This e-Vita COPD study has several strengths. To our knowledge, it is the first to combine different study designs thereby enabling simultaneous investigation of the effect of eHealth and the effects of different organizational implementation methods on health status. Randomization was performed for the level of assistance provided to patients, allowing comparison of patient groups with high and low levels of assistance. Because the care groups 1, 2, and 3 were not randomized, no analysis of the differences between these groups can be made.

An advantage of the ITS design is that it detects changes that are delayed or intermittent and can determine whether the change is permanent or temporary. The design, including the 3 datapoints of CCQ before the intervention, also allows evaluation of variables which are changing by comparing slopes of trend lines before and after the intervention.

This study also has some limitations. Development of the platforms was relatively difficult due to lack of experience in this field. Also, decisions made during the design phase were beyond the influence of our group but affected the usability of the platforms. Self-management skills imply behavioral changes which require some time, whereas the present study period was restricted to 15 months. Furthermore, patients in a primary care setting have a low burden of disease (in this study, a median score on the CCQ of 1.2) and motivation to use the platform might be negatively influenced by this fact. In respiratory medicine there is a lack of research on patients with mild-to-moderate COPD despite that over 80% of COPD patients suffer from this stage of disease and are often treated in primary care [59]. Other projects among care groups (eg, patient education, start different IDMs) might influence the speed and thoroughness of the implementation of our platforms.

This study also has limitations typically associated with eHealth trials. For example, as GPs and patients were free to volunteer, bias might have occurred in our study groups. Users were self-selected and were, presumably, motivated to use the Web-based platform as would be expected in a real-life setting. Also, the patients selected to be invited by the GPs might differ from other patient groups. Furthermore, GPs excluded 25.5% of the COPD patients from this study. Of the 702 eligible patients, 30.6% were willing to participate and provided informed consent, and 61.4% of the participants dropped out during follow-up. Even though the nonparticipants did not differ in age or gender from the participants [32], caution is required when generalizing these results to general practice.

Like most Internet outcome studies, there were 2 types of attrition in our study; attrition from the intervention itself (lack of site utilization) and attrition from the follow-up assessments. This law of attrition (the phenomenon of participants stopping usage) is a common finding in eHealth evaluations and one of the fundamental and methodological challenges in the evaluation of eHealth applications [60]. To prevent both types of attrition,

email reminders were sent by the platform to fill in the questionnaires. All users received urgent and repeated requests to fill in questionnaires by email and by telephone. The attrition curve was analyzed earlier and depicts the push factors that are required to remind participants to use the platform [32]. The loss to follow-up is high with a risk of biased results due to user bias; therefore, these results are only applicable for users of eHealth.

The study aimed to be inclusive rather than exclusive to achieve higher external validity. Patients were excluded if they were unable to fill in questionnaires, had no access to the Internet, had a terminal illness, were immobile, or were severe substance abusers. During the inclusion of patients, we found that patients did not want to start the study for several different reasons: no computer skills, old age, no problems with COPD, and other reasons. These reasons are typical for eHealth research in a primary care setting with a low burden of disease; participants were self-selected and were, presumably, motivated to use the Web-based platform as would be expected in a real-life setting. Therefore, the results of the study are not generalizable to all COPD patients but to those who are willing, motivated, and able to use eHealth. Nevertheless, we believe that this study is inclusive rather than exclusive, since there are almost no limitations for participation for this group of motivated patients.

However, the practical applicability of our results for other primary care groups is positive (ie, the study provides practical insight into successful implementation of patient platforms). Nevertheless, primary care organizations should take into account the different aspects of the organization of blended care and quality of implementation.

Although an RCT provides the most reliable evidence on the effectiveness of interventions, this was not feasible for our implementation study in a real-life health care setting with 3 different care groups. After randomization in groups 1 and 2, more patients were assigned to the groups with a low level of personal assistance (group 1B and 2B). After simple randomization, some discrepancy between the numbers in the comparison groups would be expected [61]. Such unpredictability reflects the essence of randomness. Moreover, the baseline characteristics did not differ significantly between groups with high and low assistance. Therefore, we expect no significant influence on the results.

To measure a significant difference in health status, 45 patients were needed in each subgroup; although these numbers were not met within each subgroup, analysis on the combined groups should be sufficiently powered to detect relevant differences. In addition, the number of data points collected before the intervention has a substantial impact on the strength of an ITS design. It is necessary to collect enough data points to be convinced that a stable estimate of the underlying secular trend has been obtained [62]. In our study, the 3 data points before the intervention represent a minimum number of data and may have influenced the effective power of our study.

Conclusion

There is growing interest in the potential of Web-based self-management platforms to deliver more individually tailored self-management support integrated into the everyday lives of COPD patients to improve their quality of life. In this study, the e-Vita eHealth-supported COPD programs had no significant impact on the health status of COPD patients, health status showed no significant change before or after the introduction of the eHealth-supported programs, and no differences were found between the patient groups receiving different levels of personal assistance.

Acknowledgments

The authors thank Pim Assendelft for his valuable contribution to the design and Anneke Vass for her research assistance. The authors thank the Zorgdraad Foundation for its assistance on the Zorgdraad platform and the acquisition of data. The authors wish to thank all patients, coordinators, caregivers, and participants within the care groups PreventZorg, Zorg op Noord, and Leidse Rijn Julius Gezondheidscentra. This work was supported by the Foundation Care Within Reach (in Dutch, Stichting Zorg Binnen Bereik).

Authors' Contributions

ET is the principle investigator and contributed to all aspects of the research. NV assisted on all aspects. MK assisted on the statistical analysis and is responsible for revising the manuscript several times. LH is responsible for revising the manuscript several times. IT is responsible for the acquisition of data and revising the manuscript several times. IL is responsible for revising the manuscript. NvG assisted on the statistical analysis and is responsible for revising the manuscript several times. MN is responsible for revising the manuscript. NC is responsible for the concept and design of the study and for revising the manuscript. All authors read and approved the concept and design and the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Homepage e-Vita.

[PNG File, 197KB - jmir_v19i8e291_app1.png]

Multimedia Appendix 2

Homepage Zorgdraad.

[PNG File, 194KB - jmir_v19i8e291_app2.png]

Multimedia Appendix 3

SQUIRE checklist.

[PDF File (Adobe PDF File), 448KB - jmir_v19i8e291_app3.pdf]

References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2012 Dec;380(9859):2095-2128. [doi: 10.1016/S0140-6736(12)61728-0]
2. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. Clin Chest Med 2014;35(1):7-16. [doi: 10.1016/j.ccm.2013.10.0022]
3. Word Health Organization. Chronic obstructive pulmonary disease (COPD) fact sheet. 2017. URL: http://www.who.int/mediacentre/factsheets/fs315/en/ [accessed 2017-06-22] [WebCite Cache ID 6rPMg7fia]
4. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2017 report). URL: http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/ [accessed 2017-06-22] [WebCite Cache ID 6rPOYOCL2]
5. Bellamy D, Bouchard J, Henrichsen S, Johansson G, Langhammer A, Reid J, et al. International Primary Care Respiratory Group (IPCRG) Guidelines: management of chronic obstructive pulmonary disease (COPD). Prim Care Respir J 2006 Feb;15(1):48-57 [FREE Full text] [doi: 10.1016/j.pcrj.2005.11.003] [Medline: 16701758]
6. van den Bemt L, Schermer T, Smeele I, Bischoff E, Jacobs A, Grol R, et al. Monitoring of patients with COPD: A review of current guidelines and recommendations. Resp Med 2008 May;102(5):633-641. [doi: 10.1016/j.rmed.2007.12.014]
7. Boland MRS, Kruis AL, Huygens SA, Tsiachristas A, Assendelft WJJ, Gussekloo J, et al. Exploring the variation in implementation of a COPD disease management programme and its impact on health outcomes: a post hoc analysis of the RECODE cluster randomised trial. NPJ Prim Care Resp Med 2015 Dec 17;25(1):1. [doi: 10.1038/npjpcrm.2015.71]
8. Sundh J, Montgomery S, Hasselgren M, Kämpe M, Janson C, Ställberg B, et al. Change in health status in COPD: a seven-year follow-up cohort study. NPJ Prim Care Resp Med 2016 Oct 20;26(1):1. [doi: 10.1038/npjpcrm.2016.73]
9. Kruis AL, Smidt N, Assendelft WJJ, Gussekloo J, Boland MRS, Rutten-van Molken M, et al. Integrated disease management interventions for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2013 Oct 10(10):CD009437. [doi: 10.1002/14651858.CD009437.pub2] [Medline: 24108523]
10. Kruis AL, Smidt N, Assendelft WJJ, Gussekloo J, Boland MRS, Rutten-van Molken M, et al. Cochrane corner: is integrated disease management for patients with COPD effective? Thorax 2014 Nov;69(11):1053-1055 [FREE Full text] [doi: 10.1136/thoraxjnl-2013-204974] [Medline: 24415716]
11. Zwerink M, Brusse-Keizer M, van der Valk PDLPM, Zielhuis GA, Monninkhof EM, van der Palen J, et al. Self management for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2014 Mar 19(3):CD002990. [doi: 10.1002/14651858.CD002990.pub3] [Medline: 24665053]
12. Bourbeau J, Collet J, Schwartzman K, Ducruet T, Nault D, Bradley C. Economic benefits of self-management education in COPD. Chest 2006 Dec;130(6):1704-1711. [doi: 10.1378/chest.130.6.1704] [Medline: 17166985]
13. Taylor S, Pinnock H, Epiphaniou E, Pearce G, Parke H, Schwappach A. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS ? Practical systematic RevIew of Self-Management Support for long-term conditions. Health Serv Deliv Res 2014;2(53):1. [Medline: 25642548]
14. Effing TW, Vercoulen JH, Bourbeau J, Trappenburg J, Lenferink A, Cafarella P, et al. Definition of a COPD self-management intervention: International Expert Group consensus. Eur Respir J 2016 Jul;48(1):46-54. [doi: 10.1183/13993003.00025-2016] [Medline: 27076595]
15. Effing T, Kerstjens H, van der Valk P, Zielhuis G, van der Palen J. (Cost)-effectiveness of self-treatment of exacerbations on the severity of exacerbations in patients with COPD: the COPE II study. Thorax 2009 Nov;64(11):956-962. [doi: 10.1136/thx.2008.112243] [Medline: 19736179]
16. Maltais F, Bourbeau J, Shapiro S, Lacasse Y, Perrault H, Baltzan M, et al. Effects of home-based pulmonary rehabilitation in patients with chronic obstructive pulmonary disease: a randomized trial. Ann Intern Med 2008 Dec 16;149(12):869-878. [Medline: 19075206]
17. van Houtum L, Heijmans M, Rijken M, Groenewegen P. Perceived quality of chronic illness care is associated with self-management: results of a nationwide study in the Netherlands. Health Policy 2016 Apr;120(4):431-439. [doi: 10.1016/j.healthpol.2015.11.006] [Medline: 27017047]
18. Mishali M, Omer H, Heymann AD. The importance of measuring self-efficacy in patients with diabetes. Fam Pract 2011 Feb;28(1):82-87. [doi: 10.1093/fampra/cmq086] [Medline: 21047940]

19. McLean S, Protti D, Sheikh A. Telehealthcare for long term conditions. BMJ 2011;342:d120. [Medline: 21292710]
20. Finkelstein J, Friedman RH. Potential role of telecommunication technologies in the management of chronic health conditions. Dis Manage Health Outcome 2000;8(2):57-63. [doi: 10.2165/00115677-200008020-00001]
21. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
22. Beenkens F. Acceptance of E-Health Technology: A Patient Perspective [doctoral thesis]. Delft: Delft University of Technology; 2011.
23. Bischoff EWMA, Hamd DH, Sedeno M, Benedetti A, Schermer TRJ, Bernard S, et al. Effects of written action plan adherence on COPD exacerbation recovery. Thorax 2011 Jan;66(1):26-31. [doi: 10.1136/thx.2009.127621] [Medline: 21037270]
24. Bucknall CE, Miller G, Lloyd SM, Cleland J, McCluskey S, Cotton M, et al. Glasgow supported self-management trial (GSuST) for patients with moderate to severe COPD: randomised controlled trial. BMJ 2012;344:e1060 [FREE Full text] [Medline: 22395923]
25. KPMG International. Accelerating innovation: the power of the crowd: global lessons in eHealth implementation. 2017. URL: https://assets.kpmg.com/content/dam/kpmg/nz/pdf/Dec/AcceleratingInnovationGloballessonsineHealthimplementationReport-kpmg-nz.pdf [accessed 2017-06-22] [WebCite Cache ID 6rPMyx6Vf]
26. Hosein N, Chan P. An analysis of consumer adoption rates in Internet banking. Eur J Manage 2010:10.
27. Kasteleyn MJ, Bonten TN, Taube C, Chavannes NH. Coordination of care for patients with COPD: clinical points of interest. Int J Care Coord 2015 Dec 08;18(4):67-71. [doi: 10.1177/2053434515620223]
28. In't Veen JCCM, Mennema TH, van Noort EMJ. Online self-management in COPD or asthma: with or without the healthcare provider? Eur Respir J 2012;40(Suppl 56):1284.
29. Flynn D, Gregory P, Makki H, Gabbay M. Expectations and experiences of eHealth in primary care: a qualitative practice-based investigation. Int J Med Inform 2009 Sep;78(9):588-604. [doi: 10.1016/j.ijmedinf.2009.03.008] [Medline: 19482542]
30. Hofstede J, de Bie J, van Wijngaarden B, Heijmans M. Knowledge, use and attitude toward eHealth among patients with chronic lung diseases. Int J Med Inform 2014 Dec;83(12):967-974. [doi: 10.1016/j.ijmedinf.2014.08.011] [Medline: 25269992]
31. Nijland N, Gemert-van Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
32. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJ, Numans ME, et al. High level of integration in integrated disease management leads to higher usage in the e-Vita study: self-management of chronic obstructive pulmonary disease with Web-based platforms in a parallel cohort design. J Med Internet Res 2017 May 31;19(5):e185 [FREE Full text] [doi: 10.2196/jmir.7037] [Medline: 28566268]
33. Talboom-Kamp EPWA, Verdijk NA, Blom CMG, Harmans LM, Talboom IJSH, Numans ME, et al. e-Vita: design of an innovative approach to COPD disease management in primary care through eHealth application. BMC Pulm Med 2016 Aug 16;16(1):121 [FREE Full text] [doi: 10.1186/s12890-016-0282-5] [Medline: 27530775]
34. Cochrane. EPOC resources for review authors: interrupted time series (ITS) analyses. URL: http://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/interrupted_time_series_analyses.docx [accessed 2017-06-22] [WebCite Cache ID 6rPNQpZDY]
35. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ 2015 Jun 09;350(jun09 5):h2750. [doi: 10.1136/bmj.h2750]
36. Smeele I, van Weel C, van Schayck C. Dutch College of General Practitioners guideline for COPD diagnosis. Huisarts Wet 2007;50(8):362-379.
37. VitalHealth Software Corp. Foundation Care Within Reach implements the interactive e-Vita care platform. 2017. URL: http://www.vitalhealthsoftware.com/news/2012/10/29/foundation-care-within-reach-implements-the-interactive-e-vita-care-platform [accessed 2017-06-22] [WebCite Cache ID 6rPNZLkxj]
38. van der Molen T, Willemse BWM, Schokker S, ten Hacken NHT, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. Health Qual Life Outcomes 2003 Apr 28;1:13 [FREE Full text] [Medline: 12773199]
39. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax 1999 Jul;54(7):581-586 [FREE Full text] [Medline: 10377201]
40. Scholz U, Gutiérrez Doña B, Sud S, Schwarzer R. Is general self-efficacy a universal construct? Eur J Psychol Assess 2002 Sep 1;18(3):242-251. [doi: 10.1027//1015-5759.18.3.242]

41. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [Medline: 11491192]
42. Chavannes NH, Grijsen M, van den Akker M, Schepers H, Nijdam M, Tiep B, et al. Integrated disease management improves one-year quality of life in primary care COPD patients: a controlled clinical trial. Prim Care Respir J 2009 Sep;18(3):171-176 [FREE Full text] [doi: 10.3132/pcrj.2009.00003] [Medline: 19142557]
43. Kruis AL, van Adrichem J, Erkelens MR, Scheepers H, Muris JW, Chavannes NH. Sustained effects of integrated COPD management on health status and exercise capacity in primary care patients. Int J Chron Obstruct Pulmon Dis 2010 Nov 25;5:407-413 [FREE Full text] [doi: 10.2147/COPD.S9654] [Medline: 21191435]
44. Kruis AL, Boland MRS, Assendelft WJJ, Gussekloo J, Tsiachristas A, Stijnen T, et al. Effectiveness of integrated disease management for primary care chronic obstructive pulmonary disease patients: results of cluster randomised trial. BMJ 2014 Sep 10;349:g5392. [doi: 10.1136/bmj.g5392]
45. Donaldson GC, Wedzicha JA. The causes and consequences of seasonal variation in COPD exacerbations. Int J Chron Obstruct Pulmon Dis 2014 Oct;9:1101-1110 [FREE Full text] [doi: 10.2147/COPD.S54475] [Medline: 25336941]
46. Tsiligianni I, Kocks J, Tzanakis N, Siafakas N, van der Molen T. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. Prim Care Respir J 2011 Sep;20(3):257-268 [FREE Full text] [doi: 10.4104/pcrj.2011.00029] [Medline: 21472192]
47. Chavannes NH, Sont JK, van der Boog JM, Assendelft WJJ. e-Health bij chronische ziekten e-Health in chronic diseases. Ned Tijdschr Geneeskd 2012;156:A5345.
48. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. JAMA 2016 Sep 20;316(11):1161-1171. [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
49. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van Houten H, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Arch Intern Med 2012 May 28;172(10):773-779 [FREE Full text] [doi: 10.1001/archinternmed.2012.256] [Medline: 22507696]
50. Chavannes NH. [Effectiveness of lifestyle apps not yet proven]. Ned Tijdschr Geneeskd 2016;160:D936. [Medline: 27702408]
51. Alma H, de Jong C, Jelusic D, Wittmann M, Schuler M, Flokstra-de Blok B, et al. Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. NPJ Prim Care Respir Med 2016 Dec 01;26:16041 [FREE Full text] [doi: 10.1038/npjpcrm.2016.41] [Medline: 27597571]
52. Jonkman NH, Westland H, Trappenburg JC, Groenwold RH, Bischoff EW, Bourbeau J, et al. Do self-management interventions in COPD patients work and which patients benefit most? An individual patient data meta-analysis. Int J Chron Obstruct Pulmon Dis 2016 Aug;11:2063-2074 [FREE Full text] [doi: 10.2147/COPD.S107884] [Medline: 27621612]
53. Ahmed S, Ernst P, Bartlett SJ, Valois M, Zaihra T, Paré G, et al. The effectiveness of Web-based asthma self-management system, My Asthma Portal (MAP): a pilot randomized controlled trial. J Med Internet Res 2016 Dec 01;18(12):e313 [FREE Full text] [doi: 10.2196/jmir.5866] [Medline: 27908846]
54. Ford I, Norrie J. Pragmatic Trials. N Engl J Med 2016 Aug 04;375(5):454-463. [doi: 10.1056/NEJMra1510059] [Medline: 27518663]
55. Siminerio LM. The role of technology and the chronic care model. J Diabetes Sci Technol 2010 Mar;4(2):470-475 [FREE Full text] [Medline: 20307409]
56. 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;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
57. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011 Jul;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
58. Jonkman NH, Westland H, Trappenburg JCA, Groenwold RHH, Bischoff EWMA, Bourbeau J, et al. Characteristics of effective self-management interventions in patients with COPD: individual patient data meta-analysis. Eur Respir J 2016 Jul;48(1):55-68. [doi: 10.1183/13993003.01860-2015] [Medline: 27126694]
59. Kruis AL, Boland MRS, Schoonvelde CH, Assendelft WJJ, Rutten-van Molke MPMH, Gussekloo J, et al. RECODE: design and baseline results of a cluster randomized trial on cost-effectiveness of integrated COPD management in primary care. BMC Pulm Med 2013 Mar 23;13:17 [FREE Full text] [doi: 10.1186/1471-2466-13-17] [Medline: 23522095]
60. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
61. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. Lancet 2002 Mar 16;359(9310):966-970. [doi: 10.1016/S0140-6736(02)08029-7] [Medline: 11918933]
62. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. Qual Saf Health Care 2003 Feb;12(1):47-52 [FREE Full text] [Medline: 12571345]

Abbreviations

CCQ: Clinical COPD Questionnaire
COPD: chronic obstructive pulmonary disease
EQ-5D: EuroQol 5-dimension questionnaire
FEV: forced expiratory volume
FVC: forced vital capacity
GP: general practitioner
GSES: Generalized Self-Efficacy Scale
IDM: integrated disease management
IQR: interquartile range
ITS: interrupted time series
MRC: Modified Medical Research Council Dyspnea scale
RCT: randomized controlled trial
QoL: quality of life

Edited by G Eysenbach; submitted 22.06.17; peer-reviewed by A Lenferink; comments to author 15.07.17; revised version received 24.07.17; accepted 27.07.17; published 16.08.17.

Please cite as:
Talboom-Kamp EPWA, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJSH, Looijmans-van den Akker I, van Geloven N, Numans ME, Chavannes NH
The Effect of Integration of Self-Management Web Platforms on Health Status in Chronic Obstructive Pulmonary Disease Management in Primary Care (e-Vita Study): Interrupted Time Series Design
J Med Internet Res 2017;19(8):e291
URL: http://www.jmir.org/2017/8/e291/
doi:10.2196/jmir.8262
PMID:28814380

Original Paper

Sugarsquare, a Web-Based Patient Portal for Parents of a Child With Type 1 Diabetes: Multicenter Randomized Controlled Feasibility Trial

Emiel Boogerd[1], MSc; Nienke M Maas-Van Schaaijk[1,2], PhD; Theo C Sas[3], PhD, MD; Agnes Clement-de Boers[4], MSc, MD; Mischa Smallenbroek[5], BSc; Roos Nuboer[6], MSc, MD; Cees Noordam[2,7], PhD, MD; Chris M Verhaak[1], PhD

[1]Department of Medical Psychology, Radboud University Medical Center, Nijmegen, Netherlands
[2]Children's Diabetes Center Nijmegen, Nijmegen, Netherlands
[3]Department of Pediatrics, Albert Schweitzer Hospital, Dordrecht, Netherlands
[4]Department of Pediatrics, Juliana Childrens' Hospital, The Hague, Netherlands
[5]Department of Pediatrics, Medical Center Leeuwarden, Leeuwarden, Netherlands
[6]Department of Pediatrics, Meander Medical Center, Amersfoort, Netherlands
[7]Department of Pediatrics, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Emiel Boogerd, MSc
Department of Medical Psychology
Radboud University Medical Center
Geert Groote Plein Zuid 10
PO Box 9101
Nijmegen, 6500 BH
Netherlands
Phone: 31 0243655232
Fax: 31 0243610865
Email: emiel.boogerd@radboudumc.nl

Abstract

Background: Raising a child with type 1 diabetes (T1D) means combining the demands of the disease management with everyday parenting, which is associated with increased levels of distress. A Web-based patient portal, Sugarsquare, was developed to support parents, by providing online parent-professional communication, online peer support and online disease information.

Objective: The first aim of this study was to assess the feasibility of conducting a multicenter, randomized controlled trial in Dutch parents of a child with T1D. The second aim was to assess the feasibility of implementing Sugarsquare in clinical practice.

Methods: The parents of 105 children (N=105) with T1D below the age of 13 participated in a 6-month multicenter randomized controlled feasibility trial. They were randomly assigned to an experimental (n=54, usual care and Sugarsquare) or a control group (n=51, usual care). Attrition rates and user statistics were gathered to evaluate feasibility of the trial and implementation. To determine potential efficacy, the parenting stress index (PSI-SF) was assessed at baseline (T0) and after 6 months (T1).

Results: Of a potential population of parents of 445 children, 189 were willing to participate (enrollment refusal=57.5%, n=256), 142 filled in the baseline questionnaire (baseline attrition rate=25%, n=47), and 105 also filled in the questionnaire at T1 (post randomization attrition rate during follow-up=26%, n=32). As such, 24% of the potential population participated. Analysis in the experimental group (n=54) revealed a total of 32 (59%) unique users, divided into 12 (38%) frequent users, 9 (28%) incidental users, and 11 (34%) low-frequent users. Of the total of 44 professionals, 34 (77%) logged in, and 32 (73%) logged in repeatedly. Analysis of the user statistics in the experimental group further showed high practicability and integration in all users, moderate acceptability and demand in parents, and high acceptability and demand in health care professionals. Baseline parenting stress index scores were related to the parents' frequency of logging on (ρ=.282, *P*=.03) and page-views (ρ=.304, *P*=.01). No significant differences in change in parenting stress between experimental and control group were found ($F_{3,101}$=.49, *P*=.49).

Conclusions: The trial can be considered feasible, considering the average enrollment refusal rate, baseline attrition rate and postrandomization attrition rate, compared to other eHealth studies, although lower than hypothesized. Implementing Sugarsquare

in clinical practice was partly feasible, given moderate demand and acceptability in parent users and lack of potential efficacy. Parents who reported higher levels of parenting stress used Sugarsquare more often than other parents, although Sugarsquare did not reduce parenting stress. These results indicate that Web-based interventions are a suitable way of providing parents of children with T1D with additional support. Future studies should determine how Sugarsquare could reduce parenting stress, for instance by adding targeted interventions. Factors potentially contributing to successful implementation are suggested.

Trial Registration: Nederlands Trial Register Number: NTR3643; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3643 (Archived by WebCite at http://www.webcitation.org/6qihOVCi6)

(J Med Internet Res 2017;19(8):e287) doi:10.2196/jmir.6639

KEYWORDS

diabetes mellitus, type 1; parenting; health communication; peer group; telemedicine; Internet

Introduction

Background

Type 1 diabetes (T1D) is a chronic metabolic disorder with a complex daily treatment regime, requiring patients to carry out a variety of health-related self-care behaviors, such as monitoring blood glucose levels, administering insulin, adhering to a diet, and exercising. In case of young children, parents are responsible for ensuring that these disease management tasks are performed. Having to combine these complex self-management tasks with regular parenting tasks in everyday life can have a profound impact on parents [1-10], indicated by elevated levels of stress and depressive symptoms in parents of a child with T1D [3,7,9,11], especially in those with young children and with children with a more recent diagnosis [2-7,12,13]. Family and parental functioning are related to well-being, self-care skills, and glycemic control in children, which makes it important that diabetes teams are aware of the impact of the disease and its treatment on parents [1,6,14-19]. Studies show that parents need easy access to their diabetes care team [8,20,21], local peer support [22-26], and tailored information about the disease and its management provided by their own diabetes team [8,27-30]. This positively affects their quality of life [8,23,26] and helps them adequately cope with the disease.

New technologies such as the Internet can help diabetes teams in delivering these aspects [8,25,26,29,31-40]. Despite the great potential of the Internet and parents' positive attitude toward using Internet in care, there has been little research into the efficacy and feasibility of Internet interventions for the parents of chronically ill children, especially interventions that combine multiple aspects of care [38,39,41]. This is unfortunate, considering that chronically ill patients and their parents can benefit from using the Internet, because it facilitates the exchange of knowledge and information between patients and health care professionals.

There are several challenges, when it comes to implementing and testing an Internet intervention in a clinical research context. eHealth studies are specifically subject to low retention rates (evaluation dropout), which are often the result of study-specific factors and low adherence rates (nonintervention usage) that are mostly intervention specific. These rates can lead to a loss of participants and thus to lack of statistical power [34,42-46]. Achieving successful recruitment is particularly problematic when multiple practices are involved, as practices often differ at an organizational level and local recruiters often have limited resources for recruitment [47,48].

Randomized Controlled Trial

To gain knowledge about the feasibility of conducting a randomized controlled trial (RCT) and implementing an Internet intervention in usual care for parents of a child with T1D, we developed a Web-based patient portal, called Sugarsquare [40]. Sugarsquare was specifically developed according to parents' needs and preferences [8,31] and is hypothesized to enable diabetes teams to improve their accessibility, facilitate local peer support, and provide tailored information [31]. An explorative, multicenter study was conducted to answer the following research questions:

1. Is conducting an RCT concerning Sugarsquare feasible in a population of parents of a child with T1D in terms of:
   - potential participants: what is the number of eligible parents?
   - enrollment refusal rate: what is the proportion of parents who refuse participation?
   - baseline attrition rate: what is the proportion of parents who drop out before baseline?
   - follow-up attrition rate: what is the proportion of parents who drop out during the trial?
2. Is implementation of Sugarsquare in daily clinical practice feasible in a population of parents of a child with T1D in terms of:
   - practicability: are recipients able to use Sugarsquare?
   - acceptability: do recipients use Sugarsquare?
   - demand: do recipients continue to use Sugarsquare?
   - integration: is Sugarsquare consistent with international guidelines for pediatric diabetes care?
   - potential efficacy: is usage associated with change in parenting stress?

Methods

Design and Setting of the Study

The participants for this study were recruited from 7 medical centers in the Netherlands, with a potential of 445 parents, from May 2012 to January 2013. Eligible participants were the parents of a child with T1D (one parent per child) younger than 13 years of age, had access to the Internet at home, and were able to

comprehend the Dutch language. The children had to be treated in one of the participating centers during the entire course of the study. Participants were randomly assigned to one of two conditions: (1) an intervention condition and (2) a usual care control condition. Participants in the intervention condition had access to the intervention for 6 months in addition to care as usual. Participants in the control group received care as usual during that period. An extensive report of the offline recruitment of participants, the randomization and the procedure of the data collection is described in the Sugarsquare study protocol [31]. The study described in this study was part of a larger project [31], of which all procedures were approved by the Ethics Committees of Human Experimentation of the Radboud University Medical Center and the participating hospitals and are in accordance with the Declaration of Helsinki.

Intervention

The final version of Sugarsquare consists of a Web-based patient portal providing online parent-professional communication, peer support, and disease information. Sugarsquare was developed at parents' explicit request and is based on a previous comparable intervention for adolescents with T1D [8,31,40]. Seven focus group interviews with parents [8,31] and a questionnaire for health care professionals affiliated to the cooperating centers were used to tailor the intervention to the preferences of both parents and health care professionals. In a series of pilots, involving parents and professionals participated, the intervention was further fine-tuned and facilitators and barriers were identified. The test phase ended when bugs were repaired and both parents and professionals felt the intervention was ready for use. In accordance with parents' preferences, the intervention was organized locally, so that each center for diabetes care has its own secured portal, which is only accessible to health care professionals of that particular center and to the parents of the children treated at that clinic. Sugarsquare is accessible through the Internet and has the following two main sections.

Section I: General

The first section provides online peer support and disease information and is accessible to all users (parents and health care professionals). Peer support is facilitated through a chat application, a forum application, and a blog application. Disease information is provided by means of downloadable documents and Web links.

Section II: Personal

The second section is specific to individual patients and can only be accessed by the parents of that particular patient and their diabetes team. The section contains an overview of treatment goals and an application for easily accessible private contact between parents and health care professionals. This application is only used for nonurgent matters.

The intervention has been described in the study protocol [31]. In the final version of Sugarsquare, disease information is incorporated in Section I, instead of Section II as described in the study protocol. Sugarsquare is secured by means of a 2-factor authentication, requiring a username-password combination and a personalized SMS code in the login procedure. Health care professionals of the local diabetes teams were appointed as coordinators for the local recruitment of participants and the local implementation of Sugarsquare. Screenshots of Sugarsquare for parents are displayed in Multimedia Appendix 1 [49].

Care as Usual

All children received care as usual, according to International Guidelines for Pediatric Diabetes Care [18,50], provided by a multidisciplinary team of pediatric diabetologists, diabetes nurse practitioners, dietitians, and psychologists. Parents and children were invited to visit the outpatient center for consultations with the pediatric diabetologist and nurse practitioner 4 times a year. Dieticians and psychologists were available on request by parents, children, or physicians. The diabetes care team could be contacted during business hours by telephone and email. An emergency telephone number could be accessed outside office hours to guarantee continuous access to care. Children of participants in both conditions (experimental and control) received care as usual during the entire study period. As such, Sugarsquare was used in addition to care as usual. During the study period, the parents in the experimental group could contact the diabetes care team via the portal instead of by telephone or email in case of nonurgent matters. The telephone number for emergencies was maintained.

Measures

Feasibility of the RCT was assessed in terms of the number of potential participants, the proportion of parents who refused participation, and the attrition rates. Demographic data of all the participants who were included in the final analyses were gathered at baseline.

For assessment of feasibility of the intervention, expressed in terms of practicability, acceptability, and demand [40,51], individual user data of all participants in the experimental group, such as frequency of logins and number of messages posted on the forum, were logged digitally. For feasibility in terms of integration, we assessed whether Sugarsquare was of added value for working according to International ISPAD (International Society for Pediatric and Adolescent Diabetes) and or IDF (International Diabetes Federation) and ADA (American Diabetes Association) Guidelines for Diabetes Care [18,50], by checking 9 key-elements for diabetes care, derived from these guidelines. For feasibility in terms of potential efficacy, parenting stress was assessed by means of the Dutch version of the parenting stress index-short form (PSI-SF) [52] on T0, T1, and T2. The reliability and criterion validity of the Dutch PSI-SF are shown to be good [52]. The PSI-SF consists of 25 items answered on a 6-point Likert scale, ranging from "totally agree" to "totally disagree." An example of an item on the PSI-SF is "it is not always easy to accept my child the way he or she is." The sum score on the PSI-SF can be categorized into normal, subclinical, and clinical based on standardized cutoff scores described in the manual [52]. Parenting stress was assessed at the start of the study (T0=baseline), at 6 months after the start of the study (T1), and at 12 months after the start of the study (T2=follow-up). Also, at the end of the study we asked the local Sugarsquare coordinators, who were health care professionals and part of the local diabetes teams, to evaluate

the study and identify facilitators and limitations for the implementation.

Information about the child's glycemic control (HbA1c) and the number of hospital admissions (lasting over 24 h) for keto-acidosis or severe hypoglycemia were used to explore the potential efficacy of the portal. These data were taken from the child's medical files.

Questionnaires for demographics and parenting stress were administered by means of a Web-based, secured survey program, called Radquest, which generates a closed survey system. The registered participants received an email with a Web link to the survey, which was paired with a unique user id. All items had to be answered and participants were able to change the answers until the participant submitted the completed survey. The data generated from the survey were stored on a secured server.

Some participants preferred filling in a hardcopy questionnaire, which was sent to them by post. For an elaborate overview of all measures, see Table 1.

Analyses

Demographic data were analyzed descriptively, and differences at baseline between the experimental group and the control group were assessed using an analysis of variance (ANOVA).

For feasibility, user data were analyzed by means of descriptive statistics. To compare differences in change in parenting stress between the experimental group and the control group, an analysis of covariance (ANCOVA) was performed on T1 data, using T0 data as covariate and the condition (experimental vs control) as fixed factor. A sensitivity analysis was conducted by means of a multiple imputation analysis (based on HbA1c scores at T1) to account for missing data. To test robustness of the results, a conservative analysis based on a Last Observation Carried Forward (LOCF) imputation was performed. Associations between user data and parenting stress at baseline were explored using Spearman ρ for nonparametric correlation due to high skewness of user data and a univariate ANOVA. Data on T2 were regarded as follow-up and were not analyzed in this study.

Power Calculation

We calculated that the data of 180 parents would be needed for the final analysis in order to reach a medium effect size (d=0.50), with a Cronbach alpha of .05 (two-tailed test) and a beta of .10 [31]. On the basis of recent literature, a declination rate of 25% (n=80) and a dropout rate of 25% (n=60) was hypothesized [31,34]. As such, we would need to approach 320 parents in order to reach a minimum of 240 parents at the start of the study to have data for 180 participants in the final analysis [31].

Table 1. Variables used in the Sugarsquare study.

Outcome	Measures
Demographics	
	Age and gender of the child
	Age of onset and duration of diabetes
	Pen or pump treatment
	Age, gender, and educational level of the primary parent
	Social economic status of the parents
Feasibility of the trial	
Potential population	Total population of parents, N (%)
Enrollment refusal	Participants who consented (total population of parents), n (%)
Baseline attrition	Participants who completed T0 (participants enrolled), n (%)
Postrandomization attrition (during follow-up)	Participants who completed T1 (randomized participants), n (%)
Feasibility of intervention	
Practicability (can they use it?)	Inventory of difficulties logging in and downtime (inaccessibility)
Acceptability (do they use it?)	Percentage of users who logged in at least once and used all applications
Demand or adherence (do they continue to use it?)	Percentage of users who logged in repeatedly
Integration (does it fit with the treatment?)	Evaluation of international guidelines for diabetes care (ISPAD or IDF and ADA) when using Sugarsquare
Potential efficacy (is usage associated with change in parenting stress?)	Parenting stress index-short form (PSI-SF [44])
Exploration of change in medical parameters	
Medical parameters	HbA1c
	Hospitals admissions due to glycemic disruptions

Results

Feasibility of the Randomized Controlled Trial: Enrollment and Dropout

All the parents of children with T1D, who were treated in 1 of the 7 cooperating centers for pediatric diabetes care, were invited by mail to participate in the study. The total population consisted of the parents of 445 children. A total of 189 parents of 189 children were willing to participate. The remaining 256 potential participants refused participation (enrollment refusal rate=57.5%). Frequently mentioned reasons for not participating were a lack of time, no interested in additional care and having to temporarily increase the focus on diabetes. A number of 142 parents filled in the baseline questionnaire. As such, 47 parents (baseline attrition rate=25%) dropped out before filling out the first questionnaire. Mentioned reasons for dropping out were a loss of interest and a lack of time. Subsequently, 105 parents also filled in the questionnaire at T1, meaning that 32 (postrandomization attrition rate during follow-up=26%) participants dropped out during the course of the study. Participants dropped out due to losing interest, a lack of time or because they changed from treatment center. As such, 23.6% (n=105) of the *potential population* successfully participated in the study (see also Figure 1).

Demographics

The demographic statistics of the 105 participants are displayed in Table 2. A one-way, between-group ANOVA revealed no significant differences in any of the variables at baseline between the centers.

Figure 1. Flowchart of inclusion of participants.

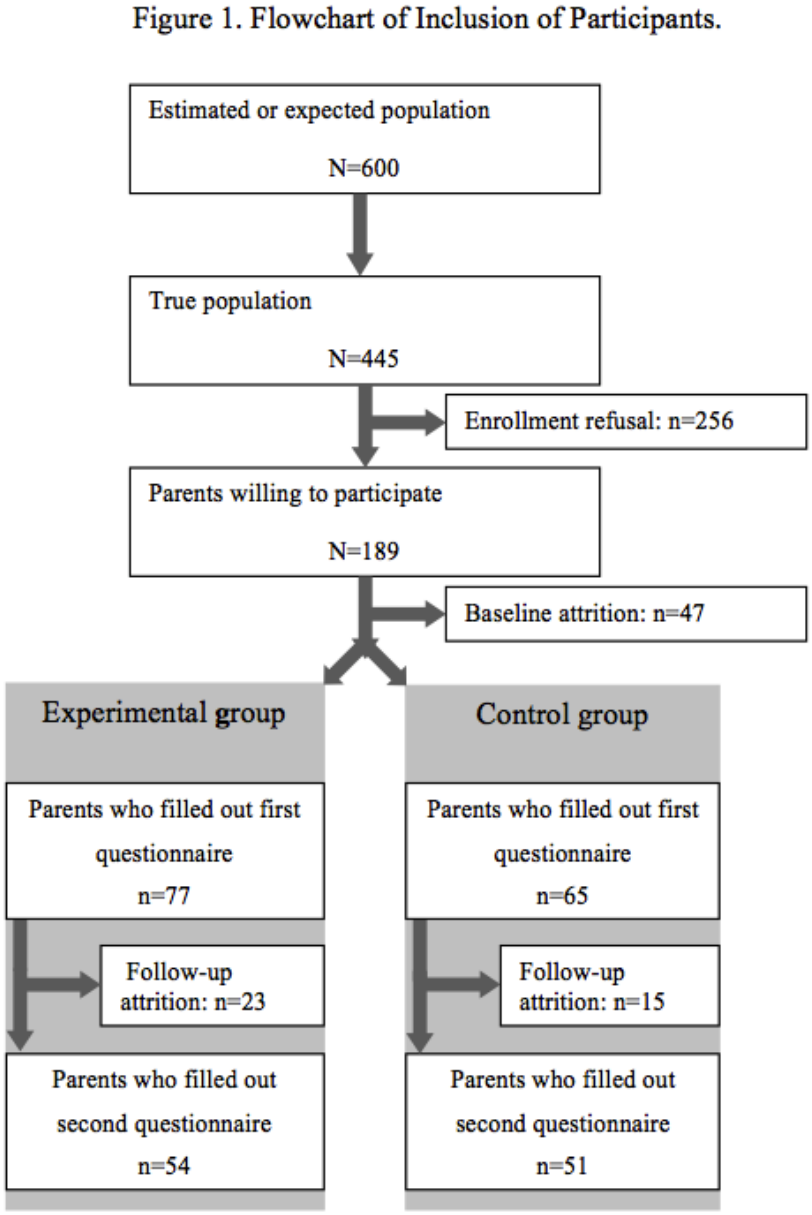


Table 2. Demographics and baseline scores of the participants.

Demographic variables	Experimental group	Control group	Total group
Parents (n)	54	51	105
Gender (male; female; filled in together)	49; 5	44; 5; 2	93; 10; 2
Educational level			
Lower secondary education, n (%)	2 (4)	4 (8)	6 (6)
Middle secondary education, n (%)	3 (5)	4 (8)	7 (7)
Higher secondary education, n (%)	24 (44)	19 (37)	43 (41)
Middle tertiary education, n (%)	9 (17)	2 (4)	11 (11)
Higher tertiary education, n (%)	9 (17)	19 (37)	28 (27)
Academia, n (%)	7 (13)	3 (6)	10 (10)
Child			
Age in years, mean (SD[a])	9,1 (2.9)	8,9 (2.5)	9 (2.7)
Gender (female; male)	30; 24	27; 24	57; 48
HbA1c in mmol/mol, mean (SD)	64 (13.77)	62 (7.77)	63 (10.62)
HbAc in %, mean (SD)	7,98 (1.17)	7,86 (0.71)	7,92 (0.97)
Insulin therapy			
Injections, n (%)	10 (19)	15 (29)	25 (24)
Pump, n (%)	44 (82)	36 (71)	80 (76)

[a]SD: standard deviation.

Feasibility of the Intervention

Data from the 54 participants in the experimental group and who therefore had access to Sugarsquare were used for the feasibility analysis and for the analysis relating user data and baseline scores on questionnaires. A proportion of 59% (n=32) of the parents who had access, used Sugarsquare during the trial (Table 3). Of the 32 unique parent users, 11 (34%) logged in repeatedly, at least once every 2 weeks and 9 (28%) logged in incidentally (3 times or more, but under once every 2 weeks), and 16 (41%) logged in once or twice during the study period. Table 3 also shows that 34 (77%) of 44 professionals who had received access at the start of the study, logged in and 32 (94%) logged in again. Thus, overall, 73% (n=32) of the professionals accessed Sugarsquare more than once. All users (parents and professionals) viewed all applications at least once when they logged in. The applications for forum (#page views=2838) and contact with the treatment team (#page views=2795) were viewed more often than the applications for information (#page views=415) and chat (#page views=683). Users reported no downtime, although 2 users reported that they sometimes could not access Sugarsquare, due to technical problems with the users' telecom providers. Some parents (n=8) said that the two-step security procedure as a hassle. Sugarsquare attributed to provision of care according to all 9 key elements, derived from the Global IDF or ISPAD and ADA Guidelines for Diabetes care in Childhood and Adolescence (see also Multimedia Appendix 2) [18,50]. According to the Sugarsquare coordinators, there were 3 factors that limited implementation. These factors were the two-step login procedure, the lack of customized instructions for health care professionals and the randomization on individual level. The local Sugarsquare coordinators and the multidisciplinary approach of the team were suggested as 2 factors that positively affected implementation.

Potential Efficacy

With regard to parenting stress, 82 (78%) parents (control and experimental condition) reported average or below average levels of parenting stress compared with Dutch healthy controls, 19 (18%) reported slightly elevated levels, and 4 (4%) reported very high levels of parenting stress (see also Table 4).

The analysis revealed no significant differences in change in parenting stress over time between the two groups ($F_{3,101}$=.49, P=.49), or between centers ($F_{3,101}$=.31, P=.91), and nor was there an interaction between groups and centers ($F_{3,101}$=1.16, P=.34). Similar results were obtained in an ANCOVA (Table 5) without the factor center and a sensitivity analysis, conducted by means of a multiple imputation analysis. Since no change was found, a conservative analysis using LOCF was not conducted. We also found no significant differences in change over time in HbA1c levels between the experimental group and the control group ($F_{3,101}$=.040, P=.84).

Baseline Parenting Stress Levels and Portal Usage

The analysis revealed that parenting stress at baseline was significantly correlation with the frequency of logging in (ρ=.282, P=.03 Table 6) and the number of pages viewed (ρ=.304, P=.02). It seems that the greater stress parents experienced, the more parents logged in and the more pages they viewed.

Table 3. Sugarsquare usage during the first phase (6 months) of the study period.

User statistics	Parents	Professionals	Parents and professionals
Parents			
n (experimental group)	54	44	
Unique visitors, n (%)	32 (59)	34 (77)	
Log-ins			
High frequent users, n (%)	12 (38)	12 (35)	
Moderate users, n (%)	9 (28)	20 (59)	
Low frequent users, n (%)	11 (34)	2 (6)	
#logins (n)	419	505	
#logins, mean (SD[a])	7,8 (13)	11,5 (16)	
Page views			
#page views (n)	5690	8006	
#mean page views, mean (SD)	105,4 (175)	182 (253)	
Information			
#Documents visits (n)			415
#Web links visits (n)			213
Patient-professional contact			
#Questions visits (n)			2795
#Questions input (n)			344
#Treatment visits (n)			674
#Treatment input (n)			29
Peer support			
#Forum visits (n)			2838
#Forum input (n)			147
#Chat visits (n)			683
#Chat input (n)			1653

[a]SD: standard deviation.

Table 4. Distribution of parenting stress index (PSI) scores for the total group.

PSI[a]-scores	n (%)
Normal stress scores	82 (78)
Elevated stress scores	19 (18)
High stress scores	4 (4)

[a]PSI: parenting stress index.

Table 5. Results of the analysis of covariance (ANCOVA) in parenting stress and HbA1c.

Efficacy variables	Experimental group		Control group		*F*
	T0	T1	T0	T1	
	Mean (SD[a])	Mean (SD)	Mean (SD)	Mean (SD)	
PSI[b]	48.13 (19.46)	51.35 (22.32)	44.61 (17.60)	44.45 (17.89)	.49
HbA1c	63.74 (12.77)	63.06 (8.98)	62.41 (7.77)	62.54 (8.64)	.04

[a]SD: standard deviation.

[b]PSI: parenting stress index.

Table 6. Correlations of parenting stress at baseline and frequency of log-ins and page views.

Efficacy variables	#log-ins	#page views
Parenting stress (baseline)	ρ=.282, *P*=.03	ρ=.304, *P*=.02

Discussion

Principal Findings

This study investigated the feasibility of conducting a trial and implementing an Internet intervention in a population of parents of children with T1D, in daily clinical practice. It revealed that eHealth has the potential to create a platform for shared, daily disease management between professionals and parents. Sugarsquare seems to attract parents with relatively high stress levels. The participation rate and dropout rate in the RCT were average, compared with other trial studies and results indicated that conducting a trial concerning Sugarsquare was feasible. The implementation of Sugarsquare in clinical practice was partly feasible, given the high practicability in all users, moderate acceptability and demand in parent users, high acceptability and demand in professional users, high level of integration and lack of potential efficacy.

It is interesting to note that parents reporting higher levels of parenting stress were more likely to use Sugarsquare compared with parents reporting lower levels. This is consistent with a recent study by Balkhi and colleagues [26], who reported that parents with higher stress levels more frequently visited diabetes-related online forums than did parents with lower stress levels. As no association between HbA1c and usage was found, it is assumed that general parenting stress is associated with usage and not stress related to medical condition of the child. However, it is quite possible that the parents who did not use Sugarsquare might do so if they have a temporary need for additional support or information, for instance if their child becomes ill, at onset of puberty or if they are planning a trip abroad.

Our enrollment refusal rate (57.5%) and baseline attrition rate (25%) fell within the ranges described in the review by Karlson and Rapoff (2009), who found the refusal rates in eHealth studies to be ranging from 0% to 75% (mean 37%) and baseline attrition rates ranging from 0% to 35% (mean 4%) [53]. From this perspective, the rates in this study are reasonable. Still, we expected a lower enrollment refusal rate, since the intervention was requested by parents and fitted to their preferences by means of focus group interviews. It could be that the questionnaires, which had to be filled in by the parents on several occasions, discouraged potential participants [54]. It is also possible that, due to the research context, parents perceived this study as an externally driven project, which conflicted with their preference for a center-driven intervention [8] and might have negatively influenced their willingness to cooperate [55]. Our study was further confronted with an average postrandomization attrition rate during follow-up (26% vs 0-54%, mean 20% in Karlson and Rapoff) [34,53]. The eHealth studies are subject to low enrollment and high dropout rates. In order to resolve the issue of low enrollment, Lernmark and colleagues [56] suggested that clarity should be provided about what participants are expected to invest and about the potential added value of the study results for the individual participant, their clinic or care in general. Baxter and colleagues [57] suggested that interaction between researchers and participants is vital for keeping participants committed after they decide to participate.

During the study, possibilities to improve the trial and implementation were identified. First of all, customized instructions for when and how to use Sugarsquare, would have helped them fit Sugarsquare into their daily workflow and encourage parents to use Sugarsquare [58-60]. Also, Sugarsquare was used in a research context and randomization took place on an individual level. As such, only a part of the population in each center participated in this study. This meant that health care professionals had to work using two procedures simultaneously, making their work very complex and intensive and complicating the integration of Sugarsquare in their workflow of everyday [61,62]. The research context also had a negative effect on the amount of interaction on Sugarsquare, since only a relatively small population of parents had access to the platform. Implementation would have been more successful if randomization was conducted on center level, which would have meant that a center would have used Sugarsquare for its entire population or not at all.

Factors that might have contributed to the success of the trial and the implementation were also identified. The teams all appointed a team member dedicated to Sugarsquare, who coordinated local recruitment and implementation, and monitored Sugarsquare usage. This might have supported the teams in integrating the intervention in usual care, since studies in the past reported that this lead to increased awareness in the team for usage of innovative interventions [44,59,62]. Also, the multidisciplinary approach of the Diabetes teams in our study might have contributed to the implementation of Sugarsquare, since literature shows that members of multidisciplinary teams are used to working toward shared, organizational goals, which makes it easier to implement changes into their workflow [58,59].

Sugarsquare has a broad focus and consists of multiple, general, potentially feasible applications. These characteristics fit to the needs of the parents, as expressed in the focus groups [8]. However, because of this broad focus, it is difficult to establish which applications (information, peer contact, contact with staff) contributed to usage and to potential effect. As such, mechanisms of change could not be identified. Future studies could apply multiple study arms to adequately assess the value of single applications, which would increase the number of participants required. [63,64]. Another way of identifying potential working mechanisms and the value of single applications would be to collect qualitative data. This is expected to provide more insight into both and future researchers should consider collecting qualitative data in their study. In this study, we used a generic questionnaire to assess parenting stress, considering its broad use in pediatrics and the lack of a diabetes-specific one. Although generic parenting stress

measures can be helpful for assessing stressors and distress, they might not be sensitive to issues specific to the parents of children with an illness or specific disease-related issues and, as such, failed to properly assess potential change in those domains [65]. Future studies could consider using an instrument designed for parents of a child with T1D or, in case this is lacking, an instrument for parents of pediatric patients, such as the Pediatric Inventory for Parents (PIP) or the recently validated pediatric parenting stress index (PPSI). The direct effect of the small sample size in this study is expected to be limited, since the sensitivity analyses did not show different outcomes compared with the completer analysis. However, indirectly, the limited number of participants in the local centers may have decreased the interaction on the local Sugarsquares and, with that, generalizability of the results. Future studies can avoid this by using randomization on center level.

Sugarsquare can be considered as a promising tool for diabetes teams, virtually extending their diabetes center. It contributes to usual care, because it offers parents and professionals a secured, Web-based platform for parent-health care professional communication, moderated peer support, and tailored disease information. In addition, it especially attracts parents who experience higher parenting stress levels. Given the complications that arose when Sugarsquare was used together with conventional communication tools, it is recommended that Sugarsquare be used as the sole medium for regular communication between parents and diabetes team. Appointing a dedicated Sugarsquare manager and using adequate instructions for the involved professionals are also hypothesized to contribute to the integration of Sugarsquare in care as usual. In order to increase usage by parent users and to improve their acceptance of Sugarsquare in daily care, diabetes teams could continuously add new content to Sugarsquare. This is expected to keep Sugarsquare interesting and to invite parent users to post information as well. It is also important that all team members post information, which shows parent users that Sugarsquare is accepted by the whole team. This might lower the threshold for parent users to use and accept Sugarsquare. This has been found to be workable in 9 centers for diabetes care in the Netherlands, which have implemented Sugarsquare in usual care.

In a recent study on the implementation of an eHealth intervention regarding online assessment of quality of life, it was noticed that successful implementation is affected by many factors acting on different aspects of implementing an intervention [66]. In general, they distinguish between factors on the level of the existing IT-structures (eg, usability, compatibility), organization (eg, support, expectations of management for usage), and the intervention itself (eg, easy to use, technical problems). As attrition rates as well as limited implementation are general challenges in eHealth, future studies should pay more attention to these factors. Another issue in the field of eHealth is that the financial costs of maintenance of interventions have yet to be included in systems for health care costs. The main problem that arises from this issue is the high number of interventions that are not implemented after a trail.

When starting an intervention study, we advise researchers to start with a single center trial for exploration of feasibility and potential efficacy. When feasibility and potential efficacy are demonstrated, a multicenter implementation could be conducted, potentially combined with assessment of efficacy using a historic design.

Conclusions

This study concerned a generic intervention, based on parents' preferences and needs, serving different aims, especially regarding shared disease management between parents and professionals. Our next step is to further develop the potential of Sugarsquare to serve as a platform for provision of more mechanism-focused interventions, targeted to reduce parenting stress, for instance, by providing online information or online cognitive behavior therapy. More generally, eHealth has possibilities to support monitoring of physical and psychosocial well-being, facilitate peer contact, interaction between patients and health care professionals and exchange of data. Sugarsquare can serve as central portal through which these applications or interventions can be accessed.

Acknowledgments

This study was performed by a grant from the Dutch Innovation Funds of Health Insurers (grant number B11-064; file 2250). Clearance for performing this study was provided by the local medical ethical commission (Commission for Human Research, CMO) and the medical ethical commissions of all cooperating hospitals.

Conflicts of Interest

All authors declare no competing interests for this study. EB, KN, and CV were involved in the process of the development of the intervention. They were not the primary developers and ownership of the intervention lies not with the authors.

Multimedia Appendix 1

Screenshots of Sugarsquare for parents.

[PDF File (Adobe PDF File), 435KB - jmir_v19i8e287_app1.pdf]

Multimedia Appendix 2

Check of the International Guideline and Standards for Diabetes Care.

[PDF File (Adobe PDF File), 19KB - jmir_v19i8e287_app2.pdf]

Multimedia Appendix 3

CONSORT-EHEALTH checklist (v1.6.1).

[PDF File (Adobe PDF File), 540KB - jmir_v19i8e287_app3.pdf]

References

1. Helgeson VS, Becker D, Escobar O, Siminerio L. Families with children with diabetes: implications of parent stress for parent and child health. J Pediatr Psychol 2012 May;37(4):467-478 [FREE Full text] [doi: 10.1093/jpepsy/jsr110] [Medline: 22267104]
2. Hilliard ME, Monaghan M, Cogen FR, Streisand R. Parent stress and child behaviour among young children with type 1 diabetes. Child Care Health Dev 2011 Mar;37(2):224-232. [doi: 10.1111/j.1365-2214.2010.01162.x] [Medline: 21083686]
3. Hullmann SE, Wolfe-Christensen C, Ryan JL, Fedele DA, Rambo PL, Chaney JM, et al. Parental overprotection, perceived child vulnerability, and parenting stress: a cross-illness comparison. J Clin Psychol Med Settings 2010 Dec;17(4):357-365. [doi: 10.1007/s10880-010-9213-4] [Medline: 21086027]
4. Mitchell SJ, Hilliard ME, Mednick L, Henderson C, Cogen FR, Streisand R. Stress among fathers of young children with type 1 diabetes. Fam Syst Health 2009 Dec;27(4):314-324 [FREE Full text] [doi: 10.1037/a0018191] [Medline: 20047355]
5. Powers SW, Byars KC, Mitchell MJ, Patton SR, Standiford DA, Dolan LM. Parent report of mealtime behavior and parenting stress in young children with type 1 diabetes and in healthy control subjects. Diabetes Care 2002 Feb;25(2):313-318. [Medline: 11815502]
6. Streisand R, Mackey ER, Elliot BM, Mednick L, Slaughter IM, Turek J, et al. Parental anxiety and depression associated with caring for a child newly diagnosed with type 1 diabetes: opportunities for education and counseling. Patient Educ Couns 2008 Nov;73(2):333-338. [doi: 10.1016/j.pec.2008.06.014] [Medline: 18692342]
7. Whittemore R, Jaser S, Chao A, Jang M, Grey M. Psychological experience of parents of children with type 1 diabetes: a systematic mixed-studies review. Diabetes Educ 2012;38(4):562-579 [FREE Full text] [doi: 10.1177/0145721712445216] [Medline: 22581804]
8. Boogerd EA, Maas-van Schaaijk NM, Noordam C, Marks HJ, Verhaak CM. Parents' experiences, needs, and preferences in pediatric diabetes care: suggestions for improvement of care and the possible role of the Internet. A qualitative study. J Spec Pediatr Nurs 2015 Jul;20(3):218-229. [doi: 10.1111/jspn.12118] [Medline: 26076888]
9. Haugstvedt A, Wentzel-Larsen T, Rokne B, Graue M. Perceived family burden and emotional distress: similarities and differences between mothers and fathers of children with type 1 diabetes in a population-based study. Pediatr Diabetes 2011 Mar;12(2):107-114. [doi: 10.1111/j.1399-5448.2010.00661.x] [Medline: 20522171]
10. Nurmi MA, Stieber-Roger K. Parenting children living with type 1 diabetes: a qualitative study. Diabetes Educ 2012;38(4):530-536. [doi: 10.1177/0145721712446636] [Medline: 22609760]
11. Patton SR, Dolan LM, Smith LB, Thomas IH, Powers SW. Pediatric parenting stress and its relation to depressive symptoms and fear of hypoglycemia in parents of young children with type 1 diabetes mellitus. J Clin Psychol Med Settings 2011 Dec;18(4):345-352 [FREE Full text] [doi: 10.1007/s10880-011-9256-1] [Medline: 21643962]
12. Lowes L, Gregory JW, Lyne P. Newly diagnosed childhood diabetes: a psychosocial transition for parents? J Adv Nurs 2005 May;50(3):253-261. [doi: 10.1111/j.1365-2648.2005.03388.x] [Medline: 15811104]
13. Cousino MK, Hazen RA. Parenting stress among caregivers of children with chronic illness: a systematic review. J Pediatr Psychol 2013 Sep;38(8):809-828 [FREE Full text] [doi: 10.1093/jpepsy/jst049] [Medline: 23843630]
14. Schwartz DD, Cline VD, Axelrad ME, Anderson BJ. Feasibility, acceptability, and predictive validity of a psychosocial screening program for children and youth newly diagnosed with type 1 diabetes. Diabetes Care 2011 Feb;34(2):326-331 [FREE Full text] [doi: 10.2337/dc10-1553] [Medline: 21216856]
15. Davis CL, Delamater AM, Shaw KH, La Greca AM, Eidson MS, Perez-Rodriguez JE, et al. Parenting styles, regimen adherence, and glycemic control in 4- to 10-year-old children with diabetes. J Pediatr Psychol 2001 Mar;26(2):123-129 [FREE Full text] [Medline: 11181888]
16. Tsiouli E, Alexopoulos EC, Stefanaki C, Darviri C, Chrousos GP. Effects of diabetes-related family stress on glycemic control in young patients with type 1 diabetes: systematic review. Can Fam Physician 2013 Feb;59(2):143-149 [FREE Full text] [Medline: 23418238]
17. Williams LB, Laffel LM, Hood KK. Diabetes-specific family conflict and psychological distress in paediatric type 1 diabetes. Diabet Med 2009 Sep;26(9):908-914 [FREE Full text] [doi: 10.1111/j.1464-5491.2009.02794.x] [Medline: 19719712]
18. International Diabetes Federation, International Society for Pediatric and Adolescent Diabetes. Global IDF/ISPAD guideline for Diabetes in Childhood and Adolescence. Brussels, Belgium: International Diabetes Federation; 2013.

19. Pihoker C, Forsander G, Wolfsdorf J, Klingensmith GJ. The delivery of ambulatory diabetes care to children and adolescents with diabetes. Pediatr Diabetes 2009 Sep;10 Suppl 12:58-70. [doi: 10.1111/j.1399-5448.2009.00585.x] [Medline: 19754619]
20. Ronksley PE, Sanmartin C, Campbell DJ, Weaver RG, Allan GM, McBrien KA, et al. Perceived barriers to primary care among western Canadians with chronic conditions. Health Rep 2014 Apr;25(4):3-10 [FREE Full text] [Medline: 24744042]
21. Beck JK, Logan KJ, Hamm RM, Sproat SM, Musser KM, Everhart PD, et al. Reimbursement for pediatric diabetes intensive case management: a model for chronic diseases? Pediatrics 2004 Jan;113(1 Pt 1):e47-e50. [Medline: 14702494]
22. Creedy DK, Ludlow T, Beggs J, Collis D, Irvine D, Price D, et al. Perceptions of psychosocial support groups by parents who have a child with diabetes: a needs analysis. Collegian 2005 Apr;12(2):28-32. [Medline: 16619910]
23. Sullivan-Bolyai S, Bova C, Leung K, Trudeau A, Lee M, Gruppuso P. Social support to empower parents (STEP): an intervention for parents of young children newly diagnosed with type 1 diabetes. Diabetes Educ 2010;36(1):88-97. [doi: 10.1177/0145721709352384] [Medline: 20016058]
24. Doherty FM, Calam R, Sanders MR. Positive parenting program (triple P) for families of adolescents with type 1 diabetes: a randomized controlled trial of self-directed teen triple P. J Pediatr Psychol 2013 Sep;38(8):846-858 [FREE Full text] [doi: 10.1093/jpepsy/jst046] [Medline: 23823086]
25. Merkel RM, Wright T. Parental self-efficacy and online support among parents of children diagnosed with type 1 diabetes mellitus. Pediatr Nurs 2012;38(6):303-8; quiz 309. [Medline: 23362628]
26. Balkhi AM, Reid AM, McNamara JP, Geffken GR. The diabetes online community: the importance of forum use in parents of children with type 1 diabetes. Pediatr Diabetes 2014 Sep;15(6):408-415. [doi: 10.1111/pedi.12110] [Medline: 24372986]
27. Hawthorne K, Bennert K, Lowes L, Channon S, Robling M, Gregory JW. The experiences of children and their parents in paediatric diabetes services should inform the development of communication skills for healthcare staff (the DEPICTED Study). Diabet Med 2011 Sep;28(9):1103-1108. [doi: 10.1111/j.1464-5491.2011.03292.x] [Medline: 21418092]
28. Howe CJ, Ayala J, Dumser S, Buzby M, Murphy K. Parental expectations in the care of their children and adolescents with diabetes. J Pediatr Nurs 2012 Apr;27(2):119-126. [doi: 10.1016/j.pedn.2010.10.006] [Medline: 22341190]
29. Holtslander L, Kornder N, Letourneau N, Turner H, Paterson B. Finding straight answers: identifying the needs of parents and service providers of adolescents with type 1 diabetes to aid in the creation of an online support intervention. J Clin Nurs 2012 Sep;21(17-18):2419-2428. [doi: 10.1111/j.1365-2702.2012.04182.x] [Medline: 22889443]
30. Swift Peter G F. Diabetes education in children and adolescents. Pediatr Diabetes 2009 Sep;10(Suppl 12):51-57. [doi: 10.1111/j.1399-5448.2009.00570.x] [Medline: 19754618]
31. Boogerd EA, Noordam C, Verhaak CM. The Sugarsquare study: protocol of a multicenter randomized controlled trial concerning a web-based patient portal for parents of a child with type 1 diabetes. BMC Pediatr 2014;14:24. [doi: 10.1186/1471-2431-14-24] [Medline: 24472527]
32. Nordfeldt S, Ängarne-Lindberg T, Nordwall M, Krevers B. Parents of adolescents with type 1 diabetes--their views on information and communication needs and internet use. A qualitative study. PLoS One 2013;8(4):e62096 [FREE Full text] [doi: 10.1371/journal.pone.0062096] [Medline: 23626772]
33. Scharer K. Internet social support for parents: the state of science. J Child Adolesc Psychiatr Nurs 2005;18(1):26-35. [doi: 10.1111/j.1744-6171.2005.00007.x] [Medline: 15701096]
34. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
35. Shaw RJ, Ferranti J. Patient-provider internet portals--patient outcomes and use. Comput Inform Nurs 2011 Dec;29(12):714-8; quiz 719. [doi: 10.1097/NCN.0b013e318224b597] [Medline: 21697705]
36. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
37. 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;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
38. Grey M, Whittemore R, Jeon S, Murphy K, Faulkner MS, Delamater A, et al. Internet psycho-education programs improve outcomes in youth with type 1 diabetes. Diabetes Care 2013 Sep;36(9):2475-2482 [FREE Full text] [doi: 10.2337/dc12-2199] [Medline: 23579179]
39. Harris MA, Hood KK, Mulvaney SA. Pumpers, skypers, surfers and texters: technology to improve the management of diabetes in teenagers. Diabetes Obes Metab 2012 Nov;14(11):967-972. [doi: 10.1111/j.1463-1326.2012.01599.x] [Medline: 22443170]
40. Boogerd EA, Noordam C, Kremer JA, Prins JB, Verhaak CM. Teaming up: feasibility of an online treatment environment for adolescents with type 1 diabetes. Pediatr Diabetes 2014 Aug;15(5):394-402. [doi: 10.1111/pedi.12103] [Medline: 24350732]
41. Nicholas DB, Gutwin C, Paterson B. Examining preferences for website support to parents of adolescents with diabetes. Soc Work Health Care 2013;52(9):862-879. [doi: 10.1080/00981389.2013.827144] [Medline: 24117033]
42. 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]

43. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
44. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
45. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
46. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
47. Dean SC, Harper CE, Cappuccio FP, Rink E, Dirckx C, Arnout J, European Collaborative Group of the IMMIDIET Project. The challenges of cross-national research in primary health care across Europe. Fam Pract 2005 Jun;22(3):341-346 [FREE Full text] [doi: 10.1093/fampra/cmi008] [Medline: 15778236]
48. Mastellos N, Bliźniuk G, Czopnik D, McGilchrist M, Misiaszek A, Bródka P, et al. Feasibility and acceptability of TRANSFoRm to improve clinical trial recruitment in primary care. Fam Pract 2016 Apr;33(2):186-191. [doi: 10.1093/fampra/cmv102] [Medline: 26711958]
49. Suikerplein. 2016. URL: https://www.suikerplein.nl/ [accessed 2016-11-25] [WebCite Cache ID 6mHcYJYQf]
50. American Diabetes Association. Diabetesjournals. 2016. Standards of medical care in diabetes URL: http://care.diabetesjournals.org/content/diacare/suppl/2016/12/15/40.Supplement_1.DC1/DC_40_S1_final.pdf [accessed 2017-06-26] [WebCite Cache ID 6rVJxqGxr]
51. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
52. De BA, Vermulst A, Gerris J, Veerman J, Abidin R. NOSI-R, Nijmeegse Ouderlijke Stress Index. Handleiding (NOSI-R, the Nijmegen Parenting Stress Index manual). Lisse, the Netherlands: Harcourt; 2006.
53. Karlson CW, Rapoff MA. Attrition in randomized controlled trials for pediatric chronic conditions. J Pediatr Psychol 2009 Aug;34(7):782-793 [FREE Full text] [doi: 10.1093/jpepsy/jsn122] [Medline: 19064607]
54. McCambridge J, Kalaitzaki E, White IR, Khadjesari Z, Murray E, Linke S, et al. Impact of length or relevance of questionnaires on attrition in online trials: randomized controlled trial. J Med Internet Res 2011;13(4):e96 [FREE Full text] [doi: 10.2196/jmir.1733] [Medline: 22100793]
55. Franciscus M, Nucci A, Bradley B, Suomalainen H, Greenberg E, Laforte D, et al. Recruitment and retention of participants for an international type 1 diabetes prevention trial: a coordinators' perspective. Clin Trials 2014 Apr;11(2):150-158 [FREE Full text] [doi: 10.1177/1740774513510070] [Medline: 24216218]
56. Lernmark B, Johnson SB, Vehik K, Smith L, Ballard L, Baxter J, et al. Enrollment experiences in a pediatric longitudinal observational study: the Environmental Determinants of Diabetes in the Young (TEDDY) study. Contemp Clin Trials 2011 Jul;32(4):517-523 [FREE Full text] [doi: 10.1016/j.cct.2011.03.009] [Medline: 21419878]
57. Baxter J, Vehik K, Johnson SB, Lernmark B, Roth R, Simell T, TEDDY Study Group. Differences in recruitment and early retention among ethnic minority participants in a large pediatric cohort: the TEDDY Study. Contemp Clin Trials 2012 Jul;33(4):633-640 [FREE Full text] [doi: 10.1016/j.cct.2012.03.009] [Medline: 22484339]
58. Aarts JW, Faber MJ, Cohlen BJ, Van OA, Nelen WL, Kremer JA. Lessons learned from the implementation of an online infertility community into an IVF clinic's daily practice. Hum Fertil (Camb) 2015;18(4):238-247. [doi: 10.3109/14647273.2015.1057901] [Medline: 26167659]
59. Atkins S, Lewin S, Ringsberg KC, Thorson A. Provider experiences of the implementation of a new tuberculosis treatment programme: a qualitative study using the normalisation process model. BMC Health Serv Res 2011 Oct 17;11:275 [FREE Full text] [doi: 10.1186/1472-6963-11-275] [Medline: 22004533]
60. Nijland N, van GJ, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008 May 15;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
61. Sharma U, Reed J, Doyle C, Bell D. Challenges in evaluating telehealth through RCT-the problem of randomization. Stud Health Technol Inform 2012;180:323-327. [Medline: 22874205]
62. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: literature review and Delphi study. Int J Qual Health Care 2004 Apr;16(2):107-123. [doi: 10.1093/intqhc/mzh030] [Medline: 15051705]
63. Murray DM, Pennell M, Rhoda D, Hade EM, Paskett ED. Designing studies that would address the multilayered nature of health care. J Natl Cancer Inst Monogr 2010;2010(40):90-96 [FREE Full text] [doi: 10.1093/jncimonographs/lgq014] [Medline: 20386057]
64. West SG, Duan N, Pequegnat W, Gaist P, Des JD, Holtgrave D, et al. Alternatives to the randomized controlled trial. Am J Public Health 2008 Aug;98(8):1359-1366. [doi: 10.2105/AJPH.2007.124446] [Medline: 18556609]
65. Devine KA, Heckler CE, Katz ER, Fairclough DL, Phipps S, Sherman-Bien S, et al. Evaluation of the psychometric properties of the Pediatric Parenting Stress Inventory (PPSI). Health Psychol 2014 Feb;33(2):130-138 [FREE Full text] [doi: 10.1037/a0032306] [Medline: 23544994]

66. Schepers SA, Sint NS, Haverman L, Wensing M, Schouten van MA, Veening MA, et al. Real-world implementation of electronic patient-reported outcomes in outpatient pediatric cancer care. Psychooncology 2017 Jul;26(7):951-959. [doi: 10.1002/pon.4242] [Medline: 27502744]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance
LOCF: last observation carried factor
PSI-SF: parenting stress index-short form
T1D: type 1 diabetes

Edited by CL Parra-Calderón; submitted 10.09.16; peer-reviewed by A Ranchhod, T Kool; comments to author 15.10.16; revised version received 25.11.16; accepted 18.12.16; published 22.08.17.

Please cite as:
Boogerd E, Maas-Van Schaaijk NM, Sas TC, Clement-de Boers A, Smallenbroek M, Nuboer R, Noordam C, Verhaak CM
Sugarsquare, a Web-Based Patient Portal for Parents of a Child With Type 1 Diabetes: Multicenter Randomized Controlled Feasibility Trial
J Med Internet Res 2017;19(8):e287
URL: http://www.jmir.org/2017/8/e287/
doi:10.2196/jmir.6639
PMID:28830853

Original Paper

A Web-Based and Print-Based Computer-Tailored Physical Activity Intervention for Prostate and Colorectal Cancer Survivors: A Comparison of User Characteristics and Intervention Use

Rianne Henrica Johanna Golsteijn[1], MSc; Catherine Bolman[1], PhD; Denise Astrid Peels[1], PhD; Esmee Volders[1], MSc; Hein de Vries[2], PhD; Lilian Lechner[1], PhD

[1]Department of Psychology and Educational Sciences, Open University of the Netherlands, Heerlen, Netherlands

[2]Department of Health Promotion, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Rianne Henrica Johanna Golsteijn, MSc
Department of Psychology and Educational Sciences
Open University of the Netherlands
PO Box 2960
Heerlen, 6401 DL
Netherlands
Phone: 31 455762539
Email: rianne.golsteijn@ou.nl

Abstract

Background: Physical activity (PA) is beneficial in improving negative physical and psychological effects of cancer. The rapidly increasing number of cancer survivors, resulting from aging and improved cancer care, emphasizes the importance to develop and provide low cost, easy accessible PA programs. Such programs could be provided through the Internet, but that could result in the exclusion of cancer survivors not familiar with the Internet. Therefore, we developed a computer-tailored PA intervention for prostate and colorectal cancer survivors in which both Web-based and print materials are provided, and participants can choose their own preferred delivery mode.

Objective: The aim of this study was to assess participants' characteristics related to delivery mode and use of intervention materials.

Methods: We studied characteristics of participants using Web-based and printed intervention materials in a randomized controlled trial (RCT). Prostate and colorectal cancer survivors recruited from hospitals were randomized to OncoActive (computer-tailored PA intervention) or a usual-care control group. OncoActive participants received both Web-based and printed materials. Participants were classified into initial print- or Web-based participants based on their preferred mode of completion of the first questionnaire, which was needed for the computer-tailored PA advice. Intervention material use during the remainder of the intervention was compared for initial print- or Web-based participants. Additionally, participants were classified into those using only print materials and those using Web-based materials. Differences in participant characteristics and intervention material use were studied through analysis of variance (ANOVAs), chi-square tests, and logistic regressions.

Results: The majority of the participants in the intervention group were classified as initial Web-based participants (170/249, 68.3%), and 84.9% (191/249) used Web-based intervention materials. Dropout was low (15/249, 6.0%) and differed between initial Web-based (4/170, 2.4%) and print-based (11/79, 14%) participants. Participants were less likely to start Web-based with higher age (odds ratio [OR]=0.93), longer time since last treatment (OR=0.87), and higher fatigue (OR=0.96), and more likely with higher education (OR=4.08) and having completed treatments (OR=5.58). Those who were older (OR=0.93) and post treatment for a longer time (OR=0.86) were less likely to use Web-based intervention materials. Initial print-based participants predominantly used print-based materials, whereas initial Web-based participants used both print- and Web-based materials.

Conclusions: To our knowledge, this is one of the first studies that assessed participant characteristics related to delivery mode in an intervention in which participants had a free choice of delivery modes. Use of print-based materials among the initial Web-based participants was substantial, indicating the importance of print-based materials. According to our findings, it may be important to offer Web- and print-based materials alongside each other. Providing Web-based materials only may exclude older, less educated, more fatigued, or currently treated participants; these groups are especially more vulnerable and could benefit most from PA interventions.

(J Med Internet Res 2017;19(8):e298) doi:10.2196/jmir.7838

KEYWORDS

eHealth; web-based intervention; print-delivered intervention; computer tailoring; intervention usage; physical activity; prostate cancer; colorectal cancer; cancer survivorship

Introduction

Cancer and cancer treatment coincide with short- and long-term effects on both physical and mental health, eventually decreasing quality of life of cancer patients and survivors (CPS) [1-6]. A healthy lifestyle, and especially physical activity (PA), is known to be beneficial for cancer survivors in improving treatment-related side effects and thereby health-related quality of life (HRQoL) [7-10]. Additionally, PA is a preventive factor for the development of other chronic diseases and comorbidities for which cancer survivors are at risk (eg, obesity, coronary heart disease, and diabetes), as well as for secondary or new cancer or cancer recurrence [10-15]. Therefore, effective PA programs for CPS are of major importance, especially since studies regarding supportive care needs have shown that CPS themselves express a substantial need for healthy lifestyle information and programs including PA [16-18].

In light of the rapidly growing population living with or after cancer, because of advances in early detection and treatment [19,20], there is a clear need for easily accessible and affordable programs aimed at self-management. Web-based interventions may be a cost-effective method since they have a large potential reach for low cost and have proven to be effective in increasing PA in both healthy and diseased populations [21-24]. A frequently used and proven effective method for Web-based interventions is computer-tailoring [23-26] where participants receive personalized feedback generated automatically using computer-based data-driven decision rules and data collected from questionnaires (eg, individual characteristics, beliefs, and behavior) [27].

With rapid increases in Internet access in recent years, preconditions for the use of Web-based interventions have improved substantially. In 2016, 94% of the Dutch population had Internet access and electronic health (eHealth) applications were increasingly used, especially by adults aged over 65 years and adults with a chronic disease [28,29]. Therefore, with a median age of 65 years at diagnosis [30], the use of eHealth for CPS seems promising. However, Internet access decreases substantially from the age of 75 years (60% compared with 90% among those aged 65-75 years in 2016), and frequency of Internet use is also substantially lower with increasing age [29]. eHealth literacy, that is, the ability to seek, find, understand, and appraise health information from electronic resources and apply that knowledge to solving a health problem or making a health-related decision [31], is important for eHealth interventions to be successful. Studies showed that older age and lower socioeconomic status (SES) are related to lower eHealth literacy [32] and that older adults may lack skills and knowledge for the use of eHealth interventions [33]. Interventions that are only provided through the Internet may therefore be less useful in a population of CPS (who are generally older aged) and may even exclude the elderly or those of lower SES from its benefits.

Alternatively, computer-tailored interventions can be delivered both through the Internet and in print. A Web-based version and a print-based version were offered alongside each other in the OncoActive intervention, a computer-tailored PA program to stimulate and maintain PA in prostate and colorectal CPS. As a result, CPS could choose their preferred delivery mode: every participant received log-in details for the OncoActive website to fill out the assessment questionnaire, as well as an additional (identical) paper-and-pencil version. After completion of the questionnaire of their own choice, participants received their tailored advice both Web-based and by normal mail, enabling them to use either one or both. Providing the ability to use the preferred method for accessing intervention materials can increase intervention reach and adherence and may eventually result in larger behavior change effects in the target population. Therefore, it is important to determine which participant characteristics (eg, demographics, disease related-factors, and health-related factors) are associated with the preference for a certain delivery mode and with the use of intervention materials. As providing the printed delivery mode alongside the Web-based intervention is associated with higher costs, it is also important to gain insight into the actual use of these materials.

Research relating participant characteristics to delivery mode preference is scarce. To our knowledge, there is only one study in which participants from a general adult population could freely choose between print-based and Web-based intervention materials. Factors associated with choosing printed materials were being older, less educated, and of poorer health status [34]. Another study examining participant characteristics of adults aged over 50 years cluster-randomized to either a print- or Web-based PA intervention found that there was a higher percentage of males in the Web-based intervention and that participants in the Web-based intervention were younger, had a higher body mass index (BMI), and a lower intention to be physically active [35]. A study from Short et al [36] regarding PA intervention preferences of the general adult population (comparing face-to-face, group-, print-, and Web-based delivery mode) revealed that factors positively associated with preference for a Web-based intervention were being middle aged, living in a rural area, and high Internet use. Web-based preference was negatively associated with female gender, obesity, and high PA participation. Preference for a print-based intervention was positively associated with older age and negatively associated with female gender and obesity [36]. A positive attitude toward eHealth interventions in a population of cancer survivors was associated with lower age, higher income, higher quality of life, having completed cancer treatment, and having prostate cancer [16].

The aim of this study was to provide insight into the characteristics of participants who initially chose to participate Web-based versus those who initially chose to participate in the print-delivered intervention. As participants could use both Web-based and printed materials or a combination after the initial choice, we also examined intervention material use and participant characteristics related to this. On the basis of findings in previous studies, we expected that age and education would be important predictors of initial Web-based participation and using Web-based intervention materials. Analyses with regard to PA and disease-related factors were exploratory.

Information regarding participant characteristics related to the initial choice for a delivery mode and the delivery mode and material use during the complete intervention would aid further implementation, as it could provide insight into the feasibility of using Web-based interventions in a population of CPS, which often is elderly. This information could also help future researchers to choose the appropriate delivery mode for their audience and provide insight in which persons may be hard to reach when providing only a Web-based intervention.

Methods

Study Design

This study is part of a randomized controlled trial (RCT) in which participants were randomized to either the OncoActive intervention group or a usual-care waiting-list control group to assess the effectiveness. Since this study only examines the intervention delivery mode, control group participants were excluded from the analyses. The RCT was approved by the Medical Ethics Committee of the Zuyderland hospital (NL47678.096.14) and is registered in the Dutch Trial Register (NTR4296). All participants provided written informed consent.

Participants

CPS (≥18 years) diagnosed with colorectal or prostate cancer could participate in the trial if they were undergoing treatment with a curative intent or if they successfully completed primary treatment (surgery, chemotherapy, or radiation) up to 1 year ago. There was no restriction for patients currently undergoing hormonal therapy. By selecting only two cancer types, we could better fine-tune the intervention to the specific needs and capabilities in relation to cancer type. Prostate cancer and colorectal cancer were selected because they are among the most common cancer types in the Netherlands. Furthermore, survival rates are good, indicating a large population possibly benefiting from a PA intervention [30,37].

Participants should have had surgery at least 6 weeks before the start of the study. Those suffering from severe medical, psychiatric, or cognitive illnesses (eg, Alzheimer disease and mobility limitations) that could interfere with participation in a PA program were not invited to participate. Proficient Dutch reading and speaking skills were required for completing questionnaires and reading the tailored advice. Lack of Internet access and Internet skills were not a reason for exclusion.

Procedure

Prostate and colorectal CPS were recruited from the urology or oncology departments of 17 hospitals in 2015 and 2016. Eligible CPS were identified by hospital staff, verbally informed (either in person or by telephone) about the study, and invited to participate. Written information was handed over or sent by mail if the patient agreed to receive an information package. Additionally, CPS were recruited via other channels (eg, calls in local newspapers, on relevant websites, discussion groups, and flyers in hospitals). Participants responding to these messages were informed by the researchers and were also sent an information package by mail.

The information package included a letter with information, a time schedule, an informed consent form, and a prepaid return envelope. If there was no response to the initial information package, 3 weeks later one postal reminder was sent. CPS who agreed to participate were randomized into either the intervention group or the control group. Subsequently, all participants wore an accelerometer (ActiGraph GT3X-BT, ActiGraph, Pensacola, FL) to objectively assess PA. Immediately after wearing the accelerometer for 7 days, every participant received an email with log-in details for the OncoActive website together with an invitation to fill out the Web-based questionnaire and an identical paper-and-pencil version of the questionnaire by normal mail, enabling them to fill out the version of their preference. After completing this baseline questionnaire (T0), the intervention group received the OncoActive intervention that is outlined below. Both groups had to fill out follow-up questionnaires at three time points: 3 (T1), 6 (T2), and 12 (T3) months after baseline. At each time point, participants could choose whether to fill out the Web-based questionnaire or the paper-based questionnaire. The T1 questionnaire was used to provide ipsative feedback in the form of tailored advice (see below). The questionnaires at 6 and 12 months were administered for efficacy and process evaluation purposes and were thus not considered part of the intervention. The T3 questionnaire is not part of this study.

The OncoActive Intervention

The OncoActive intervention is a computer-tailored intervention aimed at awareness, initiation, and maintenance of PA behavior in prostate and colorectal CPS. The intervention was based on a proven effective evidence-based intervention to stimulate PA in adults over 50 years [38,39] and adapted for prostate and colorectal CPS using the intervention mapping protocol [40].

Participants in the intervention group received tailored PA advice at three time points. The content of the first and second tailored advice was based on information gathered with the baseline questionnaire. Both the baseline (T0) and the second questionnaire (T1) provided input for the third tailored advice and allowed for the provision of ipsative feedback. The content of the advice is based on behavior change theories and targets premotivational constructs (eg, awareness and knowledge), motivational constructs (eg, self-efficacy, attitude, and intrinsic motivation), and postmotivational constructs (eg, goal setting, action and coping planning, and self-regulation) [40-42]. In addition to the tailored advice, every participant received a pedometer and access to interactive content on the website (eg,

role model videos, home exercise instruction videos, a module for goal setting using a pedometer, the option to consult a physical therapist, and additional information). A more detailed description of the intervention content can be found elsewhere [40], and some screenshots can be found in Multimedia Appendix 1.

As previously mentioned, every participant received the first questionnaire Web-based and on paper. After completion of the questionnaire of their own choice, participants received their tailored advice. If the questionnaire was completed on the website, advice was immediately available on the website, and participants were made aware that they would receive a printed version of their advice within 3 days. If participants completed the paper-and-pencil questionnaire, the advice text was available (Web-based and print-based) within 2 weeks after receiving the questionnaire, after uploading participant data by research staff. Participants were emailed (if they provided their email address) that their advice was available on the website and that they would receive a printed version of the advice within 3 days. The tailored text was exactly the same for both modalities, but the Web-based version contained more interactive content (eg, videos). All participants were made aware that they could find additional interactive content on the website. The tailored advice was displayed on a distinct section of the website.

For the second provision of advice (2 months after the start), participants received an email to notify them that their advice was available on the website and that they would receive a printed version within a few days. For the third provision of advice (within 2 weeks of completing the T1 questionnaire), participants again received 2 versions (Web and print) of a questionnaire, with a procedure similar to the first advice.

Measurements

Several demographic variables, cancer-related characteristics, PA behavior, PA determinants, and health-related outcomes were measured in the baseline questionnaire of the RCT [40]. For this study, we used the following demographic variables: age, gender, height, weight, highest educational level, and household income. Educational level was categorized into low (ie, primary, basic vocational, or lower general school), moderate (ie, medium vocational school, higher general secondary education, and preparatory academic education), or high (ie, higher vocational school or university level) according to the Dutch educational system. Height and weight were used to calculate BMI (ie, weight in kilograms divided by height in meters squared). Participants were classified as being overweight (BMI >24.9 kg/m^2) or not. Cancer-related characteristics included type of cancer, which was either prostate or colorectal in this study; treatment status; and date of their last treatment.

PA was measured in two ways. Self-reported PA was measured using the validated Short Questionnaire to Assess Health Enhancing Physical Activity (SQUASH) [43], assessing activities regarding commuting, household, occupation, and leisure time. Total minutes of PA were classified into light (metabolic equivalent [MET] <3.0), moderate (MET 3.0-5.9), and vigorous (MET >6) [44]. Minutes of moderate to vigorous PA (MVPA) were calculated by adding up total time in moderate and vigorous PA. The SQUASH questionnaire has shown to have reasonable reliability (ρ=.58) and validity against an accelerometer (ρ=.45) [43].

Additionally, objective PA was measured using the ActiGraph GT3X-BT. Participants wore the accelerometer on an elastic belt on their right hip for 7 days. Data were downloaded and analyzed using ActiLife software (ActiGraph, Pensacola, FL). Measurements were considered valid if there were at least 3 days with at least 10 hours of wear time [45-47]. Nonwear periods were excluded from the analyses and were identified according to Choi et al [48]: intervals of at least 90 consecutive min of zero counts with allowance of a maximum of 2 min of nonzero counts during a nonwear interval. MVPA was calculated using Freedson-VM cut-off points based on 60 s epochs [49].

Intention to be sufficiently physically active was assessed using a scale of three items (alpha=.91) on a 10-point scale (eg, “To what extent do you intend to be sufficiently physically active?”) [35,39]. The score of the three items was averaged, resulting in a total score ranging from 1 to 10, with a high score indicating a high intention to be physically active.

HRQoL was measured using the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30 (EORTC QLQ-C30) [50]. The questionnaire comprises several scales, with the global health status scale providing an overview of general quality of life. Global health status was measured with two items (alpha=.85) on a 7-point scale. Scores were converted to scores ranging from 0 to 100, with a high score indicating a high HRQoL.

Fatigue was assessed using the Checklist Individual Strength (CIS) [51]. The subjective fatigue subscale assesses the experience of fatigue of participants. The eight items (alpha=.89) of the subscale are scored on a scale from 1 to 7, resulting in a total score in the range of 8 to 56.

Intervention material use was assessed with two questions per advice specifically aimed at the tailored advice: “Did you read your advice on paper?” and “Did you read your advice on the website?”; participants could identify whether they read the advice “completely,” “partly,” or “not.”

Statistical Analysis

Dropout Analysis

Multiple logistic regression was performed to determine whether participants’ characteristics were predictors of dropout during the intervention (ie, at the 3-month follow-up questionnaire). Choice for the initial delivery mode was added as a variable to identify if one of the groups was more likely to drop out of the intervention. All predictors were forced into the model simultaneously (method Enter in Statistical Package for the Social Sciences [SPSS]).

Initial Choice Intervention Delivery Mode

For the analysis regarding the choice of the initial intervention delivery mode, we analyzed data from all participants who completed the baseline questionnaire. Classification into groups for the initial preferred intervention delivery mode was based on the way participants chose to complete the baseline

questionnaire. In the accompanying information letter, participants were informed that they would immediately receive their first PA advice on the website if they completed the baseline questionnaire (used for the tailored advice) through the Internet. Participants completing the first questionnaire on the website were therefore classified as "initial Web-based participants," and participants completing the first questionnaire on paper were classified as "initial print-based participants."

Descriptive statistics on demographic factors (ie, age, sex, educational level, and household income), cancer-related factors (ie, type of cancer, treatment phase, and time since last treatment), PA-related factors (ie, self-reported and objective PA behavior and intention to be physically active), and health-related factors (ie, BMI, HRQoL, and fatigue) were calculated for the complete intervention group and split for "initial Web-based participants" and "initial print-based participants."

Univariate one-way analysis of variance (ANOVA) and chi-square tests were used to determine significant differences between both groups. Multiple logistic regression (Enter method) was performed to determine differences in participant characteristics for initial intervention delivery mode choice.

Both educational level and household income are regarded as indicators of SES. We decided to include only educational level in the logistic regression as a previous study showed that compared with household income, education was more consistently predictive of eHealth use [52].

Linking Delivery Preference to Intervention Use

Use of the different tailored advice texts was assessed with self-report questions. Chi-square tests were performed to determine differences between the "initial Web-based participants" and the "initial print-based participants" with regard to the use of tailored advice. Additionally, differences regarding mode of completion of the second questionnaire (T1), which was part of the intervention, were assessed.

Continued Intervention Use Delivery Mode

On the basis of self-report regarding the use of the three sets of advice, participants were classified as "exclusively print-based participants," "participants who used both Web-based and print-based materials," and "exclusively Web-based participants." As there were only 2 participants classified as "exclusively Web-based participants," we chose to dichotomize this classification into "exclusively print-based participants" and "participants using Web-based materials."

Multiple logistic regression (Enter method) was performed to determine differences in participant characteristics between both groups.

All analyses were performed using SPSS version 22 (IBM Corp., Armonk, NY). To check for the influence of uneven groups in the multivariate logistic regression analyses, nonparametric bootstrapping with 5000 replications was applied.

Results

Dropout Analysis

Within the intervention group, 232 participants out of the 249 enrolled at baseline completed the second questionnaire and received their final advice. Two participants who did not complete the second questionnaire missed just one questionnaire, whereas 15 participants opted out of the study, resulting in a dropout rate of 6.0%. Although dropout was limited, logistic regression analyses revealed that initial print-based participants were more likely to drop out (odds ratio [OR] 4.32, 95% CI 1.15–16.25). Among the initial Web-based participants, the dropout rate was 2.4% (4/170), and among the initial print-based participants, the dropout rate was 14% (11/79).

Participant Characteristics and Initial Choice Intervention Delivery Mode

In total, 510 prostate and colorectal CPS provided informed consent and were randomized into the intervention or the control group. For this study, we only used the data from the intervention condition, as this study aims to identify individual predictors of intervention delivery mode. In total, 249 participants were randomized into the intervention condition (Figure 1). Baseline characteristics for the complete intervention group are shown in Table 1.

The majority of the participants in the intervention group (n=249) were classified as initial Web-based participants (170/249, 68.3%). Significant differences between the initial Web-based participants and the initial print-based participants were found. Initial Web-based participants were significantly younger (P<.001) and higher educated (P=.002). Furthermore, initial Web-based participants had a higher income (P=.003), were more often post cancer treatment (P=.046), and were less fatigued (P=.04) (see Table 1).

Figure 1. Flow diagram of the study. Note: the control condition was not included in this study.

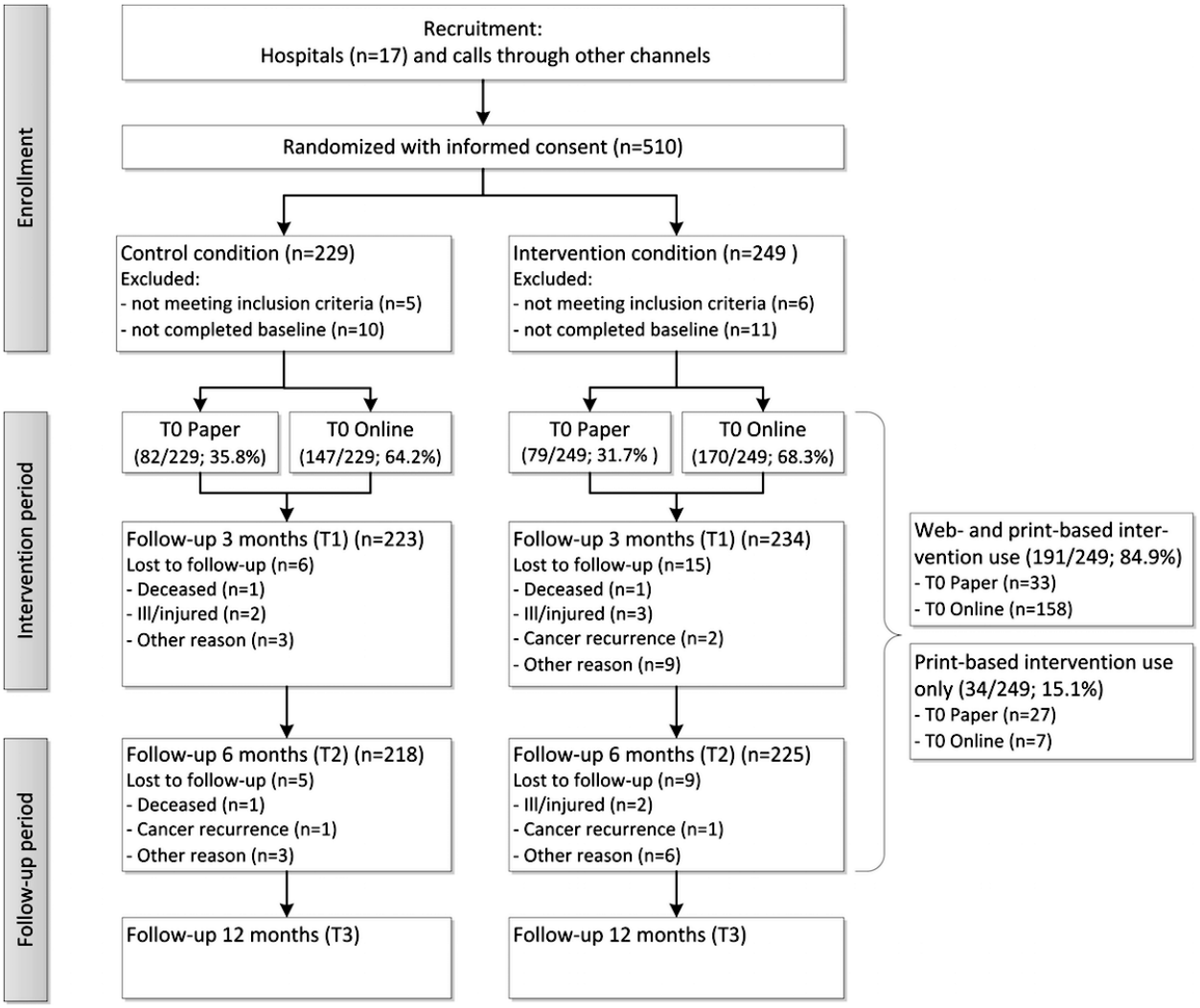


Table 1. Baseline participant characteristics of the total intervention group and split for the initial Web-based participants and the initial print-based participants.

Characteristics	Total intervention (n=249)	Initial Web-based participants (n=170)	Initial print-based participants (n=79)	*P* value
Demographic factors				
Age in years, mean (SD[a])	66.38 (8.22)	65.08 (7.84)	69.18 (8.37)	<.001
Gender, n (%)				.39
Male	212 (85.1)	147 (86.5)	65 (82)	
Female	37 (14.9)	23 (13.5)	14 (18)	
Education, n (%)				.004
Low	109 (44.0)	64 (37.6)	45 (58)	
Middle	70 (28.2)	49 (28.8)	21 (27)	
High	69 (27.8)	57 (33.5)	12 (15)	
Household income, n (%)				.005
Low	25 (12.8)	12 (8.1)	13 (23)	
Middle	79 (38.5)	56 (37.6)	23 (41)	
High	101 (49.3)	81 (54.4)	20 (36)	
Cancer-related factors				
Type of cancer, n (%)				.08
Prostate	149 (59.8)	108 (63.5)	41 (52)	
Colorectal	100 (40.2)	62 (36.5)	38 (48)	
Treatment phase, n (%)				.046
During treatment	19 (7.6)	9 (5.3)	10 (13)	
After treatment	230 (92.4)	161 (94.7)	69 (87)	
Time since last treatment in months, mean (SD)	5.64 (3.84)	5.42 (3.65)	6.13 (4.22)	.18
PA[b] -related factors				
MVPA[c] SQUASH[d], mean (SD)	798 (721)	831 (765)	727 (617)	.29
MVPA ActiGraph, mean (SD)	270 (211)	280 (199)	249 (233)	.30
PA intention, mean (SD)	7.61 (1.35)	7.71 (1.26)	7.38 (1.52)	.07
Health-related factors				
BMI[e] category, n (%)				.47
Normal weight	89 (36.2)	59 (34.7)	30 (40)	
Overweight	157 (63.8)	111 (65.3)	46 (60)	
General HRQoL[f], mean (SD)	80.01 (16.81)	80.34 (16.53)	79.28 (17.51)	.65
Fatigue, mean (SD)	24.01 (11.58)	23.04 (11.22)	26.48 (12.18)	.04

[a]SD: standard deviation.

[b]PA: physical activity.

[c]MVPA: moderate to vigorous physical activity.

[d]SQUASH: Short Questionnaire to Assess Health Enhancing Physical Activity.

[e]BMI: body mass index.

[f]HRQoL: health-related quality of life.

Table 2. Logistic regression to study relation between participant characteristics and the initial choice to participate Web-based (Nagelkerke R^2=.25). Initial Web-based participation coded as 1. Additional nonparametric bootstrap analysis led to similar results.

Characteristics	Odds ratio (95% CI)	*P* value
Demographic factors		
Age (years)	0.93 (0.89-0.98)	.005
Gender (male=Ref)	1.62 (0.52-5.03)	.40
Education (low=Ref)		
Middle	0.95 (0.42-2.15)	.90
High	4.08 (1.58-10.56)	.004
Cancer-related factors		
Type of cancer (prostate=Ref)	0.55 (0.26-1.18)	.13
Treatment phase (during treatment=Ref)	5.58 (1.36-22.82)	.02
Time since last treatment (months)	0.87 (0.79-0.96)	.007
PA[a] -related factors		
PA intention	1.08 (0.80-1.44)	.62
MVPA[b] ActiGraph	0.99 (1.0-1.0)	.33
Health-related factors		
BMI[c] (normal weight=Ref)	1.48 (0.73-3.03)	.28
General HRQoL[d]	1.00 (0.97-1.03)	.99
Fatigue	0.96 (0.93-1.00)	.04

[a]PA: physical activity.

[b]MVPA: moderate to vigorous physical activity.

[c]BMI: body mass index.

[d]HRQoL: health-related quality of life.

Multiple logistic regression (see Table 2) revealed that participants were less likely to initially start Web-based with higher age (OR=0.93, 95% CI 0.89-0.98), longer time since last treatment (OR=0.87, 95% CI 0.79-0.96), and higher levels of fatigue (OR=0.96, 95% CI 0.93-1.0). Although time since last treatment is negatively associated with initially participating Web-based, participants who had completed cancer treatment were more likely to participate Web-based than those who were still under active treatment (OR=5.58, 95% CI 1.36-22.82). Furthermore, those with a high level of education were more likely to initially participate Web-based compared with those with a low level of education (OR=4.08, 95% CI 1.58-10.56).

Linking Delivery Preference to Intervention Use

When examining intervention material use in relation to the initial choice for delivery mode (see Table 3), it can be noticed that a significantly higher percentage of initial print-based participants did not read (all three) Web-based advice (advice 1 and 2: *P*<.001; advice 3: *P*=.005). Furthermore, initial print-based participants were very consistent in their intervention material use throughout the intervention: 95% to 98% (partly) read the print-based advice and 56% to 62% did not read the Web-based advice (see Table 3). Web-based participants were more variable in the way they read their advice: completeness per advice decreases from the first advice to the final advice, with the final print-based advice being read significantly less completely (*P*<.001) by initial Web-based participants compared with initial print-based participants. Additional analyses showed that intervention completeness considering both versions of the advice was not lower for the initial Web-based participants. Percentages of participants reporting not having read any advice completely ranged from 0.9% (2/223) to 5.8% (13/225) per advice with no statistical differences between both groups.

With regard to completion of the second questionnaire (T1), it was noticed that the majority chose the same delivery mode for this questionnaire: 89.0% (146/164) of the initial Web-based participants completed the questionnaire on the website, and 86.8% (59/68) of the initial print-based participants completed the questionnaire on paper.

Table 3. Intervention material use and completeness compared for initial Web-based (n=170) and print-based (n=79) participants.

Reading of computer-tailored advice	Initial Web-based participants		Initial print-based participants		*P* value
	n	%	n	%	
Advice 1 Print-based					.36
Completely	118	72.0	51	78	
Partly	38	23.2	12	19	
Not	8	4.9	1	2	
Advice 1 Web-based					<.001
Completely	96	58.9	12	23	
Partly	44	27.0	10	19	
Not	23	14.1	31	59	
Advice 2 Print-based					.06
Completely	102	62.2	50	77	
Partly	39	23.8	12	19	
Not	23	14.0	3	5	
Advice 2 Web-based					<.001
Completely	83	51.2	11	21	
Partly	40	24.7	9	17	
Not	39	24.1	32	62	
Advice 3 Print-based					
Completely	74	48.1	47	78	<.001
Partly	58	37.7	10	17	
Not	22	14.3	3	5	
Advice 3 Web-based					.005
Completely	65	42.2	12	24	
Partly	42	27.3	10	20	
Not	47	30.5	28	56	

Table 4. Logistic regression to study relation between participant characteristics and the continued use of Web-based intervention materials (Nagelkerke R^2=.21). Use of Web-based materials coded as 1. Additional nonparametric bootstrap analysis led to similar results.

Characteristics	Odds ratio (95% CI)	*P* value
Demographic factors		
Age (years)	0.93 (0.86-1.0)	.04
Gender (male=Ref)	0.82 (0.18-3.63)	.79
Education (low=Ref)		
Middle	1.22 (0.39-3.79)	.73
High	3.52 (0.98-12.61)	.05
Cancer-related factors		
Type of cancer (prostate = Ref)	0.98 (0.33-2.87)	.96
Treatment phase (during treatment = Ref)	0.99 (0.08-12.77)	.99
Time since last treatment (months)	0.86 (0.75-0.98)	.02
PA[a] -related factors		
PA intention	1.13 (0.78-1.65)	.51
MVPA[b] ActiGraph	1.00 (1.0-1.0)	.53
Health-related factors		
BMI[c] (normal weight=Ref)	2.34 (0.92-5.95)	.08
General HRQoL[d]	1.00 (0.97-1.04)	.76
Fatigue	0.98 (0.93-1.03)	.43

[a]PA: physical activity.

[b]MVPA: moderate to vigorous physical activity.

[c]BMI: body mass index.

[d]HRQoL: health-related quality of life.

Continued Intervention Use Delivery Mode

With regard to the selected delivery mode for using the intervention materials, we noticed that the majority (191/225; 84.9%) used Web-based most often in combination with print-based materials (see Figure 1). Results of the logistic regression identifying participant characteristics concerning the delivery mode for the use of intervention materials (print-only vs using Web-based materials; see Table 4) were similar to the results for initial choice of delivery mode for age (OR=0.93, 95% CI 0.86-1.00) and time since last treatment (OR=0.86, 95% CI 0.75-0.98). Highly educated participants were not significantly more likely to use Web-based intervention materials than less educated participants, but the OR (3.52) and its 95% CI (0.98-12.61) indicate that educational level may still be an important predictor. Fatigue and treatment phase were not identified as predictors for using Web-based materials (or a combination) instead of using only print-based materials.

Discussion

This study was aimed at investigating participant characteristics in relation to initial choice of intervention delivery mode in a population of prostate and colorectal CPS. Additionally, intervention material use and participant characteristics in relation to intervention delivery mode were examined. Analyses provide insight into the feasibility of Web-based interventions in an older population of cancer patients and thereby aid further implementation of the intervention.

Participant Characteristics and Delivery Mode

Age and education level were two participant characteristics which were consistently related to intervention delivery mode both for initial choice and follow-up delivery mode (ie, Web-based vs print-based). Higher age was associated with a lower likelihood of using Web-based intervention materials. A lower educational level, although not significant in all analyses, was also associated with lower Web-based participation. This corresponds with our expectations. Previous studies also revealed that eHealth literacy is lower for older adults and those with lower education [33,53]. Kontos et al [52] found younger age and higher education to be predictors for searching health information through the Internet and using websites for diet, weight, and PA. In addition, participants in an adult population selecting print-based materials were older and had a lower level of education than those selecting Web-based materials [34].

This finding may imply that when implementing Web-based interventions in a population of prostate and colorectal CPS, but probably also in a general older population (ie, 61% of colorectal and 64% of prostate cancer patients is aged over 70 years at the time of diagnosis [30]), those who are older and those with a lower level of education may not be reached. Statistics in the Netherlands showed that Internet access and

frequency of Internet use decrease substantially from the age of 75 years and are also lower among those with a lower educational level [29]. Thus, besides lower Internet access, older and less educated participants may also have lower Internet experience and self-efficacy. As a result, they may choose to use the print-based materials, requiring less effort in comparison with the Web-based materials. This is acknowledged through the unified theory of acceptance and use of technology (UTAUT) by Venkatesh. This theory states that use of technology is influenced by (among others) facilitating conditions and performance and effort expectancy, which are moderated by age and experience [54]. Additionally, a study among colorectal CPS revealed that older patients do perceive Web-based health information tools as highly useful and indicate a willingness to use such tools but are not always able to use them optimally [55]. It may be recommended to provide both Web-based and print-based materials, especially among those aged over 75 years, to prevent exclusion of a vulnerable group of older or less educated participants and to have the most optimal use of the intervention

Besides age and education, time since treatment was also consistently related to participating in the Web-based intervention. CPS who finished their treatment longer ago were less likely to participate through the Internet. CPS who received their most recent treatment longer ago are probably already further in their recovery process, may perceive less need for PA advice and may be less committed to becoming physically active. As a result, they probably chose the delivery mode for which they needed the least effort. Using print-based materials may be perceived as easier, as all materials are delivered at home and can be completed at any time. Although accessing Web-based materials was made as easy as possible (eg, emails linked participants to the website without log-in), for Web-based participation, it is still necessary to start up a computer or tablet and go on the Web before being able to complete questionnaires [56]. Our results suggest that it may be important to provide print-based materials to also include those who completed their treatment longer ago. However, no other studies considered time since treatment, and therefore, additional research is necessary to explore the role of time since treatment.

Having finished treatment and lower levels of fatigue predicted the initial choice to participate Web-based but did not predict the use of Web-based intervention materials. Possibly, participants' Internet frequency decreases during treatments and while feeling fatigued. Going on the Web to start the intervention may be perceived as more effortful and may explain the initial choice to participate in the print-based intervention. During the course of the intervention, treatments may be finished and fatigue may decrease. As a result, CPS may decide to visit the Web-based content of the intervention during continued use. It may be important to provide both delivery modes at invitation for those who are still undergoing treatment or suffering from fatigue, a group that may benefit most from the intervention. Future research needs to confirm these findings.

In this study, gender was not predictive for intervention delivery mode. The precise role of gender differences regarding Internet access, eHealth use, and delivery mode has been ambiguous: whereas some studies found a link with gender [35,36,52], others did not find differences between males and females [32,34,57,58]. Additionally, it should be noted that there was only a small portion of women in this study, as a result of part of the intended target population being prostate CPS, which may have influenced the power to detect differences. Future research should provide more insight regarding the influence of gender on delivery mode.

It is also interesting that PA behavior and intention to be physically active are not related to intervention delivery mode preference. PA behavior has proven to be a predictor of delivery mode preference according to studies examining self-reported intervention modality preference. Studies in the general population and in a cancer population found that those with lower PA levels may have a preference for Web-based or computer-based interventions [36,57]. Others argued that those with a risk behavior (eg, low PA behavior) may prefer the instant availability and interactivity of Web-based materials [34]. However, both in this study as well as in the study of Greaney et al [34], the actual choice of delivery mode was not predicted by PA behavior. Possibly, reporting a certain preference is different from the actual choice. Therefore, additional research is necessary to examine the role of health behavior in intervention delivery mode.

Intervention Material Use

It is promising that in an older population (ie, mean age of 66 years), approximately two-thirds of the participants initially chose to participate through the Internet and that even a larger proportion (ie, almost 85%, 191/225) used the Web-based intervention content. This indicates that for a large part of our population, going on the Web was not a barrier.

Since the Web-based content of the OncoActive intervention could be accessed using a computer or tablet, participants were able to visit the website in the manner they were most familiar with (eg, computer or tablet). Providing OncoActive through different platforms may have increased the usage of the Web-based intervention materials [59].

We also examined whether intervention material use differed between initial print-based participants and initial Web-based participants. The majority of initial print-based participants predominantly used the print-based tailored advice. Significantly less participants in this group used the Web-based advice (ie, 38%-44%, Table 3), compared with the initial Web-based participants (ie, 69.5%-85.9%, Table 3). Additionally, the majority (ie, 87%, 59/68) of the initial print-based participants also completed the second questionnaire (which was part of the intervention) on paper. These findings indicate that it may be important to provide print-based intervention materials for participants who start the intervention print-based. However, since Web-based materials can be provided without additional costs, it is recommended to provide both.

The majority of the initial Web-based participants also completed the second questionnaire on the website (ie, 89.0%, 146/164) and used both Web-based (ie, 69.5%-85.9%, Table 3) and print-based (ie, 85.8%-95.2%, Table 3) intervention materials, indicating that initial Web-based participants used a mixture of both advice texts and that providing print-based

tailored advice in addition to the Web-based advice may be advantageous. Studies among elderly found similar results, indicating that even among those using the Internet, a preference for print-based or nondigital information persists [33,60]. Although this may be a temporary phenomenon (eg, rapid technology development and aging of adults more familiar with the Internet), Vandelanotte et al [24] also suggested that having access to Web-based material might not be sufficient in itself. Therefore, future research should also focus on reasons for not using Web-based materials or why there is a preference for print-based materials.

As mentioned, information regarding use of print-based materials is important, as offering Web-based and print-based materials alongside each other is associated with higher costs. A version that is only Web-based would be less costly. With regard to the questionnaires, it was noticed that a majority started the intervention with a Web-based questionnaire and also continued to complete additional questionnaires on the website. Consequently, it may be feasible to initially invite participants to complete the questionnaires on the website, while explicitly explaining that it is also possible to participate in the intervention if they do not have Internet access or are not able or willing to participate through the Internet. Print-based questionnaires can then be provided on request or with a reminder. Nevertheless, it may be advisable to offer both delivery modes with the invitation for those who are older or for those still undergoing treatment, as these participant characteristics are easy to administer at intake, and results of this study showed that they are predictors of initial print-based participation.

With regard to the computer-tailored advice, the majority used a combination of both Web-based and printed materials. Although providing print-based materials complementary to the Web-based advice is associated with higher costs (eg, printing and postage costs), it may be best to provide both, as print materials are used by all participants. Providing both delivery modes alongside each other may be more costly than providing the intervention in a singular delivery mode. Additionally, intervention efficacy in relation to delivery mode should also be considered, as information processing may depend on the delivery mode [61]. Therefore, future research should also focus on the relation between delivery mode and (cost) effectiveness.

Strengths and Limitations of This Study

Providing participants with the ability to select their own preferred intervention delivery mode is regarded as a strength of this study. As indicated by previous studies, this may enlarge intervention engagement and thereby the impact of the intervention [34]. Additionally, this study had a very low dropout rate. Only 6.0% (15/249) of the participants opted out of the study during the intervention period. This is a remarkable finding, as high dropout rates are common in Web-based interventions [62,63]. Providing a combination of Web-based and print-based materials might have prevented participants from dropping out of the study. If a specific delivery mode did not meet participants' expectations, they were able to use only the materials that were most appealing to them.

As described in the methods, the preferred initial delivery mode was based on completion of the first questionnaire, as it contained the questions to build the tailored advice. However, it should be acknowledged that because of the evaluation of the intervention in an RCT, the questionnaire was longer than the actual questionnaire needed for tailoring. As a result, the current findings may reflect the preference for completing a research survey rather than the actual intervention delivery mode. Future implementation without the research component would be necessary to confirm the current findings.

In this study, we did not collect any information regarding Internet access and previous experience with the Internet. This information might have been valuable, as other studies found those factors to be predictors of using Web-based intervention materials [33,34,52].

Participants in this study were offered both Web-based and print-based materials complementary to each other. Therefore, we were not able to discriminate whether offering only one delivery mode would yield the same results. This could be studied in future research.

In our study, recall bias was possible. The use of tailored advice was self-reported, and evaluation took place up to 3 months after participants received advice. For future studies, it is recommended to incorporate some evaluation regarding the use of Web-based material immediately after providing the materials and preferably objective usage data. Objective usage data is not available in a print-based version and incorporating an additional questionnaire immediately after provision is more complicated and burdensome for the print-based material. However, objective usage data for Web-based material can be used to validate self-report items to assess the probability of recall bias in future studies.

Conclusions

Intervention reach may be better, and interventions may possibly even be more effective if participants are able to use their own preferred delivery mode [34]. Information regarding participant characteristics related to intervention delivery mode can provide important cues for implementing computer-tailored interventions. To our knowledge, this is one of the first studies that assessed the relationship between participant characteristics and choice of delivery mode in an intervention in which both delivery modes were offered alongside each other, thereby providing participants a free choice of delivery mode.

Use of print-based materials among the initial Web-based participants was substantial, indicating that print-based materials are also important for those using Web-based materials. In contrast, by using only print-based materials, the intervention may be less attractive and useful for younger CPS, as it is known that younger CPS frequently use the Internet with regard to finding health-related information [64]. This study provides indications that Web-based and print-based materials could best be offered alongside each other. Providing Web-based materials only would exclude some of those who are older, less educated, more fatigued, or are currently undergoing treatment. Especially these participants are often more vulnerable and could benefit most from PA interventions.

Acknowledgments

The authors would like to thank the hospitals who helped with the recruitment of participants for the RCT: Admiraal de Ruyter Hospital (Goes/Vlissingen), Amphia Hospital (Breda), Albert Schweitzer Hospital (Dordrecht), Bernhoven (Uden), Catharina Hospital (Eindhoven), Canisius Wilhelmina Hospital (Nijmegen), Ikazia Hospital (Rotterdam), LangeLand Hospital (Zoetermeer), Maasstad Hospital (Rotterdam), MAASTRO clinic (Maastricht), Maastricht UMC+ (Maastricht), Rijnstate (Arnhem), St. Jans Gasthuis (Weert), Slingeland Hospital (Doetinchem), St. Anna Hospital (Geldrop), VieCuri Medical Centre (Venlo/Venray), and Zuyderland Hospital (Sittard/Heerlen).

This research was funded by the Dutch Cancer Society (KWF Kankerbestrijding, grant number NOU2012-5585).

Conflicts of Interest

Hein de Vries is the Scientific Director of Vision2Health, a company that licenses evidence-based, innovative computer-tailored health communication tools.

Multimedia Appendix 1

Screenshots of the OncoActive intervention.

[PDF File (Adobe PDF File), 4MB - jmir_v19i8e298_app1.pdf]

References

1. Harrington CB, Hansen JA, Moskowitz M, Todd BL, Feuerstein M. It's not over when it's over: long-term symptoms in cancer survivors--a systematic review. Int J Psychiatry Med 2010;40(2):163-181. [Medline: 20848873]
2. Wu H, Harden JK. Symptom burden and quality of life in survivorship: a review of the literature. Cancer Nurs 2015;38(1):E29-E54. [doi: 10.1097/NCC.0000000000000135] [Medline: 24831042]
3. Bourke L, Boorjian SA, Briganti A, Klotz L, Mucci L, Resnick MJ, et al. Survivorship and improving quality of life in men with prostate cancer. Eur Urol 2015 Sep;68(3):374-383. [doi: 10.1016/j.eururo.2015.04.023] [Medline: 25941049]
4. Denlinger CS, Barsevick AM. The challenges of colorectal cancer survivorship. J Natl Compr Canc Netw 2009 Sep;7(8):883-93; quiz 894 [FREE Full text] [Medline: 19755048]
5. El-Shami K, Oeffinger KC, Erb NL, Willis A, Bretsch JK, Pratt-Chapman ML, et al. American cancer society colorectal cancer survivorship care guidelines. CA Cancer J Clin 2015;65(6):428-455 [FREE Full text] [doi: 10.3322/caac.21286] [Medline: 26348643]
6. Skolarus TA, Wolf AM, Erb NL, Brooks DD, Rivers BM, Underwood W, et al. American cancer society prostate cancer survivorship care guidelines. CA Cancer J Clin 2014;64(4):225-249 [FREE Full text] [doi: 10.3322/caac.21234] [Medline: 24916760]
7. Cormie P, Zopf EM, Zhang X, Schmitz KH. The impact of exercise on cancer mortality, recurrence, and treatment-related adverse effects. Epidemiol Rev 2017 Jan 01;39(1):71-92. [doi: 10.1093/epirev/mxx007] [Medline: 28453622]
8. Buffart LM, Kalter J, Sweegers MG, Courneya KS, Newton RU, Aaronson NK, et al. Effects and moderators of exercise on quality of life and physical function in patients with cancer: An individual patient data meta-analysis of 34 RCTs. Cancer Treat Rev 2017 Jan;52:91-104 [FREE Full text] [doi: 10.1016/j.ctrv.2016.11.010] [Medline: 28006694]
9. Schmitz KH, Holtzman J, Courneya KS, Mâsse LC, Duval S, Kane R. Controlled physical activity trials in cancer survivors: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev 2005 Jul;14(7):1588-1595 [FREE Full text] [doi: 10.1158/1055-9965.EPI-04-0703] [Medline: 16030088]
10. Szymlek-Gay EA, Richards R, Egan R. Physical activity among cancer survivors: a literature review. N Z Med J 2011 Jun 24;124(1337):77-89. [Medline: 21946880]
11. Denlinger CS, Engstrom PF. Colorectal cancer survivorship: movement matters. Cancer Prev Res (Phila) 2011 Apr;4(4):502-511 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0098] [Medline: 21464030]
12. Mishra SI, Scherer RW, Geigle PM, Berlanstein DR, Topaloglu O, Gotay CC, et al. Exercise interventions on health-related quality of life for cancer survivors. Cochrane Database Syst Rev 2012(8):CD007566. [doi: 10.1002/14651858.CD007566.pub2] [Medline: 22895961]
13. Mishra SI, Scherer RW, Snyder C, Geigle PM, Berlanstein DR, Topaloglu O. Exercise interventions on health-related quality of life for people with cancer during active treatment. Cochrane Database Syst Rev 2012(8):CD008465. [doi: 10.1002/14651858.CD008465.pub2] [Medline: 22895974]
14. Schmitz KH, Courneya KS, Matthews C, Demark-Wahnefried W, Galvão DA, Pinto BM, et al. American college of sports medicine roundtable on exercise guidelines for cancer survivors. Med Sci Sports Exerc 2010 Jul;42(7):1409-1426. [doi: 10.1249/MSS.0b013e3181e0c112] [Medline: 20559064]

15. Speck RM, Courneya KS, Mâsse LC, Duval S, Schmitz KH. An update of controlled physical activity trials in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2010 Jun;4(2):87-100. [doi: 10.1007/s11764-009-0110-5] [Medline: 20052559]
16. Jansen F, van Uden-Kraan CF, van Zwieten V, Witte BI, Verdonck-de Leeuw IM. Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. Support Care Cancer 2015 Jun;23(6):1679-1688. [doi: 10.1007/s00520-014-2514-7] [Medline: 25424520]
17. Willems RA, Bolman CA, Mesters I, Kanera IM, Beaulen AA, Lechner L. Cancer survivors in the first year after treatment: the prevalence and correlates of unmet needs in different domains. Psychooncology 2016 Jan;25(1):51-57. [doi: 10.1002/pon.3870] [Medline: 26110652]
18. Playdon M, Ferrucci LM, McCorkle R, Stein KD, Cannady R, Sanft T, et al. Health information needs and preferences in relation to survivorship care plans of long-term cancer survivors in the American Cancer Society's Study of Cancer Survivors-I. J Cancer Surviv 2016 Dec;10(4):674-685. [doi: 10.1007/s11764-015-0513-4] [Medline: 26744339]
19. Meulepas JM, Kiemeney LA. Kanker in Nederland tot 2020: trends en prognoses [Cancer in the Netherlands up to 2020: trends and forecasts]. Amsterdam: KWF Kankerbestrijding; 2011.
20. Siesling SS, Sonke GS, de Raaf DH, Jansen-Landheer ML. [Cancer care in the picture]. Utrecht: Integraal Kankercentrum Nederland (IKNL); 2014.
21. Joseph RP, Durant NH, Benitez TJ, Pekmezi DW. Internet-based physical activity interventions. Am J Lifestyle Med 2014 Dec;8(1):42-68 [FREE Full text] [doi: 10.1177/1559827613498059] [Medline: 25045343]
22. 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]
23. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
24. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of ehealth and mhealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
25. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
26. Lustria ML, 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]
27. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4 [FREE Full text] [Medline: 10788196]
28. Krijgsman J, Swinkels I, Van Lettow B, De Jong J, Out K, Friele R, et al. Meer dan techniek: eHealth-monitor 2016 [More than technology: eHealth-monitor 2016]. Den Haag and Utrecht: Nictiz and Nivel; 2016.
29. Centraal Bureau voor de Statistiek. 2017. Internet: toegang, gebruik en faciliteiten [Internet: access, use and facilities] URL: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=83429NED&D1=0,8-14,39&D2=0,4-6,8-14&D3=a&D4=l&HDR=T&STB=G1,G2,G3&VW=T [WebCite Cache ID 6pWvl2bUU]
30. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016 Jul;66(4):271-289 [FREE Full text] [doi: 10.3322/caac.21349] [Medline: 27253694]
31. 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]
32. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
33. 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]
34. Greaney ML, Puleo E, Bennett GG, Haines J, Viswanath K, Gillman MW, et al. Factors associated with choice of web or print intervention materials in the healthy directions 2 study. Health Educ Behav 2014 Feb;41(1):52-62 [FREE Full text] [doi: 10.1177/1090198113486803] [Medline: 23720532]
35. Peels DA, Bolman C, Golsteijn RHJ, De Vries H, Mudde AN, van Stralen MM, et al. Differences in reach and attrition between web-based and print-delivered tailored interventions among adults over 50 years of age: clustered randomized trial. J Med Internet Res 2012;14(6):e179 [FREE Full text] [doi: 10.2196/jmir.2229] [Medline: 23246790]
36. Short CE, Vandelanotte C, Duncan MJ. Individual characteristics associated with physical activity intervention delivery mode preferences among adults. Int J Behav Nutr Phys Act 2014;11(1):25 [FREE Full text] [doi: 10.1186/1479-5868-11-25] [Medline: 24568611]

37. Integraal Kankercentrum Nederland. 2015. Nederlandse Kankerregistratie [Dutch Cancer Registery] URL: http://cijfersoverkanker.nl/ [WebCite Cache ID 6pWtzuAPZ]
38. Peels DA, Bolman C, Golsteijn RHJ, de Vries H, Mudde AN, van Stralen MM, et al. Long-term efficacy of a printed or a Web-based tailored physical activity intervention among older adults. Int J Behav Nutr Phys Act 2013;10:104 [FREE Full text] [doi: 10.1186/1479-5868-10-104] [Medline: 24228862]
39. 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]
40. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Development of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: OncoActive. BMC Cancer 2017 Jun 26;17(1):446 [FREE Full text] [doi: 10.1186/s12885-017-3397-z] [Medline: 28651586]
41. van Stralen MM, Kok G, de Vries H, Mudde AN, Bolman C, Lechner L. The active plus protocol: systematic development of two theory- and evidence-based tailored physical activity interventions for the over-fifties. BMC Public Health 2008 Dec 04;8:399 [FREE Full text] [doi: 10.1186/1471-2458-8-399] [Medline: 19055806]
42. Peels DA, van Stralen MM, Bolman C, Golsteijn RHJ, de Vries H, Mudde AN, et al. Development of web-based computer-tailored advice to promote physical activity among people older than 50 years. J Med Internet Res 2012 Mar 02;14(2):e39 [FREE Full text] [doi: 10.2196/jmir.1742] [Medline: 22390878]
43. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [Medline: 14680666]
44. Haskell WL, Lee I, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American college of sports medicine and the American heart association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
45. Broderick JM, Ryan J, O'Donnell DM, Hussey J. A guide to assessing physical activity using accelerometry in cancer patients. Support Care Cancer 2014 Apr;22(4):1121-1130. [doi: 10.1007/s00520-013-2102-2] [Medline: 24389829]
46. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act 2011 Jun 16;8:62 [FREE Full text] [doi: 10.1186/1479-5868-8-62] [Medline: 21679426]
47. Skender S, Schrotz-King P, Böhm J, Abbenhardt C, Gigic B, Chang-Claude J, et al. Repeat physical activity measurement by accelerometry among colorectal cancer patients--feasibility and minimal number of days of monitoring. BMC Res Notes 2015 Jun 06;8:222 [FREE Full text] [doi: 10.1186/s13104-015-1168-y] [Medline: 26048683]
48. Choi L, Ward SC, Schnelle JF, Buchowski MS. Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. Med Sci Sports Exerc 2012 Oct;44(10):2009-2016 [FREE Full text] [doi: 10.1249/MSS.0b013e318258cb36] [Medline: 22525772]
49. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Sep;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
50. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 3;85(5):365-376. [Medline: 8433390]
51. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994 Jul;38(5):383-392. [Medline: 7965927]
52. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the health information national trends survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
53. 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]
54. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
55. Bolle S, Romijn G, Smets EMA, Loos EF, Kunneman M, van Weert JCM. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. J Med Internet Res 2016 Jul 25;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
56. Kroeze W, Oenema A, Campbell M, Brug J. Comparison of use and appreciation of a print-delivered versus CD-ROM-delivered, computer-tailored intervention targeting saturated fat intake: randomized controlled trial. J Med Internet Res 2008 Apr 29;10(2):e12 [FREE Full text] [doi: 10.2196/jmir.940] [Medline: 18487136]
57. Martin EC, Basen-Engquist K, Cox MG, Lyons EJ, Carmack CL, Blalock JA, et al. Interest in health behavior intervention delivery modalities among cancer survivors: a cross-sectional study. JMIR Cancer 2016 Feb 11;2(1):e1. [doi: 10.2196/cancer.5247]

58. Kerr C, Murray E, Noble L, Morris R, Bottomley C, Stevenson F, et al. The potential of Web-based interventions for heart disease self-management: a mixed methods investigation. J Med Internet Res 2010;12(4):e56 [FREE Full text] [doi: 10.2196/jmir.1438] [Medline: 21156471]
59. Granger D, Vandelanotte C, Duncan MJ, Alley S, Schoeppe S, Short C, et al. Is preference for mHealth intervention delivery platform associated with delivery platform familiarity? BMC Public Health 2016 Jul 22;16:619 [FREE Full text] [doi: 10.1186/s12889-016-3316-2] [Medline: 27450240]
60. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
61. Smit ES, Linn AJ, Van Weert JC. 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.
62. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
63. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
64. Chou WS, Liu B, Post S, Hesse B. Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003-2008. J Cancer Surviv 2011 Sep;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]

Abbreviations

ANOVA: analysis of variance
BMI: body mass index
CIS: Checklist Individual Strength
CPS: cancer patients and survivors
eHealth: electronic health
EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30
HRQoL: health-related quality of life
MET: metabolic equivalent
MVPA: moderate to vigorous physical activity
OR: odds ratio
PA: physical activity
RCT: randomized controlled trial
SD: standard deviation
SES: socioeconomic status
SPSS: Statistical Package for the Social Sciences
SQUASH: Short Questionnaire to Assess Health enhancing physical activity
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by G Eysenbach; submitted 07.04.17; peer-reviewed by C Short, E Martin; comments to author 04.05.17; revised version received 01.06.17; accepted 16.06.17; published 23.08.17.

Please cite as:
Golsteijn RHJ, Bolman C, Peels DA, Volders E, de Vries H, Lechner L
A Web-Based and Print-Based Computer-Tailored Physical Activity Intervention for Prostate and Colorectal Cancer Survivors: A Comparison of User Characteristics and Intervention Use
J Med Internet Res 2017;19(8):e298
URL: http://www.jmir.org/2017/8/e298/
doi:10.2196/jmir.7838
PMID:28835353

Original Paper

Text Messaging Interventions on Cancer Screening Rates: A Systematic Review

Catherine Uy[1*], MD; Jennifer Lopez[1*], B.S.; Chau Trinh-Shevrin[1,2], DrPH; Simona C Kwon[1,2], DrPH, MPH; Scott E Sherman[1,2,3], MD, MPH; Peter S Liang[1,3], MD, MPH

[1]Department of Medicine, NYU School of Medicine, New York, NY, United States

[2]Department of Population Health, NYU School of Medicine, New York, NY, United States

[3]Department of Medicine, VA New York Harbor Manhattan Medical Center, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Peter S Liang, MD, MPH
Department of Medicine
VA New York Harbor Manhattan Medical Center
11N, Gastroenterology
423 E 23rd St
New York, NY, 10010
United States
Phone: 1 212 686 7500 ext 5745
Email: Peter.Liang@nyumc.org

Abstract

Background: Despite high-quality evidence demonstrating that screening reduces mortality from breast, cervical, colorectal, and lung cancers, a substantial portion of the population remains inadequately screened. There is a critical need to identify interventions that increase the uptake and adoption of evidence-based screening guidelines for preventable cancers at the community practice level. Text messaging (short message service, SMS) has been effective in promoting behavioral change in various clinical settings, but the overall impact and reach of text messaging interventions on cancer screening are unknown.

Objective: The objective of this systematic review was to assess the effect of text messaging interventions on screening for breast, cervical, colorectal, and lung cancers.

Methods: We searched multiple databases for studies published between the years 2000 and 2017, including PubMed, EMBASE, and the Cochrane Library, to identify controlled trials that measured the effect of text messaging on screening for breast, cervical, colorectal, or lung cancers. Study quality was evaluated using the Cochrane risk of bias tool.

Results: Our search yielded 2238 citations, of which 31 underwent full review and 9 met inclusion criteria. Five studies examined screening for breast cancer, one for cervical cancer, and three for colorectal cancer. No studies were found for lung cancer screening. Absolute screening rates for individuals who received text message interventions were 0.6% to 15.0% higher than for controls. Unadjusted relative screening rates for text message recipients were 4% to 63% higher compared with controls.

Conclusions: Text messaging interventions appear to moderately increase screening rates for breast and cervical cancer and may have a small effect on colorectal cancer screening. Benefit was observed in various countries, including resource-poor and non-English-speaking populations. Given the paucity of data, additional research is needed to better quantify the effectiveness of this promising intervention.

(J Med Internet Res 2017;19(8):e296) doi:10.2196/jmir.7893

KEYWORDS

text messaging; early detection of cancer; breast neoplasms; colorectal neoplasms; lung neoplasms; mHealth; uterine cervical neoplasms

Introduction

Cancer is a leading cause of death worldwide. The global burden of cancer is increasing in both developed and developing nations as the world population ages and established risk factors associated with economic development and urbanization become more prevalent, including smoking, obesity, sedentary lifestyle, and lower parity [1,2]. Worldwide, the most common malignancies are lung, colorectal, and prostate cancers for men and lung, colorectal, and breast cancers for women [3]. These cancers account for the vast majority of the cancer-related deaths and were responsible for more than 8 million deaths in 2012 [4].

In the United States alone, an estimated 1.7 million new cancer cases were diagnosed in 2016, resulting in nearly 600,000 deaths [5]. The cancer mortality rate in the United States has dropped by 23% since 1991. This trend is, in part, because of population-level screening tests, which detect early cancers and facilitate treatment before the disease becomes clinically apparent [6]. To date, screening tests for only four cancers have been shown to reduce mortality and the US Preventive Services Task Force (USPSTF) gives either an A or B grade recommendation for each. A grade A recommendation indicates a high certainty of substantial net benefit, and a grade B recommendation indicates a moderate to high certainty of at least moderate net benefit. Mammography reduces breast cancer mortality, and the USPSTF gives a B recommendation for biennial testing in average-risk women between 50 and 74 years of age [7]. Papanicolaou (Pap) testing and combined Pap and human papillomavirus (HPV) testing reduce the risk of death from cervical cancer in women aged between 21 and 65 years, which carries an A recommendation [8]. Colonoscopy, sigmoidoscopy, and fecal occult blood testing (FOBT) are the three tests that reduce colorectal cancer mortality, and the USPSTF gives an A recommendation to screen average-risk individuals in the age group of 50 to 75 years [9]. Finally, low-dose computed tomography (CT) has been shown to reduce lung cancer deaths in heavy smokers, and screening for these individuals aged 55 to 80 years carries a B recommendation [10]. Despite evidence-based guidelines for breast, cervical, colorectal, and lung cancers, there are challenges to increasing screening rates, particularly among segments of the population vulnerable to health disparities such as racial minorities and individuals from lower socioeconomic status. Tailored and targeted interventions to improve adoption and uptake of these guidelines are important to reducing overall mortality and disparities in outcome for these cancers.

As global economies and technology advance, more individuals have access to mobile phones. Text messaging, also known as short message service (SMS), is already being integrated into health care practices to improve adherence to contraception, smoking cessation, and weight loss programs, and it has been shown to increase attendance at primary care visits [11,12]. Several studies have demonstrated both patient interest in SMS reminders for cancer screening appointments and a positive impact on screening rates [13,14]. However, a cumulative analysis of the overall effect of text messaging on cancer screening rates has not been performed. The purpose of this systematic review was to assess the effect of text messaging interventions on increasing patient adherence to screening recommendations for breast, cervical, colorectal, and lung cancers.

Methods

Eligibility Criteria

We searched for clinical trials published from January 2000 to January 2017 that studied the effect of text messaging on screening for the four cancers of interest. We included all languages of publication. Types of screening methods were limited to mammography for breast cancer; Pap test and HPV test for cervical cancer; colonoscopy, FOBT, and sigmoidoscopy for colorectal cancer; and low-dose CT for lung cancer. Research subjects of all ages were considered. The outcome measures were absolute screening rates or relative screening rates in the text message group compared with a control group.

Search Strategy

Studies were identified by searching several electronic databases and additionally searching references of relevant papers. Papers published in a language other than English were translated for review. The search was performed in PubMed, EMBASE, the Cochrane Library, Clinicaltrials.gov, Inspec, HSRProj (Health Services Research Projects in Progress), and NIH Reporter (Textbox 1). Attempts to identify gray literature were made by searching the BIOSIS Citation Index and Proquest Dissertations & Theses Global database, as well as by a manual search.

Textbox 1. Search terms for PubMed and comparable search terms for other databases.

“Text Messaging”[MeSH] OR texting[tiab] OR “text messaging”[tiab] OR “text message”[tiab] OR “SMS message”[tiab] OR ((mobile OR cell OR cellular OR smart OR app OR application) AND phone[tiab]) AND (“Early Detection of Cancer”[MeSH] OR (cancer OR neoplasm OR neoplasms) AND screening) OR “Colonoscopy”[MeSH] OR “Colonoscopy”[tiab] OR “Colorectal Neoplasms/diagnosis”[MeSH] OR “Occult Blood”[MeSH] OR “fecal occult blood test”[tiab] OR “fecal occult blood testing”[tiab] OR “colon cancer screening” OR “colorectal cancer screening” OR “Mammography”[MeSH] OR mammography OR mammogram OR “Breast Neoplasms/diagnosis”[MeSH] OR “breast cancer screening” OR “Human Papillomavirus DNA Tests”[MeSH] OR “hpv test” OR “hpv testing” OR “Papanicolaou Test”[MeSH] OR “pap smear” OR “pap test” OR “Uterine Cervical Neoplasms/diagnosis”[MeSH] OR “Cervical Intraepithelial Neoplasia/diagnosis”[MeSH] OR “cervical cancer screening” OR “Lung Neoplasms/diagnosis”[MeSH] OR “Tomography, X-Ray Computed”[MeSH] OR “Tomography, Emission-Computed”[MeSH] OR “low dose CT” OR “low dose computed tomography” OR “low dose computed tomographic” OR “lung cancer screening”

Selection and Quality Assessment of Studies

Citations were independently assessed by 2 investigators (CU and JL). Discrepancies were resolved by consensus and, when needed, with input from a third investigator (PSL). For studies requiring further clarification, the corresponding authors were contacted. The quality of the studies was evaluated using the Cochrane risk of bias tool. Extracted information included the following: (1) characteristics of the participants (including age and sex); (2) type of intervention (including test, duration of follow-up, and number of text messages sent); (3) outcome measure (including screening rates, odds ratio [OR], hazard ratio [HR], and confidence interval [CI] for unadjusted and adjusted results); (4) secondary endpoints; and (5) study limitations.

Results

Study Selection

Our search yielded 2238 citations. After removing duplicates and screening abstracts, 31 papers underwent full review, 22 of which did not meet our prespecified criteria. A total of nine papers were included in the systematic review (Figure 1). These studies were based in England, Spain, Malaysia, Israel, and the United States. Five studies analyzed rates of breast cancer screening by mammography, one analyzed rates of cervical cancer screening by Pap test, and three studies measured rates of colorectal cancer screening (two by FOBT alone and one by a combination of colonoscopy, flexible sigmoidoscopy, or FOBT). No studies were found for lung cancer screening. Together, the included studies enrolled 77,099 participants, most of whom were women aged between 20 and 75 years (Tables 1 and 2). The length of follow-up from intervention delivery to outcome measurement ranged from 2 days to 6 months. Seven studies used a single or set of text messages delivered in 1 day [15-21]. The remaining trials delivered text messages over differing periods: one delivered several text messages over 7 days, and the other delivered individual texts at 1-month intervals until screening occurred (up to three texts maximum; [22,23]). Absolute screening rates were reported in all studies, and we compared the rates in intervention and control groups using the chi-square test (Table 3). Six studies also reported relative risk estimates using OR or HR (Table 4; [17-21,23]).

Figure 1. Flowchart of study selection process.

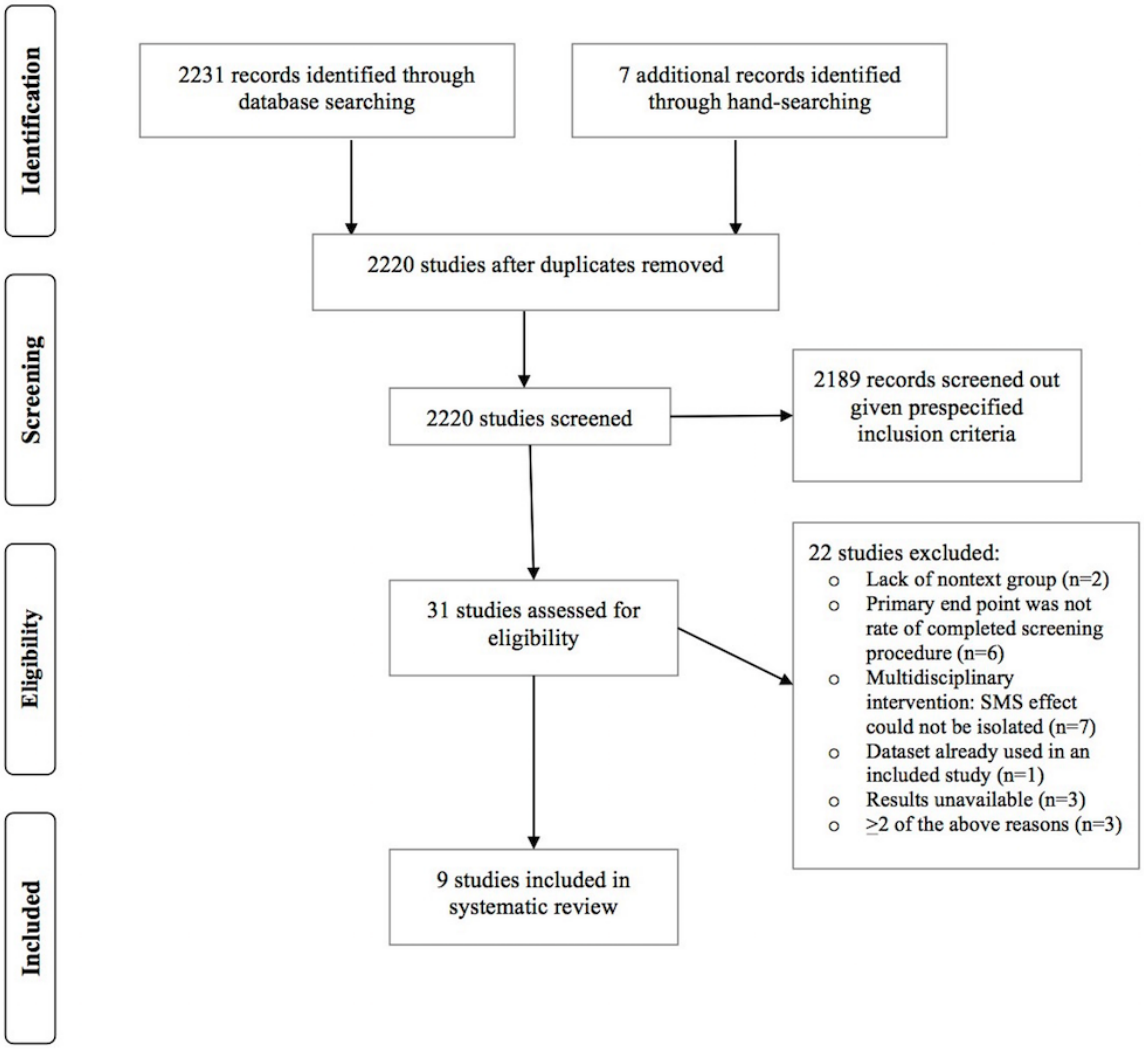


Table 1. Characteristics of studies included in the review of effect of text messaging interventions on cancer screening.

Author	Year	Cancer type	Screening test	Country	Study design	Publication type
Arcas [15]	2014	Breast	Mammogram	Spain	Randomized	Journal article
Icheku [16]	2015	Breast	Mammogram	England	Nonrandomized	Journal article
Kerrison [17]	2015	Breast	Mammogram	England	Randomized	Journal article
Lee [22]	2016	Breast	Mammogram	United States (Minnesota)	Randomized	Conference abstract
Vidal [18]	2014	Breast	Mammogram	Spain	Nonrandomized	Journal article
Abdul Rashid [19]	2013	Cervical	Pap test	Malaysia	Randomized	Journal article
Hagoel [20]	2016	Colorectal	FOBT	Israel	Randomized	Journal article
Hirst [21]	2016	Colorectal	FOBT	England	Randomized	Conference abstract
Muller [23]	2016	Colorectal	FOBT, flexible sigmoidoscopy, colonoscopy	United States (Alaska)	Randomized	Journal article

Table 2. Additional characteristics of studies included in the review of effect of text messaging interventions on cancer screening

Author	Screening test	N (text/control)	Age range	Intervention vs control groups	Follow-up
Arcas [15]	Mammogram	703 (470/233)	50-69	1 text message plus letter vs letter only	2 mo
Icheku [16]	Mammogram	2004 (552/1452)	50-70	1 text message plus letter vs letter only	1 wk
Kerrison [17]	Mammogram	2240 (1122/1118)	47-53	1 text message vs no reminder	2 d
Lee [22]	Mammogram	120 (60/60)	>40	Individualized text messages sent over 7 days vs informational brochure	6 mo
Vidal [18]	Mammogram	12,786 (3719/9067)	50-69	1 text message plus letter vs letter only	4.5 mo
Abdul Rashid [19]	Pap test	500 (250/250)	20-65	1 text message vs letter (two nonpertinent interventions were excluded from this review)	2 mo
Hagoel [20]	FOBT	48,091 (38,489/9602)	50-74	1 of 4 types of text message (I[a], I+SC[b], NI[c], NI+SC) plus letter vs letter only	6 mo
Hirst [21]	FOBT	8269 (4134/4135)	60-74	1 text message if no screening occurred at 8 weeks vs usual care	4.5 mo
Muller [23]	FOBT, flexible sigmoidoscopy, colonoscopy	2386 (1193/1193)	40-75	Individual texts sent 1 month apart until screening occurred (3 texts maximum) vs usual care	6 mo

[a]I: interrogative.

[b]SC: social context.

[c]NI: noninterrogative.

Table 3. Absolute screening rates in text messaging versus control groups.

Author	Screening test	N (text/control)	Screening rate in text group, %	Screening rate in control group, %	Absolute increase in screening with text intervention, % (*P* value[a])
Arcas [15]	Mammogram	703 (470/233)	81.3 (353/434)	76.8 (159/207)	+4.5 (*P*=.18)
Icheku [16]	Mammogram	2004 (552/1452)	68.1 (376/552)	60.47 (878/1452)	+7.6 (*P*=.002)
Kerrison [17]	Mammogram	2240 (1122/1118)	64.35 (722/1122)	59.12 (661/1118)	+5.3 (*P*=.01)
Lee [22]	Mammogram	120 (60/60)	40 (24/60)	25 (15/60)	+15.0 (*P*=.08)
Vidal [18]	Mammogram	12,786 (3719/9067)	74.91 (2786/3719)	65.00 (5894/9067)	+9.9 (*P*<.001)
Abdul Rashid [19]	Pap test	500 (250/250)	32.9 (54/164)	23.9 (47/197)	+9.1 (*P*=.05)
Hagoel [20]	FOBT	48,091 (38,489/9602)	9.78 (942/9631)	8.44 (817/9602)	+1.2 (*P*<.001)
Hirst [21]	FOBT	8269 (4134/4135)	40.49 (1674/4134)	39.85 (1648/4135)	+0.6 (*P*=.55)
Muller [23]	FOBT, flexible sigmoidoscopy, colonoscopy	2386 (1193/1193)	15.17 (181/1193)	11.90 (142/1193)	+3.3 (*P*=.02)

[a]Two-tailed *P* values based on chi-square test calculations using raw data from each study. Differences between reported (shown in text) and calculated (shown in table) *P* values are explained by differences in testing assumptions.

Table 4. Relative screening rates in text messaging versus control groups.

Author	Screening test	N (text/control)	Unadjusted OR[a]/HR[b] (95% CI)	Adjusted OR/HR (95% CI)	Adjusted variables
Kerrison [17]	Mammogram	2240 (1122/1118)	1.26 (1.05-1.48)	1.25 (1.05-1.48)	Age, socioeconomic status
Vidal [18]	Mammogram	12,786 (3719/9067)	1.63 (1.49-1.78)	1.56 (1.43-1.70)	Age
Abdul Rashid [19]	Pap test	500 (250/250)	1.20 (0.76-1.87)	-	-
Hagoel [20]	FOBT	48,091 (38,489/9602)	-	I[c]: 1.17 (1.06-1.29) I+SC[d]: 1.24 (1.12-1.36) NI[e]: 1.09 (0.99-1.21) NI+SC: 1.14 (1.04-1.26)	Type of text message, age, gender, socioeconomic status
Hirst [21]	FOBT	8269 (4134/4135)	1.04 (0.95-1.13)	-	-
Muller [23]	FOBT, flexible sigmoidoscopy, colonoscopy	2386 (1193/1193)	1.30 (1.04-1.62)	-	-

[a]OR: odds ratio.

[b]HR: hazards ratio.

[c]I: interrogative.

[d]SC: social context

[e]NI: noninterrogative.

Results and Quality of Individual Studies

Breast Cancer Studies

Arcas et al [15] assessed the impact of a single text message reminder on mammogram uptake in a single-blinded randomized control trial (RCT) over a 2-month period. The study found that women who received the SMS intervention were 4.5% more likely to undergo breast cancer screening (81.3%, 353/434, vs 76.8%, 159/207; *P*=.21). In subgroup analyses, those in the low secondary education level who received text messaging reported significantly higher screening, whereas results by age group were inconsistent. Women who were enrolled in the study were 3 times more likely to have participated in previous screening than women who were not enrolled (75.1%, 528/703, vs 24.55%, 488/1988), which may limit the generalizability of the findings. The high baseline screening rate may have also reduced the ability to detect a significant effect from the intervention.

Icheku and Arowobusoye [16] assessed how a single SMS sent 1 week before screening affected mammogram uptake. This non-RCT found that women who received text message reminders underwent mammograms 7.6% more than controls (68.1%, 376/552, vs 60.47%, 878/1452). The study determined the intervention and control groups based on whether individuals had a valid mobile number. Additionally, baseline

socioeconomic characteristics of the two groups were not reported. The definition of screening was stringent and excluded any mammograms that occurred earlier or later than the scheduled appointment. Although the authors only reported descriptive statistics, we show the results of formal statistical testing in Table 3.

Kerrison et al [17] studied the effect of a single text message sent 2 days before a scheduled breast cancer screening appointment and measured the outcome using an electronic system. This single-blinded RCT found that women in the SMS group had a 5.3% absolute increase in mammogram uptake compared with controls (64.35%, 722/1122, vs 59.12%, 661/1118) in the intention-to-treat (ITT) analysis, corresponding to a 26% increase in the relative screening rate (OR 1.26, 95% CI 1.05-1.48). Of note, only about 40.64% (456/1122) of the participants in either group had a valid mobile phone number, and the per-protocol analysis of individuals with valid mobile numbers showed an 11.9% difference in screening uptake (71.7%, 327/456, vs 59.8%, 260/435). In multivariable logistic regression analysis, individuals residing in the most socioeconomically deprived areas were half as likely to attend screening as those from the least deprived areas (OR 0.53, 95% CI 0.35-0.80). This study was conducted at a single site and was limited to women who were screening-naïve, and therefore, it may not be generalizable to all women.

Lee et al [22] tested the impact of a 7-day course of individualized text messages on breast cancer screening rates. This single-blinded RCT found that after 6 months, there was a 15% higher absolute screening rate in participants who received the SMS intervention compared with those who received an informational brochure on breast cancer screening (40%, 24/60, vs 25%, 15/60). This difference had a reported $P<.05$ using a one-tailed test, whereas our two-tailed test showed $P=.08$ (Table 3). This trial had the smallest sample size of all studies included in this review. Additionally, text messages were individualized and culturally tailored for Korean-American women, which may limit the generalizability of the findings.

Vidal et al [18] assessed the effect of a single SMS on mammography rates after a follow-up of 4.5 months. Outcome was determined using a program database. Similar to the study by Icheku and Arowobusoye, this was also a non-RCT that allocated women to the intervention versus control group based on availability of mobile phone numbers. The authors found that text message recipients had a 9.9% higher absolute screening rate compared with controls (74.91%, 2786/3719, vs 65.00%, 5894/9067), corresponding to a 63% higher relative screening in the intervention group (OR 1.63, 95% CI 1.49-1.78). The odds were 56% higher after adjusting for age (OR 1.56, 95% CI 1.43-1.70). In a stratified analysis based on previous screening behavior and geography, the effect of SMS reminders was greatest among women who had not undergone previous screening and were living in sparsely populated areas where postal service is unreliable. A limitation of the nonrandomized study design is that women with and without registered mobile numbers may have baseline differences in health care access and health literacy.

Cervical Cancer Studies

Abdul Rashid et al [19] assessed the effect of a text message intervention on Pap testing after a 2-month follow-up in a single-blinded RCT. In the per-protocol analysis, those who received SMS had a 9.1% increase in absolute screening rate compared with those who received a letter containing the same information (32.9%, 54/164, vs 23.9%, 47/197; $P=.05$). In the ITT analysis, there was a 20% higher relative screening rate (OR 1.20, 95% CI 0.76-1.87). The study only included women who had a previously normal Pap and therefore may not be generalizable to women who are screening-naïve.

Colorectal Cancer Studies

Hagoel et al [20] assessed how the language and content of SMS reminders affected FOBT completion. Participants were randomized to one of five groups. The four intervention groups received one of four types of text reminders: messages were either framed as a question (interrogative) or a statement (noninterrogative) and additionally either included or excluded a reference to peer screening behavior (social context). Controls did not receive a text intervention. Outcomes were assessed after 6 months using a national database. The authors found that 9.78% (942/9631) of patients who received any text intervention completed FOBT, compared with 8.44% (817/9602) of those who did not receive a text message. This corresponded to a 9% to 24% higher odds of screening in the intervention groups, which were statistically significant for all groups except the one that received noninterrogative messages without social context. In multivariable logistic regression analysis, participants who were older than 60 years (OR 1.13, 95% CI 1.06-1.20), female (OR 1.21, 95% CI 1.13-1.28), and of low socioeconomic status (OR 1.19, 95% CI 1.10-1.30) were more likely to undergo screening. This study excluded patients who were screening-naïve, which limits its generalizability. In addition, participants received an invitation to order or pick up a FOBT kit rather than the kit itself, which may present an additional barrier to screening.

Hirst et al [21] also assessed colorectal cancer screening using mailed FOBT kits. The intervention group received a single SMS reminder if the FOBT kit was not returned after 8 weeks from the initial invitation. The ITT analysis at 18 weeks showed that text messaging reminders did not significantly increase the number of individuals screened, with FOBT kit return rates of 40.49% (1674/4134) in the intervention group and 39.85% (1648/4135) in the control group (OR 1.04, 95% CI 0.95-1.13). However, mobile phone numbers were available for only 49.44% (4089/8269) of the study participants in either group, even though all participants were included in the analysis. Of those who actually received the SMS reminder, 18.42% (189/1026) participated in screening. Additionally, a subgroup analysis of the first-time screening group found a 5.6% higher absolute screening rate in the intervention group compared with controls (34.9%, 282/809, vs 40.5%, 297/733; OR 1.31, 95% CI 1.00-1.71). No difference in uptake was observed by sex, index of deprivation, or age.

Finally, Muller et al [23] examined text messaging and colorectal cancer screening in Alaskan Natives. This single-blinded RCT sent individuals up to three text messages

at 1-month intervals and assessed the outcome from electronic health records after 6 months of follow-up. The study found a 3.3% absolute increase in colorectal cancer screening, using colonoscopy, sigmoidoscopy, or FOBT. This corresponded to a 30% increase in the relative screening rate (HR 1.30; 95% CI 1.04-1.62). Notably, higher screening rates were only observed in women who received the intervention, and no statistically significant change was observed in men. The single site and specific population of this study may limit its generalizability.

A common limitation of the included studies was a relatively short duration of follow-up, which may have resulted in an underestimation of the text messaging effect, especially if there were substantial delays in the scheduling of screening tests. Overall, potential biases in study design and the focus on specific and often homogeneous populations for many of the studies may limit the generalizability of the results. Two studies were published as abstracts and may not have undergone a similar peer review process as the full manuscripts [21,22]. Two other studies were nonrandomized trials that assigned individuals to intervention or control groups based on the availability of a mobile phone number, which raises the concern that the groups may be different in other characteristics relevant to screening [16,18]. None of the studies formally assessed health literacy. Finally, with the exception of the Hagoel et al study [20], none of the included trials used a theoretical framework in the text messaging interventions.

Discussion

Principal Findings

In this systematic review of text messaging interventions and evidence-based cancer screening, we found that SMS can moderately increase screening rates for breast and cervical cancers and may improve colorectal cancer screening to a lesser degree. Across all studies, text messaging interventions led to increases in absolute screening rates of 0.6% to 15% and relative screening rates of 4% to 63%. Given the small number of studies and heterogeneity in design, we did not attempt to quantify the overall effect of SMS on cancer screening using meta-analysis.

Of the six studies that examined text messaging in breast and cervical cancer screening, increases in absolute screening rates ranged from 4.5% to 15%, and the three studies that reported relative screening rates found improvements of 20% to 63%. Although the smallest reported change in absolute screening rate was not statistically significant (Arcas et al, 4.5%), both the overall direction and the magnitude of absolute effect for SMS seem consistent for breast and cervical cancers. However, as only one cervical cancer screening study met our inclusion criteria, confirmatory investigations are clearly needed.

For colorectal cancer screening, the three included studies found much smaller effects on absolute screening rates, with increases ranging from 0.6% to 3.3%. Hirst et al [21] found a 0.6% difference in absolute screening and a nonsignificant 4% increase in relative screening with FOBT, but this almost certainly underestimates the true impact of SMS because only 24.81% (1026/4134) of the intervention group actually received a text message reminder. Hagoel et al [20] conducted the largest study included in this review, and the results suggest that how the message is phrased has important implications for how it is received: messages posed as a question were more effective than those phrased as a statement, and the most effective messages were questions accompanied by social context. Although the 1.2% increase in absolute screening rate is small, it should be noted that participants in the study had to first order and then complete the FOBT kit, and it has been shown that directly mailing kits to patients significantly improves screening rates [24]. The study of Alaskan Natives by Muller et al [23] was the only colorectal cancer screening study to include colonoscopy and flexible sigmoidoscopy as well as FOBT, and the 3.3% increase in absolute screening rate is more impressive after factoring in that every member of the health care system already receives telephone and letter reminders as well as in-office physician referrals for screening. Whether SMS interventions are more or less effective for colonoscopy than FOBT is an important question, because colonoscopy has become the predominant screening test in the United States and remains the requisite follow-up test for all abnormal FOBTs.

There are at least two reasons why the effectiveness of SMS may differ for colorectal cancer screening and breast or cervical cancer screening. The first may be a difference in the complexity or acceptability of the screening test. Although mammography and the Pap test simply require patients to attend a clinic appointment, patients must be willing to collect one or more stool specimens to complete a FOBT. Colonoscopy is even more complicated, as it requires patients to restrict their diet for several days, drink a large volume of fluids for bowel preparation, and finally undergo an invasive examination. A few trials have studied the impact of text messaging on the quality of bowel preparation, but none has measured the effect of bowel preparation instructions via SMS on screening rates themselves [25,26]. If the problem lies with a higher barrier to overcome for colorectal versus breast and cervical cancer screening, then the strategies used in some included studies may help to maximize the effect of SMS. Providing culturally tailored messages, as both Lee et al and Muller et al did, may be more impactful than generic messages. Similarly, Hagoel et al demonstrated the additional value of using interrogative and contextualized messages. The language of text messaging will continue to be an important area of research, especially for complex screening procedures such as colonoscopy that may require instructions as well as reminders.

A second reason may be the gender difference in the study populations, as the cervical and breast cancer study participants were all women. Indeed, perhaps the most striking finding from the Muller et al study was that only women benefited from the intervention. Hagoel et al also found that women who received the text messages had a statistically significant 21% greater odds of FOBT uptake than men (OR 1.21; 95% CI 1.13-1.28 [20]). However, Hirst et al found no significant difference in FOBT uptake between men and women. It is unclear whether true gender differences exist in how text messages are received or how it motivates behavioral change. Additional studies on the influence of SMS on colorectal cancer screening will help to better quantify its effect for each of the various screening

tests, as well as any differences that may exist with respect to gender.

There are many clear advantages to using text messaging to increase cancer screening. According to the United Nations, in 2013, nearly 6 billion of the world's 7 billion people possessed mobile phones, compared with the 4.5 billion who had access to toilets or latrines [27]. These statistics have clear implications for the potential of using mobile technology to reach underserved communities, including those living in rural and underdeveloped areas [28,29]. A 2016 survey by the National Center for Health Statistics found that 49% (18,074/36,885) of all US adults live in households with only wireless telephones, and those living in (63%, 23,274/36,885) or near (54%, 19,918/36,885) poverty were more likely to live in wireless-only households than those with higher income (48.2%, 17,779/36,885) [30]. In addition, all minority racial groups were more likely to live in wireless-only households, with the greatest difference being between whites and those of Hispanic or Latino descent (45.0%, 16,598/36,885 vs 63.7%, 23,496/36,885). For populations with low health literacy, the short and simple format of text messages may also improve comprehension and reduce a formidable barrier to access. Although the studies included in this review did not specifically report results in participants with low health literacy, text messaging has been shown to improve medication adherence in this population [31].

Beyond the prevalence of mobile phone usage and its reach in low-income settings, another benefit of utilizing text messages for patient outreach is the ability to directly send discreet information in real time, which can then be accessed at the patient's convenience. This delivery method does not require patients to be available at a particular time or place to receive information, as a telephone call would. Texting also does not rely on an updated or even stable address, as a letter would. However, SMS does require a stable phone number and also may raise concerns about confidentiality if multiple family members share a mobile phone. The ability of SMS to connect with remote and socioeconomically disadvantaged populations is supported by several of the included studies. The mammography study by Vidal et al [18] showed that women who lived in areas with less reliable postal service benefited more from the SMS intervention. Hagoel et al found higher rates of FOBT uptake in patients of low versus high socioeconomic status (OR 1.19; 95% CI 1.10-1.30 [20]). Finally, Kerrison et al [17] found that text messaging was particularly effective at increasing screening rates for patients from the lowest socioeconomic category, who saw a 13.6% absolute increase in screening from the control to intervention group. Vidal et al also demonstrated that adding text messages to an invitation letter was cost-effective for breast cancer screening, and similar conclusions have been reached about the cost-effectiveness of text messaging to improve other health care outcomes [32,33].

Text messaging has been found to be effective for promoting a variety of beneficial health behaviors. Whittaker et al [34] conducted a meta-analysis of text messaging interventions for smoking cessation and found a 67% higher likelihood of successful cessation at 6 months compared with controls. Arambepola et al [35] analyzed 13 trials of diabetic patients and found a statistically significant 0.53% decrease in hemoglobin A1c as well as a nonsignificant 0.25 kg/m^2 reduction in body mass index in patients who received an SMS intervention. Two meta-analyses have shown that text messaging also improves attendance at primary care and hospital outpatient clinic appointments by 14% to 48% compared with controls [36,37]. Together, these studies consistently convey text messaging's efficacy as an intervention for health behavioral change, which is further supported by our analysis of cancer screening.

Some studies did not meet our inclusion and exclusion criteria but merit mentioning. Oakley-Girvan et al [38] found that text messaging decreases time to follow-up after abnormal mammograms in a Spanish-speaking Latina population, suggesting that text messaging may be useful to increase rates of both regular screening and follow-up for management of abnormal results. This study also suggests the utility of text messaging for communicating in patients' preferred language, potentially overcoming communication barriers between physicians and patients. For colorectal cancer, a multifaceted intervention that included text messaging in addition to a mailed letter, instructions on using the stool test, paid return envelopes, and phone messages increased rates of FOBT screening for colorectal cancer from 37% (84/225) to 82% (185/225) [39]. However, we excluded this study because the effect of text messaging could not be isolated.

Table 5. Ongoing studies evaluating the impact of text messaging interventions on cancer screening rates.

Author	Cancer type	Country	Study title
Huf S	Cervical cancer	England	Can Text Reminders Improve Uptake of Cervical Screening?
Palafox N	Cervical cancer	United States (Hawaii)	Pilot Project 1: Reducing Cervical Cancer Screening Health Disparities Among Pacific Islanders Living in Guam (GU) and Hawaii (HI)
Baker D	Colorectal cancer	United States (Illinois)	Improving Rates of Repeat Colorectal Cancer Screening
Baker D	Colorectal cancer	United States (Illinois)	Improving Rates of Colorectal Cancer Screening Among Never Screened Patients
Ma G	Colorectal cancer	United States (Pennsylvania)	A Multilevel CBPR[a] Intervention to Improve Colorectal Cancer Screening in Underserved Vietnamese Americans
Smith J	Colorectal cancer	United States (Michigan, New Mexico, Washington)	Evaluation of an Intervention to Increase Colorectal Cancer Screening in Primary Care Clinics

[a]CPBR: Community-based participatory research.

Notably, our search did not find any studies that assessed the effect of text messaging on screening for lung cancer, a leading cause of cancer death in the United States that affects both men and women. This is not entirely surprising, given the relatively new guidelines that support screening for the subset of the population with a substantial smoking history. However, it highlights a knowledge gap in the relationship between an effective behavioral intervention and a screening test that has been proven to save lives in heavy smokers, and further studies on the subject are warranted.

Several trials are currently under way that may further quantify the relationship between text messaging and cancer screening (Table 5). In addition, as seven of the nine studies we found were conducted in Europe and Asia, more domestic research is needed to ensure that the results are applicable to the US population.

Limitations

Although we developed a comprehensive search strategy to query the relevant literature, it is possible that some studies were not identified. There was a paucity of data that met our inclusion criteria, and we found only one study on cervical cancer and three on colorectal cancer. In addition, there was substantial heterogeneity in study design, content of the text message intervention, as well as time to follow-up in the included studies. For these reasons, we did not perform a meta-analysis.

Conclusions

A systematic review of the literature suggests that text messaging interventions can increase screening rates for breast and cervical cancers to a moderate degree and for colorectal cancer to a smaller degree. Implementation of text messaging for cancer screening may be an effective method to increase screening rates and thereby decrease cancer-related mortality, even in resource-poor and non-English-speaking populations. However, additional research is needed to determine whether these results apply to all cancer screening tests at the population level.

Acknowledgments

We thank Joey Nicholson and Aileen McCrillis for assistance with the search strategy and Laura Wyatt for feedback on an early draft. This work is supported, in part, by the National Institutes of Health (National Institute on Minority Health and Health Disparities award number P60MD000538 and National Center for Advancing Translational Sciences award number UL1TR001445) and with resources and the use of facilities at the VA New York Harbor Manhattan Medical Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Conflicts of Interest

None declared.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015 Mar;65(2):87-108 [FREE Full text] [doi: 10.3322/caac.21262] [Medline: 25651787]
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61(2):69-90 [FREE Full text] [doi: 10.3322/caac.20107] [Medline: 21296855]
3. CDC. 2012. CDC - Global Cancer Statistics Internet URL: https://www.cdc.gov/cancer/international/statistics.htm [accessed 2017-04-14] [WebCite Cache ID 6pjCTwHye]

4. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends--an update. Cancer Epidemiol Biomarkers Prev 2016 Jan;25(1):16-27 [FREE Full text] [doi: 10.1158/1055-9965.EPI-15-0578] [Medline: 26667886]
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66(1):7-30 [FREE Full text] [doi: 10.3322/caac.21332] [Medline: 26742998]
6. U.S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2008 Nov 04;149(9):627-637. [Medline: 18838716]
7. Siu AL, U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2016 Feb 16;164(4):279-296. [doi: 10.7326/M15-2886] [Medline: 26757170]
8. Katki HA, Kinney WK, Fetterman B, Lorey T, Poitras NE, Cheung L, et al. Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: a population-based study in routine clinical practice. Lancet Oncol 2011 Jul;12(7):663-672 [FREE Full text] [doi: 10.1016/S1470-2045(11)70145-0] [Medline: 21684207]
9. US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. J Am Med Assoc 2016 Jun 21;315(23):2564-2575. [doi: 10.1001/jama.2016.5989] [Medline: 27304597]
10. Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. Ann Intern Med 2013 Sep 17;159(6):411-420. [doi: 10.7326/0003-4819-159-6-201309170-00690] [Medline: 23897166]
11. Smith A. Pewinternet. U.S. Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [WebCite Cache ID 6pjClYhAb]
12. ITU. 2016. ICT Facts and Figures 2016 URL: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf [accessed 2017-04-14] [WebCite Cache ID 6pjEvt0uj]
13. Kratzke C, Wilson S, Vilchis H. Reaching rural women: breast cancer prevention information seeking behaviors and interest in Internet, cell phone, and text use. J Community Health 2013 Feb;38(1):54-61. [doi: 10.1007/s10900-012-9579-3] [Medline: 22706805]
14. Dang CM, Estrada S, Bresee C, Phillips EH. Exploring potential use of internet, E-mail, and instant text messaging to promote breast health and mammogram use among immigrant Hispanic women in Los Angeles County. Am Surg 2013 Oct;79(10):997-1000. [Medline: 24160786]
15. Arcas MM, Buron A, Ramis O, Esturi M, Hernández C, Macià F. [Can a mobile phone short message increase participation in breast cancer screening programmes?]. Rev Calid Asist 2014;29(4):188-196. [doi: 10.1016/j.cali.2014.02.003] [Medline: 25002239]
16. Icheku V, Arowobusoye N. Evaluation of a service intervention to improve uptake of breast cancer screening in a London borough with many hard to reach communities. Univers J Public Health 2015 Mar;3(2):97-102. [doi: 10.13189/ujph.2015.030207]
17. Kerrison RS, Shukla H, Cunningham D, Oyebode O, Friedman E. Text-message reminders increase uptake of routine breast screening appointments: a randomised controlled trial in a hard-to-reach population. Br J Cancer 2015 Mar 17;112(6):1005-1010 [FREE Full text] [doi: 10.1038/bjc.2015.36] [Medline: 25668008]
18. Vidal C, Garcia M, Benito L, Milà N, Binefa G, Moreno V. Use of text-message reminders to improve participation in a population-based breast cancer screening program. J Med Syst 2014 Sep;38(9):118. [doi: 10.1007/s10916-014-0118-x] [Medline: 25073694]
19. Abdul RR, 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] [Medline: 24289597]
20. Hagoel L, Neter E, Stein N, Rennert G. Harnessing the question-behavior effect to enhance colorectal cancer screening in an mHealth experiment. Am J Public Health 2016 Nov;106(11):1998-2004. [doi: 10.2105/AJPH.2016.303364] [Medline: 27631750]
21. Hirst Y, Skrobanski H, Kerrison R, Kobayashi LC, Counsell N, Djedovic N, et al. Text-message reminders in colorectal cancer screening: a non-clinical randomised controlled trial. The Lancet 2016 Nov;388:S55. [doi: 10.1016/S0140-6736(16)32291-7]
22. Lee H, Le C, Ghebre R, Yee D. Mobile phone multimedia messaging intervention for breast cancer screening. Abstract P3-08-03: Mobile phone multimedia messaging intervention for breast cancer screening. Cancer Res American Association for Cancer Research; 2016 Presented at: Proceedings of the Thirty-Eighth Annual CTRC-AACR San Antonio Breast Cancer Symposium; 2015 Dec 8-12; San Antonio, TX p. 3-8 URL: http://cancerres.aacrjournals.org/content/76/4_Supplement/P3-08-03 [doi: 10.1158/1538-7445.SABCS15-P3-08-03]
23. Muller CJ, Robinson RF, Smith JJ, Jernigan MA, Hiratsuka V, Dillard DA, et al. Text message reminders increased colorectal cancer screening in a randomized trial with Alaska Native and American Indian people. Cancer 2017 Apr 15;123(8):1382-1389. [doi: 10.1002/cncr.30499] [Medline: 28001304]

24. Church TR, Yeazel MW, Jones RM, Kochevar LK, Watt GD, Mongin SJ, et al. A randomized trial of direct mailing of fecal occult blood tests to increase colorectal cancer screening. J Natl Cancer Inst 2004 May 19;96(10):770-780. [Medline: 15150305]
25. Walter BM, Klare P, Neu B, Schmid RM, von DS. Development and testing of an automated 4-day text messaging guidance as an aid for improving colonoscopy preparation. JMIR Mhealth Uhealth 2016 Jun 21;4(2):e75 [FREE Full text] [doi: 10.2196/mhealth.5289] [Medline: 27329204]
26. Lee YJ, Kim ES, Choi JH, Lee KI, Park KS, Cho KB, et al. Impact of reinforced education by telephone and short message service on the quality of bowel preparation: a randomized controlled study. Endoscopy 2015 Nov;47(11):1018-1027. [doi: 10.1055/s-0034-1392406] [Medline: 26182387]
27. UN. United Nations-DPI/NMD - UN News Service Section; 2013 URL: http://www.un.org/apps/news/story.asp?NewsID=44452 [accessed 2017-04-14] [WebCite Cache ID 6pjD6uV2R]
28. Sankaranarayanan J, Sallach RE. Rural patients' access to mobile phones and willingness to receive mobile phone-based pharmacy and other health technology services: a pilot study. Telemed J E Health 2014 Feb;20(2):182-185. [doi: 10.1089/tmj.2013.0150] [Medline: 24205837]
29. DeSouza SI, Rashmi MR, Vasanthi AP, Joseph SM, Rodrigues R. Mobile phones: the next step towards healthcare delivery in rural India? PLoS One 2014;9(8):e104895 [FREE Full text] [doi: 10.1371/journal.pone.0104895] [Medline: 25133610]
30. Blumberg S, Luke JV. CDC. 2016. Wireless Substitutionarly Release of Estimates From the National Health Interview Survey, January-June 2016 URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201612.pdf [WebCite Cache ID 6pjFZDjWw]
31. Olives TD, Patel RG, Thompson HM, Joing S, Miner JR. Seventy-two-hour antibiotic retrieval from the ED: a randomized controlled trial of discharge instructional modality. Am J Emerg Med 2016 Jun;34(6):999-1005. [doi: 10.1016/j.ajem.2016.02.046] [Medline: 26969079]
32. Guerriero C, Cairns J, Roberts I, Rodgers A, Whittaker R, Free C. The cost-effectiveness of smoking cessation support delivered by mobile phone text messaging: Txt2stop. Eur J Health Econ 2013 Oct;14(5):789-797 [FREE Full text] [doi: 10.1007/s10198-012-0424-5] [Medline: 22961230]
33. Chen Y, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess 2012;16(38):1-205, iii [FREE Full text] [doi: 10.3310/hta16380] [Medline: 23046909]
34. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
35. Arambepola C, Ricci-Cabello I, Manikavasagam P, Roberts N, French DP, Farmer A. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. J Med Internet Res 2016;18(4):e86 [FREE Full text] [doi: 10.2196/jmir.5425] [Medline: 27095386]
36. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]
37. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013;12:CD007458. [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
38. Oakley-Girvan I, Londono C, Canchola A, Watkins DS. Text messaging may improve abnormal mammogram follow-up in Latinas. Oncol Nurs Forum 2016 Jan;43(1):36-43. [doi: 10.1188/16.ONF.36-43] [Medline: 26679443]
39. Baker DW, Brown T, Buchanan DR, Weil J, Balsley K, Ranalli L, et al. Comparative effectiveness of a multifaceted intervention to improve adherence to annual colorectal cancer screening in community health centers: a randomized clinical trial. JAMA Intern Med 2014 Aug;174(8):1235-1241. [doi: 10.1001/jamainternmed.2014.2352] [Medline: 24934845]

Abbreviations

CT: computed tomography
FOBT: fecal occult blood test
HPV: human papillomavirus
HR: hazard ratio
ITT: intention-to-treat
I: interrogative
NI: noninterrogative
OR: odds ratio
RCT: randomized controlled trial
SC: social context
SMS: short message service

USPSTF: US Preventive Services Task Force

Edited by G Eysenbach; submitted 20.04.17; peer-reviewed by C Muller, L Carter-Harris, R Robinson; comments to author 28.06.17; revised version received 11.07.17; accepted 12.07.17; published 24.08.17.

Please cite as:
Uy C, Lopez J, Trinh-Shevrin C, Kwon SC, Sherman SE, Liang PS
Text Messaging Interventions on Cancer Screening Rates: A Systematic Review
J Med Internet Res 2017;19(8):e296
URL: http://www.jmir.org/2017/8/e296/
doi:10.2196/jmir.7893
PMID:28838885

Original Paper

Web-Based Mindfulness Interventions for People With Physical Health Conditions: Systematic Review

Kirsti I Toivonen[1], BA (Hons); Kristin Zernicke[2], PhD; Linda E Carlson[3,4], PhD

[1]Department of Psychology, University of Calgary, Calgary, AB, Canada

[2]Shared Mental Health Care, Alberta Health Services, Sheldon M. Shumir Health Centre, Calgary, AB, Canada

[3]Department of Psychosocial Resources, Holy Cross Site, Tom Baker Cancer Centre, Calgary, AB, Canada

[4]Department of Oncology, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Linda E Carlson, PhD
Department of Psychosocial Resources
Holy Cross Site
Tom Baker Cancer Centre
Holy Cross Centre, 2nd Fl.
2202 2nd St SW
Calgary, AB, T2S 3C1
Canada
Phone: 1 403 355 3209
Fax: 1 403 355 3206
Email: l.carlson@ucalgary.ca

Abstract

Background: Mindfulness-based interventions (MBIs) are becoming increasingly popular for helping people with physical health conditions. Expanding from traditional face-to-face program delivery, there is growing interest in Web-based application of MBIs, though Web-based MBIs for people with physical health conditions specifically have not been thoroughly reviewed to date.

Objective: The objective of this paper was to review Web-based MBIs for people with physical health conditions and to examine all outcomes reported (eg, efficacy or effectiveness for physical changes or psychological changes; feasibility).

Methods: Databases PubMed, PsycINFO, Science Direct, CINAHL Plus, and Web of Science were searched. Full-text English papers that described any Web-based MBI, examining any outcome, for people with chronic physical health conditions were included. Randomized, nonrandomized, controlled, and uncontrolled trials were all included. Extracted data included intervention characteristics, population characteristics, outcomes, and quality indicators. Intervention characteristics (eg, synchronicity and guidance) were examined as potential factors related to study outcomes.

Results: Of 435 publications screened, 19 published papers describing 16 studies were included. They examined Web-based MBIs for people with cancer, chronic pain or fibromyalgia, irritable bowel syndrome (IBS), epilepsy, heart disease, tinnitus, and acquired brain injury. Overall, most studies reported positive effects of Web-based MBIs compared with usual care on a variety of outcomes including pain acceptance, coping measures, and depressive symptoms. There were mixed results regarding the effectiveness of Web-based MBIs compared with active control treatment conditions such as cognitive behavioral therapy. Condition-specific symptoms (eg, cancer-related fatigue and IBS symptoms) targeted by treatment had the largest effect size improvements following MBIs. Results are inconclusive regarding physical variables.

Conclusions: Preliminary evidence suggests that Web-based MBIs may be helpful in alleviating symptom burden that those with physical health conditions can experience, particularly when interventions are tailored for specific symptoms. There was no evidence of differences between synchronous versus asynchronous or facilitated versus self-directed Web-based MBIs. Future investigations of Web-based MBIs should evaluate the effects of program adherence, effects on mindfulness levels, and whether synchronous or asynchronous, or facilitated or self-directed interventions elicit greater improvements.

(J Med Internet Res 2017;19(8):e303) doi:10.2196/jmir.7487

KEYWORDS

Internet; mindfulness; review

Introduction

Mindfulness-Based Interventions

Given the increased public interest and research in both mindfulness-based interventions (MBIs) [1] and Internet-delivered therapies [2], this paper summarizes the research on Web-based MBIs for people with physical health conditions. Mindfulness practice involves moment-to-moment nonjudgmental awareness, applied by purposely attending to one's own thoughts and bodily sensations with attitudes of openness and acceptance [3,4]. Although meditation practice originates from centuries of Buddhist tradition, it is often incorporated into structured, secular MBIs that are applied to a variety of clinical populations [1]. Several systematic reviews and meta-analyses have examined the effects of face-to-face group MBIs in heterogeneous clinical populations [5-8] as well as specific medical populations, including patients with chronic pain [9], fibromyalgia [10], multiple sclerosis [11], vascular disease [12], human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) [13], and cancer [14-18]. These reviews have consistently described reduced anxiety, depressive symptoms, and distress; improved mood; and improved quality of life [5-7,11,12,14,16-18]. Theoretically, improvements in psychological well-being are driven by improved emotion regulation abilities, which in turn result in decreased rumination about the past, worry about the future, and experiential avoidance of difficult feelings (L Labelle, unpublished data, 2012). Ultimately, mindfulness practice can allow people to view their illness from a new perspective and result in the improvements that have been observed across a broad range of psychological and physical outcomes [8].

Web-Based Interventions

There has been a surge in Web-based delivery of therapeutic interventions because of factors such as increased acceptability of the Internet as a social tool and continuous improvement in computer hardware and software (particularly concerning ease of use, privacy protection, and facilitating communication) [19]. Web-based therapies delivered through real time such as instant messaging platforms, telephone, or videoconferencing are categorized as *synchronous*, whereas delayed delivery methods such as email or message boards are categorized as *asynchronous*. Several reasons underlie the appeal of both types of Web-based therapies, including (1) ease and speed of accessibility by reducing wait-list times, (2) convenience of 24-hour availability for individual schedules (particularly with asynchronous therapies), (3) the ability for participants to work at their own pace in the comfort of their own homes, (4) allowing anonymity, and (5) reduced cost [20,21]. For those with physical health conditions in particular, home delivery can make participation in an intervention possible when it otherwise would not have been. For example, on-site attendance can be a barrier for those who experience limitations in mobility and energy levels or those who may have lower immunity to contagious diseases because of treatments such as chemotherapy and are avoiding groups of people. Additionally, the flexibility of Web-based delivery can be appealing for people who are busy managing appointments and treatment.

Whereas there is theoretical rationale for the utility of Web-based delivery of various therapies, and there have been hundreds of randomized controlled trials (RCTs) evaluating Internet-delivered health interventions [2], it remains unclear as to what extent people with physical health conditions currently use Web-based interventions, or Web-based MBIs specifically. Further complicating the question of use are the numerous mindfulness-based mobile apps that are widely available but have not been studied [22]. Although the actual uptake of Web-based therapies in this population is unknown, results from a recent Web-based survey suggest that people may prefer Internet interventions to in-person therapy. Wahbeh and colleagues [23] surveyed 500 people with a high prevalence of posttraumatic (70.6%, 353/500) and depressive (76.2%, 381/500) symptoms from the United States regarding their preferred delivery format of a mindfulness meditation intervention. Most (71.2%, 356/500) expressed interest in a Web-based format, and Internet delivery was the first choice of format for the greatest proportion of participants (42.7%, 212/496), followed by individual (37.8%, 187/496) and group (19.6%, 97/496). As participants were members of the general public and not people suffering from any specific medical or psychiatric conditions, the generalizability of findings to these groups is likely to be high but not certain. Despite the appeal of Web-based therapies, some potential drawbacks necessitate consideration. First, Web-based delivery may offer less interpersonal interaction and social support than in-person delivery. Second, compliance may be more difficult to determine if those delivering the intervention cannot directly observe those receiving it. Third, particularly with asynchronous interventions, it could take longer for participants to receive feedback and have questions answered. Fourth, monitoring of participants for adverse reactions may be more difficult in the Web-based environment, when personal contact may be absent. Referrals to appropriate mental health or medical services for any identified adverse reactions or need for further individualized treatment may also be more difficult in the Web-based environment where providers may be very geographically distant from patients. Finally, ethical issues around responsibility for care may also arise [24]. Thus, the benefits and drawbacks of Web-based therapies as well as treatment preferences should be considered when devising treatment plans. Ultimately, if effective, Web-based MBIs can be a more practical option for many people with physical health conditions and may remove barriers that would have otherwise precluded participation in face-to-face interventions.

Prior Reviews of Web-Based MBIs

Given the popularity of Internet interventions and their potential for widespread application, it is important to evaluate their effectiveness as they are being developed and disseminated. A 2016 review and meta-analysis by Spijkerman et al [25] examined 15 RCTs of Web-based MBIs for improving mental

health in a heterogeneous sample, including studies of healthy adults, students, and employees (*n*=7); people with psychiatric disorders or symptoms (*n*=3); and people with chronic pain or other physical illnesses (*n*=5). Of the 15 interventions, nine were guided (including real-time group or individual sessions or individual email correspondence) and six were unguided. Only three interventions were delivered as a virtual classroom, whereas 12 were delivered through websites or mobile phone apps. Most studies (*n*=10) compared treatment groups with wait-list controls. Five studies compared a treatment group with an online discussion forum, psychoeducation, or behavioral activation. Interventions typically comprised weekly sessions and were 2 to 12 weeks in duration. Web-based MBIs resulted in small effect size improvements in depression (Hedge *g*=0.29), anxiety (*g*=0.22), well-being (*g*=0.23), and mindfulness (*g*=0.32), as well as moderate effect size decreases in stress (*g*=0.51). Exploratory subgroup analyses showed that Web-based MBIs with therapist guidance resulted in greater effect size improvements in stress and mindfulness than those without guidance. The authors concluded that Web-based MBIs could be beneficial for mental health outcomes, particularly stress [25].

This review expands upon the review by Spijkerman et al, focusing on chronic physical health conditions alone and looking at both mental and physical health outcomes. Populations with primary psychological disorders were excluded from this review to prevent redundancy with Spijkerman et al [25], which focused on mental health. Furthermore, a narrower focus on populations presenting primarily with physical health conditions was selected because examination of a relatively homogeneous population may lead to more cohesive results specific to these groups.

Methods

This review was conducted in accordance with the 2009 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for systematic reviews [26]. The literature search, paper screening (title and abstract, and full-text), and data extraction were undertaken by KT.

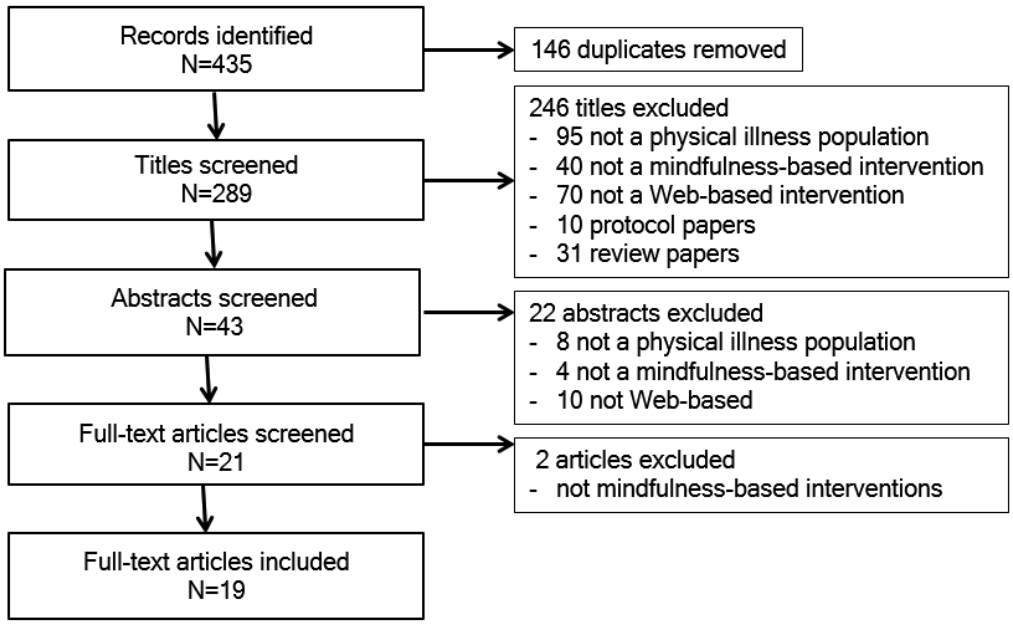
Eligibility Criteria

Inclusion criteria were studies examining individuals with chronic physical conditions that received a Web-based MBI. All types of physical health conditions were included. Studies of acceptance and commitment therapy (ACT) and dialectical behavior therapy were included as MBIs because they contain elements of mindfulness practice (eg, sustaining attention in the present and nonjudgmental observation of emotion), though they involve less formal mindfulness meditation training than other structured MBIs [27]. RCTs, non-RCTs, and uncontrolled trials were included; and all outcomes reported by studies were described. Studies that described qualitative research or program evaluation, described populations with psychiatric disorders, did not have full-text in English available, or described interventions that included a small component of mindfulness but did not emphasize it (eg, multiweek programs where mindfulness practice is addressed only in one or two sessions) were excluded.

Search Strategy

A literature search was conducted in PubMed, PsycINFO, Science Direct, CINAHL Plus with Full Text, and Web of Science in the fields all fields, all fields, keywords, abstract, and topic, respectively. All searches used the terms “online OR Internet AND mindfulness AND intervention” (see Multimedia Appendix 1 for an example of the search strategy for PubMed). All papers published before and during November 2016 were included. Reference lists of included studies were searched to identify additional studies. Titles and abstracts and full-texts were screened simultaneously, such that when a paper may have met inclusion criteria based on the title and abstract screen, the full-text was immediately screened. See Figure 1 for a flowchart of study selection and inclusion.

Figure 1. Study selection and inclusion flowchart.



Data Extraction

For each study, the following data were extracted: first author, year of publication, country, population characteristic (physical health condition, age, and sex), number of participants per group, study design (eg, randomized and nonrandomized), intervention characteristics (type of intervention, synchronous or asynchronous, delivery mode [eg, Web-based and videoconference], guided or unguided [where any correspondence with a therapist, including email, was considered guided], number of sessions, and duration), assessment times (pre, post, follow-up), all outcomes (see Table 1), and adherence and dropout (see Table 2). Note that reporting of adherence and dropout was inconsistent across studies—it is described in this review as it was reported in the original studies. Data were extracted by KT according to a predetermined form.

Methodological quality was assessed based on potential sources of bias outlined in the Cochrane Handbook for Systematic Reviews of Interventions [28]. Sources of bias that were assessed at the study level included random sequence generation (ie, whether participants were randomly assigned to groups), allocation concealment (ie, whether group assignment was unknown when participants were recruited), blinding of participants and personnel (ie, whether participants and people involved in the study were unaware of participant group assignment), and blinding of outcome assessment (ie, whether the individual rating the outcome measure was aware of group assignment). Sources of bias assessed at the outcome level were complete outcome data or intention-to-treat (ITT) analysis used (where outcome data were considered complete if ≥90% of those randomized were included), and whether all the outcomes that were described in the Methods section were reported. Note that studies were coded as yes when they met the criteria and no when they either did not meet criteria or when it was ambiguous as to whether they met the criteria.

Results

Overall, 19 published papers describing 16 studies of Web-based MBIs for physical health conditions were identified through the literature search and are subsequently described [29-47]. Three of the included papers describe long-term, follow-up results or secondary analyses that were separately published [32,37,46]. Thus, whereas all 19 papers are described in text where relevant, only the original 16 are presented in Tables 1-3 and described subsequently. Interventions with the greatest emphasis on mindfulness included Mindfulness-Based Stress Reduction (MBSR; n=1) [44], Mindfulness-Based Cancer Recovery (MBCR; n=1) [45], Mindfulness-Based Cognitive Therapy (MBCT; n=4) [30,41,42,47], Mindful Socioemotional Regulation (MSER; n=1) [33], Mindfulness-Based Chronic Pain Management (MBCPM; n=1) [34], or general mindfulness training (n=1) [35]. Interventions that included mindfulness as a component of a multimodal intervention were ACT (n=3) [29,31,43] and cognitive behavioral therapy (CBT) plus mindfulness (n=4) [36,38-40]. Most studies employed a randomized controlled design (n=13) and nearly all (n=15) had at least one comparison group. Seven studies had an active comparison: the same MBI delivered in person (n=2) [34,44], a control group receiving another Web-based therapy (n=4) [31,38,40,43], or a walking control group (n=1) [44]. The remaining studies had a psychoeducation control (n=2) [30,33], an online discussion forum control (n=4) [29,36,39,43], or wait-list or treatment as usual (n=6) [31,34,35,41,42,45]. Note that four studies had two comparison groups [31,34,43,44]. Most interventions included a form of guidance: either self-facilitated interventions including email correspondence with a therapist (n=8) [29,31,36,38-40,43,47] or therapist-facilitated sessions (n=5) [34,41,42,44,45]. Five studies were conducted with populations with chronic pain [29-31,34] or fibromyalgia [33] (which has been grouped with chronic pain in this review, as it is characterized by widespread chronic pain). Other populations examined included people with heart disease (n=1) [35], irritable bowel syndrome (IBS; n=4) [36,38-40], epilepsy (n=2) [41,42], tinnitus (n=1) [43], acquired brain injury (n=1) [44], and cancer (n=2) [45,47]. The largest number of studies were conducted in Sweden (n=7) [29,36,38-40,43,44], followed by the United States (n=3) [33,41,42], the Netherlands (n=3) [31,35,47], Canada (n=2) [34,45], and Ireland (n=1) [30]. See Table 1 for study characteristics.

Regarding primary outcomes, five studies assessed pain or related constructs (eg, pain acceptance and pain interference) [29-31,33,34], four studies assessed IBS symptoms or symptom severity [36,38-40], two assessed depressive symptoms [41,42], two assessed fatigue [44,47], one assessed exercise capacity [35], one assessed tinnitus distress severity [43], and one assessed feasibility [45]. Four studies [30,33,34,36] assessed more than one primary outcome, including outcomes such as affect, social measures, and health-related quality of life (HRQoL). Table 1 outlines a complete list of primary and secondary outcomes assessed per study. More than half (*n*=9) of the studies included follow-up assessments [29-31,36,38-40,42,43] that ranged from 3 months [36] to 12 months post intervention [39,43]. See Table 2 for summary of primary outcomes, including the mean number of sessions completed, whether primary outcomes improved over time and relative to control, and whether improvements were maintained at follow-up, where applicable. Note that if a study had both an active and wait-list control, Table 2 presents improvement over the wait-list control.

Table 1. Summary of study characteristics.

First author (year)	Population, country	Mean age in years (% F)[a]	Intervention type	Number of sessions, duration	Control condition	Outcomes[b]
Gardner-Nix (2008) [34]	Chronic pain, Canada	51-54 (75-88)	Synchronous videoconference MBCPM[c], guided (n=57)	10 sessions, 10 weeks	On-site program (n=99), wait-list (n=59)	*Health-related quality of life (physical and mental), pain catastrophizing, pain*
Buhrman (2013) [29]	Chronic pain, Sweden	49 (59)	Asynchronous Web-based ACT[d], guided (n=38)	7 sessions, 7 weeks	Online discussion forum (n=38)	*Pain acceptance*, anxiety and depressive symptoms, coping, pain consequences, relationship with pain, quality of life
Davis (2013) [33]	Fibromyalgia, United States	46 (98)	Asynchronous Web-based MSER[e], unguided (n=39)	12 sessions, 6 weeks	Web-based health behavior information (n=40)	*Pain, pain coping efficacy, positive affect, negative affect, social engagement, loneliness, family stress, family enjoyment, stress coping efficacy*
Dowd (2015) [30]	Chronic pain, Ireland	45 (90)	Asynchronous Web-based MBCT[f], unguided (n=62)	12 sessions, 6 weeks	Web-based psychoeducation (n=62)	*Pain interference, psychological distress*, pain intensity, catastrophizing, pain acceptance, mindfulness, life satisfaction, impression of change
Trompetter (2015) [31]	Chronic pain, the Netherlands	52-53 (75-77)	Asynchronous Web-based ACT[d], guided (n=82)	9 modules, 9-12 weeks	Web-based expressive writing (n=79), wait-list (n=77)	*Pain interference*, anxiety and depressive symptoms, pain intensity, pain disability, positive mental health, psychological inflexibility, mindfulness, engaged living, pain catastrophizing
Younge (2015) [35]	Heart disease, the Netherlands	43 (44-51)	Asynchronous Web-based mindfulness training, unguided (n=215)	4 parts, 12 weeks	TAU[g] (n=109)	*Exercise capacity*, heart rate, blood pressure, respiration rate, NT-proBNP[h], health status, perceived stress, psychological distress, social support, and composite "endpoint" score
Ljótsson (2010) [36]	IBS[i], Sweden	35 (85)	Asynchronous Web-based mindfulness and exposure CBT[j], guided (n=42)	5 steps, 10 weeks	Online discussion forum (n=43)	*IBS[i] symptoms and symptom severity*, quality of life, gastrointestinal-specific anxiety, depression, disability, treatment credibility
Ljótsson (2011) [38]	IBS[i], Sweden	38 (79)	Asynchronous Web-based mindfulness and exposure CBT[j], guided (n=98)	5 steps, 10 weeks	Web-based stress management program (n=97)	*IBS[i] symptom severity*, quality of life, cognitive symptoms for bowel disorders, perceived stress, anxiety and depression, symptom relief
Ljótsson (2011) [39]	IBS[i], Sweden	35 (74)	Asynchronous Web-based mindfulness and exposure CBT[j], guided (n=30)	5 steps, 10 weeks	Online discussion forum (n=31)	*IBS[i] symptom severity*, health economic data, quality of life, gastrointestinal-specific anxiety, disability
Ljótsson (2014) [40]	IBS[i], Sweden	42 (80)	Asynchronous Web-based mindfulness and exposure CBT[j], guided (n=153)	5 steps, 10 weeks	Same mindfulness CBT[j] without exposure (n=156)	*IBS[i] symptom severity*, quality of life, gastrointestinal-specific anxiety, cognitive symptoms for bowel disorders, anxiety and depressive symptoms
Thompson (2010) [41]	Epilepsy, United States	36 (81)	Synchronous MBCT[f], phone-based or Web-based, guided (n=26)	8 sessions, 8 weeks	TAU[g] wait-list (n=27)	*Depressive symptoms*, knowledge and skills, self-efficacy, satisfaction with life, quality of life, self-compassion

First author (year)	Population, country	Mean age in years (% F)[a]	Intervention type	Number of sessions, duration	Control condition	Outcomes[b]
Thompson (2015) [42]	Epilepsy, United States	41 (65)	Synchronous MBCT[f], phone-based or Web-based, guided (n=64)	8 sessions, 8 weeks	TAU[g] wait-list (n=64)	*Depressive symptoms*, knowledge and skills, self-efficacy, satisfaction with life, quality of life, self-compassion
Hesser (2012) [43]	Tinnitus, Sweden	49 (43)	Asynchronous Web-based ACT[d], guided (n=35)	8 sessions, 8 weeks	CBT[j] (n=32), online discussion forum (n=32)	*Tinnitus distress severity*, anxiety and depressive symptoms, insomnia severity symptoms, perceived stress, quality of life
Johansson (2015) [44]	Acquired brain injury, Sweden	46-51 (82)	Synchronous Web-based MBSR[k], guided (n=13)	8 sessions plus 7-hr retreat, 8 weeks	On-site MBSR[k] (n=12), group walking control (n=9)	*Mental fatigue*, anxiety and depressive symptoms, self-compassion, measures of attention and processing speed
Zernicke (2014) [45]	Cancer, Canada	58 (73)	Synchronous Web-based MBCR[l], guided (n=30)	8 sessions plus 6-hr online retreat, 8 weeks	TAU[g] wait-list (n=32)	*Feasibility*, mood, stress, posttraumatic growth, spirituality and well-being, mindfulness
Bruggeman-Everts (2015) [47]	Cancer, the Netherlands	50 (76)	Asynchronous Web-based MBCT[f], guided (n=257)	9 sessions, 9 weeks	None	*Fatigue severity*, psychological distress

[a]Mean age in years and % F (% female) are presented as ranges when original studies reported this information per group rather than for the whole sample.

[b]Primary outcomes italicized.

[c]MBCPM: Mindfulness-Based Chronic Pain Management.

[d]ACT: acceptance and commitment therapy.

[e]MSER: Mindful Socioemotional Regulation.

[f]MBCT: Mindfulness-Based Cognitive Therapy.

[g]TAU: treatment as usual.

[h]NT-proBNP: N-terminal probrain natriuretic peptide.

[i]IBS: irritable bowel syndrome.

[j]CBT: cognitive behavioral therapy.

[k]MBSR: Mindfulness-Based Stress Reduction.

[l]MBCR: Mindfulness-Based Cancer Recovery.

Table 2. Summary of outcomes.

First author (year)	Sample diagnosis	Mean sessions completed, %[a]	Primary outcome	Improvement[b]		
				Over time	Greater than comparison	Maintained at follow-up
Gardner-Nix (2008) [34]	Chronic pain	Unclear	Quality of life—physical health	No	No	N/A[c]
			Quality of life—mental health	Yes	Yes[d]	N/A
			Pain catastrophizing	Yes	Yes[d]	N/A
			Usual pain	Yes	Yes[d]	N/A
Buhrman (2013) [29]	Chronic pain	60	Pain acceptance	Yes	Yes	Yes (6 mo[e])
Davis (2013) [33]	Fibromyalgia	69	Pain	No	No	N/A
			Pain coping efficacy	Yes	Yes	N/A
			Positive affect	Yes	Yes	N/A
			Negative affect	Yes	No	N/A
			Social engagement	Yes	Yes	N/A
			Loneliness	Yes	Yes	N/A
			Family stress	Yes	No	N/A
			Family enjoyment	Yes	Yes	N/A
			Stress coping efficacy	Yes	Yes	N/A
Dowd (2015) [30]	Chronic pain	94[f]	Pain interference	Yes	No	Yes (6 mo)
			Psychological distress	No	No	No (6 mo)
Trompetter (2015) [31]	Chronic pain	Unclear	Pain interference	Yes	No[g]	No[g] (6 mo)
Younge (2015) [35]	Heart disease	Unclear	Exercise capacity	Yes	Marginally (P=.05)	N/A
Ljótsson (2010) [36]	IBS[h]	Unclear	IBS symptoms	Yes	Yes	N/A
			IBS symptom severity	Yes	Yes	Yes (3 mo)
Ljótsson (2011) [38]	IBS	Unclear	IBS symptom severity	Yes	Yes	Yes (6 mo)
Ljótsson (2011) [39]	IBS	Unclear	IBS symptom severity	Yes	Yes	Yes (12 mo)
Ljótsson (2014) [40]	IBS	Unclear	IBS symptom severity	Yes	Yes	Yes (6 mo)
Thompson (2010) [41]	Epilepsy	75[i]	Depressive symptoms	Yes	Yes	N/A
Thompson (2015) [42]	Epilepsy	83	Depressive symptoms	Yes	Yes	Yes (4.5-5 mo)
Hesser (2012) [43]	Tinnitus	Unclear	Tinnitus distress severity	Yes	Yes[d]	Yes (12 mo)
Johansson (2015) [44]	Acquired brain injury	Unclear	Mental fatigue	Yes	Yes	N/A
Zernicke (2014) [45]	Cancer	67	Feasibility	N/A	N/A	N/A
Bruggeman-Everts (2015) [47]	Cancer	70	Fatigue severity	Yes	N/A	N/A

[a]Mean intervention completion refers to the mean % of modules the participants participated in.

[b]Improvement—Over time and Improvement—Greater than comparison refer to postintervention assessments.

[c]N/A: not applicable.

[d]Greater than wait-list control but not greater than on-site comparison group.

[e]mo: months.

[f]Of those who completed follow-up.

[g]Greater than active comparison group but not wait-list control.

[h]IBS: irritable bowel syndrome.

[i]Median reported.

Chronic Pain and Fibromyalgia

Five of the identified studies examined a Web-based MBI for populations with chronic pain or fibromyalgia, an illness characterized by chronic pain. Of these, four studies were RCTs [29-31,33], whereas one study was not randomized [34]. Gardner-Nix and colleagues [34] compared MBCPM delivered via videoconference or on-site to wait-list control in chronic pain patients. Most participants (70%, 40/57) remained in the videoconference group. Physical HRQoL improved more in the on-site group, and distance delivery was no better than control, whereas mental HRQoL improved in both on-site and distance groups relative to control. Pain catastrophizing improved in both on-site and distance groups relative to control groups, and actual pain improved more in the distance group than the control group but more in the on-site group than the distance group.

Burhman and colleagues [29] compared an asynchronous Web-based ACT program with a wait-list control with an online discussion group for 76 patients with chronic pain. Less than half (40%, 15/38) of the treatment group completed all seven sections. The primary outcome, chronic pain acceptance, was higher in the treatment group than control (d=0.41) and did not change at a 6-month follow-up. Of secondary outcomes, anxiety and depressive symptoms improved more in the treatment group than the control (d=0.44 and d=0.18, respectively) and did not change at a 6-month follow-up. Two of eight subscales on a measure of coping (d=0.28-0.51) and two of eight subscales on a measure of pain symptoms (d=0.30-0.56) improved more in the treatment group and were maintained at follow-up. There were no effects on quality of life, pain impairment, or relationship with pain.

Davis and Zautra [33] compared Web-based MSER with a Web-based health education attention control group in an RCT of 79 fibromyalgia patients. Approximately half (49%, 19/39) completed all 12 MSER modules. Those who participated in MSER experienced greater improvement in pain coping efficacy, positive affect, social activity engagement, loneliness, family enjoyment, and stress coping relative to control (effect sizes, representing within-group variance accounted for by group assignment, ranged from .01 to .06). Both groups resulted in decreased negative affect, and neither group experienced decreased pain symptoms.

Dowd and colleagues [30] compared Web-based MBCT with Web-based pain management psychoeducation for adults (n=124) with chronic pain. Whereas 45% (28/62) and 37% (23/62) participants were retained at post and follow-up assessments, respectively, most of those who completed the follow-up assessment (74%, 17/23) viewed all sessions. Additionally, 6-month follow-up measures were included in the primary analyses. Of the primary outcomes, pain interference improved across both groups (d=0.76), but psychological distress did not improve for either group. Of secondary outcomes measured, satisfaction with life improved more in the mindfulness group than psychoeducation (d=0.59); pain acceptance, mindfulness, and catastrophizing all improved in both groups over time (d=0.42-0.58). No significant improvements were observed in average pain experienced, and a trend toward less “pain right now” across both groups was observed. Impression of change was measured with three subscales (ability to manage emotions, dealing with stressful situations, ability to enjoy pleasant events)—all of which improved more in the treatment group (d=0.41-0.62) and were maintained at 6 months, except for ability to enjoy pleasant events, which improved post intervention but was not maintained.

Trompetter and colleagues [31] compared Web-based ACT with an expressive writing control or wait-list control group for 238 people with chronic pain. Nearly half (48%, 39/82) adhered to ACT according to the author’s definition of participating for ≥3 hours per week. The primary outcome, pain interference, was more improved in the ACT group than the expressive writing group at the 3-month (post intervention; d=0.33) and 6-month follow-ups (d=0.47), but not the wait-list group. Regarding secondary outcomes, at post treatment, ACT outperformed expressive writing for improving pain intensity, outperformed wait-list for improving pain catastrophizing, and outperformed both groups for improving psychological inflexibility (d=0.23-0.60). At the 6-month follow-up, ACT outperformed expressive writing for improving pain disability, outperformed wait-list for improving mindfulness, and outperformed both groups for improving depressive symptoms, pain intensity, psychological inflexibility, and catastrophizing (d=0.28-0.54). No differences were observed between groups on anxiety, positive mental health, or engaged living outcomes. In follow-up analyses of their results, Trompetter and colleagues concluded that changes in psychological flexibility mediated the observed changes in pain interference, psychological distress, and pain intensity and suggested that pain catastrophizing served as an indirect mechanism of change through its effect on psychological flexibility [32].

Taken together, results support improvements in psychological outcomes and pain coping efficacy following Web-based MBIs for populations experiencing chronic pain but are mixed regarding the effectiveness of MBIs for decreasing actual symptoms of pain. Specifically, two studies reported improved pain over control groups, one study reported a trend toward pain improvement across treatment and control groups, and two reported no improvement in pain symptoms. As only one of the studies examining pain was synchronous [34], no conclusions can be drawn regarding the relative effectiveness of synchronous or asynchronous interventions for people with chronic pain. There appeared to be no difference between guided or unguided interventions. Only one study [34] compared a Web-based MBI to the same intervention delivered in person. Although its results suggest that the Web-based MBI may be just as effective as an on-site MBI for outcomes such as pain catastrophizing and mental HRQoL, the study included methodological issues such as lack of randomization or allocation concealment (see Table 3).

Table 3. Risk of bias assessment. Outcome data are considered complete if ≥90% of those randomized were included in outcome data.

First author (year)	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Complete outcome data or intention-to-treat analysis used	All outcomes reported
Gardner-Nix (2008) [34]	No	No	No	No	No	Yes
Buhrman (2013) [29]	Yes	Yes	No	No	Yes	Yes
Davis (2013) [33]	Yes	Yes	No	No	Yes	Yes
Dowd (2015) [30]	Yes	Yes	No	No	Yes	Yes
Trompetter (2015) [31]	Yes	Unclear	No	No	Yes	Yes
Younge (2015) [35]	Yes	Yes	No	Yes	No	No
Ljótsson (2010) [36]	Yes	Yes	No	No	Yes	Yes
Ljótsson (2011) [38]	Yes	Yes	No	Unclear	Yes	Yes
Ljótsson (2011) [39]	Yes	No	No	No	Yes	Yes
Ljótsson (2014) [40]	Yes	Yes	No	Unclear	Yes	Yes
Thompson (2010) [41]	Yes	Yes	No	No	No	Yes
Thompson (2015) [42]	Yes	Yes	No	No	Yes	Yes
Hesser (2012) [43]	Yes	Yes	No	No	Yes	Yes
Johansson (2015) [44]	No	No	No	No	No[a]	Yes
Zernicke (2014) [45]	Yes	Yes	No	No	Yes	Yes
Bruggeman-Everts (2015) [47]	No	No	No	No	Yes	Yes

[a]Intention-to-treat analysis conducted with a subset of participants but not all.

Heart Disease

One large RCT, conducted by Younge and colleagues [35], compared unguided, Web-based mindfulness training with usual care in 324 patients with heart disease. The completion rate (defined by the authors as completing at least 50% of the intervention) was 53.5% (115/215) of those who were randomized to intervention. On the primary outcome, a measure of exercise tolerance (the 6-min walk test), the mindfulness group performed better than control at a difference bordering significance, and the mindfulness group showed lower resting heart rate (d=0.20). There were no differences between the groups on blood pressure, blood levels of N-terminal probrain natriuretic peptide (NT-proBNP), mental or physical HRQoL, subjective health status, anxiety and depressive symptoms, perceived stress, perceived social support, and adverse events (all-cause mortality, heart failure, symptomatic arrhythmia, cardiac surgery, or percutaneous cardiac intervention). An as-treated analysis showed small effect size improvements in exercise tolerance, heart rate, systolic blood pressure, and stress (d=0.19-0.21). Changes in weight or blood levels of creatinine were not reported. On the basis of this one large trial the intervention appears only marginally better than usual care.

Irritable Bowel Syndrome

Ljótsson and colleagues conducted a series of RCTs examining Web-based mindfulness plus exposure-based CBT for people with IBS. Described in Ljótsson et al [36], the intervention was tested in 85 people with IBS randomized to the intervention or an online discussion forum control group. Most participants (69%, 29/42) in the intervention group completed all modules. The intervention group experienced large effect size reductions in primary outcome measures, a composite score of IBS symptoms (d=1.19) and IBS symptom severity (d=1.21) relative to the control group. All secondary outcome measures, IBS-related quality of life, gastrointestinal-specific anxiety, depressive symptoms, and perceived disability, improved more in the intervention group than control (d=0.43-0.93). IBS quality of life continued to improve at a 3-month follow-up for the intervention group, and IBS symptom severity and gastrointestinal-specific anxiety did not change. Long-term, follow-up results 15 to 18 months following treatment, after the wait-list group also received the treatment, showed that the improvements in IBS symptom severity, IBS quality of life and gastrointestinal-related anxiety were maintained [37].

Ljótsson et al [38] then tested the same Web-based mindfulness plus CBT treatment in an RCT against an active control—a Web-based stress management group [38]. The primary outcome, IBS symptom severity, was more improved in the intervention group than the active control (d=0.38). Primary analyses included pre, post, and 6-month follow-up assessments. Improvements in the intervention group relative to control were observed for secondary outcomes IBS quality of life, gastrointestinal-related anxiety, and an IBS-related cognitions scale (including negative thoughts about bowel function and personality characteristics thought to be linked to IBS; d=0.33-0.52). Perceived stress and anxiety and depressive symptoms improved over time for both conditions. At post, those in the intervention group did not feel significantly more symptom relief than control (69%, 68/98, vs 58%, 56/97) but

at a 6-month follow-up the difference was significant (65%, 64/98, vs 44%, 43/97).

Ljótsson et al [39] conducted another study comparing Web-based mindfulness plus CBT with a wait-list online discussion forum for IBS patients recruited from a gastrointestinal clinic. The primary outcome, IBS symptom severity, had greater reductions in the intervention group relative to control (d=0.77). All secondary outcomes—IBS-related quality of life, gastrointestinal-specific anxiety, and perceived disability—improved more in the intervention group relative to the control group (d=0.19-0.79). Scores for IBS quality of life further improved and all other outcomes were maintained at a 12-month follow-up in the intervention group.

Most recently, Ljótsson et al [40] investigated whether the systematic exposure component had incremental effects over the other components of the mindfulness plus CBT therapy in a randomized controlled dismantling design. Participants with IBS (n=309) received the usual intervention or a version of it without the systematic exposure component. The primary outcome, IBS symptom severity, improved more in the intervention group with exposure post treatment (d=0.47) and was maintained at a 6-month follow-up. Other secondary outcomes (gastrointestinal-specific anxiety, IBS-related quality of life, and anxiety and depressive symptoms) improved more in the group with exposure (d=0.18-0.36). There was no difference regarding cognitions related to IBS between the two groups. The authors concluded that systematic exposure had incremental benefits over the other components of the intervention. In sum, this line of research has consistently demonstrated the beneficial effects of a Web-based mindfulness plus CBT treatment for IBS symptoms, quality of life, and psychological distress among people with IBS. Although it remains unclear as to what extent therapist guidance contributed to the positive results, results suggest that the active exposure component is an important implicated process.

Epilepsy

Thompson and colleagues conducted two RCTs to examine the effectiveness of a distance delivery version of MBCT for depressive symptoms in people with epilepsy. First, in 2010, they randomly assigned 53 people with epilepsy to receive distance MBCT (via the Internet or telephone) or a wait-list control in a stratified cross-over design [41]. The primary outcome, depressive symptoms, improved more in the intervention group than control. Whereas results were slightly better for the phone group, there was no significant difference between telephone or Internet delivery. Regarding secondary outcomes, knowledge and skills (about epilepsy and depression) improved more in the MBCT group compared with control, but no significant differences between groups were observed for the outcomes self-efficacy, satisfaction with life, or quality of life. Once all individuals received the intervention, only 30% (13/44) participated in every session. Effect sizes were not reported.

Most recently, Thompson et al [42] examined distance-delivered MBCT for the prevention of depressive symptoms in people with epilepsy. Participants (n=128) were randomized to the same MBCT intervention or a wait-list condition. The primary outcome, depressive symptoms, improved more in the treatment group than control and remained lower than baseline at an 18 to 20 week follow-up. There was reduced incidence of depressive episodes in the treatment group relative to control. Regarding the secondary outcomes, knowledge and skills and satisfaction with life both improved more in the treatment group than control. Effect sizes were not reported. There were no significant changes in the treatment group over control in the outcomes depression coping self-efficacy, self-compassion, or quality of life.

Considered together, results suggest that Web-based MBCT is effective for reducing depressive symptoms among patients with epilepsy as well as for educating them about depression in the context of their illness but does not improve self-efficacy or quality of life. As the Web-based MBCT delivered in both studies was guided, synchronous and compared with treatment as usual, it is difficult to determine to what extent the intervention accounted for the observed changes in depressive symptoms, or which other factors (eg, receipt of treatment in general and social support) may have contributed.

Tinnitus

Hesser and colleagues [43] conducted an RCT comparing ACT with CBT and an online discussion forum control condition for people with tinnitus experiencing significant distress (n=99). The primary outcome, tinnitus distress severity, improved more in the ACT condition than control (d=0.68). This improvement did not differ between ACT and CBT conditions and was maintained at a 1-year follow-up. Regarding secondary outcomes, depressive symptoms, anxiety, and perceived stress all improved more in the ACT group than the control group (d=0.59-0.69). There were no differences between ACT and control on outcomes insomnia symptom severity and quality of life. Authors concluded that for coping with tinnitus, ACT may be as effective as CBT, which they noted was currently the most supported psychological intervention for tinnitus management. Data describing rates of module completion among those who started the study were not reported.

Acquired Brain Injury

A small nonrandomized trial compared Web-based MBSR with a control condition that comprised weekly walking sessions for those with acquired brain injuries (either from traumatic brain injury or stroke) experiencing mental fatigue (n=34) [44]. Of the outcomes examined, only mental fatigue reduced significantly more in the Web-based MBSR condition than both the face-to-face and walking control conditions. Depression, anxiety, and attention and processing speed measures improved within the Web-based MBSR group, whereas self-compassion did not change. Mental fatigue also significantly decreased for the walking control group when they were subsequently given Web-based MBSR. The mean attendance rates were 81% in the initial Web-based MBSR group, 82% when the walking control was subsequently given Web-based MBSR, and 94% in the face-to-face MBSR group. Authors concluded that Web-based MBSR can be helpful in improving mental fatigue, though this study included several methodological issues such as lack of randomization or allocation concealment (see Table 3).

Cancer

Two published studies have examined Web-based MBIs for cancer survivors. Zernicke and colleagues [45] compared a Web-based synchronous MBCR program with wait-list control for cancer survivors (n=62) in an RCT. The program was considered feasible (the primary outcome) because of exceeded target numbers for interest and recruitment. Regarding secondary outcomes, participants in the MBCR group reported greater improvements in mood, stress symptoms, spirituality, and acting with awareness relative to the control group (*d*=0.37-0.50). Both groups reported improvements in posttraumatic growth, and in four subscales of a mindfulness measure (mindfulness observing, describing, nonreacting, and nonjudging). After the wait-list control group subsequently received the Web-based MBCR intervention, Zernicke et al [46] published exploratory analyses of outcomes and associations between outcomes and age, sex, and cancer stage. Mood and mindfulness subscales, awareness, observing, describing, and nonjudging, improved overall following the intervention. Stress symptoms, spirituality, and the mindfulness subscale nonreacting improved more among younger than older participants. Posttraumatic growth improved more for men than women, though men had significantly worse posttraumatic growth scores at baseline. There were no differential effects based on cancer stage.

In a large trial using a one-group pre-post design, Bruggeman-Everts and colleagues [47] examined the effectiveness of Web-based MBCT for improving cancer-related fatigue in 257 patients. The primary outcome, fatigue severity, significantly decreased following participation in the intervention (*d*=1.45). Regarding secondary outcomes, psychological distress decreased (*d*=0.71), clinically significant improvement in fatigue severity occurred in 35% (89/257) of patients, and most endorsed satisfaction with the intervention.

Taken together, these results suggest that Web-based MBIs for cancer survivors are effective for improving fatigue symptoms, mood, and psychological distress. It is unclear whether guidance is related to improved outcomes as both interventions included some form of guidance. It is also unclear whether synchronous or asynchronous interventions may be more effective for cancer survivors. As Zernicke et al [45] compared MBSR with a wait-list control group and Bruggeman-Everts et al [47] had no comparison group, it is also unclear whether Web-based MBIs would be effective for fatigue, mood, and distress outcomes beyond an active control condition.

Overall, of 28 primary outcomes (excluding feasibility) from the 16 studies, 25 of the primary outcomes improved over time, and 20 of 27 improved more in intervention groups than in control groups (one primary outcome was from a study that did not have a control group). Of the 10 primary outcomes assessed at follow-up, eight showed maintained improvements (see Table 2).

Risk of Bias

Table 3 summarized the risk of bias assessment based on potential sources of bias outlined in the Cochrane Handbook for Systematic Reviews of Interventions [28]. Of the 16 studies, 15 reported all outcomes [29-31,33,34,36,38-45,47], 13 used random sequence generation [29-31,33,35,36,38-43,45], 11 had allocation concealment [29,30,33,35,36,38,40-43,47], and 12 had complete outcome data or used ITT analysis [29-31,33,36,38-40,42,43,45,47]. Only one study had blinded outcome assessments [35] and none blinded participants or personnel.

Discussion

Principal Findings

Nineteen published papers describing 16 studies examining Web-based MBIs for people with chronic physical health conditions were reviewed. Overall, most primary outcomes improved over time, but results were mixed as to whether they improved more in the intervention group relative to the control group. Specifically, outcomes including pain acceptance, stress coping efficacy, family enjoyment, social engagement, depressive symptoms, and fatigue all improved more after mindfulness interventions than control conditions, whereas other outcomes such as psychological distress, pain interference, and negative affect did not. Furthermore, among the studies including both active and wait-list control groups, some symptoms improved more in the Web-based MBI group than the wait-list control but not the active comparison group [34,43], and one study reported greater improvement in the Web-based MBI group than the active comparison but not the waitlist control [31]. Among all studies, anxiety and depressive symptoms were most frequently included as outcome variables and typically improved to a greater degree in mindfulness groups than control groups. This is consistent with reviews describing improvements in anxiety and depressive symptoms following participation in face-to-face MBIs for people with physical health conditions [5-7]. Evidence was mixed regarding the effectiveness of Web-based MBIs for improving the quality of life, but improvements in quality of life following interventions relative to control were reported more often than no improvements were reported. Although few studies examined mood, positive affect, stress, and social outcomes, improvements in these areas following intervention relative to control were reported. Among the few studies that examined distress, there was no evidence suggesting that MBIs caused improvement over a control condition. The mixed results for improvements and distress and quality of life are consistent with review by Goyal et al [5].

Whereas none of the studies included blinding of participants and personnel, this is not always possible in an RCT where a mindfulness intervention is being delivered. However, this criterion was not removed from the bias assessment in recognition that expectations surrounding the effectiveness of MBIs from participants and those who deliver interventions may represent sources of positive bias. Furthermore, lack of blinding of participants precluded blinded outcome assessment for many of these studies, as most outcomes were assessed with self-report measures. Beyond these inherent limitations, just over half (9/16) of the studies in Table 3 met all other quality indicators of randomization, allocation concealment, complete data or ITT analyses, and reporting all outcomes.

Some research groups showed great success implementing Web-based MBIs for specific populations. Ljótsson and colleagues investigated mindfulness plus CBT for IBS in a series of studies [36-40], showing consistent improvements in IBS symptom severity and quality of life in treatment groups relative to control groups. Thompson and colleagues [41,42] consistently demonstrated the effectiveness of Web-based MBIs for reducing depressive symptoms among those with epilepsy, and the effectiveness of MBCT for improving cancer-related fatigue reported by Bruggeman-Everts et al [47] is promising. It may follow that Web-based MBIs could be most effective when tailored and targeted for specific symptoms in specific populations. Indeed, the largest effect sizes of all studies reviewed were those reported for reduction of IBS symptom severity (d=1.21) [36] and cancer-related fatigue severity (d=1.45) [47], though the latter effect size represents the pre-post change for the intervention group only, not relative to a control, so it would not likely be as large in a controlled study. Both of these studies were tailored to address specific symptoms, included therapist guidance through correspondence, and included both mindfulness and CBT components. As the evidence for MBCT for cancer-related fatigue though would be more compelling if improvement was found to be greater relative to an active control group, this research group plans to investigate MBCT for cancer-related fatigue in a 3-armed RCT [48].

Younge et al [35] examined a variety of physical outcomes (eg, exercise tolerance, heart rate, blood pressure, and NT-proBNP), reporting only significant improvements in heart rate and marginally significant improvement in exercise tolerance. Although there is strong evidence to support reductions in blood pressure in cardiovascular disease populations following in-person MBIs [8], this was not replicated in the 2015 study by Younge and colleagues [35]. However, as there were a considerable number of dropouts from this study, further studies examining physical outcomes should be conducted before conclusions regarding the effectiveness of Web-based MBIs for objective physical outcomes such as blood pressure can be drawn. Whereas pain is not an objective physical outcome as it can only be ascertained by self-report, it can arguably be considered, at least in part, a physical outcome. In this review, there were mixed results regarding the effectiveness of Web-based MBIs for reducing pain specifically, but improvements occurred among outcomes reflecting abilities to regulate and deal with pain such as pain acceptance, catastrophizing, coping, and perceived pain-related disability. This is consistent with early work from Jon Kabat-Zinn et al [49] showing that people with chronic pain learned through MBSR to relate to their pain in a different and less distressing way, although the pain itself did not diminish significantly. Thus, whereas reductions in actual pain may not follow participation in Web-based MBIs per se, participants develop skills to regulate pain symptoms that may ultimately lead to less pain-related distress. Overall, the stronger evidence for impact on psychological than physical outcomes observed in this review is consistent with prior reviews of MBIs, which describe higher effect sizes for psychological outcomes than for physical outcomes [6].

The review of Web-based MBIs for improving mental health outcomes by Spijkerman et al [25] reports small effect size improvements for depression, anxiety, well-being, and mindfulness and moderate effect size for stress (though the authors cautioned that one outlier may inflate the effectiveness of MBIs for stress in their review). The authors concluded that the effect sizes they observed for psychological outcomes following Web-based MBIs were generally lower than the medium to large effect sizes found for the same outcomes in studies of face-to-face MBIs. They considered that the inclusion of healthy populations may have contributed to a floor effect, wherein smaller improvements were noted because there was less room to improve. This review did not include healthy participants, and although a meta-analytic synthesis of effect sizes was not conducted, small to moderate effect sizes were still reported most often. Spijkerman et al [25] also considered that poorer adherence to Web-based MBIs may contribute to lower effect sizes than face-to-face interventions, as Web-based interventions provide more anonymity and result in less accountability. They described adherence rates varying from 35% to 92% (which also varied somewhat because of differing definitions of adherence among different studies). Likewise, this study reported the average number of intervention sessions completed to range from 60% to 94%, and the proportion of those who participated in every session ranged from 30% to 69%. Given that regular practice is considered a necessity for development of mindfulness skills, and dose-response relationships have been found between amount of time practicing and degree of improvement [50], poor adherence can hamper the effectiveness of Web-based MBIs. It could also be that the relative absence of social support typical of Web-based interventions may partially account for the smaller effect sizes observed in outcomes related to psychological well-being such as anxiety and depressive symptoms.

Among the interventions that included mindfulness training as the core of the program [30,33-35,41-43,45,47], most primary outcomes improved over time and approximately half improved more in the intervention than control group. Among the interventions that included mindfulness as one component of a multimodal intervention [29,31,36,38-40,43], all primary outcomes improved over time and most improved more in the intervention than control group. However, the studies of interventions with mindfulness as a component of a multimodal intervention had over twice as many primary outcomes, were all asynchronous, and were all guided (whereas approximately half of the studies with mindfulness as the core of the intervention were synchronous or guided). Furthermore, primary outcomes were highly variable. Thus, it is difficult to determine to what extent the emphasis on mindfulness training alone in an intervention may have influenced observed improvements.

The results of this review give no clear indication of whether synchronous versus asynchronous Web-based MBIs or facilitated versus self-directed Web-based MBIs are more effective. Theoretically, synchronous and facilitated group interventions include more potentially therapeutic components such as social support and most closely align with traditional face-to-face program delivery. Furthermore, the presence of a facilitator could improve treatment adherence by ensuring that

the participants actually understand and engage in the therapeutic processes, and they may feel more accountable to attend and practice at home. However, because of the small number of studies included in this review and relative heterogeneity of the specific interventions and outcomes investigated, any conclusion regarding the effectiveness of synchronous versus asynchronous or facilitated versus self-directed therapies would be premature.

Limitations

There are some limitations that necessitate consideration. First, only published studies were included; thus, it is possible that results may be biased in favor of positive trial results. Second, only a small number of studies were included. Although patterns of results emerged, more cumulative findings need to be analyzed before definitive conclusions can be drawn, particularly among physical outcomes. Furthermore, although a population of those with a chronic physical condition is more homogeneous than mixed samples that have been reported in prior studies [25,51], there is still variability between people with physical health conditions, and thus different groups may be differently affected by MBIs. Analyses of results varying by condition are not possible with a small number of studies (especially when a specific population may only be represented in one study); thus, it is difficult to determine for which populations MBIs may be more or less effective. Third, as noted by Spijkerman et al [25], although all therapies included were unified by an emphasis on mindfulness, subtle differences exist between the therapies that may differentially affect outcomes. For example, there was differing relative emphasis on meditation practice, and some therapies included goal setting whereas others emphasized on nonstriving [25]. Finally, relevant studies may be missing from this review as studies not available in full-text or English language were excluded, and screening and data extraction were conducted by only 1 author.

Future Work

As the delivery of Web-based MBIs for people with physical health conditions represents a new area of research, more high-quality studies are needed to establish the effectiveness of Web-based MBIs and to consider for whom they are most effective, for what outcomes, and using which specific MBIs. Whereas this review did not include studies where mindfulness was not the primary emphasis of the program, there are interventions with mindfulness components that appear to similarly benefit people with physical health conditions [52,53]. Studies comparing Web-based MBIs to active controls, as well as noninferiority studies directly comparing Web versions of MBIs with the face-to-face interventions on which they are based will be helpful in determining the utility of Web-based MBIs and whether they represent an equally effective delivery method. Adherence should also be consistently measured and described (eg, studies should report both the mean number of sessions completed and the proportion of participants who completed all sessions), and methods to improve adherence to Web-based MBIs should be investigated. Although all studies employed treatments in which mindfulness was a central component, few of the reviewed studies actually assessed changes in mindfulness levels from pre- to posttreatment. Future studies may consider including measures of mindfulness to investigate whether actual changes in dispositional mindfulness or mindfulness skills drive the observed improvements in symptoms. Finally, further research should examine whether synchronous interventions and the inclusion of therapist guidance (or the amount of guidance) result in greater improvement in outcomes.

Conclusions

Web-based MBIs may be helpful for improving depressive symptoms, pain acceptance, fatigue, stress coping efficacy, family enjoyment, and social engagement. Furthermore, Web-based MBIs may be particularly effective when they are tailored for specific symptoms. Future studies should continue to compare Web-based MBIs to traditional delivery methods, and examine which features of Web-based MBIs are more or less effective for reduction of symptom burden.

Acknowledgments

L Carlson holds the Enbridge Research Chair in Psychosocial Oncology, cofunded by the Canadian Cancer Society Alberta/NWT and the Alberta Cancer Foundation, and receives salary support from Alberta-Innovates Health Solutions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[PDF File (Adobe PDF File), 13KB - jmir_v19i8e303_app1.pdf]

References

1. Marchand WR. Mindfulness-based stress reduction, mindfulness-based cognitive therapy, and Zen meditation for depression, anxiety, pain, and psychological distress. J Psychiatr Pract 2012 Jul;18(4):233-252. [doi: 10.1097/01.pra.0000416014.53215.86] [Medline: 22805898]

2. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
3. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of your Body and Mind to Face Stress, Pain, and Illness. New York, NY: Delacourt; 1996.
4. Shapiro SL, Carlson LE, Astin JA, Freedman B. Mechanisms of mindfulness. J Clin Psychol 2006 Mar;62(3):373-386. [doi: 10.1002/jclp.20237] [Medline: 16385481]
5. Goyal M, Singh S, Sibinga EMS, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. JAMA Intern Med 2014 Mar;174(3):357-368 [FREE Full text] [doi: 10.1001/jamainternmed.2013.13018] [Medline: 24395196]
6. Gotink RA, Chu P, Busschbach JJV, Benson H, Fricchione GL, Hunink MG. Standardised mindfulness-based interventions in healthcare: an overview of systematic reviews and meta-analyses of RCTs. PLoS One 2015;10(4):e0124344 [FREE Full text] [doi: 10.1371/journal.pone.0124344] [Medline: 25881019]
7. Bohlmeijer E, Prenger R, Taal E, Cuijpers P. The effects of mindfulness-based stress reduction therapy on mental health of adults with a chronic medical disease: a meta-analysis. J Psychosom Res 2010 Jun;68(6):539-544. [doi: 10.1016/j.jpsychores.2009.10.005] [Medline: 20488270]
8. Carlson LE. Mindfulness-based interventions for physical conditions: a narrative review evaluating levels of evidence. ISRN Psychiatry 2012;2012:651583 [FREE Full text] [doi: 10.5402/2012/651583] [Medline: 23762768]
9. Veehof MM, Oskam M, Schreurs KM, Bohlmeijer ET. Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542 [FREE Full text] [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
10. Lauche R, Cramer H, Dobos G, Langhorst J, Schmidt S. A systematic review and meta-analysis of mindfulness-based stress reduction for the fibromyalgia syndrome. J Psychosom Res 2013 Dec;75(6):500-510. [doi: 10.1016/j.jpsychores.2013.10.010] [Medline: 24290038]
11. Simpson R, Booth J, Lawrence M, Byrne S, Mair F, Mercer S. Mindfulness based interventions in multiple sclerosis--a systematic review. BMC Neurol 2014 Jan 17;14:15 [FREE Full text] [doi: 10.1186/1471-2377-14-15] [Medline: 24438384]
12. Abbott RA, Whear R, Rodgers LR, Bethel A, Thompson Coon J, Kuyken W, et al. Effectiveness of mindfulness-based stress reduction and mindfulness based cognitive therapy in vascular disease: a systematic review and meta-analysis of randomised controlled trials. J Psychosom Res 2014 May;76(5):341-351 [FREE Full text] [doi: 10.1016/j.jpsychores.2014.02.012] [Medline: 24745774]
13. Riley KE, Kalichman S. Mindfulness-based stress reduction for people living with HIV/AIDS: preliminary review of intervention trial methodologies and findings. Health Psychol Rev 2015;9(2):224-243. [doi: 10.1080/17437199.2014.895928] [Medline: 26209210]
14. Musial F, Büssing A, Heusser P, Choi KE, Ostermann T. Mindfulness-based stress reduction for integrative cancer care: a summary of evidence. Forsch Komplementmed 2011;18(4):192-202 [FREE Full text] [doi: 10.1159/000330714] [Medline: 21934319]
15. Matchim Y, Armer JM, Stewart BR. Mindfulness-based stress reduction among breast cancer survivors: a literature review and discussion. Oncol Nurs Forum 2011 Mar;38(2):E61-E71. [doi: 10.1188/11.ONF.E61-E71] [Medline: 21356643]
16. Zainal NZ, Booth S, Huppert FA. The efficacy of mindfulness-based stress reduction on mental health of breast cancer patients: a meta-analysis. Psychooncology 2013 Jul;22(7):1457-1465. [doi: 10.1002/pon.3171] [Medline: 22961994]
17. Cramer H, Lauche R, Paul A, Dobos G. Mindfulness-based stress reduction for breast cancer-a systematic review and meta-analysis. Curr Oncol 2012 Oct;19(5):e343-e352 [FREE Full text] [doi: 10.3747/co.19.1016] [Medline: 23144582]
18. Piet J, Würtzen H, Zachariae R. The effect of mindfulness-based therapy on symptoms of anxiety and depression in adult cancer patients and survivors: a systematic review and meta-analysis. J Consult Clin Psychol 2012 Dec;80(6):1007-1020. [doi: 10.1037/a0028329] [Medline: 22563637]
19. 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]
20. 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]
21. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
22. Plaza I, Demarzo MM, Herrera-Mercadal P, García-Campayo J. Mindfulness-based mobile applications: literature review and analysis of current features. JMIR Mhealth Uhealth 2013 Nov 01;1(2):e24 [FREE Full text] [doi: 10.2196/mhealth.2733] [Medline: 25099314]
23. Wahbeh H, Svalina MN, Oken BS. Group, one-on-one, or internet? preferences for mindfulness meditation delivery format and their predictors. Open Med J 2014;1:66-74 [FREE Full text] [doi: 10.2174/1874220301401010066] [Medline: 27057260]
24. Childress CA. Ethical issues in providing online psychotherapeutic interventions. J Med Internet Res 2000;2(1):E5 [FREE Full text] [doi: 10.2196/jmir.2.1.e5] [Medline: 11720924]

25. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Baer RA. Mindfulness-based treatment approaches: Clinician's guide to evidence base and applications. Burlington, MA: Elsevier; 2014.
28. Higgins JPT, Altman DG, Sterne JAC. Cochrane. 2011. Assessing risk of bias in included studies URL: http://handbook-5-1.cochrane.org/ [accessed 2017-08-23] [WebCite Cache ID 6svURitRL]
29. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. Behav Res Ther 2013 Jun;51(6):307-315. [doi: 10.1016/j.brat.2013.02.010] [Medline: 23548250]
30. Dowd H, Hogan MJ, McGuire BE, Davis MC, Sarma KM, Fish RA, et al. Comparison of an online mindfulness-based cognitive therapy intervention with online pain management psychoeducation: a randomized controlled study. Clin J Pain 2015 Jun;31(6):517-527. [doi: 10.1097/AJP.0000000000000201] [Medline: 25565584]
31. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KMG. Internet-based guided self-help intervention for chronic pain based on acceptance and commitment therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
32. Trompetter HR, Bohlmeijer ET, Fox JP, Schreurs KM. Psychological flexibility and catastrophizing as associated change mechanisms during online Acceptance & Commitment Therapy for chronic pain. Behav Res Ther 2015 Nov;74:50-59. [doi: 10.1016/j.brat.2015.09.001] [Medline: 26409158]
33. Davis MC, Zautra AJ. An online mindfulness intervention targeting socioemotional regulation in fibromyalgia: results of a randomized controlled trial. Ann Behav Med 2013 Dec;46(3):273-284. [doi: 10.1007/s12160-013-9513-7] [Medline: 23670111]
34. Gardner-Nix J, Backman S, Barbati J, Grummitt J. Evaluating distance education of a mindfulness-based meditation programme for chronic pain management. J Telemed Telecare 2008;14(2):88-92. [doi: 10.1258/jtt.2007.070811] [Medline: 18348755]
35. Younge JO, Wery MF, Gotink RA, Utens EM, Michels M, Rizopoulos D, et al. Web-based mindfulness intervention in heart disease: a randomized controlled trial. PLoS One 2015;10(12):e0143843 [FREE Full text] [doi: 10.1371/journal.pone.0143843] [Medline: 26641099]
36. Ljótsson B, Falk L, Vesterlund AW, Hedman E, Lindfors P, Rück C, et al. Internet-delivered exposure and mindfulness based therapy for irritable bowel syndrome--a randomized controlled trial. Behav Res Ther 2010 Jun;48(6):531-539. [doi: 10.1016/j.brat.2010.03.003] [Medline: 20362976]
37. Ljótsson B, Hedman E, Lindfors P, Hursti T, Lindefors N, Andersson G, et al. Long-term follow-up of internet-delivered exposure and mindfulness based treatment for irritable bowel syndrome. Behav Res Ther 2011 Jan;49(1):58-61. [doi: 10.1016/j.brat.2010.10.006] [Medline: 21092934]
38. Ljótsson B, Hedman E, Andersson E, Hesser H, Lindfors P, Hursti T, et al. Internet-delivered exposure-based treatment vs. stress management for irritable bowel syndrome: a randomized trial. Am J Gastroenterol 2011 Aug;106(8):1481-1491. [doi: 10.1038/ajg.2011.139] [Medline: 21537360]
39. Ljótsson B, Andersson G, Andersson E, Hedman E, Lindfors P, Andréewitch S, et al. Acceptability, effectiveness, and cost-effectiveness of internet-based exposure treatment for irritable bowel syndrome in a clinical sample: a randomized controlled trial. BMC Gastroenterol 2011 Oct 12;11:110 [FREE Full text] [doi: 10.1186/1471-230X-11-110] [Medline: 21992655]
40. Ljótsson B, Hesser H, Andersson E, Lackner JM, El Alaoui S, Falk L, et al. Provoking symptoms to relieve symptoms: a randomized controlled dismantling study of exposure therapy in irritable bowel syndrome. Behav Res Ther 2014 Apr;55:27-39. [doi: 10.1016/j.brat.2014.01.007] [Medline: 24584055]
41. Thompson NJ, Walker ER, Obolensky N, Winning A, Barmon C, Dilorio C, et al. Distance delivery of mindfulness-based cognitive therapy for depression: project UPLIFT. Epilepsy Behav 2010 Nov;19(3):247-254. [doi: 10.1016/j.yebeh.2010.07.031] [Medline: 20851055]
42. Thompson NJ, Patel AH, Selwa LM, Stoll SC, Begley CE, Johnson EK, et al. Expanding the efficacy of project UPLIFT: distance delivery of mindfulness-based depression prevention to people with epilepsy. J Consult Clin Psychol 2015 Apr;83(2):304-313 [FREE Full text] [doi: 10.1037/a0038404] [Medline: 25495361]
43. Hesser H, Gustafsson T, Lundén C, Henrikson O, Fattahi K, Johnsson E, et al. A randomized controlled trial of Internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. J Consult Clin Psychol 2012 Aug;80(4):649-661. [doi: 10.1037/a0027021] [Medline: 22250855]
44. Johansson B, Bjuhr H, Karlsson M, Karlsson JO, Rönnbäck L. Mindfulness-based stress reduction (MBSR) delivered live on the internet to individuals suffering from mental fatigue after an acquired brain injury. Mindfulness 2015 Apr 10;6(6):1356-1365. [doi: 10.1007/s12671-015-0406-7]

45. Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. Psychosom Med 2014 May;76(4):257-267. [doi: 10.1097/PSY.0000000000000053] [Medline: 24804884]
46. Zernicke KA, Campbell TS, Speca M, McCabe Ruff K, Flowers S, Tamagawa R, et al. The eCALM trial: eTherapy for cancer applying mindfulness. exploratory analyses of the associations between online mindfulness-based cancer recovery participation and changes in mood, stress symptoms, mindfulness, posttraumatic growth, and spirituality. Mindfulness 2016 May 14;7(5):1071-1081. [doi: 10.1007/s12671-016-0545-5]
47. Bruggeman-Everts FZ, van der Lee ML, de Jager Meezenbroek E. Web-based individual mindfulness-based cognitive therapy for cancer-related fatigue - a pilot study. Internet Interv 2015;2(2):200-213. [doi: 10.1016/j.invent.2015.03.004]
48. Wolvers MD, Bruggeman-Everts FZ, Van der Lee ML, Van de Schoot R, Vollenbroek-Hutten MM. Effectiveness, mediators, and effect predictors of internet interventions for chronic cancer-related fatigue: the design and an analysis plan of a 3-armed randomized controlled trial. JMIR Res Protoc 2015 Jun 23;4(2):e77 [FREE Full text] [doi: 10.2196/resprot.4363] [Medline: 26104114]
49. Kabat-Zinn J, Lipworth L, Burney R. The clinical use of mindfulness meditation for the self-regulation of chronic pain. J Behav Med 1985 Jun;8(2):163-190. [Medline: 3897551]
50. Carmody J, Baer RA. Relationships between mindfulness practice and levels of mindfulness, medical and psychological symptoms and well-being in a mindfulness-based stress reduction program. J Behav Med 2008 Feb;31(1):23-33. [doi: 10.1007/s10865-007-9130-7] [Medline: 17899351]
51. Cavanagh K, Strauss C, Forder L, Jones F. Can mindfulness and acceptance be learnt by self-help?: a systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. Clin Psychol Rev 2014 Mar;34(2):118-129. [doi: 10.1016/j.cpr.2014.01.001] [Medline: 24487343]
52. Cheung EO, Cohn MA, Dunn LB, Melisko ME, Morgan S, Penedo FJ, et al. A randomized pilot trial of a positive affect skill intervention (lessons in linking affect and coping) for women with metastatic breast cancer. Psychooncology 2016 Nov 16:- Epub ahead of print. [doi: 10.1002/pon.4312] [Medline: 27862646]
53. Cohn MA, Pietrucha ME, Saslow LR, Hult JR, Moskowitz JT. An online positive affect skills intervention reduces depression in adults with type 2 diabetes. J Posit Psychol 2014 Jan 01;9(6):523-534 [FREE Full text] [doi: 10.1080/17439760.2014.920410] [Medline: 25214877]

Abbreviations

ACT: acceptance and commitment therapy
AIDS: acquired immunodeficiency syndrome
CBT: cognitive behavioral therapy
HRQoL: health-related quality of life
HIV: human immunodeficiency virus
IBS: irritable bowel syndrome
MBCPM: Mindfulness-Based Chronic Pain Management
MBCR: Mindfulness-Based Cancer Recovery
MBCT: Mindfulness-Based Cognitive Therapy
MBIs: Mindfulness-Based Interventions
MBSR: Mindfulness-Based Stress Reduction
MSER: Mindful Socioemotional Regulation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
NT-proBNP: N-terminal probrain natriuretic peptide
RCTs: randomized controlled trials
TAU: treatment as usual

Edited by G Eysenbach; submitted 10.02.17; peer-reviewed by B Ljotsson, M Spijkerman; comments to author 30.03.17; revised version received 15.06.17; accepted 30.06.17; published 31.08.17.

Please cite as:
Toivonen KI, Zernicke K, Carlson LE
Web-Based Mindfulness Interventions for People With Physical Health Conditions: Systematic Review
J Med Internet Res 2017;19(8):e303
URL: https://www.jmir.org/2017/8/e303/
doi:10.2196/jmir.7487
PMID:28860106

Original Paper

The Use of Mobile Apps and SMS Messaging as Physical and Mental Health Interventions: Systematic Review

Amy Leigh Rathbone[1], BSc; Julie Prescott[1], PhD

School of Education and Psychology, University of Bolton, Bolton, United Kingdom

Corresponding Author:
Amy Leigh Rathbone, BSc
School of Education and Psychology
University of Bolton
Deane Road
Bolton, BL3 5AB
United Kingdom
Phone: 44 01204903676
Fax: 44 01204903676
Email: alr3wss@bolton.ac.uk

Abstract

Background: The initial introduction of the World Wide Web in 1990 brought around the biggest change in information acquisition. Due to the abundance of devices and ease of access they subsequently allow, the utility of mobile health (mHealth) has never been more endemic. A substantial amount of interactive and psychoeducational apps are readily available to download concerning a wide range of health issues. mHealth has the potential to reduce waiting times for appointments; eradicate the need to meet in person with a clinician, successively diminishing the workload of mental health professionals; be more cost effective to practices; and encourage self-care tactics. Previous research has given valid evidence with empirical studies proving the effectiveness of physical and mental health interventions using mobile apps. Alongside apps, there is evidence to show that receiving short message service (SMS) messages, which entail psychoeducation, medication reminders, and links to useful informative Web pages can also be advantageous to a patient's mental and physical well-being. Available mHealth apps and SMS services and their ever improving quality necessitates a systematic review in the area in reference to reduction of symptomology, adherence to intervention, and usability.

Objective: The aim of this review was to study the efficacy, usability, and feasibility of mobile apps and SMS messages as mHealth interventions for self-guided care.

Methods: A systematic literature search was carried out in JMIR, PubMed, PsychINFO, PsychARTICLES, Google Scholar, MEDLINE, and SAGE. The search spanned from January 2008 to January 2017. The primary outcome measures consisted of weight management, (pregnancy) smoking cessation, medication adherence, depression, anxiety and stress. Where possible, adherence, feasibility, and usability outcomes of the apps or SMS services were evaluated. Between-group and within-group effect sizes (Cohen *d*) for the mHealth intervention method group were determined.

Results: A total of 27 studies, inclusive of 4658 participants were reviewed. The papers included randomized controlled trials (RCTs) (n=19), within-group studies (n=7), and 1 within-group study with qualitative aspect. Studies show improvement in physical health and significant reductions of anxiety, stress, and depression. Within-group and between-group effect sizes ranged from 0.05-3.37 (immediately posttest), 0.05-3.25 (1-month follow-up), 0.08-3.08 (2-month follow-up), 0.00-3.10 (3-month follow-up), and 0.02-0.27 (6-month follow-up). Usability and feasibility of mHealth interventions, where reported, also gave promising, significant results.

Conclusions: The review shows the promising and emerging efficacy of using mobile apps and SMS text messaging as mHealth interventions.

(J Med Internet Res 2017;19(8):e295) doi:10.2196/jmir.7740

KEYWORDS

mHealth; smartphone; health; review; systematic; short message service; treatment efficacy; portable electronic applications; intervention study

Introduction

The initial introduction of the World Wide Web in 1990 brought around the biggest change in information acquisition that the modern world has ever seen [1]. Less than three decades later, statistics show that 3.2 billion people, over half of the population, can access the Internet with ease [2]. During any given second, there are estimated to be over 8 billion devices simultaneously connected to the World Wide Web, using mobile phones as a medium [3]. In modern day society, the ubiquity of mobile phones has become the norm. In 2012, it was estimated that 91% of the population were in possession of a mobile phone [4]. Due to the abundance of devices and ease of access they subsequently allow, the utility of mobile health (mHealth) has never been more endemic.

mHealth is the practice in, and support of, public health interests, which is reinforced and sustained by mobile devices [5,6]. It has become an extensive platform for the promotion and patient-led continuity of self-care [7], improving patient-centered care (PCC) [8] and prompting the progression of health literacy to positively skew societal view of diagnoses [9,10].

As of June 2016, Android users and those with access to Google Play had access to the download of 2.2 million apps, and Apple's app store offered 2 million apps [11]. However, it is salient to take into consideration that this is not a precise quantifier of relevant apps due to duplicate, nonfunctional, and alternatively topical apps. Studies have found that 31% of mobile phone owners use them to access health information; 19% have also installed a mobile app that relates to a current medical condition or to manage their health and well-being [12]. Another study has also found that over 56% of health care proviso settings inclusively use mHealth to aid clinical practice [13].

A substantial amount of interactive and psychoeducational apps are readily available to download concerning a wide range of health issues. There are informative fitness apps to tackle obesity; journal type apps that can help when managing a chronic illness such as diabetes; tracking apps for menstruation, ovulation, and fertility; and even apps that have introduced touch sensitive methods, which when used, provide detailed analytics of an individual's pulse and heart rate [14-16].

The physical outcomes for the mHealth apps largely focus on weight management and physical activity, smoking cessation, and medication adherence. Although it is apparent that each of these issues causes pejorative symptoms, the apps aim to reduce or avoid these via the medium of behavior modification. Psychoeducation within these apps, such as the disadvantages of obesity, monetary loss due to smoking, and the side effects of missed medication convey presumed consequences to amplify desirable behaviors. Behavior modification apps have proven effective in reducing the risk of obesity [17], treating eating disorders [18], and reducing anxiety [19].

An emergent number of these available apps focus on supporting individuals with mental health issues. This is inclusive of, but not exclusive to, disorders such as stress, anxiety, depression, post-traumatic stress disorder (PTSD), and obsessive-compulsive disorder (OCD). In 2014 alone, surveys suggested that 1 in 10 individuals in the United Kingdom waited longer than 12 months for a mental health assessment [20], and it is projected that by 2030, the United Kingdom will see an additional surplus of over 2 million people experiencing mental health issues [21]. mHealth has the potential to reduce waiting times for appointments and eradicate the need to meet in person with a clinician. Consequently, mHealth interventions have the potential to diminish the workload of mental health professionals, be more cost effective to practices, and encourage self-care tactics [22].

One previous study was carried out by Bakker et al [23]. This was a review that focused upon the development and validation of mental health apps (MHapps). It aimed to review current MHapps to guide future development and provide recommendations to MHapp developers for optimization of features. The study aided the authors to formulate sixteen recommendations for development, such as the inclusion of cognitive behavioral therapy (CBT), automated tailoring of the app, and coping skills training.

Donker et al [24] have meticulously reviewed the efficacy of mobile phone apps for the management of mental health. This comprehensive review performed qualitative synthesis on 8 trials that studied the effects of mHealth apps. Each study had a pre- and posttest design, or the app was administered alongside a control group. The results of this review showed the efficacy of the apps in reducing symptoms of stress, depression, and substance abuse. Within-group and between-group intention-to-treat effect sizes oscillated from 0.29-2.28 and 0.01-0.48 at the posttest and follow-ups. This was a stringent and precise review that excluded any other mHealth methods to allow for a rigorous review of apps. This literature review builds upon and varies from previous research by being inclusive of SMS text messaging (short message service, SMS) as a mHealth intervention.

Previous research and empirical studies has given valid evidence of the effectiveness of mental health interventions using mobile apps [25-27]. Alongside apps, there is evidence to show that receiving SMS messages that entail psychoeducation, medication reminders, and links to useful informative Web pages can also be advantageous to a patient's mental and physical well-being. As SMS messages can be sent directly to a patient's mobile phone, they are deemed just as convenient and as easy to use as an app [28-32].

SMS services have shown positive results, when used as mHealth interventions, for both physical and mental health issues. Studies have shown that SMS services, when used as reminders, are highly effective in increasing adherence to prescription medication [33-36]. This is also true for SMS services that prompt patients to attend their health care appointments by acting as a reminder, consisting of time and location [37]. Interactive SMS services have been found to reduce the likelihood of binge drinking in young adults and also had sustainable reductions up to 6 months after the interaction with the intervention [38,39]. One of the most salient applications for SMS as an intervention is the dispersion of psychoeducation for mental health ailments. There are services that provide information on mental illnesses such as schizophrenia [40], bipolar [41], psychosis [42], and other

common mental health disorders such as depression, anxiety, and stress [43,44]. SMS are sent to primary, private inboxes of the participant and can easily be received and disposed of; this could be one such reason for their efficacy as mHealth interventions. SMS mHealth interventions are also deemed to be more anonymous and therefore, break down certain barriers to accessing health care and eradicate stigma [45].

The inclusion of such an expansive assemblage of health issues can appear complicated due to their incongruent symptomology and treatment. However, this review aims to focus upon the general efficacy, usability, and feasibility of an mHealth intervention, as opposed to focusing upon specific health issues.

With the ever increasing use and pervasiveness of mobile phones comes an even larger market place for apps and SMS services. To our knowledge, there is no monitoring of, or stringent guidelines, which these mHealth interventions must adhere to, so effectiveness is yet to be confirmed by repeated replicability of studies. However, this area is still deemed to be in its early stages. Available mHealth apps and SMS services and their ever improving quality necessitate a systematic review in the area in reference to reduction of symptomology, adherence to intervention, and usability. The aim of this paper was to systematically review the existing empirical studies and mHealth literature, focusing on the efficacy, feasibility, and usability of apps as mHealth tools for physical and mental health.

Methods

Search Strategy and Selection of Studies

A comprehensive literature search in relevant bibliographic, Web-based databases was carried out (JMIR, PubMed, PsychINFO, PsychARTICLES, Google Scholar, MEDLINE, and SAGE). The search terms used were not restricted to the title only. They were found within the title, abstract, full paper text, or keywords. Words searched were ones such as “mHealth,” “physical,” “mental,” “mobile,” “application,” “SMS,” “internet,” “smartphone,” and “technology.” The conjunction “AND” and the disjunction “OR” logical operators were also used in the search terms (eg, mHealth AND technology, smartphone AND health, and physical OR mental health technology). The term mHealth was coupled with the words application and SMS as frequently as possible so as to avoid other mHealth interventions, such as those that were Web based (Multimedia Appendix 1).

The search terms used for physical mHealth interventions were quite specific and largely based around diet, physical inactivity, and obesity [46,47], as these health issues account for some of the most salient expenditures within the public health sector. So too does the issue of medication adherence, as missed or incorrectly taken medication can lead to increased hospital admissions and further medical issues.

The search terms used for mental mHealth interventions were broader and less specific. This was as to be careful that no mental health illness was ruled out of the search.

Of the papers the search produced, the abstracts were reviewed for eligibility. If the paper was deemed irrelevant from the abstract alone, the paper was retracted from further analyses. Successively, the remaining full text articles were further screened for relevance to the review. If they met the exclusion criterion they were also discarded. Additionally, the references of full text articles that met the inclusion criterion were also screened in an attempt to find other relevant papers. No gray literature was searched or included in the review, neither were dissertations or unpublished studies.

The first author carried out the study search, excluded searches, and elected which studies were included. The second author then considered and concurred with the final included studies. The first author carried out the risk of bias assessment.

Inclusion Criteria

The inclusion criterion for the review was the study has to be documented in English; included studies ranged from January 2008 to January 2017. The reason for this was that a more expansive date range may have seemingly diluted the prevalence of mobile apps in mHealth as they were only released in 2008 [48,49]. It was deemed necessary that the study reflect the high volume of mobile apps used in mHealth currently. All studies needed to include a mHealth intervention, which was app or SMS based. There were no demographic restrictions on the inclusive studies. There were no restrictions on the field of the studies, but due to the nature of the review most could be classified as mHealth, eHealth, and telemedicine.

Exclusion Criteria

Studies were excluded if they did not include an actual intervention. For example, a large amount of the search studies that appeared relevant were actual proposals or theoretical interventions with no technological development. Studies that used quasi-experimental designs were excluded, as were reviews. Conference papers and case studies met the exclusion criteria. Studies that did not report outcomes on physical and mental health and adherence to intervention and usability were also excluded. All studies that focused on the perceived detrimental effects of mobile phones or the Internet itself, such as radiation or addiction were excluded from the study.

Primary Outcome Measures

This review looked at the primary outcome measure of mHealth interventions that targeted physical health and well-being, such as weight loss and management, (pregnancy) smoking cessation, and medication adherence. It also reviewed mHealth interventions that aimed to improve mental health and well-being, such as depression, anxiety, stress, major depressive disorder (MDD), schizophrenia, and other common mental health disorders, as evaluated by validated mental health measures.

Analyses of Effect Sizes

Within included studies where data was attainable, between-group and within-group effect sizes (Cohen *d*) for the mHealth intervention method group were determined by extracting the variance between the pre- and posttest results (within-group effect size) or the variance between the control and intervention group posttest results (between-group effect size) and dividing by the pooled standard deviation (SD). Effect

sizes of 0.2 are relatively small, effect sizes of 0.5 are deemed to be moderate, and those of 0.8 or higher can assume to be associated with large effect sizes [50,51].

Quality Assessment

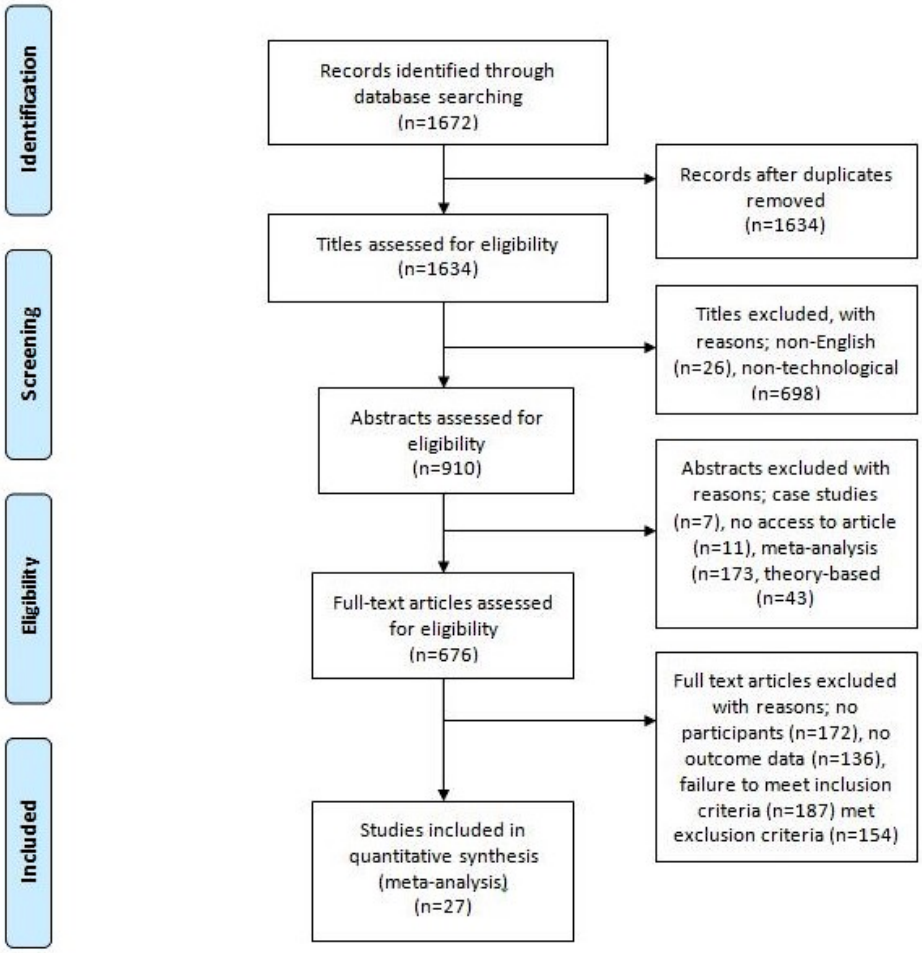
The qualities of all the studies included in this review were assessed by using the six basic criteria of the Cochrane Risk of Bias Assessment Tool [52]. This consists of screening for biases such as random sequence generation, allocation concealment, incomplete outcome data, selective reporting, and other biases (Multimedia Appendix 2). Blinding criteria was excluded from scrutiny as the criteria is nigh on impossible to adhere to in mental health interventions [24].

Results

Included Papers

An amassed total of 1672 records were considered upon their title alone. After non-English articles and duplicates were extracted, abstracts were then reviewed for applicability (n=910). Studies in which the abstract specified the use of meta-analyses, case studies, or theory-based content analysis were removed. The remaining articles (n=676) were deemed eligible for full text review. Upon review, a further 648 articles were excluded for being pilot protocols or proposals for prototype apps, having no participant, theoretical content, no provided outcome data, failure to meet the inclusion criterion, or they met the criterion for exclusion. There were 27 studies eligible for inclusion as presented in Figure 1.

Figure 1. Flow diagram of study inclusion.



Study and Intervention Characteristics

Across all studies there were an overall total of 4658 participants. Of all studies, 5 focused on the use of SMS text messaging as an mHealth intervention [53-58], 21 addressed mobile app interventions [27,59-78], and 1 study combined the two [79].

Some mHealth interventions have remained unnamed within the studies (n=5), whereas others have been coined (n=22). There are more interventions than studies due to the fact that 1 study accounts for 3 different eHealth interventions [67].

There were many physical and psychological primary outcome measures for the study, such as smoking cessation (n=4), physical activity (n=3), medication adherence (n=3), weight management (n=3), depression (n=6), anxiety (n=2), stress (n=2), schizophrenia (n=1), MDD (n=2), and other common mental health disorders (n=1). Some studies measured more than one primary outcome measure.

Within the sample used, there were within-group studies (n=7), within-group with qualitative aspects study (n=1), and randomized controlled trials (RCT) (n=19). Although a lot of the studies were of RCT design, overall, they were deemed to be of low to moderate quality when assessed against Cochrane Risk of Bias Assessment Tool [50]. Some studies were within-group types and therefore had more margin for bias as they had no control group or differing intervention to be used as a comparison [53,55,68,69,71,73,74]. One study was an unblended RCT [76]. Another study recruited their participants from a sample who had engaged in a similar, previous study [53]. Two studies failed to report their study's random sequence generation, so it was unclear whether adequate procedures had been carried out [57,61]. The risk of bias assessment aided to identify other biases in the inclusive studies. It was found that 2 of the studies proffered monetary rewards, which increased with adherence [54,62], and 1 study offered a different form of rewards [69]. In 2 studies, there was a significant unequal gender split in participants [61,67]. Some studies were found to contain elements of selective reporting. One study did not report the limitations of their study [78,] and 1 produced a positive generalization of results from a relatively small sample size [68]. A large amount of the inclusive studies (n=7) had low response rates and low engagement [27,53,57,61,66,68,76].

Data from all of the included studies were amassed into a concise table (Multimedia Appendix 3), which presents the characteristics of both the studies and their intervention.

Effects of mHealth as an Intervention and Feasibility as Treatment

In the review, there were 5 RCTs that used mobile apps as mHealth interventions [59-61,65,66] and 1 RCT that used an app combined with SMS text messaging [78]. One within-group study [53] and 3 RCTs [54,58,62] were carried out using interventions for smoking cessation. Three RCTs evaluated the efficacy of two mobile apps (FORA device [63], 2013; ALICE [64]) and 1 unnamed SMS text messaging intervention [56] for the improvement of medication adherence. There were 2 within-group design studies that used apps to target depression. Five RCTs that aimed to intervene with depression used apps [27,67,70,72,76], and 1 study used SMS text messaging [55]. Overall, across all inclusive studies, the range of the within-group and between-groups effect sizes were 0.05-3.37 immediately posttest, 0.05-3.25 at the 1-month follow-up, 0.08-3.08 at the 2-month follow-up, 0.00-3.10 at the 3-month follow-up, and 0.02-0.27 at the 6-month follow-up.

Mobile Application Intervention

Weight Management and Physical Activity

When studying the issue of inactivity in pregnant women, Choi et al [59] foound that their unnamed supportive app, which was used alongside the FitBit hardware, had a significant small effect on increased steps taken per day at the 2-month follow-up (d=0.16). This progressed to a moderate to large effect after 3-month use (d=0.48, P<.05). The intervention also reduced the prevalence of depression symptomology at the 3-month follow-up (Center for Epidemiologic Studies Depression Scale [CES-D]: d=0.44). The control group, which was the use of the FitBit hardware without the supportive app, provided a lesser effect size with no significance (CES-D: d=0.2).

Harries et al [61] found that bActive instigated a significant increase in step count. Laing et al [62] evaluated MyFitnessPal and found a significant increase in a participants self-monitoring of calorie intake; however, neither study provided sufficient data to calculate an effect size.

Smoking Cessation

Buller et al [58] found that using a mobile app (REQ-Mobile) as an intervention was more effective than a control group of SMS text messaging for point prevalence abstinence (d=0.45, P<.05).

Medication Adherence

The FORA device [63], when used as an intervention within standard care as a control, had a moderate to large between-group effect on medication adherence 1 month posttest (d=0.77), 2 month posttest (d=0.88), and continued to stay just as effective, if not become even more so, at the 3-month follow-up (d=1.02).

Mira et al [64] also used a mobile app (ALICE) as an intervention as opposed to the control group consisting of treatment as usual. There was a small significant effect on medication adherence (Morisky Medication Adherence Scale [MMAS]-4: d=0.12, P<.001). Immediately posttest, ALICE had also been effective in increasing a participants self-perceived health status (d=0.30).

Depression

Torous et al [74] reported that the app Mindful Moods was an effective tool for assessing the symptoms of depression. Results show that the app significantly increased higher rates of disclosure on Patient Health Questionnaire-9 (PHQ-9) symptomology.

Kinderman et al [68] created and evaluated the effectiveness of an app they named "Catch It," which used the basic principles of CBT. They found that the app had small to moderate effect on positive moods (d=0.17) but had a moderate to large effect on negative moods (d=0.69).

Arean et al [67] found that cognitive control and problem solving therapy had a greater effect on depression than information control did immediately posttest (d=012), 1 month posttest (d=0.20), and at the 3-month follow-up (d=0.32). The results show that the effect size was becoming consistently larger over an extended period of time.

Proudfoot et al [70] used an app named myCompass and compared it to a control group that used an attention control method. There was large within-group effect immediately posttest on depression (Depression Anxiety Stress Scale [DASS] depression: d=0.50), which dropped to moderate at the 1-month follow-up (d=0.34). The control group that used attention control gave a small effect immediately posttest (d=0.13) but continued to rise at the 1-month follow-up (d=0.27). There was also a moderate to large between-group effect (d=0.46).

The app, MEMO [57], had a large, significant between-group effect on increased positivity when using CBT as an SMS text messaging intervention (d=1.19, P<.001).

Watts et al [27] used the Get Happy app and found a significant within-group reduction in MDD from posttest to the 1-month follow-up (PHQ-9: d=1.56, P<.001; Beck Depression Inventory-II [BDI-II]: d=1.90, P<.001). At the 3-month follow-up, there were no differences between the control and the intervention group for depression. The results shown low effect sizes and insignificance (PHQ-9: d=−0.14, P=.34; BDI-II: d=−0.11, P=.52).

Anxiety and Stress

Three apps found a significant decrease in anxiety and stress [70,71,77]. Proudfoot et al [70] used myCompass for anxiety and stress as well as depression. The app had a significant small effect, immediately posttest, on anxiety (DASS anxiety: d=0.25) and a significant moderate effect on stress (DASS stress: d=0.41). A follow-up test was carried out 1 month later and, after extended use of the app, the effects became larger (DASS anxiety: d=0.52, DASS stress: d=0.47). Ly et al [77] carried out an RCT using an unnamed app that produced moderate within-group effects on stress (Perceived Stress Scale [PSS]: d=0.50) and a moderate to large effect between groups (PSS: d=0.62).

SMS Intervention

Weight Management and Physical Activity

At the 2-month follow-up, SMART MOVE [60] provided a moderate to large significant effect on increased step count when using the app (d=0.42, P<.05). The control group who were told to walk 30 min per day had no effect (d=0.08). However, it is worth noting that both the control group and the intervention group had small to moderate effects on their perceived state of health (EuroQol-Visual Analogue Scale [EQ-VAS]: d=0.26, d=0.32).

Partridge et al [78] found a small to moderate significant between-group effect on participants weight (d=0.26, P<0.1) and body mass index (BMI; d=0.19, P<.05) when using both the TXT2BFiT mobile app alongside SMS text messaging. Within-group effects of SMS text messaging and psychoeducation had an insignificant effect on weight but a significant small to moderate sized effect on BMI (weight: d=0.02, BMI: d=0.12, P<.05). These results were taken 3 months post study.

Smoking Cessation

Abroms et al [53] used SMS text messaging as an intervention and found a small within-group effect on smoking cessation immediately posttest (d=0.12).

Abroms et al [54] used SMS text messaging as an intervention and found that over 6 months, biochemically confirmed abstinence favored this intervention over the control group. This study did not give enough statistically significant data to report the effect size of the intervention. Hertzberg et al [60] found that their mobile Contingency Management (mCM) app gave results of being a useful adjunctive smoking cessation treatment. However, there was no statistical significance between the intervention and control group.

As may be apparent, the papers that evaluated the efficacy of particular mHealth interventions on physical health affected by life style changes (ie, weight management, smoking cessation, and regular medication adherence) could also be inclusive of the succeeding mental health issues. All three above topics could be attributed to addiction, behavioral and cognitive distortion; however they were reviewed separately due to the analysis of somatic consequences of nonadherence to intervention.

Efficacy, Usability, and Feasibility

Of the 27 papers reviewed, 5 did not report the usability and the feasibility of the mHealth intervention they used [58,62,65,71,73].

Five of the inclusive studies evaluated the effectiveness of SMS text messaging as a mHealth intervention. Park et al [56] found that both of their experimental groups reported high satisfaction with the texting intervention. The majority of participants stated that they “strongly believed“ that the texts assisted with medication adherence. Around 88.6% agreed that SMS text messaging was a convenient and easy to use method. Partridge et al [78] found that over half of their participants (53.7%) replied to at least half of their SMS text messages. From the 110 participants, 100 (90.9%) self-reported that the SMS method, alongside the use of the app, was effective, and they utilized the texts to improve their physical health. The feedback from participants who used the Quit4baby app [53] suggested that 88% of participants were satisfied with the amount of SMS messages they received. In the study carried out by Abroms et al [54], when intervention engagement was assessed, it was found that 85.1% of participants communicated at least once. Those who messaged at least once had an average of 28.47 (SD=25.81) interactions.

Approximately 17 of the papers used mobile apps as their preferred method of mHealth intervention. In reference to feasibility of the mHealth intervention as treatment, the majority of the papers gave positive results and received satisfactory feedback. Pham et al [76] received response from 100% of their sample, stating that the app Flowy was a useful intervention for anxiety attacks. The FORA device [63] was evaluated by participants using the Likert scale (1=strongly disagree to 5=strongly agree). Results show that participants found the app easy to use from home (4.8/5), and it was useful for medication and health management (4.3/5). This same study provided such relevant data that physicians were able to use the information it provided and make medication alterations and remain assured that medication adherence had improved. Seven changes in the medication of 5 patients were made for participants in the control group. ALICE [64] improved the medication adherence of over half the participants in the intervention group (59%). In 1 study [57], female participants deemed the app to be more useful than their male counterparts.

One study focused solely on the benefits mobile phone apps can have on psychoeducation [55]. Their study provided results showing that their app, which aimed to provide women with

higher levels of information regarding cervical cancer screening, yielded a significant effect on psychoeducation (d=1.4, P<.001).

McGillicuddy et al [63] reported a high overall satisfaction rate with the app, providing an average score of 4.8/5 points on the Likert scale. Only 1 participant of Hidalgo-Mazzei et al's [75] study disagreed with the utility of their app intervention; 82% agreed it would be pertinent for the self-management of their conditions.

Most of the articles had a high retention rate. For example, Glynn et al [60] had an 86% completion rate from participants.

Engagement varied across the studies. Kinderman et al [68] had a sample of 285 participants. They found that 65% of their sample used the app once (n=186), 17% used it twice (n=49), and only 7% completed three entries into their CBT, self-guided journal (n=21). Harries et al [61] found that participants opened and utilized the app (on average) 3.9 times per day (median=3.5, SD=2.6). Of the 36 participants that Ly et al [77] included in their study, only 16 adhered to the intervention for the full time span of 6 weeks. Hidalgo-Mazzei et al [75] gained and retained a high level of engagement from participants from the very start. After the first month of use, 46 participants (94%) continued to use the app, 40 did so at 2 months (82%), and 36 continued at 3 months (74%).

Discussion

Principal Findings and Comparison With Prior Work

Overall, the app and SMS based interventions included in this systematic literature review have provided promising indication of their efficacy to improve a patient's physical and mental health state. Previous reviews of mHealth literature have gleaned similar results [24,79,80]. The usability and feasibility as mHealth interventions were also proven effectual, as previous research has shown [81].

The three apps that aimed to increase medication adherence did so with large effect sizes [61,63,64]. Although there are only three apps, the results are promising. Medication nonadherence is a common health care problem and has the ability to impact a patient's health, the doctor patient relationship, and become an increasing burden on health care settings such as general practitioners (GPs) and hospitals due to increase in attendance [82]. A previous review, with a more extensive sample size, found that medication adherence apps are inexpensive, scalable, and easily accessible and have a significant effect on adherence [83].

However, it needs to be noted that 8 of the studies did not allocate control groups due to being within-group designs. Due to this, the results yielded have no comparable counterparts. For example, Lee et al [55] used SMS text messaging to provide psychoeducation to women in regards to cervical cancer screening. This method produced a large effect size, yet was not paralleled by an alternative manner of psychoeducation deliverance. The outcome data of these studies only reinforces the theory that RCT trials are more reliable due to the ability to compare and contrast effect sizes of an intervention.

Although the inclusion and exclusion criterion for this review were stringent, studies that provided support to participants through the medium of health care professionals or other mHealth methods, such as those which are Web-based, were not excluded. Whereas Hertzberg et al [62] found an increase in smoking cessation among participants when using their app, during the study they were also provided with cessation counseling sessions. Therefore, there is no way to differentiate between the intervention and the support, to indicate which caused the cessation. Similarly, Pramana et al [71] provided participants with once weekly meeting with a CBT counselor, while simultaneously using SmartCAT.

SMS text messaging appeared to have a more significant effect when it was used for conveying psychoeducation [54,56,57]. Overall, apps proved to be more effective when used as an intervention for stress, anxiety, and depression, showing significant large main effect sizes.

Usability and Efficacy

Where assessed, the majority of all participants ranked the usability and feasibility of, and satisfaction with, their allocated mHealth intervention as satisfactory to high. These results emulate those of previous studies that present the positive aspects of such interventions [80,81]. Patients not only perceive mHealth to be effective when used as a treatment method, but they are also showing significant, positive improvements on health and well-being in empirical studies. However, regardless of the ease of use and access to mobile phone devices, there remain to be common drawbacks and obstacles to using the devices for treatment and maintaining continuity (eg, lack of Internet connection due to socioeconomic status, app conflicting operating systems, and battery failure due to high data usage). The only study that failed to yield any effect on the dependent variable (weight loss) via the use of a mobile app was the Mobile POD [65]. This could be due to the fact that whereas the intervention consisted of an app, both the intervention group and the control group were required to listen to and interact with a pod cast. Although podcasts are relatively modern and have become a popular method of gaining health information [84], it can be argued that they are not interactive enough forms of information for participants to engage with.

Limitations

This review has two notable limitations. Initially, although the literature search was extensive, the final amount of studies used in the review was low due to the strict inclusion or exclusion criteria. Therefore, any interpretations made from the review itself cannot be generalized to a larger sample size to confirm the efficacy of the mHealth interventions in question. None of the included studies collected and collated data past a time span of 6 months. Another limitation is that only studies reported in the English language were included in the review, therefore, cross cultural variations cannot be reported or even considered. The risk of bias assessment was only carried out by one author. The outcomes of the review as a whole are also incommensurable due to the heterogeneous nature. To give just one example, the proffered treatment for addiction (smoking cessation) is completely different to that of mental health issues such as depression, stress, anxiety, PTSD, and schizophrenia.

Future Research

In regards to mental health, there is an ever present stigma that shrouds the issue. Although said stigma is steadily being eradicated in the face of improved health literacy and a wider understanding in general, it still remains. Torous et al [72] evaluated a mobile phone app that successfully incited participants to be more open when disclosing symptoms of depression. This shows significant efficacy of the app itself as it is pertinent that a patient can be open about their mental health so they can access the most suitable medical help.

mHealth interventions such as apps and SMS text messaging are still new and emerging in the field of self-care. There is call for more extensive research in the area using stringent RCT designs, and clinical trials where possible, to ensure a high quality of data and to minimize the risk of bias. Future research would also benefit from longitudinal studies with mHealth interventions to study relapse rates, sustainability, and effectiveness. Conversely, one such barrier remains ever present in this topic; apps are more commercially driven than they are scientifically derived and evaluated [85]. It can take years to design, develop, and evaluate the effectiveness of mHealth apps using rigorous scientific measures, but the time needed despairingly contrasts with how rapid pioneering technology is actually developing. Due to this, scientifically validated mHealth interventions are always going to be one step behind unregulated commercially driven apps unless the field is granted more substantial funds for larger scale studies and more sophisticated software. mHealth has the potential to provide considerable cost-effectiveness for health care settings and professionals if regulated and utilized in the correct manner.

Conclusions

In summation, this systematic review provides a distinctive awareness of the feasibility of mobile apps and SMS text messaging as physical and mental health interventions, alongside the usability. The review draws attention to the fact that such mHealth interventions have the potential to positively address physical and mental health issues. Exceeding all results, it appears that apps have a more significant effect on medication adherence, and phsycoeducational information is understood more when read in SMS form. Although there is not enough provided information to extract a firm deduction, this review shows promising evidence that apps and SMS text messaging, when used as a method of self-care, have the potential to reduce the symptoms of stress, depression, and anxiety; can encourage a healthier lifestyle; and encourage participants to adhere to their prescribed medication, thus improving patient compliance and reducing visits to health care settings and professionals [86,87].These mHealth interventions also allow a patient to become active participants within their own health care proviso while in a setting in which they feel safe. Considering the prevalence of the mobile phone across the world, these interventions could benefit participants globally. More research needs to be directed into the testing and validation of apps and SMS text messaging interventions to find evidence-based proof of efficacy as a mHealth intervention rather than a stand-alone psychoeducational tool.

Acknowledgments

The authors would like to thank the authors of all data used in the review and the University of Bolton for funding this review through the Ainsworth Scholarship.

Conflicts of Interest

None declared.

Authors' Contributions

AR and JP were involved in the concept and the design of the above review. Both authors made major contributions to data analysis. Both authors made major contributions to the write-up and editing of the manuscript. Both authors read and approved the final manuscript.

Multimedia Appendix 1

Exact words searched.

[PDF File (Adobe PDF File), 13KB - jmir_v19i8e295_app1.pdf]

Multimedia Appendix 2

Cochrane Risk of Bias Assessment of all inclusive studies.

[PDF File (Adobe PDF File), 81KB - jmir_v19i8e295_app2.pdf]

Multimedia Appendix 3

Characteristics of mobile app and SMS mHealth interventions for physical and mental health.

[PDF File (Adobe PDF File), 58KB - jmir_v19i8e295_app3.pdf]

References

1. Berners-Lee T, Fischetti M. Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor. San Francisco: Harper; 1999.
2. ONS. 2016. Internet users in the UK: 2016 URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2016%23main-pointsb.%20 [accessed 2017-03-21] [WebCite Cache ID 6p7w8nkei]
3. Gsmaintelligence. Definitive data and analysis for the mobile industry URL: https://www.gsmaintelligence.com/ [accessed 2017-03-21] [WebCite Cache ID 6p7vztJh2]
4. Pearson AL, Mack E, Namanya J. Mobile phones and mental well-being: initial evidence suggesting the importance of staying connected to family in rural, remote communities in Uganda. PLoS One 2017;12(1):e0169819 [FREE Full text] [doi: 10.1371/journal.pone.0169819] [Medline: 28095427]
5. Rehalia A, Prasad S. Global harnessing of advanced mHealth for community mobilization. Mhealth 2016;2:7 [FREE Full text] [doi: 10.21037/mhealth.2016.03.02] [Medline: 28293585]
6. O'Leary DP, Zaheer A, Redmond HP, Corrigan MA. Integration of advances in social media and mHealth technology are pivotal to successful cancer prevention and control. Mhealth 2016;2:38 [FREE Full text] [doi: 10.21037/mhealth.2016.09.02] [Medline: 28293611]
7. Finkelstein J, Knight A, Marinopoulos S, Gibbons MC, Berger Z, Aboumatar H, et al. Enabling patient-centered care through health information technology. Evid Rep Technol Assess (Full Rep) 2012 Jun(206):1-1531. [Medline: 24422882]
8. Tang C, Lorenzi N, Harle CA, Zhou X, Chen Y. Interactive systems for patient-centered care to enhance patient engagement. J Am Med Inform Assoc 2016 Jan 14;23(1):2-4. [doi: 10.1093/jamia/ocv198]
9. Nelson R, Staggers N. Health informatics: An interprofessional approach. USA: Mosby; 2017.
10. 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;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
11. Statista. 2017. App stores: number of apps in leading app stores URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2017-03-21] [WebCite Cache ID 6sEzhoDhi]
12. Fox S, Duggan M. Pew Research Center. 2012. Mobile health URL: http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [WebCite Cache ID 6sF208zcA]
13. Franko OI, Tirrell TF. Smartphone app use among medical providers in ACGME training programs. J Med Syst 2012 Oct;36(5):3135-3139. [doi: 10.1007/s10916-011-9798-7] [Medline: 22052129]
14. Årsand E, Frøisland DH, Skrøvseth SO, Chomutare T, Tatara N, Hartvigsen G, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. J Diabetes Sci Technol 2012 Sep;6(5):1197-1206 [FREE Full text] [Medline: 23063047]
15. Mertz L. Ultrasound? Fetal monitoring? Spectrometer? There's an app for that!: biomedical smart phone apps are taking healthcare by storm. IEEE Pulse 2012 Mar;3(2):16-21. [doi: 10.1109/MPUL.2011.2181019] [Medline: 22481741]
16. Ho C, Fu Y, Lin M, Chan S, Hwang B, Jan S. Smartphone applications (apps) for heart rate measurement in children: comparison with electrocardiography monitor. Pediatr Cardiol 2014 Apr;35(4):726-731. [doi: 10.1007/s00246-013-0844-8] [Medline: 24259012]
17. Wadden TA, Berkowitz RI, Womble LG, Sarwer DB, Phelan S, Cato RK, et al. Randomized trial of lifestyle modification and pharmacotherapy for obesity. N Engl J Med 2005 Nov 17;353(20):2111-2120. [doi: 10.1056/NEJMoa050156] [Medline: 16291981]
18. Juarascio AS, Manasse SM, Goldstein SP, Forman EM, Butryn ML. Review of smartphone applications for the treatment of eating disorders. Eur Eat Disord Rev 2015 Jan;23(1):1-11 [FREE Full text] [doi: 10.1002/erv.2327] [Medline: 25303148]
19. Lee J, Jung H, Lee G, Kim H, Park S, Woo S. Effect of behavioral intervention using smartphone application for preoperative anxiety in pediatric patients. Korean J Anesthesiol 2013 Dec;65(6):508-518. [doi: 10.4097/kjae.2013.65.6.508] [Medline: 24427456]
20. Bennion MR, Hardy G, Moore RK, Millings A. E-therapies in England for stress, anxiety or depression: what is being used in the NHS? a survey of mental health services. BMJ Open 2017 Jan 23;7(1):e014844 [FREE Full text] [doi: 10.1136/bmjopen-2016-014844] [Medline: 28115336]
21. Cotton R, Irwin J, Wilkins A, Young C. Nhsconfed. 2014. The future's digital: mental health and technology URL: http://www.nhsconfed.org/~/media/Confederation/Files/Publications/Documents/the-futures-digital.pdf [WebCite Cache ID 6sF1rHmpP]
22. Spurgeon JA, Wright JH. Computer-assisted cognitive-behavioral therapy. Curr Psychiatry Rep 2010 Dec;12(6):547-552. [doi: 10.1007/s11920-010-0152-4] [Medline: 20872100]
23. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
24. 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;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

25. Marks IM, Mataix-Cols D, Kenwright M, Cameron R, Hirsch S, Gega L. Pragmatic evaluation of computer-aided self-help for anxiety and depression. Br J Psychiatry 2003 Jul;183:57-65 [FREE Full text] [Medline: 12835245]
26. Lippman H. How apps are changing family medicine. J Fam Pract 2013 Jul;62(7):362-367. [Medline: 23957029]
27. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
28. Chen Z, Fang L, Chen L, Dai H. Comparison of an SMS text messaging and phone reminder to improve attendance at a health promotion center: a randomized controlled trial. J Zhejiang Univ Sci B 2008 Jan;9(1):34-38 [FREE Full text] [doi: 10.1631/jzus.B071464] [Medline: 18196610]
29. Ybarra ML, Holtrop JS, Prescott TL, Rahbar MH, Strong D. Pilot RCT results of stop my smoking USA: a text messaging-based smoking cessation program for young adults. Nicotine Tob Res 2013 Aug;15(8):1388-1399 [FREE Full text] [doi: 10.1093/ntr/nts339] [Medline: 23348969]
30. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012;3:CD009756. [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
31. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
32. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
33. Sheikh Mohammed SI, Lechner A, Ferrari U, Froeschl G, Alam DS, Holle R, et al. Mobile phone intervention for increasing adherence to treatment for type 2 diabetes in an urban area of Bangladesh: protocol for a randomized controlled trial. BMC Health Serv Res 2014;14:586 [FREE Full text] [doi: 10.1186/s12913-014-0586-1] [Medline: 25424425]
34. Sheikh Mohammed SI, Niessen LW, Ferrari U, Ali L, Seissler J, Lechner A. Effects of mobile phone SMS to improve glycemic control among patients with type 2 diabetes in Bangladesh: a prospective, parallel-group, randomized controlled trial. Diabetes Care 2015 Aug;38(8):e112-e113. [doi: 10.2337/dc15-0505] [Medline: 26207059]
35. Chi BH, Stringer JS. Mobile phones to improve HIV treatment adherence. Lancet 2010 Nov 27;376(9755):1807-1808. [doi: 10.1016/S0140-6736(10)62046-6] [Medline: 21071073]
36. Lester RT, Mills EJ, Kariri A, Ritvo P, Chung M, Jack W, et al. The HAART cell phone adherence trial (WelTel Kenya1): a randomized controlled trial protocol. Trials 2009 Sep 22;10:87 [FREE Full text] [doi: 10.1186/1745-6215-10-87] [Medline: 19772596]
37. Sims H, Sanghara H, Hayes D, Wandiembe S, Finch M, Jakobsen H, et al. Text message reminders of appointments: a pilot intervention at four community mental health clinics in London. Psychiatr Serv 2012 Feb 1;63(2):161-168. [doi: 10.1176/appi.ps.201100211] [Medline: 22302334]
38. Suffoletto B, Kristan J, Chung T, Jeong K, Fabio A, Monti P, et al. An interactive text message intervention to reduce binge drinking in young adults: a randomized controlled trial with 9-month outcomes. PLoS One 2015;10(11):e0142877 [FREE Full text] [doi: 10.1371/journal.pone.0142877] [Medline: 26580802]
39. Hospital MM, Wagner EF, Morris SL, Sawant M, Siqueira LM, Soumah M. Developing an SMS intervention for the prevention of underage drinking: results from focus groups. Subst Use Misuse 2016 Jan 28;51(2):155-164. [doi: 10.3109/10826084.2015.1073325] [Medline: 26789549]
40. Bankovská ML, Dragomirecká E, Blabolová A, Španiel F, Slováková A. Psychoeducation for schizophrenia in the Czech Republic: curriculum modification based on opinions of service users and providers. Acad Psychiatry 2015 Apr;39(2):186-190. [doi: 10.1007/s40596-014-0234-4] [Medline: 25416609]
41. González-Ortega I, Ugarte A, Ruiz de Azúa S, Núñez N, Zubia M, Ponce S, et al. Online psycho-education to the treatment of bipolar disorder: protocol of a randomized controlled trial. BMC Psychiatry 2016 Dec 22;16(1):452 [FREE Full text] [doi: 10.1186/s12888-016-1159-0] [Medline: 28007034]
42. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
43. Bockting CL, Kok GD, van der Kamp L, Smit F, van VE, Schoevers R, et al. Disrupting the rhythm of depression using mobile cognitive therapy for recurrent depression: randomized controlled trial design and protocol. BMC Psychiatry 2011 Jan 14;11:12 [FREE Full text] [doi: 10.1186/1471-244X-11-12] [Medline: 21235774]
44. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013;15(8):e172 [FREE Full text] [doi: 10.2196/jmir.2258] [Medline: 23963284]
45. Naughton F, Jamison J, Sutton S. Attitudes towards SMS text message smoking cessation support: a qualitative study of pregnant smokers. Health Educ Res 2013 Oct;28(5):911-922 [FREE Full text] [doi: 10.1093/her/cyt057] [Medline: 23640985]
46. Mytton OT, Tainio M, Ogilvie D, Panter J, Cobiac L, Woodcock J. The modelled impact of increases in physical activity: the effect of both increased survival and reduced incidence of disease. Eur J Epidemiol 2017 Mar;32(3):235-250 [FREE Full text] [doi: 10.1007/s10654-017-0235-1] [Medline: 28258521]

47. Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. J Public Health (Oxf) 2011 Dec;33(4):527-535. [doi: 10.1093/pubmed/fdr033] [Medline: 21562029]
48. Apple. 2008. Apple introduces the new iPhone 3G URL: http://www.apple.com/uk/pr/library/2008/06/09Apple-Introduces-the-New-iPhone-3G.html [WebCite Cache ID 6sF0AYugU]
49. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
50. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L Erlbaum Associates; 1988.
51. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
52. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
53. Abroms LC, Johnson PR, Heminger CL, Van Alstyne J, Leavitt LE, Schindler-Ruwisch JM, et al. Quit4baby: results from a pilot test of a mobile smoking cessation program for pregnant women. JMIR Mhealth Uhealth 2015;3(1):e10 [FREE Full text] [doi: 10.2196/mhealth.3846] [Medline: 25650765]
54. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250. [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
55. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
56. Park LG, Howie-Esquivel J, Chung ML, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
57. Whittaker R, Merry S, Stasiak K, McDowell H, Doherty I, Shepherd M, et al. MEMO--a mobile phone depression prevention intervention for adolescents: development process and postprogram findings on acceptability from a randomized controlled trial. J Med Internet Res 2012;14(1):e13 [FREE Full text] [doi: 10.2196/jmir.1857] [Medline: 22278284]
58. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214 [FREE Full text] [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
59. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth Physical Activity Intervention: A Randomized Pilot Study in Physically Inactive Pregnant Women. Matern Child Health J 2016 May;20(5):1091-1101. [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
60. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391. [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
61. Harries T. Walking in the wild? Using an always-on smartphone application to increase physical activity. 2013 Sep Presented at: IFIP Conference on Human-Computer Interaction; 2013; Berlin p. 19-36.
62. Hertzberg JS, Carpenter VL, Kirby AC, Calhoun PS, Moore SD, Dennis MF, et al. Mobile contingency management as an adjunctive smoking cessation treatment for smokers with posttraumatic stress disorder. Nicotine Tob Res 2013 Nov;15(11):1934-1938 [FREE Full text] [doi: 10.1093/ntr/ntt060] [Medline: 23645606]
63. McGillicuddy JW, Gregoski MJ, Weiland AK, Rock RA, Brunner-Jackson BM, Patel SK, et al. Mobile health medication adherence and blood pressure control in renal transplant recipients: a proof-of-concept randomized controlled trial. JMIR Res Protoc 2013;2(2):e32 [FREE Full text] [doi: 10.2196/resprot.2633] [Medline: 24004517]
64. Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014;16(4):e99 [FREE Full text] [doi: 10.2196/jmir.3269] [Medline: 24705022]
65. Turner-McGrievy G, Tate D. Tweets, apps, and pods: results of the 6-month mobile pounds off digitally (mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
66. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12. [doi: 10.7326/M13-3005] [Medline: 25402403]
67. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
68. Kinderman P, Hagan P, King S, Bowman J, Chahal J, Gan L, et al. The feasibility and effectiveness of Catch It, an innovative CBT smartphone app. BJPsych Open 2016 May;2(3):204-209 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002436] [Medline: 27703777]

69. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]
70. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
71. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014 May;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
72. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
73. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253. [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
74. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]
75. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, Del Mar Bonnín C, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLe smartphone application: feasibility, acceptability and satisfaction. J Affect Disord 2016 Aug;200:58-66. [doi: 10.1016/j.jad.2016.04.042] [Medline: 27128358]
76. Pham Q, Khatib Y, Stansfeld S, Fox S, Green T. Feasibility and efficacy of an mHealth game for managing anxiety: "flowy" randomized controlled pilot trial and design evaluation. Games Health J 2016 Feb;5(1):50-67. [doi: 10.1089/g4h.2015.0033] [Medline: 26536488]
77. Ly K, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014 Jul;1(3):95-101. [doi: 10.1016/j.invent.2014.06.003]
78. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mHealth lifestyle program with telephone support (TXT2BFiT) to prevent unhealthy weight gain in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2015;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
79. 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]
80. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332. [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
81. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
82. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;11:CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
83. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers: response to Aungst. J Am Pharm Assoc (2003) 2013;53(4):345. [doi: 10.1331/JAPhA.2013.13121] [Medline: 23892801]
84. Dingfelder S. APA. 2010. Giving psychology away, one podcast at a time URL: http://www.apa.org/monitor/2010/02/podcast.aspx [WebCite Cache ID 6sF1ZXGiU]
85. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
86. Clough B, Casey L. The smart therapist: a look to the future of smartphones and mHealth technologies in psychotherapy. Prof Psychol Res Pr 2015;46(3):147-153. [doi: 10.1037/pro0000011]
87. Kelli HM, Witbrodt B, Shah A. THE future of mobile health applications and devices in cardiovascular health. Euro Med J Innov 2017 Jan;2017:92-97 [FREE Full text] [Medline: 28191545]

Abbreviations

AMI: Anxiety Management Intervention
BDI: Beck Depression Inventory
BMI: body mass index
CBT: cognitive behavioral therapy
CES-D: Center for Epidemiologic Studies Depression Scale
DASS: Depression Anxiety Stress Scale

EQ-VAS: EuroQol-Visual Analogue Scale
mCM: mobile Contingency Management
MDD: major depressive disorder
MMAS: Morisky Medication Adherence Scale
PHQ: Patient Health Questionnaire
PSS: Perceived Stress Scale
PTSD: post-traumatic stress disorder
RCT: randomized controlled trial
SMS: short message service

Edited by T Cosco; submitted 23.03.17; peer-reviewed by D Bakker, N Bragazzi, M Ashford; comments to author 19.04.17; revised version received 05.05.17; accepted 25.05.17; published 24.08.17.

Please cite as:
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;19(8):e295
URL: http://www.jmir.org/2017/8/e295/
doi:10.2196/jmir.7740
PMID:28838887

Review

Categorizing Health Outcomes and Efficacy of mHealth Apps for Persons With Cognitive Impairment: A Systematic Review

Daniel R Bateman[1,2], MD; Bhavana Srinivas[1,3], MS; Thomas W Emmett[4], MD, MLS; Titus K Schleyer[5,6], DMD, PhD; Richard J Holden[1,3,6], PhD; Hugh C Hendrie[1,2], MBChB, DSc; Christopher M Callahan[1,5], MD

[1]Indiana University Center for Aging Research, Regenstrief Institute, Inc, Indianapolis, IN, United States

[2]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, United States

[3]School of Informatics and Computing, Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States

[4]Ruth Lilly Medical Library, Indiana University School of Medicine, Indianapolis, IN, United States

[5]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

[6]Center for Biomedical Informatics, Regenstrief Institute, Inc, Indianapolis, IN, United States

Corresponding Author:
Daniel R Bateman, MD
Indiana University Center for Aging Research
Regenstrief Institute, Inc
1101 West 10th Street
Indianapolis, IN, 46202
United States
Phone: 1 317 274 9334
Fax: 1 317 274 9307
Email: darbate@iupui.edu

Abstract

Background: Use of mobile health (mHealth) apps is growing at an exponential rate in the United States and around the world. Mild cognitive impairment (MCI), Alzheimer disease, and related dementias are a global health problem. Numerous mHealth interventions exist for this population, yet the effect of these interventions on health has not been systematically described.

Objective: The aim of this study is to catalog the types of health outcomes used to measure effectiveness of mHealth interventions and assess which mHealth interventions have been shown to improve the health of persons with MCI, Alzheimer disease, and dementia.

Methods: We searched 13 databases, including Ovid MEDLINE, PubMed, EMBASE, the full Cochrane Library, CINAHL, PsycINFO, Ei Compendex, IEEE Xplore, Applied Science & Technology Source, Scopus, Web of Science, ClinicalTrials.gov, and Google Scholar from inception through May 2017 for mHealth studies involving persons with cognitive impairment that were evaluated using at least one quantitative health outcome. Proceedings of the Annual ACM Conferences on Human Factors in Computing Systems, the ACM User Interface Software and Technology Symposium, and the IEEE International Symposium on Wearable Computers were searched in the ACM Digital Library from 2012 to 2016. A hand search of JMIR Publications journals was also completed in July 2017.

Results: After removal of duplicates, our initial search returned 3955 records. Of these articles, 24 met final inclusion criteria as studies involving mHealth interventions that measured at least one quantitative health outcome for persons with MCI, Alzheimer disease, and dementia. Common quantitative health outcomes included cognition, function, mood, and quality of life. We found that 21.2% (101/476) of the fully reviewed articles were excluded because of a lack of health outcomes. The health outcomes selected were observed to be inconsistent between studies. For those studies with quantitative health outcomes, more than half (58%) reported postintervention improvements in outcomes.

Conclusions: Results showed that many mHealth app interventions targeting those with cognitive impairment lack quantitative health outcomes as a part of their evaluation process and that there is a lack of consensus as to which outcomes to use. The majority of mHealth app interventions that incorporated health outcomes into their evaluation noted improvements in the health of persons with MCI, Alzheimer disease, and dementia. However, these studies were of low quality, leading to a grade C level of evidence. Clarification of the benefits of mHealth interventions for people with cognitive impairment requires more randomized controlled trials, larger numbers of participants, and trial designs that minimize bias.

Trial Registration: PROSPERO Registration: PROSPERO 2016:CRD42016033846; http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016033846 (Archived by WebCite at http://www.webcitation.org/6sjjwnv1M)

(J Med Internet Res 2017;19(8):e301) doi:10.2196/jmir.7814

KEYWORDS

mHealth; mobile health; applications; Alzheimer disease; dementia; systematic review

Introduction

Industry analysts expect worldwide mobile phone app users to increase from 2.6 billion in 2015 to 6.1 billion users by 2020 [1,2]. Similarly, analysts forecast worldwide mobile device app downloads and usage to grow from 111.2 billion in 2015 to 284.3 billion by 2020 [3]. From a financial scope, global mobile app gross revenue for 2016 surpassed US $51 billion and by 2020 is expected to exceed US $101 billion [3]. The National Institutes of Health (NIH) Consensus Group on mHealth defines mobile health (mHealth) as "the use of mobile and wireless devices to improve health outcomes, health care services, and health research" [4,5]. The NIH Strategic Plan for 2016-2020 incorporates the study of mHealth technologies and their ability to help prevent and treat illness as a research priority [6]. Following the aforementioned NIH definition of mHealth, a mHealth app operates on either a mobile or wireless device, with an objective of improving health outcomes, health care services, or health research [4].

The mHealth technologies and apps that help persons with mild cognitive impairment (MCI), Alzheimer disease, and dementia offer a unique opportunity for intervention because there are no disease-modifying agents for Alzheimer disease and related dementias [7]. More than 5.4 million people in the United States live with Alzheimer disease, the most common type of dementia [8]. Scientists predict the number of people with Alzheimer disease in the United States to reach 8.4 million by year 2030 [8]. Until disease-modifying agents are found, innovative psychosocial interventions, including mHealth interventions, offer the greatest potential for improving quality of life for persons with dementia and their caregivers [9].

Much remains unknown about the health outcomes used in mHealth apps and the effectiveness of these apps in improving the health of persons with MCI, Alzheimer disease, and dementia. There are literally hundreds of mobile apps that persons with MCI or dementia can use. Those mHealth apps are being marketed to help persons with cognitive impairment with unclear validity to their claims. Persons with cognitive impairment are already using mHealth apps, and will continue to do so in greater numbers, yet often the risks and benefits are not fully understood [3]. Similarly, the effects of these apps on persons with MCI, Alzheimer disease, and dementia have not been adequately reviewed and summarized in a systematic fashion. Therefore, the primary aim of this systematic review seeks to catalog the types of quantitative health outcomes utilized in these mHealth app studies. The secondary aim of this review strives to evaluate the effectiveness of mHealth apps in improving the health outcomes of persons with MCI, Alzheimer disease, and dementia through a review of the current scientific literature.

Background on the types of mHealth interventions included in this review are listed subsequently. These interventions can be grouped into a number of different categories, including cognitive training and serious games, wandering and wayfinding, reminiscence therapy, prompts and multicomponent interventions, engagement interventions, and exercise interventions.

Types of Interventions

Cognitive Training and Serious Games

There exists a great deal of interest in using computerized cognitive training as an intervention to prevent and treat neurodegenerative disorders. Rebok et al [10] showed that independent older adults who underwent computerized cognitive training retained cognitive and functional benefits 10 years out from the intervention. However, the potential benefits for persons with MCI, Alzheimer disease, or persons with dementia is much less clear. A recent systematic review found that persons with MCI who received cognitive training had improvements in cognition, whereas persons with dementia had limited evidence for efficacy [11].

Serious games are games with a primary purpose other than entertainment, enjoyment, and fun [12]. They often include cognitive training or exercise training in the form of games. An example of this could be seen with a patient recovering from a stroke playing a serious game involving the activity of swinging a baseball bat in a virtual game rather than doing traditional exercises.

Wandering and Wayfinding

Wandering is a very common problem in older adults with mild-to-moderate stages of dementia. Such behavior usually occurs as a result of memory deficits and spatial disorientation, which makes persons with dementia less likely to recognize the route [13]. Navigation systems, such as satellite navigation (Global Positioning System), three-dimensional maps, and electronic maps, could provide assistance in locating the patient irrespective of the closed or outdoor environment, and could also support the person with dementia in finding their way back home [13,14]. This practice of finding one's way back home or to a preselected destination is known as "wayfinding."

Reminiscence Therapy

Reminiscence therapy works under the assumption that remote memory remains intact until later in the course of dementia and that recalling and discussing past events and life experiences can help the psychological wellness and cognition of people with dementia [15]. Often reminiscence therapy therapists will utilize music, pictures, art, and other aids in sessions. A therapist or a staff member trained in reminiscence therapy leads the session, which can take either a group or an individual therapy

format. Reminiscence therapy has been shown to improve well-being, patient-caregiver relations, global cognition, and decrease social withdrawal [16-19].

Prompts and Multicomponent Interventions

With the progression of disease, persons with dementia lose their ability to perform activities of daily living (ADL) and often require frequent support and assistance from a family member or caregiver [20]. Prompts incorporate a unique approach to support and provide assistance to persons with cognitive impairment. Studies have shown that prompts can help persons with dementia to be less dependent on caregivers [21]. The function of prompts can range from reminders to take medications, to notifications of a scheduled activity, to a verbal or visual cue to get dressed or shower. We combined the prompt category and multicomponent intervention category because there was significant overlap between these two. Multicomponent interventions typically included prompts and some notification system for the caregiver. Other examples of components involve patient location, a communication system with health care professionals, and engagement activities.

Engagement Interventions

Past studies demonstrated that recreational activities and engagement can lead to persons with dementia having more positive affect, decreased agitation, and decreased passivity [22,23].

Exercise Intervention

Exercise training has been shown to reduce behavioral and psychological symptoms of dementia [24], to slow progression of cognitive decline in MCI [25], and to lead to increased hippocampus size [26], a region of the brain responsible for short-term memory. A recent review examining the evidence for physical and cognitive interventions to improve brain health found sufficient evidence that both physical and cognitive interventions lead to enhanced neuroplasticity and prevention of pathological aging (MCI, Alzheimer disease, and dementia) [27]. Evidence also suggests that the combination of physical and cognitive interventions may amplify these positive effects on neuroplasticity [27].

Methods

Using the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) checklist, we systematically reviewed the scientific literature to find mHealth apps that sought to improve health outcomes of persons with MCI, Alzheimer disease, and dementia. The PRISMA guidelines provide a standardized structure for the design, iterative process, extraction, and synthesis that take place during the development of a systematic review [28]. Aim 1 of this systematic review attempts to catalog quantitative health outcomes used to evaluate mHealth apps, whereas aim 2 seeks to assess the effectiveness of mHealth apps for persons with cognitive impairment that incorporate at least one quantitative health outcome. The systematic review protocol was registered with PROSPERO, the international prospective register of systematic reviews, at inception to avoid duplication [29].

Data Sources and Searches

A comprehensive search of the literature was performed by a medical librarian (TWE) in Ovid, MEDLINE, PubMed, EMBASE, the full Cochrane Library, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsycINFO, Ei Compendex, IEEE Xplore, Applied Science & Technology Source, Scopus, Web of Science, ClinicalTrials.gov, and Google Scholar. All databases were searched from inception. Proceedings of the Annual Association for Computing Machinery (ACM) Conferences on Human Factors in Computing Systems, the ACM User Interface Software and Technology Symposium, and the IEEE International Symposium on Wearable Computers were searched in the ACM Digital Library from 2012 to 2016. Initial searches were conducted in February 2016 and updates were performed in May 2017. Bibliographies of relevant studies were also reviewed for additional references. A hand search of JMIR Publications journals was completed in July 2017. It should be noted that in the biomedical literature this type of separate search would be classified as a search of the “grey literature”; however, in the fields of computer science, engineering, and human computer interaction, conference proceedings are considered the primary source of scientific literature [30].

The complete search strategies for each database are reported in Multimedia Appendix 1. Database-specific subject headings and keyword variants for each of the two main concepts—dementia/cognitive impairment and mobile technology—were identified and combined. Results were limited to the English language, and animal studies were excluded.

Study Eligibility

Only empirical studies were included in this systematic review. Inclusion criteria required the study to use a mHealth app on a tablet, a mobile phone, a personal digital assistant, another handheld mHealth device, or a mHealth app accessed from a computer as an intervention for persons with MCI, Alzheimer disease, or dementia. Studies using computers were only included if they accessed a program that was also a mHealth app.

Inclusion criteria required there be at least one quantitative health outcome in the study. Persons aged 18 years or younger were not included because the focus of this study was on cognitive impairment that develops during adulthood. Case series of more than two subjects, case-control studies, cross-sectional studies, and cohort studies were included. Studies were excluded if (1) the study focused primarily on participant populations outside of those with MCI, Alzheimer disease, and dementia. This included populations with diagnoses of traumatic brain injury, human immunodeficiency virus, multiple sclerosis, serious mental illness, intellectual disabilities, or active status as a caregiver, or (2) the primary purpose of the mHealth app was to screen for illness, make an assessment, or determine diagnosis. These studies were excluded because the focus of this study was on active mHealth interventions. Caregivers were not a target of this review. However, some studies included both persons with cognitive impairment and their caregivers. Studies were ruled out if caregivers were the population focus of the study.

Study Selection and Data Extraction

Screening of records by title and abstract were completed independently by two authors (DB and BS). An adjudication process was used by the two authors, where they met face-to-face to review screened records. When the authors did not agree on a record, they came to a consensus together through discussion and re-review of the record. The same process was used when evaluating full articles for inclusion and when categorizing the final included articles with the Oxford Centre for Evidence-based Medicine (OCEBM) Levels of Evidence system. Both reviewers independently reviewed full articles and completed data extraction. Using a standard approach, they extracted study design, intervention type, technology type, population diagnoses, mean age of population, mean Mini-Mental Status Examination (MMSE) or comparable cognitive exam, health outcomes, and information on the effectiveness of the mHealth app. Study quality was assessed and categorized using the OCEBM Levels of Evidence System, where studies are categorized into one of five levels of evidence, with one being the strongest level [31]. Levels of evidence using the OCEBM system are (1) level 1: systematic reviews of randomized controlled trials (RCTs), individual RCTs, and all-or-none case series; (2) level 2: systematic reviews of cohort studies, individual cohort studies, and "outcomes" research; (3) level 3: systematic review of case-control studies and individual case-control studies; (4) level 4: case-series and poor quality cohort studies; and (5) level 5: expert opinion. Recommendation grades are listed as consistent level 1 studies ("A"), consistent level 2 or 3 studies or extrapolations from level 1 studies ("B"), level 4 studies or extrapolations from level 2 or 3 studies ("C"), and level 5 evidence or troubling inconsistent or inconclusive studies of any level ("D") [31].

Results

A total of 4752 records were identified through database searches. After removing duplicates, 3955 unique titles and abstracts were screened, and 476 full articles were reviewed (Figure 1) [32]. Division of these records according to database can be found in Table 1. A total of 24 separate articles met study inclusion criteria.

Figure 1. PRISMA flow diagram.

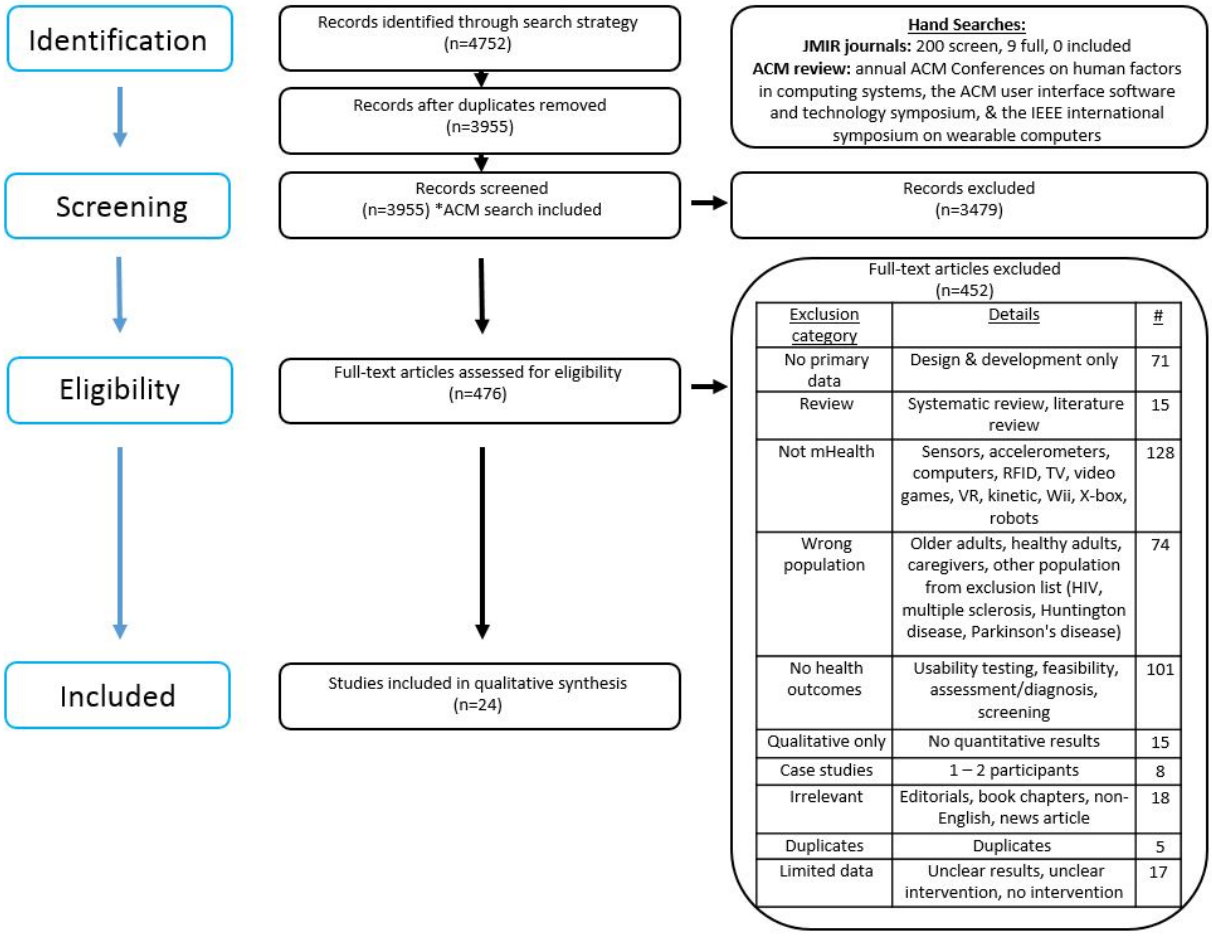


Table 1. Search results with and without duplicates.

Database	Duplicates included, n	Duplicates removed, n
ACM Digital Library	292	289
Applied Science &Technology Source	37	23
CINAHL	360	291
ClinicalTrials.gov	47	47
Cochrane Library	224	105
Ei Compendex	120	106
EMBASE	736	640
Google Scholar	67	37
IEEE Xplore	453	344
Ovid MEDLINE	1277	1276
PsycINFO	250	134
PubMed	507	507
Scopus	207	106
Web of Science	175	50
Total	4752	3955

Health outcomes were extracted and grouped by intervention type (Table 2). Of the 24 individual studies included, 14 studies lacked controls. The majority of studies were small in size. Using the Modified OCEBM Levels of Evidence rating system, four studies met criteria for level 2 evidence [33-36] and none met criteria for either level 1 or level 3 evidence.

The remaining 83% (20/24) studies were identified as meeting criteria for level 4 evidence (Tables 3-6). All studies were found in biomedical journals or were from biomedical conferences. None of the included studies came from the engineering literature. Two of the four level 2 studies showed some evidence of efficacy [33,39]. When looking at all 24 studies regardless of quality, 58% (14/24) showed some degree of efficacy.

Table 2. Health outcomes and efficacy by mHealth intervention type (N=24).

Intervention type	Studies, n[a]	Health outcomes	Studies with efficacy, n (%)
Cognitive training with no games	6	Cognition [33-35,37-39]; function [35]; mood [33]	5 (83)
Serious games	4	Cognition [36,40-42]; mood, anxiety, stress [36]	1 (25)
Wandering and wayfinding	1	Cognition [43]; unsafe walking behavior [43]	1 (100)
Reminiscence therapy	2	Cognition [19]; communication ability [19]; mood [19]; social interest [19]; psychological stability [44]	2 (100)
Prompts and multicomponent interventions	4	Cognition [20,45]; subjective report of cognition [45]; mood [45]; psychological stability [44]; perceived autonomy [20,46]; feeling of competence [20]; number of caregiver and patient unmet needs [20]; quality of life for caregiver or patient [20,45,46]; caregiver burden [45]	1 (25)
Engagement interventions	7	Cognition [47]; well-being and mood [47-49]; behavioral and psychological symptoms of dementia [50-52]; engagement in activities [47,48]; quality of life of patient [48]; helpfulness to caregiver [53]	5 (71)
Exercise intervention	1	Quality of life, self-efficacy, change in weekly steps taken, 6-min walk, Mini-Physical Performance Test [54]	0 (0)
Total	24		14 (58)

[a]Categories are not mutually exclusive; one article was counted twice.

Table 3. Cognitive training interventions (n=6).

Author, year	Study type	Intervention type	Technology type	Population[a]	Outcomes	OCEBM level[b]	Effective/Results[c]
Barnes et al, 2006 [33]	RCT	Cognitive training	Computer	36 pts w/ MCI; age: mean 74 years; RBANS total: mean 86.6	Cognition, mood	2	No improvement in cognition (RBANS total)
Chan et al, 2017 [34]	RCT-single blind	Cognitive training	Tablet	99 pts w/ MCI; age: mean 68.7 years; MOCA: mean 24.4	Cognition	2	Yes, treatment group had improvements in working memory; both treatment and control groups had improvements in immediate and delayed recall
Gooding et al, 2015 [37]	RCT	Cognitive training	Computer	74 pts w/ subclinical cognitive decline; age: mean 75.6 years; mMMSE: mean 50.6	Cognition	4	Yes, improvement in cognition (mMMSE, BSRT, and LMS)
Han et al, 2014 [38]	Pilot	Cognitive training	Tablet	10 pts w/ MCI; age mean: 69.7 years; MMSE: mean 26.7; CDR: mean 0.5	Cognition	4	Yes, significant improvement in cognition (word list memory test)
Mansbach et al, 2017 [39]	Controlled trial	Cognitive training	Mobile app from computer	38 pts w/ normal cognition, MCI, and mild dementia; age: mean 78.1 years; BCAT: mean 37.3	Cognition, subjective report of cognition	4	Yes, treatment group had greater improvements in cognition
Tarraga et al, 2006 [35]	RCT	Cognitive training	Computer	43 pts w/ MCI; age: mean 76.7 years; MMSE: mean 21.9	Cognition, function	2	Yes, improvement in cognition (ADAS-Cog, MMSE); no functional improvements

[a]BCAT: Brief Cognitive Assessment Tool; CDR: Clinical Dementia Rating scale; MCI: mild cognitive impairment: MMSE: Mini-Mental State Examination; mMMSE: Modified Mini-Mental State Examination; MOCA: Montreal Cognitive Assessment.

[b]Oxford Centre for Evidence-based Medicine's Levels of Evidence and Grades of Recommendation (1=highest quality; 5=lowest quality).

[c]ADAS-Cog: Alzheimer's disease Assessment Scale cognitive subscale; BSRT: Buschke Selective Reminding Test; LMS: Logical Memory Subtest; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status.

Table 4. Serious games with cognitive training (n=4).

Author, year	Study type	Intervention type	Technology type	Population[a]	Outcomes	OCEBM level[b]	Effective/Results
Finn and McDonald, 2011 [36]	RCT	Serious games, cognitive training	Computer	25 pts w/ MCI; age: mean 74.2 years; MMSE: mean 27.8	Cognition, Mood, Anxiety, Stress	2	No, only improvement was in visual sustained attention; no improvement in cognition, depression, or anxiety
Hsiung et al, 2009 [40]	Pilot	Serious games, cognitive training	Handheld device	17 pts total 12 w/ MCI, 2 healthy, 3 w/ subjective memory complaints; age: mean 72 years: MMSE: mean NR	Cognition	4	No improvement in cognition
Manera et al, 2015 [41]	Pilot	Serious games, cognitive training	Tablet	9 pts w/ MCI, 12 pts w/ Alzheimer disease; age mean 78.4 years; MMSE MCI: mean 27.2, MMSE Alzheimer disease: mean 18.4	Cognition	4	Yes, improvement in praxis & executive function
Merilampi et al, 2014 [42]	Pilot	Serious games, cognitive training	Tablet, computer	16 pts w/ mild-to-moderate cognitive impairment; age: mean 90 years; MMSE: mean 21.6	Cognition	4	No improvement in cognition

[a]MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; NR: not reported.

[b]Oxford Centre for Evidence-based Medicine's Levels of Evidence and Grades of Recommendation (1=highest quality; 5=lowest quality).

Table 5. Wandering and wayfinding (n=1), reminiscence therapy (n=2), and prompts and multicomponent (n=4) interventions.

Author, year	Study type	Intervention type	Technology type	Population[a]	Outcomes[b]	OCEBM level[c]	Effective/Results[b]
Hettinga et al, 2009 [43]	Pilot	Wandering and wayfinding	PDA	4 pts w/ mild dementia; age: ≥55 years; MMSE: range 17-25	Unsafe walking behaviors, working memory	4	No unsafe walking behaviors
Hattink et al, 2016 [20]	RCT	Multicomponent intervention	Early detection system-touchscreen or mobile device	42 pts w/ MCI and dementia; age: mean 78.7 years; MMSE: mean 18.1	Cognition, QOL-AD for CG or patient, perceived autonomy, feeling of competence, number of CG and patient unmet needs	4	No differences in QOL-AD for CG or patient, perceived autonomy, grade for QOL, feeling of competence, MMSE, number of caregiver and patient unmet needs.
Imbeault et al, 2016 [45]	Case series	Prompts	Mobile phone app	3 pts w/ Alzheimer disease; age: mean 69 years; MMSE: mean 28	Cognition, subjective report of cognition, depression, CG burden	4	No, almost all cognitive tests remained the same or decreased. Unclear results for depression and caregiver burden.
Meiland et al, 2012 [46]	Pilot	Prompts	Mobile device, sensors, touchscreen, & actuators	12 pts w/ MCI, dementia, or Alzheimer disease; age: range 57-84 years; MMSE: range 17-25 (mean age and MMSE NR)	Quality of life, perceived autonomy	4	No effect on QOL of Patient or CG, or perceived autonomy.
Yasuda et al, 2013 [44]	Pilot	Prompts and reminiscence therapy	Computer	4 pts w/ Alzheimer disease; age: mean 78.7 years; MMSE: mean 19.5	Psychological stability, communication ability, IADL completion	4	Yes, 3 of 4 participants had benefit in psychological stability. Improvements were also noted in communication ability and in IADL competition.
O'Rourke et al, 2011[d] [19]	Pilot	Reminiscence therapy	Mobile app from computer	6 pts w/ dementia; age: mean 72 years; MMSE: mean 17.8	Cognition, communication ability (FLCI), depression, social interest questionnaire	4	Yes, MMSE scores improved in half of pts FLCI scores improved or remained stable in all but one participant.

[a]MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; NR: not reported.

[b]CG: caregiver; FLCI: Functional Linguistic Communication Inventory; IADL: Instrumental Activities of Daily Living; QOL-AD: Quality of Life Scale in Alzheimer's Disease.

[c]Oxford Centre for Evidence-based Medicine's Levels of Evidence and Grades of Recommendation (1=highest quality; 5=lowest quality).

[d]Categories are not mutually exclusive; one article was counted in both the reminiscence therapy and prompts section.

Table 6. Engagement (n=7) and exercise (n=1) interventions.

Author, year	Study type	Intervention type	Technology type	Population[a]	Outcomes[b]	OCEBM level[c]	Effective/Results
Astell et al, 2016 [49]	Controlled trial	Engagement	Tablet	30 pts w/ dementia; age: mean 87.3 years; MOCA: mean 13.4	Enjoyment	4	Yes, 88% of patients reportedly enjoyed the games
Hsu et al, 2016 [50]	Case series	Engagement	Tablet	3 pts w/ dementia; age: mean 78 years; MOCA: mean 23.5 (1 pt refused MOCA)	Behavioral and psychological symptoms of dementia	4	Yes, decreased use of "as-needed" medications for behavioral problems
Leng et al, 2014 [47]	Pilot	Engagement	Tablet	6 pts w/ dementia; age: mean 77 years; MMSE: mean 21	Cognition, mood, engagement, well-being	4	Yes, tablet activities at least an equal positive effect on mood, engagement, and well-being vs traditional group activities
Lim et al, 2013 [53]	Pilot	Engagement	Tablet	21 dyads of people with early dementia and CGs; PWD age: mean 73.5 years; MMSE: NR	Helpfulness to caregiver	4	Yes, 47.6% of CG found tablet somewhat, moderately, or extremely helpful
Tyack et al, 2015 [48]	Pilot	Engagement	Tablet	12 dyads-PWD and CGs; age: mean 75 years; MMSE: NR	QOL-AD; Visual Analogue Scale for Happiness, Wellness, and interestedness	4	No improvement in happiness, wellness, interestedness
Vahia et al, 2016 [51]	Open-label study	Engagement	Tablet	36 pts w/ dementia; age: mean 79.9 years; MMSE: NR	Behavioral and psychological symptoms of dementia, agitation	4	Yes, agitation decreased post intervention
Van Der Ploeg et al, 2015 [52]	RCT	Engagement	Mobile app from computer	17 pts w/ dementia; age: mean 86.7 years; MMSE: mean 7.3	Behavioral and psychological symptoms of dementia, agitation	4	No significant reduction in agitation
Vidoni et al, 2016 [54]	Pilot	Exercise	Accelerometer, mobile app from computer	21 pts as normal control, 9 pt w/ Alzheimer disease; control age: mean 72.3 years, Alzheimer disease age: mean 69.6 years, MMSE: NR	QOL, self-efficacy, change in weekly steps taken, 6-min walk, Mini-Physical Performance Test	4	No improvement in outcomes

[a]CG: caregiver; MMSE: Mini-Mental State Examination; MOCA: Montreal Cognitive Assessment; NR: not reported; PWD: people with dementia.

[b]QOL: Quality of Life; QOL-AD: Quality of Life Scale in Alzheimer's Disease.

[c]Oxford Centre for Evidence-based Medicine's Levels of Evidence and Grades of Recommendation (1=highest quality; 5=lowest quality).

Cognitive Training and Serious Games

Study Characteristics

Cognitive training with and without serious games made up a large number of the studies in this review (n=10). Intervention design and duration varied significantly among studies. All studies in the two groups incorporated cognition as an outcomes measure, another included function [35] and still others used mood, stress, and anxiety [33,36] as outcomes. When examining the two groups in combination 60% (6/10) of the studies showed some degree of efficacy [34,35,37-39,41] (see Tables 3 and 4). All four studies in this review that met criteria for level 2 quality of evidence fell into one of these two categories [33-36].

Description of Apps and Technology

A number of studies in this group used commercially available cognitive training programs, such as those by the company Lumos Laboratory, marketed as Lumosity, and by the company Posit Science marketed as BrainHQ. Other studies reported on mHealth apps that had been developed through research. Many of these studies were conducted on computers; however, studies were only included if the apps were able to be accessed by mobile devices. One study that demonstrated efficacy used a tablet-based Chinese calligraphy program as a form of cognitive training [34].

Wayfinding and Wandering

Study Characteristics

Although a number of pilot studies reported on experiments involving navigation systems for persons with dementia, only one met criteria for inclusion [43]. Health outcomes for the study consisted of working memory and unsafe walking behaviors [43]. The study proved to be effective in that no unsafe walking behaviors were found for those who used the navigation system.

Description of Apps and Technology

In the study by Hettinga et al [43], software called TomTom was used for navigation support. They studied the safety of TomTom use by people with dementia and the effectiveness of familiar versus unfamiliar voice prompts. The use of navigation software was found to be safe based on observations of street-crossing behavior, response to navigation instructions, and number of occurrences of stopping at device prompts. Additionally, they observed that familiar voice prompts were more effective compared to unfamiliar ones. This was determined by measuring walking time, number of errors (route deviations and repeated instructions), and number of times assistance was requested. Warning sounds seemed to have a negative effect on wayfinding [43]. Participants were captured on video. These videos were coded for unsafe walking behaviors [43].

Reminiscence Therapy

Study Characteristics

Two studies incorporated a reminiscence therapy intervention [19,44]. Health outcomes included cognition, communication ability, mood, social interest, and psychological stability. Both studies showed some degree of efficacy [19].

Description of Apps and Technology

O'Rourke et al [19] used the Web-based video website YouTube to help facilitate reminiscence therapy in persons with dementia. Five of six participants showed improvement or stability in their communication ability over the 6-week pilot study. They concluded that the website YouTube is a suitable tool for delivering personalized computer-based reminiscence therapy [19]. YouTube also functions as a mobile app.

In their pilot study, Yasuda et al [44] used a videophone as both a remote reminiscence conversation system and as a schedule prompter system. The remote reminiscence conversation system shared reminiscence photos through a videophone triggered by the conversation partner on the other end of the phone. The study results showed following conversations with the system patients had improved psychological stability, verbal communication, and rates of instrumental ADL completion. The schedule prompter system used more than 10 different video reminders, such as prompts to take medications or to prepare meals. Individual experiments were conducted to evaluate each system using four patients with dementia. The first experiment evaluated the effectiveness of the remote reminiscence conversation system by performing two tasks: watching TV and remote video chatting. Results were measured using the Gottfries-Brane-Steen (GBS) scale that measures psychological variables such as confusion, irritability, anxiety, restlessness, reduced mood, and agony. The GBS scale is scored on a scale from zero (most stable) to six (least stable). Results showed that three of four patients obtained psychological stability as defined by the authors. The second experiment determined the effectiveness of a schedule prompter system in completing the scheduled task. Participants received three different types of video prompts: navigational prompts to move toward the computer, motivational prompts to inspire completion of tasks, and scheduled prompts to remind participants of tasks scheduled for completion. The results described the mean completion of tasks for the four patients to as 83% while using the prompter system [44].

Prompts and Multicomponent Interventions

Study Characteristics

We found four studies in these combined categories [20,44-46]. Health outcomes for this group included cognition, subjective report of cognition, mood, psychological stability, perceived autonomy, feeling of competence, the number of caregiver and patient unmet needs, caregiver burden, and the quality of life for the caregiver and patient. Of the four studies in this category, only the study by Yasuda et al [44] showed improvement in health outcomes [44].

Description of Apps and Technology

Yasuda et al [44] was previously described in the reminiscence therapy section because the study contained both reminiscence therapy and prompt components. Overall, in these studies, prompts were delivered in either an auditory or visual manner. The use of mobile devices in this group varied. Imbeault et al

[45] studied the impact of an electronic organizer "AP@LZ" on the cognition, subjective report of cognition, depression, and caregiver burden of three persons with Alzheimer disease. Unfortunately, there were no cognitive benefits. The results of the effect on depression and caregiver burden were unclear [45].

In 2009, Meiland et al [46] evaluated the use of a digital prosthetic, "COGNOW Day Navigator," by 12 persons with dementia and their caregivers. The digital prosthetic was designed to help with memory, social contacts, daily activities, and safety. Participants rated the study as useful and user-friendly. Effectiveness of the system could not be determined because of the short study duration and instability of the digital prosthetic prototype.

Lastly, one multicomponent intervention study by Hattink et al [20] met criteria for inclusion. This intervention, "Rosetta," targeted four domains of required support for persons with dementia: (1) prompts and reminders, (2) leisure, (3) communication, and (4) safety. The Rosetta intervention was tested with 42 patients with either MCI or Alzheimer disease. Contained in their measured health outcomes were cognition, perceived autonomy, feelings of competence, the number of unmet caregiver and patient needs, and quality of life of the caregiver and patient. No improvements in outcomes occurred following the intervention [20].

Engagement Interventions

Study Characteristics

Past studies demonstrated that recreational activities and engagement can lead to persons with dementia having more positive affect, decreased agitation, and decreased passivity [22,23]. Health outcomes varied by study. They consisted of mood, engagement, well-being, behavioral and psychological symptoms of dementia, agitation, and quality of life of the caregiver and patient. Five of the seven (71%) studies using a mHealth app intervention to engage people with dementia in activities had evidence of efficacy [47-53].

Description of Apps and Technology

Tablets were the technology of choice for these studies. In an adult day program, Leng et al [47] studied how iPad group activity sessions compared to traditional group engagement activities of cooking and arts and crafts for persons with dementia. The study found that for the persons with dementia who participated, iPad activities had at least an equal positive effect on mood, engagement, and well-being as compared to traditional group activities [47].

Another group, Lim et al [53], studied the usability of iPads by persons with dementia and their caregivers. In all, 95% of persons with dementia participating in the study had not previously used an iPad. Apps in the categories of art, music, simple interactive games, and relaxation were loaded onto each iPad. The patient-caregiver dyad was given the iPad to take home and use for 7 days with the recommendation that the caregiver provide 30 minutes of daily supervision and interaction while the person with dementia used the iPad. Nearly half of the caregivers indicated the iPad was somewhat, moderately, or extremely helpful.

Astell et al [49] studied 30 persons with dementia and measured the impact familiar and nonfamiliar games on a tablet had on the participants' enjoyment. A total of 90% of participants attempted to use the tablet. Regardless of familiarity, close to 90% of participants displayed enjoyment from playing games on a tablet.

Tyack et al [48] developed and tested an art viewing tablet app for persons with dementia as a well-being intervention. Twelve patient-caregiver dyads participated. Participants were asked to use the tablet and program five times over a 2-week span and were given a list of questions to facilitate conversation while using the app. There were no significant pre-post differences in any of the outcomes measured, including patient happiness, wellness, engagement, and patient and caregiver quality of life [48]. Despite this, there was a trend toward improvement in all outcome categories, indicating that a larger study might demonstrate effect.

Van der Ploeg et al [52] tested Internet video conferencing (Skype) and telephone calls with family members as an intervention to reduce agitation in persons with dementia. Preintervention and postintervention measurements showed no difference in Cohen-Mansfield Agitation Inventory scores [52].

Vahia et al [51] and Hsu et al [51] both used mobile apps on a tablet to treat symptoms of agitation and behavioral and psychological symptoms of dementia. Vahia et al found that their intervention reduced agitation and that all participants including those with severe dementia were able to use apps on the tablet [51]. In the small study by Hsu et al, those who received the tablet intervention had decreased use of "as-needed" medications to treat behavioral and psychological symptoms of dementia [50].

Exercise Intervention

Study Characteristics

The one exercise intervention that met inclusion criteria selected health outcomes of quality of life, self-efficacy, change in weekly steps taken, the 6-minute walk time, and the Mini-Physical Performance Test. None of these outcomes improved with the exercise intervention [54].

Description of App and Technology

Vidoni et al [54] sought to improve the health of persons with dementia by prescribing physical activity in conjunction with using a wearable mHealth device (Fitbit) in a tertiary medical clinic setting. The wearable device was an accelerometer that measured steps taken and communicated with the mHealth app installed on either a mobile device or computer. The randomized trial included participants with normal cognition and Alzheimer disease-related cognitive impairment. The study was designed to last 16 weeks, but only 8 weeks of data was reported. Normal controls improved on their baseline weekly steps taken, whereas those with cognitive impairment did not. Only 62% of those with cognitive impairment completed the intervention [54].

Discussion

Theoretical Implications

To our knowledge, this comprehensive review is the first to examine the efficacy of mHealth app interventions on the health outcomes of persons with MCI, Alzheimer disease, and dementia. We employed state-of-the-art methodology in identifying the relevant literature, rating the quality of studies, and extracting standardized data. Using a broad search strategy, we discovered this literature spread across the research and innovation outlets of multiple disciplines. The level of evidence supporting the use of mHealth app interventions for people with these disorders (MCI, Alzheimer disease, and dementia) was low, as reflected by a grade C level of evidence using the modified OCEBM rating system. Some degree of efficacy was seen in 58% (14/24) of all included studies. However, given the limited quality of these studies it is difficult to draw any firm conclusions. Given the potential size of the market for these interventions and their apparent potential for improving the care of persons with cognitive impairment, we uncovered a need not only for more research in this area, but also for greater agreement in study design and consensus on health outcome measures.

Unexpectedly, none of the 24 studies came from the primary computer science literature. We attribute this lack of computer science study selection to the inclusion criteria requiring at least one quantitative health outcome. In our review, we found a number of innovative studies of mHealth app interventions for people with cognitive impairment. In the computer science literature, however, the vast majority of these studies lacked any measure of patient health. This significant finding highlights the potential opportunity and need for collaboration between technology researchers and health care professionals to develop mHealth app interventions that improve the health of individuals with cognitive impairment.

Outcomes

Aim 1 was to examine and catalog the types of quantitative health outcomes used by mHealth app interventions for persons with MCI, Alzheimer disease, and dementia. We found that 101 of 476 (21.2%) articles fully reviewed were excluded because of a lack of health outcomes. There were also inconsistencies in the health outcomes selected by study investigators. There appears to be a lack of consensus on which health care outcomes should be used to evaluate mHealth app interventions targeting those with cognitive impairment. The most commonly used health outcomes were cognition, function, quality of life, mood and well-being, and behavioral and psychological symptoms of dementia. There exists a greater need for consistency in the health outcomes for persons with cognitive impairment in these studies.

Efficacy

Aim 2 of this study sought to determine the efficacy of mHealth app interventions focused on persons with cognitive impairment that included at least one quantitative health outcome in the study evaluation. This systematic review found that currently there is little evidence to support the efficacy of most mHealth app interventions improving the health of persons with cognitive impairment. The number of mHealth apps continues to grow [3]. However, there is limited oversight and rigor in the development and testing of these apps and their claims of benefit. Lumos Laboratory, doing business as Lumosity, served as the most recent well-publicized example of a company charged with unproven medical claims about the benefits received from their cognitive training mHealth app. The Federal Trade Commission alleged that the company misled consumers in claiming that their product “delays age-related mental decline and protects against dementia and Alzheimer’s disease” [55]. Without rigorous RCTs to prove efficacy of individual mHealth app interventions, mHealth apps run the risk of being the newest “snake oil” to treat dementias and associated disorders. More work needs to be done to determine which apps are effective at improving patient outcomes.

Limitations

Our review has several limitations. The review topic was broad, which creates challenges in terms of in-depth data synthesis. An adjudication process between two authors (DB and BS) was used to find consensus on article screening, full article review, and categorization with the OCEBM Levels of Evidence System. Interrater reliability or kappa was not calculated and is therefore a limitation. This review’s original search strategy did not focus explicitly on cognitive training; therefore, there may have been studies in this area that were not captured in this review. In the field of mHealth and in our review the majority of the studies are of small size and lacked sufficient quality in study design.

Future Directions

Greater need for clinical trials to test mHealth interventions necessitates involvement with the Food and Drug Administration (FDA), the government agency with the widest jurisdiction over the regulation of mobile health technologies [56]. The FDA issued its most recent nonbinding industry and staff guidance document on “mobile medical applications,” in 2013 and updated the document in 2015 [57]. In these reports, The FDA defines mobile medical apps as “a mobile app that meets the definition of device in section 201(h) of the Federal Food, Drug and Cosmetic Act (FD&C Act); and is either intended (1) to be used as an accessory to a regulated medical device; or (2) to transform a mobile platform into a regulated medical device” [57]. Section 201 (h) describes a medical device as one “...intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man...” or “...intended to affect the structure or any function of the body of man...” This definition includes software or apps on computers, websites, and handheld devices.

The FDA has made clear their intention to apply regulatory oversight to the subset of mobile medical apps that transform a mobile platform into a medical device [56,57]. They indicate that they reserve the right to enforce the guidelines at their discretion and will focus on mobile medical apps that have the potential to cause harm to patients [57].

Outside of efficacy and safety, other concerns exist with mHealth technologies, including privacy risks. In their 2016 *JAMA* letter, Blenner et al [58] showed that 81% of available

Android diabetes apps did not have privacy policies and that 48.4% of diabetes apps with privacy policies shared user information. In downloading and installing an app, health consumers often inadvertently give apps permission to collect personal information [58]. As noted by Blenner et al, there are currently no federal laws to prevent app companies from selling medical app data to third parties [58,59]. One could imagine the potentially damaging effect that could be incurred if sensitive medical information were to be shared with health or life insurance companies or with a prospective employer.

Although not the focus of this review, we found a large number of studies did not take end users' (persons with dementia or caregivers) input into consideration during the development of mHealth interventions, creating potential mismatch between the proposed solution and the participant needs. This suggests an opportunity for greater emphasis on user-centered design. User-centered design is a philosophy and methodology for designing and evaluating systems based on end-user involvement and a strong understanding of end-user characteristics, goals, tasks, needs, capabilities, and contexts [60]. This approach has the ultimate goal of optimal functioning of the human-machine system [61].

Despite these different concerns and drawbacks, mHealth apps possess great potential to improve the health and outcomes of people with cognitive impairment and offer advantages over traditional psychosocial interventions. With the growing numbers of persons with dementia, limitations of family caregivers, and work force shortage, there is an extraordinary need for engagement and social support of persons with dementia. One could imagine mHealth apps that could engage persons with dementia and help improve their mood through serious games or through a mHealth peer-support program. The aim of these apps would not be to replace caregivers or humans, but rather to allow for engagement when caregivers or others are busy or not immediately available.

The use of mHealth apps offer an opportunity to help persons with MCI, Alzheimer disease, and dementia maintain their independence longer than would be otherwise possible. Apps designed to prompt persons with dementia with necessary tasks such as taking medication, taking out the garbage, or cooking meals, could help persons with dementia complete these essential tasks rather than having to hire a caregiver or move into an assisted living to receive support. Similarly, wayfinding apps could improve persons with dementia's ability to travel safely in their community. And remote monitoring apps could allow caregivers and health care providers to supervise or receive notifications of changes in the health status of the person with dementia.

In addition, mHealth apps could improve measurement accuracy of variables pertaining to the health status of persons with dementia. Through ecological momentary assessment, the repeated sampling of a participant's experience or behaviors in real time, an app could provide up-to-date information on a person's mood status or neuropsychiatric symptoms of dementia [62], thus avoiding the pitfalls of recall bias seen with more traditional office-based questionnaires. Monitoring or prompting apps could help measure real-time decline in a person with dementia's cognition or function, offering providers a more accurate depiction of disease progression.

Virtual coaches hold promise as the next generation of mHealth apps and have the potential to help persons with cognitive impairment. Siewiorek et al [63] from Carnegie Mellon University and the National Science Foundation-funded Quality of Life Technology Center have pioneered the theory and design of virtual coaches [63,64]. A virtual coach moves beyond the rote and static reminders of a prompt system. Rather, a virtual coach adequately adapts to the needs of the user. The ideal qualities of a virtual coach as a cognitive aid, as outlined by Siewiorek et al [63], include the virtual coach reducing the number of cues as the user learns, matching the level of support to changes in the user's ability, allowing for caregivers to upload new capabilities to the virtual coach and providing consistent monitoring of adherence to a caregiver's instructions.

Conclusion

We found that many mHealth app interventions targeting those with cognitive impairment lack health outcomes as a part of their evaluation process and that there is a lack of consensus as to which health outcomes should be used. Of note, the most common health outcomes in this review were cognition, function, quality of life, mood and well-being, and behavioral and psychological symptoms of dementia. These align with outcomes for clinical trials for Alzheimer disease as described in previous systematic reviews [65-67]. From their comprehensive review of the scientific literature in 2017, Bentvelzen et al [67] distilled a list of best practice outcomes for dementia, called the Dementia Outcome Measurement Suite. Our results suggest that a best practice or at least greater consensus is needed around selection of appropriate health outcomes for mHealth app interventions targeting persons with cognitive impairment. The Dementia Outcome Measurement Suite is one best practice guideline that could be considered. The suite has six outcome domains: (1) cognition, (2) staging, (3) function, (4) behavior, (5) delirium, and (6) quality of life [67]. More in-depth discussion of these domains lies outside the scope of this review.

The evidence that use of mHealth app interventions improves the health of people with MCI, Alzheimer disease, and dementia is of limited quality. Evidence met criteria for grade C level of quality as per the OCEBM Levels of Evidence System. Study reports of efficacy were mixed with more than half of the studies (58%) showing some degree of effectiveness. More RCTs, a larger number of participants, and a design that minimizes bias are needed to better clarify the benefits of these types of interventions.

Acknowledgments

We would like to thank the support of Indiana University Center for Aging Research, Faculty of the Ruth Lilly Library at Indiana University, the Indiana University Department of Psychiatry and Malaz Boustani, MD, The Richard M Fairbanks Professor of Aging Research. This work was supported by the Richard M Fairbanks Professor of Aging Research Chair, Indiana University School of Medicine.

Authors' Contributions

All authors have read and approved the manuscript. All authors were involved in review and manuscript development.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search Strategy.

[PDF File (Adobe PDF File), 50KB - jmir_v19i8e301_app1.pdf]

References

1. Ericsson. 2016. Ericsson mobility report URL: https://www.ericsson.com/mobility-report [accessed 2016-11-16] [WebCite Cache ID 6m38yjOGD]
2. Shealy C. Solutions Review: Application Development. 2016 Jul 08. A collection of mobile application development statistics: growth, usage, revenue and adoption URL: https://solutionsreview.com/mobile-application-development/a-collection-of-mobile-application-development-statistics-growth-usage-revenue-and-adoption/ [accessed 2017-04-02] [WebCite Cache ID 6pPa7wLBU]
3. Levitas D. App Annie. 2016 Feb 10. App forecast: over $100 billion in revenue by 2020 URL: https://www.appannie.com/insights/app-annie-news/app-annie-releases-inaugural-mobile-app-forecast/ [accessed 2017-04-02] [WebCite Cache ID 6pPZSqt6S]
4. HIMSS. 2012 Jan 05. Definitions of mHealth URL: http://www.himss.org/definitions-mhealth [accessed 2016-11-06] [WebCite Cache ID 6louNg2wj]
5. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organization; 2011.
6. National Institutes of Health. NIH-Wide Strategic Plan Fiscal Years 2016-2020. 2015. URL: https://www.nih.gov/sites/default/files/about-nih/strategic-plan-fy2016-2020-508.pdf [accessed 2016-11-06] [WebCite Cache ID 6losEYpDB]
7. Hane FT, Robinson M, Lee BY, Bai O, Leonenko Z, Albert MS. Recent progress in Alzheimer's disease research, part 3: diagnosis and treatment. J Alzheimers Dis 2017;57(3):645-665 [FREE Full text] [doi: 10.3233/JAD-160907] [Medline: 28269772]
8. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. Alzheimers Dement 2016 Apr;12(4):459-509. [Medline: 27570871]
9. Oyebode JR, Parveen S. Psychosocial interventions for people with dementia: an overview and commentary on recent developments. Dementia (London) 2016 Jul 04:1. [doi: 10.1177/1471301216656096] [Medline: 27380931]
10. Rebok GW, Ball K, Guey LT, Jones RN, Kim H, King JW, ACTIVE Study Group. Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. J Am Geriatr Soc 2014 Jan;62(1):16-24 [FREE Full text] [doi: 10.1111/jgs.12607] [Medline: 24417410]
11. Hill NT, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
12. Michael D, Chen S. Serious Games: Games That Educate, Train, and Inform. Mason, OH: Course Technology; 2005.
13. Teipel S, Babiloni C, Hoey J, Kaye J, Kirste T, Burmeister OK. Information and communication technology solutions for outdoor navigation in dementia. Alzheimers Dement 2016 Jun;12(6):695-707 [FREE Full text] [doi: 10.1016/j.jalz.2015.11.003] [Medline: 26776761]
14. Lanza C, Knörzer O, Weber M, Riepe MW. Autonomous spatial orientation in patients with mild to moderate Alzheimer's disease by using mobile assistive devices: a pilot study. J Alzheimers Dis 2014;42(3):879-884. [doi: 10.3233/JAD-140063] [Medline: 24958461]
15. Cotelli M, Manenti R, Zanetti O. Reminiscence therapy in dementia: a review. Maturitas 2012 Jul;72(3):203-205. [doi: 10.1016/j.maturitas.2012.04.008] [Medline: 22607813]
16. Tadaka E, Kanagawa K. Effects of reminiscence group in elderly people with Alzheimer disease and vascular dementia in a community setting. Geriatr Gerontol Int 2007 Jun;7(2):167-173. [doi: 10.1111/j.1447-0594.2007.00381.x]

17. Lai CK, Chi I, Kayser-Jones J. A randomized controlled trial of a specific reminiscence approach to promote the well-being of nursing home residents with dementia. Int Psychogeriatr 2004 Mar;16(1):33-49. [Medline: 15190995]
18. Okumura Y, Tanimukai S, Asada T. Effects of short-term reminiscence therapy on elderly with dementia: A comparison with everyday conversation approaches. Psychogeriatrics 2008 Sep;8(3):133. [doi: 10.1111/j.1479-8301.2008.00236.x]
19. O'Rourke J, Tobin F, O'Callaghan S, Sowman R, Collins DR. 'YouTube': a useful tool for reminiscence therapy in dementia? Age Ageing 2011 Nov;40(6):742-744. [doi: 10.1093/ageing/afr100] [Medline: 21880626]
20. Hattink BJ, Meiland FJ, Overmars-Marx T, de BM, Ebben PW, van BM, et al. The electronic, personalizable Rosetta system for dementia care: exploring the user-friendliness, usefulness and impact. Disabil Rehabil Assist Technol 2016 Jul;11(1):61-71. [doi: 10.3109/17483107.2014.932022] [Medline: 24989993]
21. Cahill S, Begley E, Faulkner J, Hagen I. “It gives me a sense of independence”? Findings from Ireland on the use and usefulness of assistive technology for people with dementia. Technol Disabil 2007 Jan;19(2):133-142.
22. Kolanowski AM, Litaker M, Buettner L. Efficacy of theory-based activities for behavioral symptoms of dementia. Nurs Res 2005;54(4):219-228. [Medline: 16027564]
23. Tak S, Beck C, Hong S. Feasibility of providing computer activities for nursing home residents with dementia. Nonpharmacol Ther Dement 2013;3(1):1-10 [FREE Full text] [Medline: 25343006]
24. Abraha I, Rimland JM, Trotta FM, Dell'Aquila G, Cruz-Jentoft A, Petrovic M, et al. Systematic review of systematic reviews of non-pharmacological interventions to treat behavioural disturbances in older patients with dementia. The SENATOR-OnTop series. BMJ Open 2017 Mar 16;7(3):e012759 [FREE Full text] [doi: 10.1136/bmjopen-2016-012759] [Medline: 28302633]
25. Yoon DH, Kang D, Kim H, Kim J, Song HS, Song W. Effect of elastic band-based high-speed power training on cognitive function, physical performance and muscle strength in older women with mild cognitive impairment. Geriatr Gerontol Int 2017 May;17(5):765-772. [doi: 10.1111/ggi.12784] [Medline: 27396580]
26. Erickson KI, Voss MW, Prakash RS, Basak C, Szabo A, Chaddock L, et al. Exercise training increases size of hippocampus and improves memory. Proc Natl Acad Sci U S A 2011 Feb 15;108(7):3017-3022 [FREE Full text] [doi: 10.1073/pnas.1015950108] [Medline: 21282661]
27. Bamidis PD, Vivas AB, Styliadis C, Frantzidis C, Klados M, Schlee W, et al. A review of physical and cognitive interventions in aging. Neurosci Biobehav Rev 2014 Jul;44:206-220. [doi: 10.1016/j.neubiorev.2014.03.019] [Medline: 24705268]
28. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
29. PROSPERO: International Prospective Register of Systematic Reviews. URL: https://www.crd.york.ac.uk/PROSPERO/about.php?about=about [accessed 2017-08-19] [WebCite Cache ID 6sq33bS5e]
30. Meho LI, Rogers Y. Citation counting, citation ranking, and -index of human-computer interaction researchers: a comparison of Scopus and Web of Science. J Am Soc Inf Sci 2008 Sep;59(11):1711-1726. [doi: 10.1002/asi.20874]
31. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. Centre for Evidence-based Medicine. OCEBM Levels of Evidence URL: http://www.cebm.net/ocebm-levels-of-evidence/ [accessed 2017-08-19] [WebCite Cache ID 6spyPoWpY]
32. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-e34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
33. Barnes D, Yaffe K, Belfor N, Jagust WJ, DeCarli C, Reed BR, et al. Computer-based cognitive training for mild cognitive impairment: results from a pilot randomized, controlled trial. Alzheimer Dis Assoc Disord 2009;23(3):205-210. [doi: 10.1097/WAD.0b013e31819c6137] [Medline: 19812460]
34. Chan SC, Chan CC, Derbie AY, Hui I, Tan DG, Pang MY, et al. Chinese calligraphy writing for augmenting attentional control and working memory of older adults at risk of mild cognitive impairment: a randomized controlled trial. J Alzheimers Dis 2017 Jun;58(3):735-746. [doi: 10.3233/JAD-170024] [Medline: 28482639]
35. Tárraga L, Boada M, Modinos G, Espinosa A, Diego S, Morera A, et al. A randomised pilot study to assess the efficacy of an interactive, multimedia tool of cognitive stimulation in Alzheimer's disease. J Neurol Neurosurg Psychiatry 2006 Oct;77(10):1116-1121 [FREE Full text] [doi: 10.1136/jnnp.2005.086074] [Medline: 16820420]
36. Finn M, McDonald S. Computerised cognitive training for older persons with mild cognitive impairment: a pilot study using a randomised controlled trial design. Brain Impair 2011 Dec;12(3):187-199. [doi: 10.1375/brim.12.3.187]
37. Gooding AL, Choi J, Fiszdon JM, Wilkins K, Kirwin PD, van Dyck CH, et al. Comparing three methods of computerised cognitive training for older adults with subclinical cognitive decline. Neuropsychol Rehabil 2016 Oct;26(5-6):810-821. [doi: 10.1080/09602011.2015.1118389] [Medline: 26674122]
38. Han JW, Oh K, Yoo S, Kim E, Ahn K, Son Y, et al. Development of the ubiquitous spaced retrieval-based memory advancement and rehabilitation training program. Psychiatry Investig 2014 Jan;11(1):52-58 [FREE Full text] [doi: 10.4306/pi.2014.11.1.52] [Medline: 24605124]
39. Mansbach WE, Mace RA, Clark KM. The efficacy of a computer-assisted cognitive rehabilitation program for patients with mild cognitive deficits: a pilot study. Exp Aging Res 2017;43(1):94-104. [doi: 10.1080/0361073X.2017.1258256] [Medline: 28067610]

40. Hsiung G, Kupferschmidt A, Naus K, Feldman H, Jacova C. Cognitive and clinical outcomes in a pilot study using a handheld computer game for cognitive training of patients with Mild Cognitive Impairment (MCI). Alzheimers Dement 2009 Jul;5(4):411. [doi: 10.1016/j.jalz.2009.04.1003]
41. Manera V, Petit P, Derreumaux A, Orvieto I, Romagnoli M, Lyttle G, et al. 'Kitchen and cooking,' a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study. Front Aging Neurosci 2015 Mar;7:24 [FREE Full text] [doi: 10.3389/fnagi.2015.00024] [Medline: 25852542]
42. Merilampi S, Sirkka A, Leino M, Koivisto A, Finn E. Cognitive mobile games for memory impaired older adults. J Assist Technol 2014 Dec 09;8(4):207-223. [doi: 10.1108/JAT-12-2013-0033]
43. Hettinga M, De Boer J, Goldberg E, Moelaert F. Navigation for people with mild dementia. Stud Health Technol Inform 2009;150:428-432. [Medline: 19745347]
44. Yasuda K, Kuwahara N, Kuwabara K, Morimoto K, Tetsutani N. Daily assistance for individuals with dementia via videophone. Am J Alzheimers Dis Other Demen 2013 Aug;28(5):508-516. [doi: 10.1177/1533317513494440] [Medline: 23813611]
45. Imbeault H, Gagnon L, Pigot H, Giroux S, Marcotte N, Cribier-Delande P, et al. Impact of AP@LZ in the daily life of three persons with Alzheimer's disease: long-term use and further exploration of its effectiveness. Neuropsychol Rehabil 2016 Apr 28:1-24. [doi: 10.1080/09602011.2016.1172491] [Medline: 27126266]
46. Meiland FJ, Bouman AI, Sävenstedt S, Bentvelzen S, Davies RJ, Mulvenna MD, et al. Usability of a new electronic assistive device for community-dwelling persons with mild dementia. Aging Ment Health 2012 Jul;16(5):584-591. [doi: 10.1080/13607863.2011.651433] [Medline: 22360649]
47. Leng FY, Yeo D, George S, Barr C. Comparison of iPad applications with traditional activities using person-centred care approach: impact on well-being for persons with dementia. Dementia (London) 2014 Mar 01;13(2):265-273. [doi: 10.1177/1471301213494514] [Medline: 24339097]
48. Tyack C, Camic PM, Heron MJ, Hulbert S. Viewing art on a tablet computer. J Appl Gerontol 2015 Dec 01:733464815617287. [doi: 10.1177/0733464815617287] [Medline: 26675353]
49. Astell AJ, Joddrell P, Groenewoud H, de Lange J, Goumans M, Cordia A, et al. Does familiarity affect the enjoyment of touchscreen games for people with dementia? Int J Med Inform 2016 Jul;91:e1-e8. [doi: 10.1016/j.ijmedinf.2016.02.001] [Medline: 26897552]
50. Hsu C, Lee E, Vahia I. Tablet computer use in reducing neuropsychiatric symptoms in patients with dementia. J Am Geriatr Soc 2016 May;64(S1):S90.
51. Vahia IV, Kamat R, Vang C, Posada C, Ross L, Oreck S, et al. Use of tablet devices in the management of agitation among inpatients with dementia: an open-label study. Am J Geriatr Psychiatry 2016 Aug 04:860-864. [doi: 10.1016/j.jagp.2016.07.011] [Medline: 27746070]
52. Van der Ploeg ES, Eppingstall B, O'Connor DW. Internet video chat (Skype) family conversations as a treatment of agitation in nursing home residents with dementia. Int Psychogeriatr 2015 Nov 12;28(04):697-698. [doi: 10.1017/s1041610215001854]
53. Lim FS, Wallace T, Luszcz MA, Reynolds KJ. Usability of tablet computers by people with early-stage dementia. Gerontology 2013;59(2):174-182. [doi: 10.1159/000343986] [Medline: 23257664]
54. Vidoni ED, Watts AS, Burns JM, Greer CS, Graves RS, Van Sciver A, et al. Feasibility of a memory clinic-based physical activity prescription program. J Alzheimers Dis 2016 Apr 21;53(1):161-170. [doi: 10.3233/JAD-160158] [Medline: 27104905]
55. Federal Trade Commission. 2015 Jan 06. Lumosity to pay $2 million to settle FTC deceptive advertising charges for its "brain training" program URL: https://www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges [accessed 2017-04-02] [WebCite Cache ID 6pPYtU3fr]
56. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med 2014 Jul 24;371(4):372-379. [doi: 10.1056/NEJMhle1403384] [Medline: 25054722]
57. US Department of Health and Human Services Food and Drug Administration, Center for Devices and Radiological Health, Center for Biologics Evaluation and Research. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD: Food and Drug Administration; Feb 09, 2015.
58. Blenner SR, Köllmer M, Rouse AJ, Daneshvar N, Williams C, Andrews LB. Privacy policies of android diabetes apps and sharing of health information. JAMA 2016 Mar 08;315(10):1051-1052. [doi: 10.1001/jama.2015.19426] [Medline: 26954415]
59. Andrews L. I Know Who You Are and I Saw What You Didocial Networks and the Death of Privacy. New York: Free Press; 2012.
60. Fisk A, editor. Designing for Older Adults: Principles and Creative Human Factors Approaches. 2nd ed. Boca Raton, FL: CRC Press; 2009.
61. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
62. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]

63. Siewiorek D, Smailagic A, Dey A. Architecture and applications of virtual coaches. Proc IEEE 2012 Aug;100(8):2472-2488. [doi: 10.1109/JPROC.2012.2200560]
64. Carnegie Mellon University. Quality of Life Technology Center URL: http://www.cmu.edu/qolt/ [accessed 2017-04-02] [WebCite Cache ID 6pPYjHPQ2]
65. Schneider LS, Mangialasche F, Andreasen N, Feldman H, Giacobini E, Jones R, et al. Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. J Intern Med 2014 Mar;275(3):251-283 [FREE Full text] [doi: 10.1111/joim.12191] [Medline: 24605808]
66. Cooper C, Mukadam N, Katona C, Lyketsos CG, Blazer D, Ames D, et al. Systematic review of the effectiveness of pharmacologic interventions to improve quality of life and well-being in people with dementia. Am J Geriatr Psychiatry 2013 Feb;21(2):173-183. [doi: 10.1016/j.jagp.2012.10.018] [Medline: 23343491]
67. Bentvelzen A, Aerts L, Seeher K, Wesson J, Brodaty H. A comprehensive review of the quality and feasibility of dementia assessment measures: the dementia outcomes measurement suite. J Am Med Dir Assoc 2017 Mar 07:1. [doi: 10.1016/j.jamda.2017.01.006] [Medline: 28283381]

Abbreviations

ADL: activities of daily living
FDA: Food and Drug Administration
GBS: Gottfries-Brane-Steen
MCI: mild cognitive impairment
MMSE: Mini-Mental Status Examination
NIH: National Institutes of Health
OCEBM: Oxford Centre for Evidence-based Medicine
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
RCT: randomized controlled trials

Edited by G Eysenbach; submitted 01.04.17; peer-reviewed by P Bamidis, B Davis, SFA Shah, M Ashford; comments to author 24.04.17; revised version received 15.07.17; accepted 01.08.17; published 30.08.17.

Please cite as:
Bateman DR, Srinivas B, Emmett TW, Schleyer TK, Holden RJ, Hendrie HC, Callahan CM
Categorizing Health Outcomes and Efficacy of mHealth Apps for Persons With Cognitive Impairment: A Systematic Review
J Med Internet Res 2017;19(8):e301
URL: http://www.jmir.org/2017/8/e301/
doi:10.2196/jmir.7814
PMID:28855146

Original Paper

A Peer-Led, Social Media-Delivered, Safer Sex Intervention for Chinese College Students: Randomized Controlled Trial

Wai Han Sun[1], BSocSc, MPhil; Carlos King Ho Wong[1], BSc, MPhil, PhD; William Chi Wai Wong[1], MB, ChB, MD, MPH, FRCGP, FRACGP, honMFPH

Department of Family Medicine and Primary Care, The University of Hong Kong, Ap Lei Chau, Hong Kong, China (Hong Kong)

Corresponding Author:
William Chi Wai Wong, MB, ChB, MD, MPH, FRCGP, FRACGP, honMFPH
Department of Family Medicine and Primary Care
The University of Hong Kong
Ap Lei Chau Clinic
3/F, 161 Main Street
Ap Lei Chau, Hong Kong,
China (Hong Kong)
Phone: 852 2518 5657
Fax: 852 2814 7475
Email: wongwcw@hku.hk

Abstract

Background: The peer-led, social media-delivered intervention is an emerging method in sexual health promotion. However, no research has yet investigated its effectiveness as compared with other online channels or in an Asian population.

Objective: The objective of this study is to compare a peer-led, social media-delivered, safer sex intervention with a sexual health website. Both conditions target Chinese college students in Hong Kong.

Methods: A randomized controlled trial was conducted with a peer-led, safer sex Facebook group as the intervention and an existing online sexual health website as the control. The intervention materials were developed with peer input and followed the information-motivation-behavioral skills model; the intervention was moderated by peer educators. The participants filled out the online questionnaires before and after the 6-week intervention period. Outcome evaluations included safer sex attitudes, behavioral skills, and behaviors, while process evaluation focused on online experience, online-visiting frequency, and online engagement. The effect of online-visiting frequency and online engagement on outcome variables was investigated.

Results: Of 196 eligible participants—100 in the control group and 96 in the intervention group—who joined the study, 2 (1.0%) control participants joined the Facebook group and 24 of the remaining 194 participants (12.4%) were lost to follow-up. For the process evaluation, participants in the intervention group reported more satisfying online experiences (P<.001) and a higher level of online-visiting frequency (P<.001). They also had more positive comments when compared with the control group. For outcome evaluation, within-group analysis showed significant improvement in condom use attitude (P=.02) and behavioral skills (P<.001) in the intervention group, but not in the control group. No significant between-group difference was found. After adjusting for demographic data, increased online-visiting frequency was associated with better contraceptive use behavioral intention (P=.05), better behavioral skills (P=.02), and more frequent condom use (P=.04).

Conclusions: A peer-led, social media-delivered, safer sex intervention was found to be feasible and effective in improving attitudes toward condom use and behavioral skills, but was not significantly more effective than a website. Future research may focus on the long-term effectiveness and cost-effectiveness of this popular method, as well as the potential cultural differences of using social media between different countries.

Trial Registration: Chinese Clinical Trial Registry (ChiCTR): ChiCTR-IOR-16009495; http://www.chictr.org.cn/showprojen.aspx?proj=16234 (Archived by WebCite at http://www.webcitation.org/6s0Fc2L9T)

(J Med Internet Res 2017;19(8):e284) doi:10.2196/jmir.7403

KEYWORDS

sex education; social media; randomized controlled trial

Introduction

Youth Sexual Health and Education

The sexual and reproductive health of young people constitutes a major health concern, as they are at risk of unwanted pregnancy [1] or of contracting sexually transmitted infections (STIs) [2]. Indeed, unsafe sex and the lack of contraception were the second- and fourth-most common risk factors for cause-specific, disability-adjusted life years for young people globally [3]. In China, the prevalence of condom use among sexually active college students is low at 49.7% [4]; more alarmingly, over 40% of students reported having either impregnated a girlfriend or having an unwanted pregnancy [5]. In Hong Kong, chlamydia prevalence was high in sexually active young women and men, at 5.8% and 4.1%, respectively [6]. A study of Chinese college students also showed poor sexual health knowledge and negative attitudes toward contraception [5]. Similar to their western peers, Chinese young people appear to have less than optimal knowledge, attitudes, and behaviors in relation to safer sex practice. In Hong Kong, no mandatory sex education is required in schools, although it is sometimes provided within extracurricular activities or on an ad hoc basis at the discretion of individual institutions. Instead, most young people learn about sex from their peers, from social workers, or from the Internet [7].

Safer sex promotion via sexual health education has been found to successfully increase sexual health knowledge and change attitudes and behaviors among teens and young adults [8,9]. Among them, a number of trials of the information-motivation-behavioral (IMB) skills model found information on risk reduction, motivation, and behavioral skills to be fundamental determinants in changing HIV risk behaviors [10]. Globally, safer sex interventions targeting college students combined both IMB model and peer-led approaches, with favorable outcomes [11-13]; for instance, the frequency of keeping condoms at hand increased, as did condom use [11,14,15]. Furthermore, peer-led sex education is a popular approach in youth sex education, as it is believed that young people are more easily influenced by their peers [16]. As found in a systematic review and meta-analysis, which examined a whole spectrum of peer-led sex education studies in more-developed countries, peer-led sex education was effective in changing sexual health knowledge and attitudes among young people [17].

Social Media as a New Tool

The Internet provides a unique opportunity for safer sex promotion due to its accessibility, confidentiality, and potential for creating personalized messages [18]. As the Internet has become more popular, the amount of research relating to online youth sex education has increased. While there was research on the effectiveness of sexual health promotion through short message service (SMS) text messaging, websites, and online games targeting young people [19-22], research into the use of social media on sex education is mainly exploratory in nature and focuses on process evaluation [23,24]. As young people nowadays spend a large amount of their time online on social media [25], and sexual health has been reported as the most frequently searched topic by young people [23,26], further research is needed to evaluate the effectiveness of social media in promoting safer sex among young people. Furthermore, a systematic examination of sexual health promotional activities on the Internet identified 178 activities on social networking sites [27] and revealed that social media provided valuable peer and social support due to its interactive features [23]. In addition, trained peer educators' involvement in the development and implementation processes adds value to the interventions [28].

Youth sexual health promotion via social media has many advantages, namely, its popularity among young people; its nature of interactivity, large potential reach, and instant reaction; as well as its facilitation of the use of multimedia materials. This can increase participants' online engagement and provide a new way of learning, which could bring about safer sex behaviors effectively. Combined with the peer-led approach, this form of online sex education further allows tailor-made content for a large target population involving the consumers and the empowering of youth. These characteristics may contribute to a more successful intervention as it reaches more young people and makes the content more relevant to young people. As shown in previous research, this approach could improve offline safer sex behaviors [28,29].

To the authors' knowledge, none of the previous research has compared the effectiveness of sexual health promotion delivered on social media with its effectiveness when delivered via other online channels. An active control group is needed to investigate the relative effectiveness of a social media-delivered intervention and other existing digital interventions targeting youth, such as websites or SMS text messaging. This research compares a social media-delivered intervention with an existing sexual health website, as websites have been shown to be effective in bringing about safer sex behaviors [30]. Apart from the mode of delivery, the content of the website is expert led rather than peer led. Moreover, previous trials were conducted in developed countries, such as the United States or Australia, and none have been conducted in Asia. It is therefore unclear whether the findings could be generalized across cultures.

Study Aims

This study aimed to evaluate the effectiveness of a peer-led, safer sex, social media-delivered intervention among Chinese college students in Hong Kong. Social media has extra benefits for health promotion as compared to traditional websites; for example, it allows peer-to-peer discussion [23]. Among the few evaluative studies on safer sex interventions through social media using the peer-led approach, positive results were evident, such as increasing condom use and increasing STI-testing behaviors [28,29]. Moreover, peer-led approaches were found to be more effective than adult-led approaches in previous literature reviews [17,31]. A sexual health website with expert-led material that did not involve the use of social media was used as the control in this study.

The first hypothesis was that the intervention would have a larger positive effect than would the control on attitudes, behavioral skills, and behaviors in relation to safer sex. The second hypothesis was that the intervention group would have a more positive online experience and higher online-visiting

frequency than the control group. The third hypothesis was that the higher the online usage and engagement, the larger the effect the intervention would have on the outcome measures.

Methods

Design

A randomized controlled trial with two arms was conducted on college students using a peer-led, safer sex intervention; the trial was developed with active youth input based on the IMB model and conducted via social media in the form of a Facebook group. An existing online sexual health website was used as the control condition.

Participants and Sample Size

The participants were recruited from three different colleges in Hong Kong. The three universities include two large-scale universities and one small-scale university in different districts of Hong Kong. This enabled the recruitment from a large student base with a wide range of disciplines and geographic locations. The research was promoted to the students through mass emails, student societies, posters around campus, and social media. The eligibility criteria included being an undergraduate student and aged under 25 years.

A previous study on safer sex motivation, which used the same validated scale, showed that the effect size was medium (ie, 0.5) [32]. With the confidence level set at 95% and power as 0.9, the sample size calculation yielded a sample size of 172 [33]. As previous studies indicated that the dropout rate in online research could be as high as 15% [34,35], the final number estimated was 99 individuals in each group.

Questionnaire and Procedures

The participants signed up voluntarily through an open online survey website, SurveyMonkey, and eligibility questions were asked to screen participants. Screened participants were given information about the research online and they gave their informed consent before they filled in the self-administered survey. The information included a brief introduction to the study, a description of the targeted group, and the benefits and risks of joining the study. The participants were reminded that they could always contact the research team through email or phone if they had any questions or if they came to any harm during the study. Multiple identities were prevented by the survey software prohibiting multiple submissions from the same IP address and through email confirmation. Researchers also sent out the postintervention survey through email. Apart from process and outcome evaluations, the survey also collected sociodemographic information, including age, gender, sexual orientation, and the number of romantic relationships. There were 33 items in total on seven screen pages. A completeness check was applied, but participants were not able to review and change their answers. At the end of the survey, the online survey software randomized the participants into one of the two groups. Our online survey software did the random assignment by allowing the researcher to set a certain percentage for each group. In this study, 50% of the participants were randomized into the intervention condition, while the other 50% were randomized into the control condition. The collected data were protected by password and could only be accessed by the researchers through the institute computer. The usability and technical functionality of the online questionnaire was tested before launching. By completing the baseline and postintervention surveys, participants would receive HK $100 (US $1=HK $7.80) as an incentive.

Intervention

The intervention material was developed by a group of trained peer educators following the three elements in the IMB model: safe sex knowledge, motivation, and behavioral skills. The details of the peer input are described below. The peer-led intervention was delivered through Facebook, the most popular social media platform among Hong Kong college students. The research team from the host university collaborated with a local community-based organization, Sticky Rice Love, to develop the intervention material with their young educators. This nonprofit organization has been promoting sexual health to Hong Kong youth on social media since 2014, following a peer-led approach. They recruited young people as educators and empowered them to create online sex education material. They have thousands of followers on their Facebook and Instagram profiles.

The peer educators were recruited through Sticky Rice Love in October 2015; they went through training followed by a final assessment in April 2016. The selection criteria included a high attendance in the training sessions, ample time available to be involved in the program, passion toward sex education, accurate sexual health knowledge, and being nonjudgmental when conducting sex education. There were six training sessions covering topics relating to (1) attitude of a sex education provider, (2) sexual violence, (3) female sexual health, (4) male sexual health, (5) safe sex and STIs, and (6) communication skills in a relationship. They were conducted by the Sticky Rice Love staff, a doctor, and a social worker in the specific field. Training activities included direct teaching, interactive games, role play, question sessions, videos, and condom demonstration. There were 19 youth educators with a mean age of 22.5 years (range 19-24) and slightly more female members (11/19, 58%).

The peer educators attended a session about the IMB model in June 2016. They then created the intervention material at their monthly meetings. They had three meetings to finalize the intervention material. The educators were responsible for the selection of the intervention content, presentation, and production. The content included pregnancy, STIs, and safe sex practices (eg, abstinence, condom use, and pills) in the form of text, images, GIFs, and videos. The participants were able to "like" or comment on the Facebook posts. The researchers (WCWW and WHS) examined the information and reminded the peer educators of any missing aspect from the IMB model. One of the investigators (WCWW), who is a physician with over 15 years of sexual health research experience, provided professional input and advice on its development and ensured that the sexual health information was correct and accurate.

The intervention lasted for 6 weeks, from October to November 2016, and the frequency of posts was three to four per week, totaling 21 posts. The intervention was delivered to a "secret" Facebook group created purely for this project. A Facebook

group refers to a page in Facebook where group members can post on the wall and interact through discussion threads. As the privacy setting was set to "secret," only group members could see the group and know who else was in the group. Only people who were invited could join the group. This was used as it could provide a closed network for research purposes. In addition, a Facebook group lets the researchers record the online usage data of each participant, which could be used as an objective process measure. The participants would receive Facebook notifications for each new post. Peer educators were added to the group to moderate by posting questions and responding to comments.

Control

The participants randomized to the control group would be given the sex education website link of the Hong Kong Family Planning Association [36], the most established, government-funded sex education organization in Hong Kong. The website was set up in 2004 and targets young people in its promotion of sexual health online. A range of information related to safe sex, including contraception and STIs, is included on the website with credible sources, which is similar to the intervention in terms of content. The website is the most well-regarded existing mode of delivery for safe sex information currently available to college students. It serves as an online promotion of sexual health without the peer-led and social media elements and belongs to the Web 1.0 stage of the World Wide Web. Weekly emails were sent to the participants to remind them to visit the website.

Evaluation

Process Measures

The online experience of the participants was assessed according to the important elements of online sexual health promotion found in previous qualitative research [30,37]. These elements are (1) credibility, (2) personal relevance, (3) respect for autonomy, (4) comfort to learn, (5) engaging experience, (6) ease of use, and (7) privacy. Seven items were included on a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). An open-ended question was also included to allow the participants to express their feelings toward the safer sex intervention in their own words.

All participants were asked to self-report the frequency with which they visited their assigned online education platform on a 5-point scale, from 1 (Rarely) to 5 (Always). For the intervention group, online engagement was also recorded when a participant reacted or commented on the Facebook group content.

Outcome Measures

Following the IMB model, the measured outcomes included motivation, behavioral skills, and behavior. Although information is one of the constructs in the model, it was found to be an inconsistent construct in relation to behavior [38]. It was included at the development stage but not at the evaluation stage. Motivation change included condom use, attitude, and contraceptive use behavioral intention [10]. Outcome measures were assessed through scales adopted from an IMB model-based HIV prevention intervention and were on a scale that ranged from 1 (Strongly Disagree) to 4 (Strongly Agree) [32]. The condom use attitudes scale score included four items that assessed participants' beliefs about condom use with a reported internal consistency, alpha, at .82. The contraceptive use behavioral intention scale had three items with a reported internal consistency, alpha, at .61.

To measure changes in behavioral skills, we applied the perceived difficulty and ease of condom use scale, which was first validated in 1998 [39] and was used in other IMB model-based studies [15,40]. It is a 5-point scale ranging from 1 (Very Difficult) to 7 (Very Easy), with seven items. The reported internal reliability, alpha, was .91 among college students [15]. For behavioral changes, the frequency of condom use was measured on a 5-point scale ranging from 1 (Never) to 5 (Always). The measures are included in Multimedia Appendix 1.

Statistical Methods

Data were analyzed using SPSS version 23 (IBM Corp). Differences in demographic characteristics at baseline between the intervention and control group were assessed using chi-square tests for categorical variables—age group, gender, sexual orientation, and sexually active status. Differences in outcome variables at baseline were assessed using independent *t* tests.

For process evaluation, the aim was to examine if there were differences in online experience and frequency of online visiting between intervention and control groups. Quantitatively, independent *t* tests were used to examine the differences between online experience scores and frequency of online visiting. Qualitatively, the opinions of the participants toward the Facebook group and website were summarized to understand any differences between the two conditions.

For outcome evaluation, the aim was to examine whether the kind of promotion had a significant effect on outcome variables between intervention and control groups. Paired *t* tests were used to examine the within-group changes in outcomes. Repeated measures analysis of variance (ANOVA) was used to compare the effectiveness between the two groups using the Time X Treatment analysis. In addition, linear regressions were performed to assess the effect of frequency of online visiting and online engagement on outcome variables, with adjustment for age, gender, and sexual orientation. Participants in the intervention group were divided into low, medium, and high online engagement according to the number of their Facebook reactions, comments, and posts, but their number of views was not counted. Participants in the control group were regarded as a control for online engagement.

All significance tests were two-tailed and the level of significance was set at a *P* value of less than .05. For missing data, the missing value was imputed by an average score of the group at the respective time point. We applied the intention-to-treat principle, that is, we assumed that participants would remain in their initially allocated group until the end of trial.

Ethical Approval

This research was approved by the Institutional Review Board of the University of Hong Kong, Hospital Authority Hong Kong West Cluster, in August 2016 (No. UW 16-419).

Results

Overview

A total of 287 people initially signed up, while 196 eligible participants completed the preintervention survey and successfully joined the study (see Figure 1) in October 2016; the recruitment period was in September 2016. Participants were randomized to the intervention group (n=96) and the control group (n=100). There were 2 participants in the control group that decided to join the Facebook group; they were subsequently removed from the study. Follow-up was done immediately after the 6-week intervention; 24 participants were lost to follow-up, giving a dropout rate of 12.4% (24/194).

Baseline Characteristics

Demographic information, including age, gender, sexual orientation, and sexual experience, was requested in the preintervention survey (see Table 1). Participants were mostly aged 19-20 years (85/194, 43.8%), female (131/194, 67.5%), heterosexual (156/194, 80.4%), and sexually inactive (144/194, 74.2%), and most participants never had sex (123/194, 63.4%). The baseline level of the participants' outcome performances is presented in Table 2. No significant difference between the intervention and control groups at baseline was found.

Figure 1. Participant flow diagram.

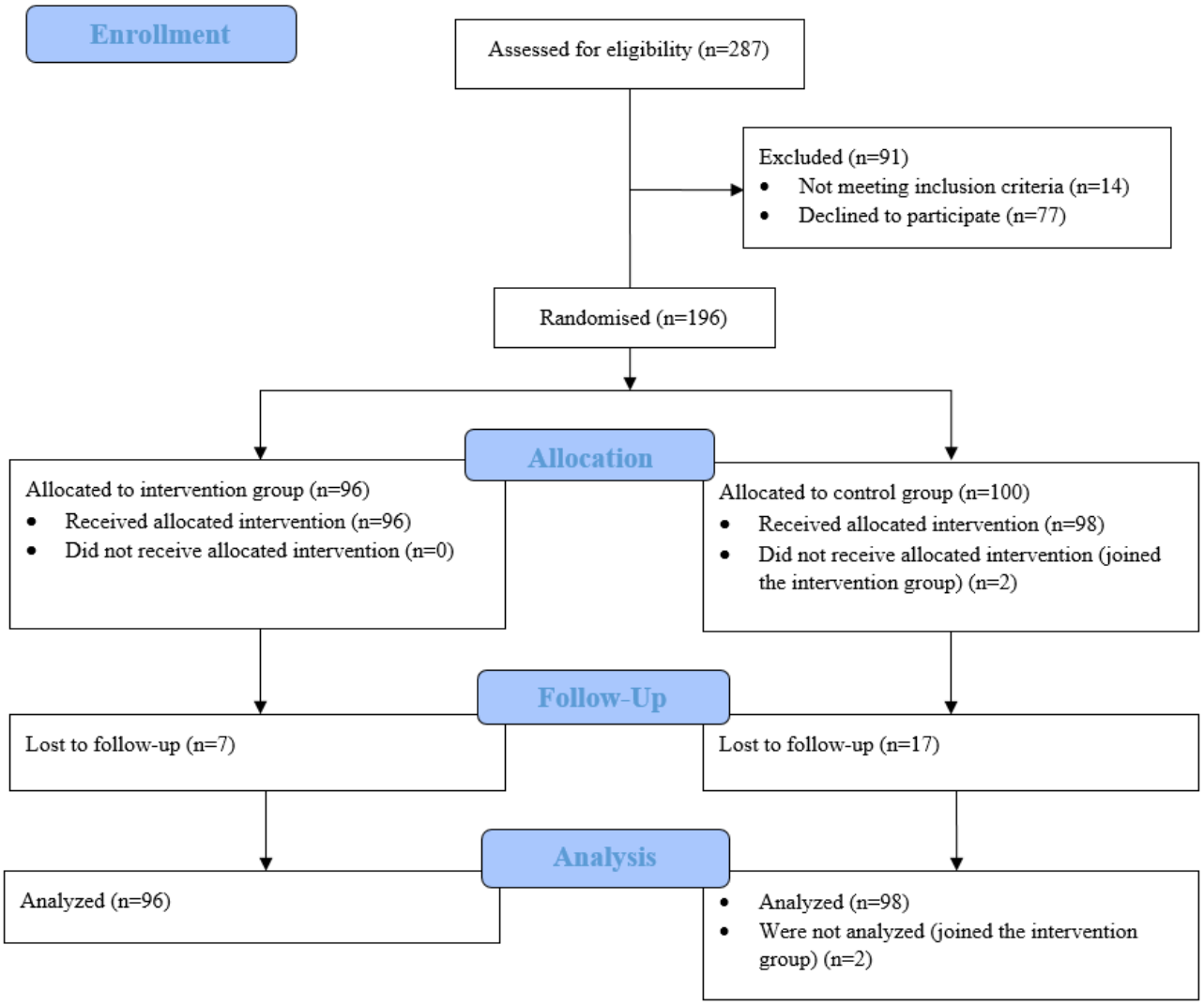


Table 1. Baseline characteristics and outcomes of participants.

Characteristics	Social media intervention group (n=96)	Website control group (n=98)	Total (N=194)	*P* [a]
Age (years), n (%)				.67
17-18	23 (24)	24 (24)	47 (24.2)	
19-20	41 (59)	44 (45)	85 (44.3)	
21-22	30 (31)	24 (24)	54 (27.8)	
23-24	2 (2)	6 (6)	8 (4.1)	
Gender, n (%)				.22
Male	27 (28)	36 (37)	63 (32.5)	
Female	69 (72)	62 (63)	131 (67.5)	
Sexual orientation, n (%)				.47
Heterosexual	79 (82)	77 (78)	156 (80.4)	
Homosexual	9 (9)	7 (7)	16 (8.2)	
Bisexual	5 (5)	11 (11)	16 (8.2)	
Uncertain	3 (3)	3 (3)	6 (3.1)	
Ever had sex, n (%)				.97
Yes	35 (37)	36 (37)	71 (36.6)	
No	61 (64)	62 (63)	123 (63.4)	
Sexually active	26 (27)	24 (24)	50 (25.8)	.77
Safe sex attitudes and behavior, mean (SD)				
Condom use attitude (range 1-4)	3.15 (0.43)	3.24 (0.47)	3.19 (0.46)	.14
Contraceptive use intention (range 1-4)	3.04 (0.52)	3.09 (0.68)	3.06 (0.61)	.62
The perceived difficulty and ease of condom use (range 1-5)	3.26 (0.58)	3.32 (0.57)	3.29 (0.58)	.48
Condom use frequency (n=50)[b], (range 1-5)	3.80 (1.38)	3.28 (1.54)	3.54 (1.46)	.22

[a]Chi-square tests or independent *t* tests between intervention and control groups.

[b]For sexually active participants only.

Table 2. Process evaluation.

Assessment measure	Social media intervention group, mean (SD)	Website control group, mean (SD)	*P* [a]	Mean difference	95% CI
Online experience score	3.86 (0.49)	3.58 (0.51)	<.001	0.27	0.13-0.41
Online-visiting frequency	3.53 (0.81)	2.47 (0.90)	<.001	1.06	0.82-1.30

[a]Independent *t* tests between intervention and control groups.

Process Evaluation

As shown in Table 2, the evidence supports the hypothesis that the intervention group would have a significantly better online experience ($P<.001$) and a significantly higher online-visiting frequency ($P<.001$) compared to the control group. The second hypothesis, that the intervention group would have a more positive online experience and more online usage than the control group, was also supported.

Qualitatively, comments were received on both intervention and control conditions. For the intervention group, 42% (5/12) of the comments were positive, 16% (2/12) were negative, and 42% (5/12) were neutral. Participants mentioned that the videos were funny, that some practical knowledge was provided, and that the promotion could be extended as a concurrent program to reach other age groups, such as primary school or secondary school students. Participants also pointed out that the information should be presented in bullet point form to improve readability. One participant expressed the view that the content was “inadequate and too superficial.” A total of 2 participants suggested including more information applicable to sexual minorities. A participant said that, since none of his friends were in the group, he felt safe to interact in the group; otherwise, it would be embarrassing. For the control group, 12% (1/8) of the comments were positive, 76% (6/8) were negative, and 12% (1/8) were neutral. A participant expressed the view that it was easy to get information. A total of 4 participants commented on the layout of the sexual health website and said it was “dull,

cluttered, and outdated." A more user-friendly design and a mobile version were both recommended. Another participant felt that the website targeted older audiences.

Outcome Evaluation

Paired *t* tests of the intervention group showed that the intervention group had a significant change in condom use attitude (*P*=.02) and behavioral skills (*P*<.001), while there was no significant change in the control group (see Table 3). There was no significant change in contraceptive use behavioral intention and contraceptive use frequency in either the intervention or the control group. The intervention group performed better in changing half of the outcome variables, while the control group did not change any outcome variables significantly. Time X Treatment analysis showed that there were insignificant differences between the two groups, although the intervention group showed a larger improvement. All of the *P* values in the repeated measures ANOVA were larger than .05. The first hypothesis, that the intervention would have a larger positive effect on outcome variables than the control, was thus not supported.

In total, 159 Facebook interactions were recorded among 96 users. During the 6-week intervention period, there were 144 "likes," 13 comments, and 2 participant-initiated posts. A total of 70% (67/96) of the users were categorized as having low online engagement, 20% (19/96) as having medium online engagement, and 10% (10/96) as having high online engagement. The posts on the Facebook group were viewed, on average, by 77% (74/96) of participants. In addition, the peer educators interacted with the participants through commenting on their blogs. They answered the questions asked by the participants and directed them to further resources or, at times, "liked" some of the posts to show encouragement. After adjusting for age, gender, and sexual orientation, regression analysis showed a significant effect of frequency of online visiting on contraceptive use behavioral intention, behavioral skills, and condom use frequency, but not on condom use attitude (see Table 4). The effect of online engagement on outcome variables was not significant.

Table 3. Outcome evaluation.

Outcomes	Social media intervention group			Website control group			Time X Treatment analysis
	Baseline, mean (SD)	Postintervention, mean (SD)	*P* [a]	Baseline, mean (SD)	Postintervention, mean (SD)	*P* [a]	*P* [b]
Condom use attitude	3.15 (0.43)	3.26 (0.44)	.02	3.24 (0.47)	3.31 (0.39)	.15	.54
Contraceptive use intention	3.04 (0.52)	3.15 (0.51)	.08	3.09 (0.68)	3.05 (0.52)	.53	.10
The perceived difficulty and ease of condom use	3.26 (0.58)	3.49 (0.52)	<.001	3.32 (0.57)	3.43 (0.57)	.07	.17
Condom use frequency (n=40)	3.90 (1.34)	4.10 (0.94)	.51	3.47 (1.58)	3.31 (1.49)	.27	.29

[a]Paired *t* test between baseline and postintervention data.

[b]Repeated measures analysis of variance (ANOVA) between the baseline and postintervention data of intervention and control groups.

Table 4. Regression analysis on the effects of grouping, online-visiting frequency, and online engagement on outcome variables with adjustment for age, gender, and sexual orientation.

Outcome variables	Beta (regression coefficient)	95% CI	*P*
Condom use attitude			
Online-visiting frequency	.02	-0.08 to 0.13	.66
Online engagement (control as reference)			
Low	.03	-0.13 to 0.19	.73
Medium	.08	-0.17 to 0.32	.55
High	.01	-0.31 to 0.33	.96
Contraceptive use behavioral intention			
Online-visiting frequency	.14	0 to 0.28	.05
Online engagement (control as reference)			
Low	.01	-0.20 to 0.22	.92
Medium	.21	-0.11 to 0.54	.19
High	-.05	-0.46 to 0.37	.83
The perceived difficulty and ease of condom use			
Online-visiting frequency	.16	0.03 to 0.30	.02
Online engagement (control as reference)			
Low	.01	-0.20 to 0.21	.96
Medium	-.06	-0.37 to 0.26	.73
High	-.09	-0.50 to 0.32	.66
Condom use frequency (n=40)			
Online-visiting frequency	.65	0.18 to 1.27	.04
Online engagement (control as reference)			
Low	-.15	-1.08 to 0.78	.74
Medium	-.38	-1.58 to 0.82	.52
High	-.59	-2.32 to 1.14	.50

Discussion

Principal Findings

This study evaluated a peer-led, social media-delivered, safer sex intervention against a control group, an existing sexual health website, over a 6-week period delivered to a group of Chinese college students. For the process evaluation, participants in the intervention group reported more satisfying online experiences and a higher online-visiting frequency when compared with the control group. Open-ended answers yielded more positive comments from the intervention group than from the control group. For outcome evaluation, the within-group analysis showed significant improvement in condom use attitude and behavioral skills in the intervention group, but not in the control group. Although the intervention group showed more improvement in outcome variables, no significant between-group difference was found. After adjusting for demographic data, only online-visiting frequency was found to have a significant effect on the three outcome variables: contraceptive use behavioral intention, behavioral skills, and condom use frequency.

Comparison With Previous Research

Previous peer-led sexual health interventions on social media have reported improved results, including increased condom use [28] and increased STI-testing behaviors [29] compared with control groups. However, each of those two interventions was compared with another Facebook group that was unrelated to sexual health: one focused on news content [28] and the other on general health information [29]. This study compared the intervention with an active control, which was a sexual health website. Furthermore, both of the previous studies targeted high-risk groups, such as men having sex with men [29] and ethnic minorities [28], while this research recruited general college students as participants. Moreover, both of the intervention periods in the previous studies were longer—2 months [28] and 12 weeks [29], respectively—than that of this study. On the other hand, a 1-month safer sex intervention compared with no intervention found a significant increase in knowledge and insignificant results on behaviors [41]. While our study had a longer duration, no statistically significant between-group results as such were found. This may be related to the fact that the other studies had longer-term assessment periods, from 1 to 3 months, while ours only included an

immediate postintervention assessment, which renders it unable to capture the attitude or behavioral change over a longer period [29,42].

The three main differences between this study and the previous ones are the comparison groups, postintervention measurement, and target populations. Indeed, previous research showed that computerized Web-based interventions are effective in bringing about safer sex behaviors [15,43,44]. A social media-delivered intervention was not found to be significantly more effective than a Web-based intervention in this study, given that the participants were general college students. Having an active control group with immediate follow-up and targeting a low-risk, general college student group might have contributed to the insignificant between-group results from this study.

Previous studies in the United States and Europe reported that the level of participation [41] and online social network ties [29] enhance the intervention effect. This was not supported by this study, as the results failed to show any effect related to the level of online engagement. The differences in the behavior in the virtual world could be due to the cultural difference between an individualistic culture and a collectivistic culture, as found in previous literature [45]. A study investigating the motivation and patterns in using social media among college students in the United States and South Korea found that there were cultural differences in terms of developing and managing social relationships [46]. US students' online social networks were nearly five times larger than those of their Korean peers, with a similar amount of time spent on social media. This might imply that the ways they view online relationships are different: US students may tend to use social media primarily for casual relationships, whereas Korean students may require a deeper involvement and commitment to obtain social support [46]. In our study, participants expressed the position that they would not join the study if it was not via a "secret" Facebook group. It showed that they tend to care about how friends view them on social media and engaging in sexual health discussion on social media may hurt their social relationships. Discussion of sexual health issues is a sensitive issue, especially in many Asian communities. Participants expressed concerns about letting others know that they were learning about sexual health online. They may be resistant to openly discussing sexual health online and to forming relevant social ties that may threaten their reputation and impose risks to their existing online relationships. Public social ties may be harder to build in Asian communities than previous research has assumed.

Social Media and Peer-Led Approaches Performed Better on Process Evaluation

While there were many factors affecting the outcomes, online-visiting frequency was found to be significantly related to the outcomes in both groups. It is important to understand what factors contribute to a higher motivation and interest in visiting the site. Comparing the intervention group to the control group, participants in the intervention group had a significantly higher online-visiting frequency and more satisfying online experience. For online-visiting frequency, the mean difference between the two groups was larger than 1, measured on a 5-point scale. Young people have constant access to their mobile phones nowadays. Fewer and fewer of them visit websites, but more are using them for social media [47]. Therefore, the use of platforms also affects online-visiting frequency. Promoting sexual health on social media facilitates a higher online-visiting frequency, which should lead to a larger intervention effect.

The online experience includes elements that were found to be important to online sexual health promotion, including credibility, personal relevance, respect for autonomy, comfort to learn, engaging experience, ease of use, and privacy. Peer-led approaches and social media delivery are better at addressing young people's needs and expectations relating to sexual health promotion, resulting in a more positive online experience. Furthermore, the content and presentation of the intervention are more up to date; participants reported that the sexual health website was dull and outdated. The intervention materials were developed with a high level of youth input and the result supports the notion that it better suits young people's needs than expert-led materials.

The importance of anonymity was mentioned by the participants, which was found to be a common concern in some exploratory research on the feasibility of sexual health promotion on social media [24,37,48]. Young people generally care about their image among peers. Although social media has different privacy settings to let users hide certain types of information, young people still have concerns about discussing sexual health on social media. It is also noted that combining peer-led approaches with social media delivery might counterbalance some of the advantages of social media delivery. For instance, training and maintaining the peer educators could increase the cost of the intervention. The sex education material created may not be applicable to other groups in the population, thus decreasing the reach.

Practical Suggestions

Firstly, from our experience, the engagement and training of the peer educators is of paramount importance. The promotional materials should be appealing to youth and have the ability to reach out to young people through various channels. Training content should be comprehensive, not only including sexual health knowledge, but also online moderating skills and ethical issues such as cyberbullying. Secondly, continuous supervision and safeguarding of the peer educators cannot be ignored. The nature of online programs is such that peer educators often work alone, so team building is highly recommended to promote bonding and prevent volunteers from dropping out. Regular face-to-face meetings or continuous engagement of peer educators would allow them to share their emotions and any issues encountered during the intervention. Some incentives, whether financial or simply a certificate, can help motivate the peer educators to get more involved and acquire new skills. In our program, the peer educators moderated the online Facebook group by providing positive and constructive feedback with frequent auditing and advice from the investigative team. Last but not least, a balance between exposure and privacy for the participants is stressed for a project delivered on social media. As for Facebook, there are different kinds of privacy settings for groups, pages, and profiles. The peer-led approach has been found to be effective in improving safe sex knowledge and

attitudes [17], while social media sexual health promotion leads to behavioral change [28,29]. This study shows that peer-led, social media sexual health promotion improves condom use attitude and condom use behavioral skills, and performs better on process evaluation than a sexual health website. Therefore, this approach is recommended, especially when there is no sexual health-related website available for young people.

Limitations

This research is subject to several limitations. Firstly, outcome variables rely on self-reported measures. Although validated scales were adopted, there is still potential for bias. Social desirability bias is possible, as the baseline of the outcome variables was high, and some participants even achieved full marks at baseline assessment. Secondly, the intervention time may not be long enough to promote a high level of change. However, a longer intervention duration may lead to a high dropout rate, and previous research did show significant results in a 1-month period [41]. Thirdly, due to constraints on resources, this study lacks follow-up assessments. Behavioral and attitudinal change may need a longer time to be realized. Previous online research targeting young people showed that the dropout rate in a 3-month follow-up measurement was nearly 90% [49]. Due to the limitations on resources, we decided at the time to conduct only the immediate follow-up to assess change from this intervention, acknowledging that some changes might occur sometime after the intervention. Lastly, there was a contamination issue between the control and intervention groups, with some of the students signing up together but then being randomized into different groups. Although the participants were asked to maintain confidentiality on the assignment of groups and the intervention content, 2 of the participants assigned to the control group joined the Facebook group and were removed from the study.

Conclusions

This randomized controlled trial provides important evidence on an emerging online sexual health approach. A peer-led, social media-delivered, safer sex intervention has been found to be effective in improving condom use attitude and behavioral skills in this study. However, when compared with a website, the improvement was not significant. Future research is recommended to further evaluate the long-term effectiveness and cost-effectiveness of a peer-led, social media-delivered approach against current online sexual health promotion for a more systematic comparison. Online-visiting frequency has been found to be significantly related to most of the outcomes: higher online-visiting frequency lead to better improvement on contraceptive use behavioral intention, behavioral skills, and condom use frequency. The intervention group was found to have a higher online-visiting frequency and better online experience. Sexual health promotion on social media is rapidly developing. More programmers and researchers are interested in this delivery method to reach youth. It is recommended that future research investigate which components of social media are important and that researchers be aware of the possible cultural differences between populations using social networking technology.

Acknowledgments

This study is supported by funding for the Common Core Course, *Sexuality and Culture* (UH9039), at the University of Hong Kong.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Measures used in this research.

[PDF File (Adobe PDF File), 37KB - jmir_v19i8e284_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist V1.6.1 [50].

[PDF File (Adobe PDF File), 890KB - jmir_v19i8e284_app2.pdf]

References

1. Singh S, Sedgh G, Hussain R. Unintended pregnancy: Worldwide levels, trends, and outcomes. Stud Fam Plann 2010 Dec;41(4):241-250. [Medline: 21465725]
2. Gottlieb SL, Low N, Newman LM, Bolan G, Kamb M, Broutet N. Toward global prevention of sexually transmitted infections (STIs): The need for STI vaccines. Vaccine 2014 Mar 20;32(14):1527-1535. [doi: 10.1016/j.vaccine.2013.07.087] [Medline: 24581979]
3. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: A systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]

4. Song Y, Ji C. Sexual intercourse and high-risk sexual behaviours among a national sample of urban adolescents in China. J Public Health (Oxf) 2010 Sep;32(3):312-321. [doi: 10.1093/pubmed/fdp123] [Medline: 20147385]
5. Zhou H, Wang X, Ye F, Gu H, Zeng XL, Wang Y. Contraceptive knowledge, attitudes and behavior about sexuality among college students in Beijing, China. Chin Med J (Engl) 2012 Mar;125(6):1153-1157. [Medline: 22613546]
6. Wong WCW, Zhao Y, Wong NS, Parish WL, Miu HYH, Yang L, et al. Prevalence and risk factors of chlamydia infection in Hong Kong: A population-based geospatial household survey and testing. PLoS One 2017 Feb;12(2):e0172561 [FREE Full text] [doi: 10.1371/journal.pone.0172561] [Medline: 28225805]
7. Family Planning Association of Hong Kong (FPAHK). The Report of Youth Sexuality Study 2011. Hong Kong, China: Family Planning Association of Hong Kong (FPAHK); 2014.
8. Becasen JS, Ford J, Hogben M. Sexual health interventions: A meta-analysis. J Sex Res 2015 Sep;52(4):433-443. [doi: 10.1080/00224499.2014.947399] [Medline: 25211119]
9. Lindberg LD, Maddow-Zimet I. Consequences of sex education on teen and young adult sexual behaviors and outcomes. J Adolesc Health 2012 Oct;51(4):332-338. [doi: 10.1016/j.jadohealth.2011.12.028] [Medline: 22999833]
10. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [Medline: 1594721]
11. Ibrahim N, Rampal L, Jamil Z, Zain AM. Effectiveness of peer-led education on knowledge, attitude and risk behavior practices related to HIV among students at a Malaysian public university: A randomized controlled trial. Prev Med 2012 Nov;55(5):505-510 [FREE Full text] [doi: 10.1016/j.ypmed.2012.09.003] [Medline: 22982947]
12. O'Grady MA, Wilson K, Harman JJ. Preliminary findings from a brief, peer-led safer sex intervention for college students living in residence halls. J Prim Prev 2009 Nov;30(6):716-731. [doi: 10.1007/s10935-009-0195-7] [Medline: 19937119]
13. Thato R, Penrose J. A brief, peer-led HIV prevention program for college students in Bangkok, Thailand. J Pediatr Adolesc Gynecol 2013 Feb;26(1):58-65. [doi: 10.1016/j.jpag.2012.09.011] [Medline: 23332197]
14. Jaworski BC, Carey MP. Effects of a brief, theory-based STD-prevention program for female college students. J Adolesc Health 2001 Dec;29(6):417-425 [FREE Full text] [Medline: 11728891]
15. Kiene SM, Barta WD. A brief individualized computer-delivered sexual risk reduction intervention increases HIV/AIDS preventive behavior. J Adolesc Health 2006 Sep;39(3):404-410. [doi: 10.1016/j.jadohealth.2005.12.029] [Medline: 16919803]
16. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
17. Sun WH, Miu HYH, Wong CKH, Tucker JD, Wong WCW. Assessing participation and effectiveness of the peer-led approach in youth sexual health education: Systematic review and meta-analysis in more developed countries. J Sex Res 2016 Nov 29:1-14. [doi: 10.1080/00224499.2016.1247779]
18. Kanuga M, Rosenfeld WD. Adolescent sexuality and the Internet: The good, the bad, and the URL. J Pediatr Adolesc Gynecol 2004 Apr;17(2):117-124. [doi: 10.1016/j.jpag.2004.01.015] [Medline: 15050988]
19. Lim MSC, Hocking JS, Aitken CK, Fairley CK, Jordan L, Lewis JA, et al. Impact of text and email messaging on the sexual health of young people: A randomised controlled trial. J Epidemiol Community Health 2011 Mar 17;66(1):69-74. [doi: 10.1136/jech.2009.100396]
20. Free C, McCarthy O, French RS, Wellings K, Michie S, Roberts I, et al. Can text messages increase safer sex behaviours in young people? Intervention development and pilot randomised controlled trial. Health Technol Assess 2016 Jul;20(57):1-82 [FREE Full text] [doi: 10.3310/hta20570] [Medline: 27483185]
21. Chu SKW, Kwan ACM, Reynolds R, Mellecker RR, Tam F, Lee G, et al. Promoting sex education among teenagers through an interactive game: Reasons for success and implications. Games Health J 2015 Jun;4(3):168-174. [doi: 10.1089/g4h.2014.0059] [Medline: 26182060]
22. Bailey JV, Murray E, Rait G, Mercer CH, Morris RW, Peacock R, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev 2010(9):CD006483. [doi: 10.1002/14651858.CD006483.pub2] [Medline: 20824850]
23. 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;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
24. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman IM. Social media and HIV: A systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
25. Statistical Report on Internet Development in China. Beijing, China: China Internet Network Information Center (CNNIC); 2014 Jan. URL: https://cnnic.com.cn/IDR/ReportDownloads/201404/U020140417607531610855.pdf [accessed 2017-07-17] [WebCite Cache ID 6s1JBPRNG]
26. Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: Use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb;29(1):147-157 [FREE Full text] [doi: 10.1093/her/cyt071] [Medline: 23861481]
27. Gold J, Pedrana AE, Sacks-Davis R, Hellard ME, Chang S, Howard S, et al. A systematic examination of the use of online social networking sites for sexual health promotion. BMC Public Health 2011;11:583 [FREE Full text] [doi: 10.1186/1471-2458-11-583] [Medline: 21777470]

28. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: A cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
29. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: A cluster randomized trial. Ann Intern Med 2013 Sep 3;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
30. Brady SS, Sieving RE, Terveen LG, Rosser BRS, Kodet AJ, Rothberg VD. An interactive website to reduce sexual risk behavior: Process evaluation of TeensTalkHealth. JMIR Res Protoc 2015 Sep 02;4(3):e106 [FREE Full text] [doi: 10.2196/resprot.3440] [Medline: 26336157]
31. Mellanby AR, Newcombe RG, Rees J, Tripp JH. A comparative study of peer-led and adult-led school sex education. Health Educ Res 2001 Aug;16(4):481-492. [Medline: 11525394]
32. Anderson ES, Wagstaff DA, Heckman TG, Winett RA, Roffman RA, Solomon LJ, et al. Information-Motivation-Behavioral Skills (IMB) Model: Testing direct and mediated treatment effects on condom use among women in low-income housing. Ann Behav Med 2006 Feb;31(1):70-79. [doi: 10.1207/s15324796abm3101_11] [Medline: 16472041]
33. Machin D, Campbell MJ, Tan SB, Tan SH. Sample Size Tables for Clinical Studies. 3rd edition. West Sussex, UK: Wiley-Blackwell; 2009.
34. Kypri K, Gallagher SJ. Incentives to increase participation in an Internet survey of alcohol use: A controlled experiment. Alcohol Alcohol 2003;38(5):437-441. [Medline: 12915520]
35. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: A randomized study of Internet versus paper-and-pencil versions. J Med Internet Res 2007 Sep 30;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
36. Family Planning Association of Hong Kong (FPAHK) Sexuality Education. URL: http://www.famplan.org.hk/sexedu/en/ [accessed 2017-05-09] [WebCite Cache ID 6qK1N59ql]
37. Selkie EM, Benson M, Moreno M. Adolescents' views regarding uses of social networking websites and text messaging for adolescent sexual health education. Am J Health Educ 2011 Jul;42(4):205-212. [doi: 10.1080/19325037.2011.10599189] [Medline: 22229150]
38. Fisher JD, Fisher WA. Theoretical approaches to individual-level change in HIV risk behavior. In: Peterson JL, DiClemente RJ, editors. Handbook of HIV Prevention. New York, NY: Springer Science+Business Media; 2000:3-55.
39. Misovich SJ, Fisher WA, Fisher JD. A measure of AIDS prevention information, motivation, behavioral skills, and behavior. In: Davis CM, Yarber WL, Bauserman R, Schreer G, Davis SL, editors. Handbook of Sexuality-Related Measures. Thousand Oaks, CA: SAGE Publications; 1998:328-337.
40. Fisher JD, Fisher WA, Bryan AD, Misovich SJ. Information-motivation-behavioral skills model-based HIV risk behavior change intervention for inner-city high school youth. Health Psychol 2002 Mar;21(2):177-186. [Medline: 11950108]
41. Moretti R, Cremaschini M, Brembilla G, Fenili F, Gambirasio M, Valoti M. Is Facebook effective in preventing HIV/STDs in scholarized adolescents? A cluster RCT. Austin J HIV AIDS Res 2015;2(2):1016 ISSN: 2380-0755.
42. Bull SS, Breslin LT, Wright EE, Black SR, Levine D, Santelli JS. Case study: An ethics case study of HIV prevention research on Facebook: The Just/Us study. J Pediatr Psychol 2011 Feb;36(10):1082-1092 [FREE Full text] [doi: 10.1093/jpepsy/jsq126] [Medline: 21292724]
43. Lightfoot M, Comulada WS, Stover G. Computerized HIV preventive intervention for adolescents: Indications of efficacy. Am J Public Health 2007 Jun;97(6):1027-1030. [doi: 10.2105/AJPH.2005.072652] [Medline: 16670219]
44. Roberto AJ, Zimmerman RS, Carlyle KE, Abner EL. A computer-based approach to preventing pregnancy, STD, and HIV in rural adolescents. J Health Commun 2007 Jan;12(1):53-76. [doi: 10.1080/10810730601096622] [Medline: 17365349]
45. Jackson LA, Wang J. Cultural differences in social networking site use: A comparative study of China and the United States. Comput Human Behav 2013 May;29(3):910-921. [doi: 10.1016/j.chb.2012.11.024]
46. Kim Y, Sohn D, Choi SM. Cultural difference in motivations for using social network sites: A comparative study of American and Korean college students. Comput Human Behav 2011 Jan;27(1):365-372. [doi: 10.1016/j.chb.2010.08.015]
47. The Kantar China Social Media Impact Report 2015: Difference Between Social Media Users and Urban Residents. 2015. URL: http://us.kantar.com/media/908826/0202-en-infographic-1950.pdf [accessed 2017-05-18] [WebCite Cache ID 6qXnAwgi1]
48. Evers CW, Albury K, Byron P, Crawford K. Young people, social media, social network sites and sexual health communication in Australia: “This is funny, you should watch it”. Int J Commun 2013;7:263-280 ISSN: 1932-8036.
49. Mak WW, Chio FH, Chan AT, Lui WW, Wu EK. The efficacy of Internet-based mindfulness training and cognitive-behavioral training with telephone support in the enhancement of mental health among college students and young working adults: Randomized controlled trial. J Med Internet Res 2017 Mar 22;19(3):e84 [FREE Full text] [doi: 10.2196/jmir.6737] [Medline: 28330831]
50. Eysenbach G. CONSORT-EHEALTH: Implementation of a checklist for authors and editors to improve reporting of Web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]

Abbreviations

ANOVA: analysis of variance
IMB: information-motivation-behavioral
SMS: short message service
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 26.01.17; peer-reviewed by B Cao, S Bull; comments to author 26.03.17; revised version received 18.05.17; accepted 13.06.17; published 09.08.17.

Please cite as:
Sun WH, Wong CKH, Wong WCW
A Peer-Led, Social Media-Delivered, Safer Sex Intervention for Chinese College Students: Randomized Controlled Trial
J Med Internet Res 2017;19(8):e284
URL: http://www.jmir.org/2017/8/e284/
doi:10.2196/jmir.7403
PMID:28793980

Original Paper

The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review

Christopher Whitaker[1], BDS; Sharon Stevelink[1], MSc, PhD; Nicola Fear[1], BSc, MSc, DPhil

Kings Centre for Military Health Research, Department of Psychological Medicine, Kings College London, London, United Kingdom

Corresponding Author:
Christopher Whitaker, BDS
Kings Centre for Military Health Research
Department of Psychological Medicine
Kings College London
Weston Eduction Centre
Cutcome Road
London, SE5 9RJ
United Kingdom
Phone: 44 2078485817
Fax: 44 2078485408
Email: cjwhitaker@hotmail.co.uk

Abstract

Background: Social media is a popular online tool that allows users to communicate and exchange information. It allows digital content such as pictures, videos and websites to be shared, discussed, republished and endorsed by its users, their friends and businesses. Adverts can be posted and promoted to specific target audiences by demographics such as region, age or gender. Recruiting for health research is complex with strict requirement criteria imposed on the participants. Traditional research recruitment relies on flyers, newspaper adverts, radio and television broadcasts, letters, emails, website listings, and word of mouth. These methods are potentially poor at recruiting hard to reach demographics, can be slow and expensive. Recruitment via social media, in particular Facebook, may be faster and cheaper.

Objective: The aim of this study was to systematically review the literature regarding the current use and success of Facebook to recruit participants for health research purposes.

Methods: A literature review was completed in March 2017 in the English language using MEDLINE, EMBASE, Web of Science, PubMed, PsycInfo, Google Scholar, and a hand search of article references. Papers from the past 12 years were included and number of participants, recruitment period, number of impressions, cost per click or participant, and conversion rate extracted.

Results: A total of 35 studies were identified from the United States (n=22), Australia (n=9), Canada (n=2), Japan (n=1), and Germany (n=1) and appraised using the *Critical Appraisal Skills Programme* (CASP) checklist. All focused on the feasibility of recruitment via Facebook, with some (n=10) also testing interventions, such as smoking cessation and depression reduction. Most recruited young age groups (16-24 years), with the remaining targeting specific demographics, for example, military veterans. Information from the 35 studies was analyzed with median values being 264 recruited participants, a 3-month recruitment period, 3.3 million impressions, cost per click of US $0.51, conversion rate of 4% (range 0.06-29.50), eligibility of 61% (range 17-100), and cost per participant of US $14.41. The studies showed success in penetrating hard to reach populations, finding the results representative of their control or comparison demographic, except for an over representation of young white women.

Conclusions: There is growing evidence to suggest that Facebook is a useful recruitment tool and its use, therefore, should be considered when implementing future health research. When compared with traditional recruitment methods (print, radio, television, and email), benefits include reduced costs, shorter recruitment periods, better representation, and improved participant selection in young and hard to reach demographics. It however, remains limited by Internet access and the over representation of young white women. Future studies should recruit across all ages and explore recruitment via other forms of social media.

(J Med Internet Res 2017;19(8):e290) doi:10.2196/jmir.7071

KEYWORDS

epidemiology; social media; review; research

Introduction

Social media is a popular Web-based tool that allows users to communicate and exchange information [1]. It allows digital content such as pictures, videos, and websites to be shared, discussed, republished, and endorsed by its users, their friends, and businesses. Adverts can be posted and promoted to specific target audiences by demographics such as region, age, or gender.

Social media has grown tremendously with Facebook, increasing from 6m to 1bn daily users from 2005 to 2015 [2]. This visibility lead to most social media sites monetizing adverts, with 92% of the private sector currently using social media as one of their employee recruitment strategies [3]. In 2014, 66% of the UK population used social media, with 96% of those users choosing Facebook [4]. It continued to grow in 2016, with 72% of the population using social media and 97% of them choosing Facebook [1].

Recruiting for health research is complex with strict requirement criteria imposed on the participants. Traditional research recruitment relies on flyers, newspaper adverts, radio and television broadcasts, letters, emails, website listings, and word of mouth. These methods are potentially poor at recruiting hard to reach demographics, can be slow, and expensive [5,6]. Recruitment via social media, in particular Facebook, may be faster and cheaper.

This paper aims to summarize the available evidence regarding Facebook as a recruitment tool for health research in terms of cost, speed, and its ability to find and represent hard to reach demographic groups (see Table 1 for common definitions). It will be compared with traditional methods and deemed successful if it shows equal or better costing and representation of target demographics. This will be the first systematic review the authors are aware of to summarize and appraise this data.

Table 1. Common definitions.

Impressions	The number of times that the ad is fetched (starts downloading to a computer or device)
Cost per click	The cost of advertising divided by the number of times the advert is clicked shown in USD ($)
Conversion rate	The number of people who click on the ad and then proceed to become paying customers, or in the case of research, participants (considered before their eligibility)
Eligibility	The percentage of participants who respond and are eligible for the trial. This reflects the specificity of ad campaigns targeting
Cost per participant	The cost of advertising divided by the eligible recruited participants

Methods

A search of six databases, namely MEDLINE, EMBASE, Web of Science, PubMed, PsycInfo, Google Scholar, and an additional hand search of reference lists was performed in March 2017. It spanned the past 12 years due to the rise of social media from a negligible entity in 2006. A combination of the following keywords was implemented as a search strategy looking within the title or abstract:

Facebook, social media*, social network* AND internet, online, web* AND recruit*, research*, volunteer*, participant*, respondent*, patient select*, stud*, epidemiology, clinical*, health communication*, survey*

All the papers identified were exported to RefWorks [7], and duplicates were removed. Subsequently, the following exclusion criteria were applied: (1) Non-English language; (2) those not using Facebook as the recruitment tool; (3) those not recruiting for health research purposes; (4) those not constituting original research; (5) conference proceedings, letters to editors, posters, comments, and dissertations (due to difficulty accessing the full text and probable lack of detail); and (6) systematic reviews (although their reference lists were examined for eligible papers).

Full papers were appraised using the *Critical Appraisal Skills Programme* (CASP) checklist [8], and those deemed invalid were excluded (scoring less than 7/9; see Multimedia Appendix 1). Results tabulated included target demographic, number recruited, recruitment length, impressions, cost per ad click, conversion rate, eligibility, and cost per participant. Data was exported to Microsoft Excel for statistical analysis. Major outliers were removed (outside three standard deviations [SDs]), after which mean, median, and interquartile range were calculated.

Results

Summary of Accepted Studies

A total of 5818 records were identified during the initial searches. Duplicates were removed (n=1239) and 4579 records were screened against the exclusion criteria (Figure 1). Additionally, 123 full papers were assessed for quality using the study design specific CASP checklist revealing 35 papers (scoring 7-9/9) to be included in the review (see Multimedia Appendix 1). Quantitative and qualitative data was tabulated (Tables 2 and 3), allowing comparison of cost and demographic recruited.

Most studies were conducted in the United States (n=22) with some in Australia (n=9) and Canada (n=2) and one in Japan and Germany, respectively. Some studies also tested interventions (n=10); three recruited for smoking cessation [6,9,10], two for *human papillomavirus* (HPV) vaccination [11,12], two for healthier lifestyle intervention [13,14]; and one each for perinatal studies [15], human immunodeficiency virus (HIV) prevention via soap opera viewing [3], and depression intervention [16]. Ten papers recruited those aged 18 years and over, 7 the age group of 18-25 years, and 16 recruited different ages (See Tables 3 and 4 for more demographic information).

Table 2. Extracted quantitative data from the 35 included papers.

Author	Number recruited	Recruitment length (months)[a]	Impressions (millions)	Cost per ad click (US $)[b]	Conversion rate (%), n numbers included where available	Eligibility (%), n numbers included where available	Cost per participant (US $)[b]
Adam LM (2016) [15]	45	0.8	0.04	0.21[b]	NR[c]	56 (n=45)	15.12[b]
Admon L (2016) [29]	1178	1.0	0.36	0.58	13.2 (n=1592)	74 (n=1178)	14.63
Akard TF (2015) [17]	106	2.0	3.90	1.08	3.0	61 (n=106)	17
Arcia, A (2014) [18]	344	4.0	10.50	0.63	6.0	50	11.11
Batterham PJ (2014) [28] stage 1	610	0.1	NR	NR	3.0	NR	9.82[b]
Batterham PJ (2014) [28] stage 2	1283	0.1	NR	NR	3.0	NR	1.51[b]
Bauermeister JA (2012) [30]	22	NR	NR	NR	NR	NR	NR
Bull S (2013) [31]	1578	36.0[d]	NR	NR	NR	NR	NR
Carlini B (2015) [19]	285	4.0	NR	NR	NR	NR	8.92
Carter-Harris L (2016) [9]	331	0.6	0.06	0.45	29.5	NR	1.51
Child RJH (2014) [20]	78	0.1	NR	NR	NR	NR	NR
Chu JL (2013) [21]	88	9.0	17.50	0.39[b]	5.0 (n=180)	49	15.35[b]
Close S (2013) [22]	39	0.2	2.50	0.61	18.0	NR	19.44
Crosier BS (2016) [32]	264	1.0	0.01	0.20	NR	NR	8.14
Fenner Y (2012) [33]	278	4.0	36.10	0.48[b]	4.0	NR	14.50[b]
Frandsen TL (2014) [6]	138	19.0	14.50	0.68[b]	NR	NR	30.48[b]
Frandsen M (2016) [10]	92	13.5	NR	NR	NR	61	74.64[b]
Harris ML (2015) [34]	NR	8.0	NR	0.51[b]	2.0	93 (n=3795)	8.55[b]
Jones R (2015) [35]	230	1.0	NR	0.36	3.0	39	37.74
Kappa JM (2013) [36]	0	0.3	0.90	0.98	3.0	78 (n=280)	NR
Miyagi E (2014) [37]	126	9.0	5.70	NR	NR	95	NR
Moreno MA (2017) [14]	8	NR	NR	NR	NR	NR	40.99
Morgan AJ (2015) [16]	35	11.0	2.00	0.45[b]	NR	NR	14.32[b]
Musiat P (2016) [38]	26	3.0	0.50	1.74[b,d]	0.1	90	76.15[b]
Nelson EJ (2014) [24]	1003	2.0	NR	NR	48.0[d](n=1003)	91	1.36
Parkinson S (2013) [39]	100	0.2	1.30	NR	15.0	83	NR
Pedersen ER (2014) [25]	1023	1.0	3.30	0.38	5.0	45	7.05
Ramo DE (2014) [40]	1548	13.0	28.70	0.45	1.0	NR	4.28
Ramo DE (2012) [41]	230	2.0	3.20	0.34	9.0	51	8.80
Raviotta JM (2016) [11]	428	6.0	21.00	1.24	NR	NR	110.00[d]
Remschmidt C (2014) [42]	1161	2.0	62.90	NR	9.0	NR	NR
Schumacher KR (2014) [26]	394	12.0	NR	NR	NR	100 (n=394)	NR
Schwinn T (2017) [43]	797	4.2	187.00[d]	0.6	2.8	43 (n=1873)	51.70
Staffileno BA (2016) [13]	23	18.0	NR	0.73	NR	17	NR
Subasinghe AK (2016) [12]	919	13.0	55.40	0.67[b]	NR	NR	17.29[b]

Author	Number recruited	Recruitment length (months)[a]	Impressions (millions)	Cost per ad click (US $)[b]	Conversion rate (%), n numbers included where available	Eligibility (%), n numbers included where available	Cost per participant (US $)[b]
Yuan P (2014) [27]	1404	4.0	NR	NR	NR	NR	3.56

[a]Calculated as a percentage of a 31-day month.

[b]AUD converted to USD with 0.72 and CD to USD with 0.75 exchange rates where appropriate.

[c]NR: not reported; not reported if data unavailable.

[d]Outliers of over 3 standard deviations excluded from statistical calculation.

Figure 1. Article selection diagram.

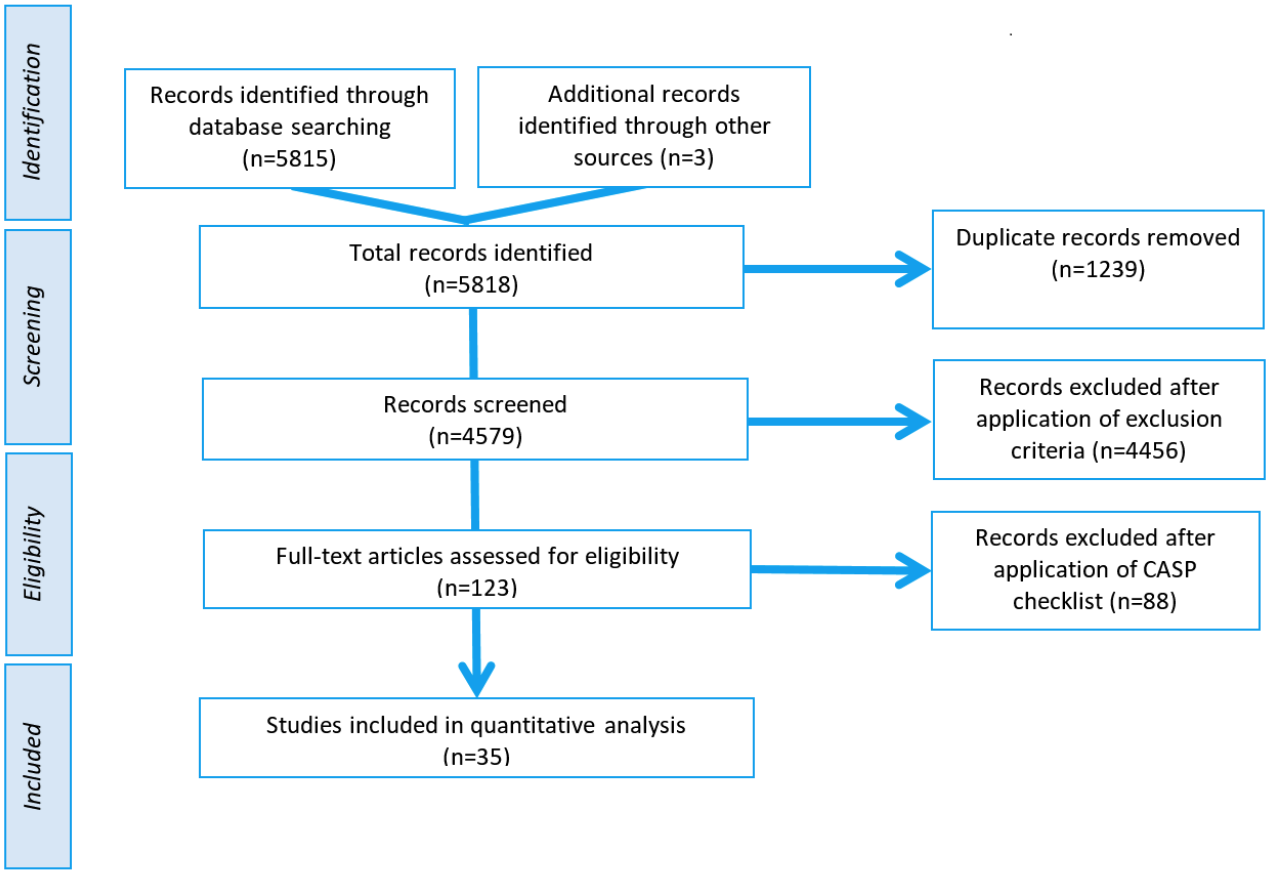


Table 3. Extracted qualitative data (Authors G-Z) from the 35 included papers.

Author	Country	Target demographic[a]	Comparison to control demographic
Harris ML (2015) [34]	Australia	18-23 years	Partly representative; higher proportion of female and tertiary education
Jones R (2015) [23]	United States	18-29 years, female, in a sexual relationship with at least one man in past 3 months	Partly representative; higher proportion of education
Kappa JM (2013) [36]	United States	35-49 years, female	No comparison made
Miyagi E (2014) [37]	Japan	16-35 years, female	Partly representative; higher proportion of 26-35 age group and a low BMI[b], and lower proportion of 16-21 age group
Moreno MA (2017) [14]	United States	14-18 years	No comparison made
Morgan AJ (2015) [16]	Australia	No other criteria	No comparison made
Musiat P (2016) [38]	Australia	18-25 years	No comparison made
Nelson EJ (2014) [24]	United States	18-30 years, lives in metropolitan area	Partly representative; higher proportion of HPV[c]vaccination
Parkinson S (2013) [39]	Australia	18-25 years	Partly representative; higher proportion of females, university education, unemployed and high income rate, and lower proportion of full time employment
Pedersen ER (2014) [25]	United States	18-34 years, previously served in the US Air Force, Army, Marine Corps, Navy	Partly representative; higher proportion of Hispanic or Latino and lower proportion of black or African American
Ramo DE (2014) [40]	United States	18-25 years, smoker	Partly representative; higher proportion of white and males
Ramo DE (2012) [41]	United States	18-25 years	Partly representative; higher proportion of white and males
Raviotta JM (2016) [11]	United States	18-25, male, student, lives in Pittsburgh	Partly representative; higher proportion of homo or bisexual and social media use
Remschmidt C (2014) [42]	Germany	18-25 years	Representative
Schumacher KR (2014) [26]	United States	15-18 years, parents of <15 years, Fontan-associated protein losing enteropathy, plastic bronchitis	Representative
Schwinn T (2017) [43]	United States	13-14 years, female	Partly representative; higher proportion of African American and less reported parents completing high school. Smoking, drinking, and drugs use was representative
Staffileno BA (2016) [13]	United States	18-45 years, prehypertension	No comparison made
Subasinghe AK (2016) [12]	Australia	18-25 years, in Victoria who had not been vaccinated against HPV	Representative
Yuan P (2014) [27]	United States	HIV[d]positive	No comparison made

[a]Assume all are over 18 years and English speaking unless otherwise stated.

[b]BMI: body mass index.

[c]HPV: *human papillomavirus.*

[d]HIV: human immunodeficiency virus.

Table 4. Extracted qualitative data (authors A-F) from the 35 included papers.

Author	Country	Target demographic[a]	Comparison with control demographic
Adam LM (2016) [15]	Canada	23-40 years, female, <25 miles from center, 8-20 weeks pregnant	No comparison made
Admon L (2016) [29]	United States	African American or Hispanic interested in pregnancy	Partly representative; higher proportion of African Americans, high income, pregnancy, and reporting fair or poor health
Akard TF (2015) [17]	United States	Parents of children or teenagers	Partly representative; higher proportion of white and female
Arcia, A (2014) [18]	United States	18-44 years, nulliparous, >20 weeks gestation	Partly representative; higher proportion of younger age groups
Batterham PJ (2014) [28] stage 1	Australia	No other criteria	Partly representative; higher proportion of education, females, young adults, and lower levels of young adolescents
Batterham PJ (2014) [28] stage 2	Australia	No other criteria	No comparison made
Bauermeister JA (2012) [30]	United States	18-24 years	Partly representative; higher proportion of white ethnicity and tertiary education and lower proportion of cigarette use
Bull S (2013) [31]	United States	15-24 years	Representative
Carlini B (2015) [19]	United States	Brazilian and Portuguese speakers	No comparison made
Carter-Harris L (2016) [9]	United States	55-77 years, current or ex-smokers	No comparison made
Child RJH (2014) [20]	United States	Emergency nurses	Representative
Chu JL (2013) [21]	Canada	15-24 years, PTSD[b]	Partly representative; higher proportion of females and younger adults
Close S (2013) [22]	United States	Any age, Klinefelter syndrome	Representative
Crosier BS (2016) [32]	United States	Self-reports auditory hallucinations	Partly representative; higher proportion females
Fenner Y (2012) [33]	Australia	16-25 years, female	Partly representative; higher proportion of increased BMI[c]
Frandsen TL (2014) [6]	Australia	Smoking >10 cigarettes per day for 3+ years, not enrolled in a cessation trial in the last 3 months	Partly representative; higher proportion of young adults
Frandsen M (2016) [10]	Australia	Smokers 10+ per day, 3 years +, no intention to quit next month, >25km from city center	No comparison made

[a]Assume all are over 18 years and English speaking unless otherwise stated.

[b]PTSD post-traumatic stress disorder.

[c]BMI: body mass index.

Other than basic demographic information including age and sex, most papers recruited participants with specific characteristics (n=18), including parents of children [17], nulliparous women at the beginning of their pregnancy [18], Brazilian and Portuguese speakers [19], emergency nurses [20], those with post-traumatic stress disorder (PTSD) [21], those with Klinefilter syndrome [22], those in sexual relationships [23], those living in a metropolitan area [24], US veterans [25], parents of children with Fontan-associated protein losing enteropathy [26], and those of positive HIV status [27]. Two papers [16,28] had no targeting features except being over 18 years old.

Summary of Quantitative Data

There were several pieces of data that outlay three SDs and so were removed from statistical analysis, namely, a recruitment length of 36 months [31], an impression count of 127 million [43], a cost per click of US $1.74 [38], a conversion rate of 48% [24], and a cost per participant of US $110.00 [11].

Table 5 shows median data: 264 recruited participants, a 3-month recruitment period, 3.3 million impressions, cost per click of US $0.51, conversion rate of 4% (range 0.06-29.50), eligibility of 61% (range 17-100), and cost per participant of US $14.41.

Table 5. Statistical analysis of extracted data with outliers removed.

Form of distribution analysis	Number recruited	Recruitment length (months)	Impressions (millions)	Cost per ad click (US $)	Conversion rate (%)	Eligibility (%)	Cost per participant (US $)
Mean	463	5.13	12.9	0.57	7	65	19.77
Median	264	3.00	3.3	0.51	4	61	14.41
Interquartile range	775	8.00	16.6	0.28	6	39	10.66

Target Population

Most studies (n=24) compared their recruited participants with either a control group recruited by traditional methods or to national data. This showed the recruited participants to be relatively representative except for some minor differences:

- There was over representation of females in 5 papers [17,21,28,32,39] and of males in 2 papers [40,41].
- Four papers reported an over representation of white ethnicity [17,30,40,41], two of African American representation [29,43], and 1 an over representation of Hispanic or Latino ethnicities [25].
- Four papers suggested over representation of a young adult group [9,18,21,28], including Frandsen M (2014), who found Web-based age to be significant younger than from the control groups recruited by mail, newspaper ads, and flyers.
- Four papers reported a higher degree of education [23,30,34,39] and two a higher rate of income [29,39] than that of the general population.
- Fenner Y (2012) reported an over representation of people with a higher body mass index (BMI) in Australia [33], whereas Miyagi, E (2014) reported an over representation of low BMI in Japan [37].
- Nelson EJ (2014) reported a higher rate of HPV vaccination [24] than predicted in Australia, whereas Remschmidt C (2014) shows it to be representative of the general population in Germany [42].
- Bauermeister JA (2012) showed the participants to be representative of the general population for alcohol consumption, marijuana, ecstasy, and cocaine use [30], with Jones R (2012) showing representation of marijuana use, sexually transmitted infection (STI) rates, and sexual relationship history [35].
- Full time employment [25] and nonsmoker status [30] where each under represented once compared with the general population.

Discussion

This paper summarizes the available evidence regarding the success of Facebook as a recruitment tool for research purposes. Some of the results can only be compared with Web-based advertising, namely, the impressions, cost per click, and conversion rate as traditional recruitment uses different markers. The remaining data on recruitment number, length of study, eligibility, and cost per participant can be compared widely with other forms of traditional recruitment.

Facebook Compared With Web-Based Advertising

Cost per click only varied slightly across studies, especially when targeting similar groups. The median cost per click from this review was US $0.51 compared with US $0.27—the mean cost per click on Facebook as a whole [44]. This shows people are less likely to interact with a health recruitment ad. The conversion rate of 4% can also be compared with the mean value of 2.4% across all Web-based advertising [45]. This suggests that although people are less interested in health research ads overall, those who do interact with them are more likely to convert. This increase in conversion rate however, doesn't appear large enough to counteract the increased cost per click with health recruitment still costing more than general advertising.

Facebook Compared With Traditional Methods

The cost per participant on Facebook was shown to be less than traditional methods. Our median value of US $14.41 compares favorably with rates suggested by Tate, D (2014) of US $1094.27 per participant for television recruitment, US $811.99 for printed media, US $635.92 for radio, and US $37.77 for email when recruiting for a survey on English language competency [5]. Carlini BH (2015) had similar findings with a mean cost per participant of US $16.22 via Google ads, and between US $13.12 and US $250.00 for other traditional methods when recruiting young adults for weight gain analysis [19].

The cost per participant values contained a major outlier; Raviotta JM (2016) reported a cost of US $110 per recruited participant [11]. The cost per click of the study (US $1.24) fell slightly outside one SD of the median and did not explain the increased cost per participant. On closer inspection, the reason for the expense became clear, "the difference in time and effort required to complete a 7-13 month study with two blood draws and three vaccine injections vs. a 30 minute survey...explains the increased cost" [11].

Facebook Compared With Other Social Media Sites

Three articles simultaneously used other social media sites to recruit participants, namely Twitter and MySpace. Bull S (2013) used MySpace but found it unsuccessful in recruiting any participants [31]. This is unsurprising considering the massive drop in MySpace primary users from 2008 to 2011 [46]. Harris, M. L (2015) implemented recruitment via Facebook and Twitter but changed to use Facebook alone due to its increased success [34]. Yuan P (2014) also used Twitter alongside Facebook. The study received 10,006 Facebook ad clicks and 161 Twitter interactions. It was found that the number of Facebook ad clicks was moderately correlated to the number recruited (*r*=.52

P<.001) but that there was little correlation between Twitter interactions and links clicked (*r*=.17, *P*=.06; *r*=.18, *P*=.06, respectively) [27]. These findings suggest Facebook is a superior recruitment tool when compared with Twitter and MySpace, although there is limited analysis across the three papers. This was an interesting albeit unintentional finding, but more research should be carried out in this area before making conclusions.

Facebook's Representation of the Population

Sociodemographic characteristics of the recruited participants were compared either with traditionally recruited participants or to available national statistics. Alcohol consumption; marijuana, ecstasy, and cocaine use [30]; STI rates; and sexual relationship history [35] were found to be representative of the total population. Those who use recreational drugs and have at risk sexual behavior tend to be found in hard to reach populations. The fact these studies mirror national statistics highlights the power of social media to target specific populations. Traditional methods tend to under represent these groups [47], meaning Facebook recruitment potentially yields more significant results.

Another point highlighting the success of Facebook recruitment would be the differing BMI results from Miyagi E (2014) and Fenner Y (2012), with the latter Australian paper showing a considerably higher average than the Japanese study. This simply shows the different obesity rates of the two countries. Australia reports 64% of its population to have a BMI above 25kg/m^{2}compared with 24% in Japan [48]. This also seems true for the differences in reported rates of HPV vaccination between Nelson EJ (2014) and Remschmidt C (2014); 39.7% of adolescent females in America [49] are vaccinated compared with 49% of Germans [50].

Other demographic data was found to be representative of the target population and comparable with traditional recruitment with only a few exceptions:

There was an over representation of white ethnicity. Facebook claims to be diverse [51], but papers suggest either these claims are not true, targeted marketing misses certain ethnicities, or that different groups have different response rates. This over representation is also shown in a review of traditional methods by Yancey AK (2005), suggesting the problem is not limited to social media recruitment [52].

Four papers also showed over representation of females. This may be due to the fact that a higher percentage of women use Facebook [1] or because fewer men respond to recruitment in general [53]. Brown WJ (1998) found similar results using traditional methods, again suggesting the problem is not limited to Facebook [54].

Education and income are often confounding factors, and it is perhaps unsurprising to find over representation in both these areas. This is comparable again with traditional recruitment methods, with Gorelick PB (1998) finding that more years in education increased the likelihood of entering and completing a clinical trial with those of lower levels "not wanting to be guinea pigs" [55].

Strengths and Limitations

Strengths of this review include the wide search ensuring all available literature was gathered and the detailed cost analysis. The main limitation is the relatively small number of studies available with numerical data on costings and population comparisons. Several papers had substantial recruitment numbers (n=1578 [30), but many were small (n=26 [13]), reducing the reliability. Most papers focused on ages in the range of 18-30 years. Carter-Harris L (2016) [9] recruited those aged 55-77 years, showing that although older people may be less likely to adopt newer technologies (of those over 65 years, 48% are active Facebook users compared with 64% for 50-64 year olds, 79% for 30-49 year olds, and 82% for 18-29 year olds [56]), recruitment can still be successful, reporting US $1.51 cost per participant. The expected barrier from lack of Internet access or experience in the older population is smaller than most think.

The percentage of people with access to the Internet is steadily increasing [1], and procedural methods can be put into place to prevent this misrepresentation of data [57]. Young, SD (2013) found that even 79% of homeless youths manage to access social media sites once per week [58]. Although Internet access currently remains to be a barrier, it seems to be smaller than barriers facing traditional methods and is set to improve in the future.

Conclusions

There is growing evidence to suggest that Facebook is a successful recruitment tool, and its use, therefore, should be considered when implementing future health research. Benefits include reduced cost, shorter recruitment periods, better representation, and improved participant selection in young and hard to reach demographics. This may spell the end for traditional methods, although currently the minor limitations of Internet access and the over representation of young white women may make its use inappropriate in some settings.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Cohort study CASP checklist scores – score out of first 9 questions.

[PDF File (Adobe PDF File), 41KB - jmir_v19i8e290_app1.pdf]

References

1. Ofcom. 2015. Adults' media use and attitudes: report 2015 URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0014/82112/2015_adults_media_use_and_attitudes_report.pdf [accessed 2017-07-29] [WebCite Cache ID 6sJrJ59f0]
2. Fb. 2015. Company info URL: http://newsroom.fb.com/company-info [accessed 2016-10-24] [WebCite Cache ID 6lVO6l8iw]
3. Johnson K, Mueller N, Gutmann D. Abstract 4847: efficacy of different recruitment methods for a neurofibromatosis type 1 online registry. Cancer Res 2014 Nov 18;73(8 Supplement):4847-4847. [doi: 10.1158/1538-7445.AM2013-4847]
4. Ofcom. 2015. Adults media use and attitudes 2014 URL: http://stakeholders.ofcom.org.uk/market-data-research/other/research-publications/adults/adults-media-lit-14/ [accessed 2016-10-24] [WebCite Cache ID 6lVOBN6rc]
5. Tate DF, LaRose JG, Griffin LP, Erickson KE, Robichaud EF, Perdue L, et al. Recruitment of young adults into a randomized controlled trial of weight gain prevention: message development, methods, and cost. Trials 2014;15:326 [FREE Full text] [doi: 10.1186/1745-6215-15-326] [Medline: 25128185]
6. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
7. Refworks. 2015. Want to learn how to get the most out of RefWorks? URL: https://www.refworks.com/refworks2/default.aspx?r=authentication::init [accessed 2016-10-24] [WebCite Cache ID 6lVODCQmY]
8. Wix. 2017. 12 questions to help you make sense of cohort study URL: http://media.wix.com/ugd/dded87_5ad0ece77a3f4fc9bcd3665a7d1fa91f.pdf [accessed 2017-07-29] [WebCite Cache ID 6sJorPhFK]
9. Bradshaw RA, Walsh KA, Neurath H. Amino acid sequence of bovine carboxypeptidase A. Isolation and characterization of selected peptic and nagarse peptides and the complete sequence of fragment F-I. Biochemistry 1971 Mar 16;10(6):961-972. [Medline: 4927805]
10. Dausset J. [The law of choice of an organ donor for transplantation]. Arch Mal Coeur Vaiss 1969 Jun;62(6):841-847. [Medline: 4997003]
11. Heilmann E, Müller KM, Westerloh K, Manz B. [Clinical picture of acute intermittent porphyria with reference to morphological findings]. Med Welt 1974 Sep 27;25(39):1538-1539. [Medline: 4427547]
12. Collins TF. Ambulatory treatment of active pulmonary tuberculosis without interruption of normal employment. S Afr Med J 1969 Jan 18;43(3):58-60. [Medline: 4975794]
13. Staffileno BA, Zschunke J, Weber M, Gross LE, Fogg L, Tangney CC. The feasibility of using facebook, craigslist, and other online strategies to recruit young African American women for a web-based healthy lifestyle behavior change intervention. J Cardiovasc Nurs 2017;32(4):365-371. [doi: 10.1097/JCN.0000000000000360] [Medline: 27428356]
14. Moreno MA, Waite A, Pumper M, Colburn T, Holm M, Mendoza J. Recruiting adolescent research participants: in-person compared to social media approaches. Cyberpsychol Behav Soc Netw 2017 Jan;20(1):64-67. [doi: 10.1089/cyber.2016.0319] [Medline: 27976951]
15. Adam LM, Manca DP, Bell RC. Can facebook be used for research? experiences using facebook to recruit pregnant women for a randomized controlled trial. J Med Internet Res 2016 Sep 21;18(9):e250 [FREE Full text] [doi: 10.2196/jmir.6404] [Medline: 27655184]
16. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
17. Akard TF, Wray S, Gilmer MJ. Facebook advertisements recruit parents of children with cancer for an online survey of web-based research preferences. Cancer Nurs 2015;38(2):155-161 [FREE Full text] [doi: 10.1097/NCC.0000000000000146] [Medline: 24945264]
18. Arcia A. Facebook advertisements for inexpensive participant recruitment among women in early pregnancy. Health Educ Behav 2013 Sep 30;41(3):237-241 [FREE Full text] [doi: 10.1177/1090198113504414] [Medline: 24082026]
19. Carlini BH, Safioti L, Rue TC, Miles L. Using Internet to recruit immigrants with language and culture barriers for tobacco and alcohol use screening: a study among Brazilians. J Immigr Minor Health 2015 Apr;17(2):553-560. [doi: 10.1007/s10903-013-9934-1] [Medline: 24563138]
20. Child RJH, Mentes JC, Pavlish C, Phillips LR. Using Facebook and participant information clips to recruit emergency nurses for research. Nurse Res 2014 Jul;21(6):16-21. [doi: 10.7748/nr.21.6.16.e1246] [Medline: 25059083]
21. Chu JL, Snider CE. Use of a social networking web site for recruiting Canadian youth for medical research. J Adolesc Health 2013 Jun;52(6):792-794. [doi: 10.1016/j.jadohealth.2012.12.002] [Medline: 23352727]
22. Close S, Smaldone A, Fennoy I, Reame N, Grey M. Using information technology and social networking for recruitment of research participants: experience from an exploratory study of pediatric Klinefelter syndrome. J Med Internet Res 2013;15(3):e48 [FREE Full text] [doi: 10.2196/jmir.2286] [Medline: 23512442]
23. Jones R, Lacroix LJ, Nolte K. "Is Your Man Stepping Out?" an online pilot study to evaluate acceptability of a guide-enhanced HIV prevention soap opera video series and feasibility of recruitment by facebook advertising. J Assoc Nurses AIDS Care 2015;26(4):368-386 [FREE Full text] [doi: 10.1016/j.jana.2015.01.004] [Medline: 26066692]
24. Nelson EJ, Hughes J, Oakes JM, Pankow JS, Kulasingam SL. Estimation of geographic variation in human papillomavirus vaccine uptake in men and women: an online survey using facebook recruitment. J Med Internet Res 2014;16(9):e198 [FREE Full text] [doi: 10.2196/jmir.3506] [Medline: 25231937]

25. 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;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]
26. Schumacher KR, Stringer KA, Donohue JE, Yu S, Shaver A, Caruthers RL, et al. Social media methods for studying rare diseases. Pediatrics 2014 May;133(5):e1345-e1353 [FREE Full text] [doi: 10.1542/peds.2013-2966] [Medline: 24733869]
27. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
28. Batterham PJ. Recruitment of mental health survey participants using internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191. [doi: 10.1002/mpr.1421] [Medline: 24615785]
29. Lusterman EA. The dental profession and the dental laboratory industry. N Y State Dent J 1965 Dec;31(10):457-458. [Medline: 5215244]
30. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838 [FREE Full text] [Medline: 22846248]
31. Bull S, Levine D, Schmiege S, Santelli J. Recruitment and retention of youth for research using social media: experiences from the Just/Us study. Vulnerable Child Youth Stud 2013 Jun;8(2):171-181. [doi: 10.1080/17450128.2012.748238]
32. Siegfried J. [Contribution of functional neurosurgery in the treatment of cerebral palsy patients]. Rev Otoneuroophtalmol 1970 Nov;42(7):412-415. [Medline: 4925933]
33. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
34. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young australian women. Am J Epidemiol 2015 Apr 15;181(10):737-746. [doi: 10.1093/aje/kwv006]
35. Jones L, Saksvig BI, Grieser M, Young DR. Recruiting adolescent girls into a follow-up study: benefits of using a social networking website. Contemp Clin Trials 2012 Mar;33(2):268-272 [FREE Full text] [doi: 10.1016/j.cct.2011.10.011] [Medline: 22101207]
36. Kapp JM, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. J Cancer Educ 2013 Mar;28(1):134-137. [doi: 10.1007/s13187-012-0443-z] [Medline: 23292877]
37. Miyagi E, Motoki Y, Asai-Sato M, Taguri M, Morita S, Hirahara F, et al. Web-based recruiting for a survey on knowledge and awareness of cervical cancer prevention among young women living in Kanagawa prefecture, Japan. Int J Gynecol Cancer 2014 Sep;24(7):1347-1355 [FREE Full text] [doi: 10.1097/IGC.0000000000000220] [Medline: 25054449]
38. Musiat P, Winsall M, Orlowski S, Antezana G, Schrader G, Battersby M, et al. Paid and unpaid online recruitment for health interventions in young adults. J Adolesc Health 2016 Sep 20;59(6):662-667. [doi: 10.1016/j.jadohealth.2016.07.020] [Medline: 27663927]
39. Parkinson S, Bromfield L. Recruiting young adults to child maltreatment research through facebook: a feasibility study. Child Abuse Negl 2013 Sep;37(9):716-720. [doi: 10.1016/j.chiabu.2013.04.009] [Medline: 23768931]
40. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64. [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
41. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
42. Remschmidt C, Walter D, Schmich P, Wetzstein M, Deleré Y, Wichmann O. Knowledge, attitude, and uptake related to human papillomavirus vaccination among young women in Germany recruited via a social media site. Hum Vaccin Immunother 2014;10(9):2527-2535 [FREE Full text] [doi: 10.4161/21645515.2014.970920] [Medline: 25483492]
43. Schwinn T, Hopkins J, Schinke SP, Liu X. Using Facebook ads with traditional paper mailings to recruit adolescent girls for a clinical trial. Addict Behav 2017 Feb;65:207-213. [doi: 10.1016/j.addbeh.2016.10.011] [Medline: 27835860]
44. Smith C. Expandedramblings. 2015. By the numbers+ amazing facebook advertising statistics URL: http://expandedramblings.com/index.php/facebook-advertising-statistics/ [accessed 2016-10-24] [WebCite Cache ID 6lVOKiatq]
45. Kim L. Searchengineland. 2015. 7 conversion rate truths that will change your landing page strategy URL: http://searchengineland.com/7-conversion-rate-truths-will-change-landing-page-optimization-strategy-191083 [accessed 2016-10-24] [WebCite Cache ID 6lVORt7Nj]
46. Wagner DA. Child development research and the third world. A future of mutual interest? Am Psychol 1986 Mar;41(3):298-301. [Medline: 3963617]
47. Dreyer K. Comscore. Infographic: myspace vs facebook URL: https://www.comscore.com/Insights/Data-Mine/Infographic-Myspace-vs-Facebook?cs_edgescape_cc=US [accessed 2017-07-29] [WebCite Cache ID 6sJrRMoX2]
48. WHO. Global health observatory (GHO) data URL: http://www.who.int/gho/en/ [accessed 2017-07-29] [WebCite Cache ID 6sJrjyMpC]

49. Elam-Evans LD, Yankey D, Jeyarajah J, Singleton JA, Curtis RC, MacNeil J, Immunization Services Division, National Center for ImmunizationRespiratory Diseases, Centers for Disease ControlPrevention (CDC). National, regional, state, and selected local area vaccination coverage among adolescents aged 13-17 years--United States, 2013. MMWR Morb Mortal Wkly Rep 2014 Jul 25;63(29):625-633 [FREE Full text] [Medline: 25055186]
50. Comporti M. Lipid peroxidation and cellular damage in toxic liver injury. Lab Invest 1985 Dec;53(6):599-623. [Medline: 3906270]
51. Marlow C. Cameronmarlow. ePluribus: ethnicity on social networks URL: http://cameronmarlow.com/media/chang-ethnicity-on-social-networks_0.pdf [accessed 2017-07-29] [WebCite Cache ID 6sJsTsXFa]
52. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27:1-28. [doi: 10.1146/annurev.publhealth.27.021405.102113] [Medline: 16533107]
53. Smith W. Ed.gov. Does gender influence online survey participation? a record-linkage analysis of university faculty online survey response behavior URL: https://eric.ed.gov/?id=ED501717 [accessed 2017-07-29] [WebCite Cache ID 6sJq7IPvs]
54. Brown WJ, Bryson L, Byles JE, Dobson AJ, Lee C, Mishra G, et al. Women's health Australia: recruitment for a national longitudinal cohort study. Women Health 1998;28(1):23-40. [Medline: 10022055]
55. Gorelick PB, Harris Y, Burnett B, Bonecutter FJ. The recruitment triangle: reasons why African Americans enroll, refuse to enroll, or voluntarily withdraw from a clinical trial. An interim report from the African-American antiplatelet stroke prevention study (AAASPS). J Natl Med Assoc 1998 Mar;90(3):141-145. [Medline: 9549977]
56. Duggan M. Pew Research Center. The demograophics of social media users URL: http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/ [accessed 2016-10-25] [WebCite Cache ID 6lVOfDiVE]
57. Lord S, Brevard J, Budman S. Connecting to young adults: an online social network survey of beliefs and attitudes associated with prescription opioid misuse among college students. Subst Use Misuse 2011;46(1):66-76 [FREE Full text] [doi: 10.3109/10826084.2011.521371] [Medline: 21190407]
58. Young SD, Rice E. Online social networking technologies, HIV knowledge, and sexual risk and testing behaviors among homeless youth. AIDS Behav 2011 Feb;15(2):253-260 [FREE Full text] [doi: 10.1007/s10461-010-9810-0] [Medline: 20848305]

Abbreviations

HIV: human immunodeficiency virus
HPV: human papillomavirus
PTSD: post-traumatic stress disorder

Edited by CL Parra-Calderón; submitted 04.02.17; peer-reviewed by A Arcia, M Ashford; comments to author 05.03.17; revised version received 06.04.17; accepted 09.05.17; published 28.08.17.

Please cite as:
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;19(8):e290
URL: http://www.jmir.org/2017/8/e290/
doi:10.2196/jmir.7071
PMID:28851679

Original Paper

Identifying Potential Norovirus Epidemics in China via Internet Surveillance

Kui Liu[1,2*], MSc; Sichao Huang[3*], MSc; Zi-Ping Miao[1,2*], BSc; Bin Chen[1,2], MSc; Tao Jiang[1], MPH; Gaofeng Cai[1], MPH; Zhenggang Jiang[1], MSc; Yongdi Chen[1], MPH; Zhengting Wang[1], MPH; Hua Gu[1], MD, PhD; Chengliang Chai[1,2], MPH; Jianmin Jiang[1,2], MD, PhD

[1]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

[2]Key Laboratory of Vaccine, Prevention and Control of Infectious Disease of Zhejiang Province, Hangzhou, China

[3]School of Medicine, Ningbo University, Ningbo, China

[*]these authors contributed equally

Corresponding Author:
Jianmin Jiang, MD, PhD
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Rd
Binjiang District
Hangzhou, 310051
China
Phone: 86 57187115009
Email: jmjiang@cdc.zj.cn

Abstract

Background: Norovirus is a common virus that causes acute gastroenteritis worldwide, but a monitoring system for norovirus is unavailable in China.

Objective: We aimed to identify norovirus epidemics through Internet surveillance and construct an appropriate model to predict potential norovirus infections.

Methods: The norovirus-related data of a selected outbreak in Jiaxing Municipality, Zhejiang Province of China, in 2014 were collected from immediate epidemiological investigation, and the Internet search volume, as indicated by the Baidu Index, was acquired from the Baidu search engine. All correlated search keywords in relation to norovirus were captured, screened, and composited to establish the composite Baidu Index at different time lags by Spearman rank correlation. The optimal model was chosen and possibly predicted maps in Zhejiang Province were presented by ArcGIS software.

Results: The combination of two vital keywords at a time lag of 1 day was ultimately identified as optimal (ρ=.924, P<.001). The exponential curve model was constructed to fit the trend of this epidemic, suggesting that a one-unit increase in the mean composite Baidu Index contributed to an increase of norovirus infections by 2.15 times during the outbreak. In addition to Jiaxing Municipality, Hangzhou Municipality might have had some potential epidemics in the study time from the predicted model.

Conclusions: Although there are limitations with early warning and unavoidable biases, Internet surveillance may be still useful for the monitoring of norovirus epidemics when a monitoring system is unavailable.

(J Med Internet Res 2017;19(8):e282) doi:10.2196/jmir.7855

KEYWORDS

norovirus; Internet surveillance; disease prediction

Introduction

Acute gastroenteritis, inflammation of the gastrointestinal tract, is defined as the sudden onset of diarrhea, with or without signs of nausea, vomiting, fever, or abdominal pain [1,2]. The known pathogenic causes of acute gastroenteritis include various infective pathogens and other noninfectious causes. Infectious acute gastroenteritis is generally caused by enteric viruses, bacteria, and protozoal pathogens. Norovirus, a single-stranded RNA virus of the *Caliciviridae* family, is a leading cause of infectious acute gastroenteritis worldwide across all age groups, particularly in health care and community settings [3,4]. In the United States, norovirus causes nearly 21 million cases of acute

gastroenteritis annually, and nearly 50% of acute gastroenteritis occurrences across Europe were attributed to the norovirus infection [5,6]. Although norovirus infection is viewed as a self-limited illness, it might be still responsible for severe dehydration, and even potential death, in children and the elderly population [7-9]. Infection might be attributable to closed bedside care and exposure to vomit of contaminated food, water and aerosol of norovirus, and other factors such as prolonged time for viral shedding and the enhanced viability in the environment might play vital roles in improving transmissibility of norovirus [4,10-12]. Additionally, no obvious evidence supports the idea of there being a specific reservoir, and only the scattered speculation from available literature viewed immunocompromised individuals, elderly, and malnourished hosts as potential norovirus reservoirs [13]. Available studies from surveillance systems demonstrate that 0.7% of reported outbreaks were foodborne, 28.5% were person-to-person, and the remaining 70.8% were unclear or not described [10]. Transmission of norovirus commonly occurs before the typical symptoms appear, which further increases the difficulty for interventions [11]. Thus, how to effectively supervise and control norovirus infection has aroused substantial concern around the world.

Norovirus was ranked the second most common etiological agent only after rotavirus in children younger than 5 years in China [14-16]. Field surveys imply that unboiled water and contaminated food are the common causes of norovirus infection [12,17]. The high contagiosity, frequent virus mutation, and limited immune protection has resulted in more frequent epidemics of norovirus outbreak on the Chinese mainland since the winter of 2014, especially in schools [18-20]. Norovirus infection is classified as “other infectious diarrhea” (excluding the illness of cholera, bacillary dysentery, amebic dysentery, and typhoid/paratyphoid) in the Chinese National Notifiable Infectious Disease Reporting System. Accordingly, norovirus cases are not reported in an independent module, and some norovirus cases manifesting the main symptom of vomiting are inevitably omitted as well. The Public Health Emergency Management Information System can only focus on some clustered epidemics; therefore, some sporadic cases or subclinical infections of norovirus may be omitted. Thus, effective interventions for norovirus at the early stage are a pressing issue for public health.

With the rapid development of Internet technology, an increasing number of researchers have tried to take advantage of the retrieval function of search engines to forecast and warn against infectious diseases [21,22]. In China, both wired and wireless networks have been booming to meet the increasing demands of the cyber citizens. The China Internet Development Statistical Report released on July 23, 2015, revealed a total of 668 million Internet users (48.8% of the population), 18.94 million more than 6 months previous [23]. As the most frequently used search engine, Baidu has a priority selection incidence of 89.1% among Chinese cyber users [24]. The Baidu Index, based on the search frequency of some keywords within the Baidu search engine, can be viewed as awareness and requirement of cyber users [25]. Evidence from previous studies of communicable diseases suggested a potential relationship between search volume and the number of infected cases [26,27]. One study also used the comprehensive Baidu Index to construct a linear regression model to predict the potential cases of epidemic erythromelalgia, suggesting that the Baidu Index may serve as a good early indicator for epidemics [22]. Given that there is no specific monitor system for norovirus in China, the purpose of this study was to determine whether Internet surveillance was a helpful supplement to traditional surveillance of norovirus epidemics in China.

Methods

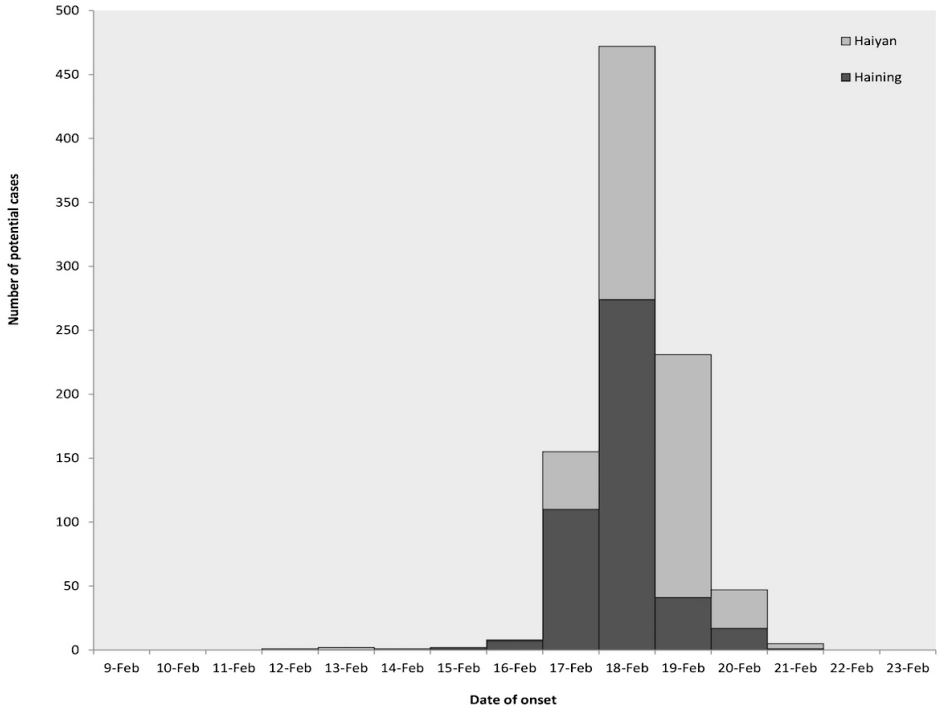
Ethics

This study was approved by the Ethics Committee of Zhejiang Provincial Center for Disease Control and Prevention. Given no privacy information involving human participants, it was granted an exemption from informed consent.

Information of Epidemic

Clustered cases in schools of diarrhea and vomiting were notified in Haining and Haiyan, two counties in the Jiaxing Municipality, on February 17, 2014. After receiving notice, a field investigation conducted by the Zhejiang Provincial Center for Disease Control and Prevention was performed to find the potential causes. Using a standard questionnaire, potential cases were searched within seven schools in Haiyan and six schools in Haining. Through the field survey, the first case was identified on February 12, 2014, and the epidemic lasted for 10 days. The inclusion criteria of possible cases were vomiting more than once or having diarrhea more than three times in one day, including probable cases and confirmed cases. There were a total of 924 cases in this outbreak with a ratio of 1 male to 1.2 females (Figure 1).

Figure 1. Daily cases of norovirus outbreak in Jiaxing Municipality from February 12 to 21, 2014, within the counties of Haiyan and Haining.



Baidu Index

The Baidu Index comprehensively reflects media exposure and users' concerns based on certain keywords used by cyber users in the past day. Although the specific algorithm of Baidu Index is not available to the public because it is proprietary information, it was proven to be similar to Google Flu Trend in identifying public behavior on the Internet for different areas on diverse days [25]. Given the potential time lag between the onset of the symptoms in cases and related Internet searching, we collected the Baidu Index of Jiaxing Municipality from February 10 to 28, 2014. The data in the same period during 2013 and 2015 were also extracted.

Keyword Screening and Data Collection

In China, the same idea can be expressed with diverse characters among different populations. That is to say, the retrieval of disease-related information may be distinctive through the search engine. Thus, how to recognize keywords specific to norovirus was vital for Internet surveillance. Furthermore, no standardized guidance was available for this issue; the disease names and main presentations had been commonly chosen as the primary keywords [21,22,28]. In this study, the primary keywords chosen (in Chinese) were "norovirus," "nausea," "emesis," "abdominal pain," and "diarrhea." More norovirus-related keywords searched by cyber users were acquired at Keywords Mining, a website that uses semantic correlation analysis [29]. All keywords obtained from the website were stemmed from search engines and also websites, blogs, and other online sources.

We retrieved from this website the top 100 keywords for each of the five initial keywords. Two individuals evaluated these 500 keywords to exclude unrelated ones. In case of discrepancy in opinion, a third person made the final decision. Then, correlations between potential case number and Baidu Index of the screened keywords with possible time lags were calculated. If $P<.05$ combined with a Spearman rank correlation coefficient (ρ) >0.4, the keyword was brought into the group with the special time lag. Similar to a previous study, we chose the onset of illness as the beginning of the study time and had five groups with time lags of 0 to 4 days [22].

Composite Baidu Index

After excluding keywords with no significance, the meaningful keywords were grouped by different time lags. In each group, the weight of each keyword ($weight_{ti}$) and composite Baidu Index (composite BDI_{ti}) were calculated as equation a and b in Figure 2, where *t* represents the potential time lag, *i* indicates the order number of present keyword, *n* is the number of keywords included in a specific time lag, ρ_{ti} is the Spearman rank correlation of included keyword (i) with specific time lag (t), and BDI_{ti} denotes the daily Baidu Index of included keywords (i) with specific time lag (t) [22].

Figure 2. Equations used in the study.

$$\text{weight}_{ti}=\frac{\rho_{ti}}{\sum_{i=1}^{n}\rho_{ti}} \quad \text{a}$$

$$\text{composite BDI}_{ti}=\sum_{i=1}^{n}\text{weight}_{ti}\,BDI_{ti} \quad \text{b}$$

$$y_{LRM}=b_0+(b_1*x) \quad \text{c}$$

$$y_{QCRM}=b_0+(b_1*x)+(b_2*x^2) \quad \text{d}$$

$$y_{CCRM}=b_0+(b_1*x)+(b_2*x^2)+(b_3*x^3) \quad \text{e}$$

$$y_{ECM}=b_0*e^{b_1*x} \quad \text{f}$$

$$y_{GCM}=e^{(b_0+(b_1*x))} \quad \text{g}$$

$$\text{OCR}=(1-\frac{|T_p-T_r|}{T_r})*100\% \quad \text{h}$$

Optimal Time Lag and Model Construction

According to the Spearman rank correlation of the composite Baidu Index and potential case number in each time lag, the time lags with the superior coefficient were determined optimal. Additionally, the mean composite Baidu Index normalized by local netizens (1/100 million) was calculated to avoid potential bias stemming from the distinction of Internet users in diverse area (Multimedia Appendix 1). Prediction models were constructed to explore the relationship of the mean composite Baidu Index and potential case number under diverse optimal lag periods, respectively.

Given the distribution characteristics of potential norovirus cases, we constructed a linear regression model and latent curve models. In this study, five possible models shown in Figure 2 (c-g) were evaluated, including linear regression model (LRM), quadratic curve regression model (QCRM), cubic curve regression model (CCRM), exponential curve model (ECM), and growth curve model (GCM).

In these models, b_1, b_2, and b_3 indicate the coefficients; x and y represent the mean composite Baidu Index and case number, respectively. The optimal model was examined by the P value of variance (ANOVA) and t test for coefficient. The overall coincidence rate (OCR) shown in Figure 2 (h) would be employed to determine the optimal model, in which T_p represents the total predicted case number from the specific prediction model during the study period, and T_r denotes the case number during the study period. The more the value of OCR trended to 100%, the better the selected model would be.

Spatial Presentation for Predicted Norovirus

The composite Baidu Index of 11 municipalities in Zhejiang Province was calculated from February 10 to 28, 2014. Based on the optimal prediction model described previously, the predicted norovirus cases in 11 municipalities were acquired and shown by ArcGIS software. The predicted norovirus cases in the same period of 2013 and 2015 were also shown on the map.

Statistical Analysis

All analyses were performed using SPSS Statistics 20.0 (SPSS Inc, Chicago, IL, USA). Spatial display was done with the Geographic Information System version 10.1 (SERI Inc, Redlands, CA, USA). Results were considered statistically significant if $P<.05$ with two sides.

Results

Epidemiological Characteristics of the Norovirus Outbreak

This epidemic was first reported in Haiyan, followed by Haining, within the Jiaxing Municipality in Zhejiang Province of China. A total of 924 cases (420 male and 504 female) involving 13 schools were detected from February 12 to 21, 2014; five were teachers and the rest were students. The clinical symptoms were mild and the main symptom was vomiting accompanied by nausea, diarrhea, fever, and abdominal pain, but no death occurred. The local departments of disease control in both counties responded rapidly to the epidemic, with such interventions as class suspension, disinfection within dormitories and classrooms, and sealing barreled water. Vomitus and anal swabs were retained from some cases. Given the potential exposure to barreled water among most cases, local centers for disease control sampled the different brands of barreled water in the schools. After 15:00 on February 21, no new cases were reported. Laboratory tests detected norovirus genogroup II in eight samples of anal swabs, five samples of opened barreled

water, and one sample of unopened barreled water, suggesting that water contaminated by norovirus caused the epidemic. Through in-depth investigation of the drinking water, we eventually deemed the occurrences at two sites as one outbreak because the contaminated water was supplied by the same supplier. This detailed information had been described in another study [30].

Optimal Time Lags of the Composite Baidu Index

Five possible time lags (0, 1, 2, 3, and 4 days) were considered to screen the considerable time lags, and details of all inclusion keywords for each time lag are listed in Table 1. The correlation coefficient peaked at the time lag of 2 days with five keywords (ρ=.945, *P*<.001). Considering the potential epidemiology significance and delicate difference at the time lag of 1 day with two keywords (ρ=.924, *P*<.001), both time lags were included to construct appropriate models.

Prediction Model for Norovirus

The composite Baidu Index was calculated with the weight and Baidu Index of the different keywords at lag times of 1 and 2 days, respectively. After the standardization of local netizens, the regression models were then constructed to predict the potential norovirus cases by the mean composite Baidu Index independent variable. Of the five candidate models considered in our study, ECM was determined the optimal model for the time lag of 1 day (Figure 3), whereas the top model for the time lag of 2 days was GCM (Table 2). Then, OCR values of both models in Jiaxing Municipality were calculated, demonstrating that OCR in ECM was 90.69% and in GCM was 66.00% (Table 3). Consequently, the optimal model was decided as ECM with 1 day lag. In this model, $y=1.809*e^{0.764*x}$, which was interpreted as a one-unit increase in the mean composite Baidu Index contributed to an increase of norovirus infections by 2.15 times during the outbreak.

Table 1. Inclusive keywords at time lags of 0 to 4 days after screening.

Time lag and keyword	Indicators for keyword			Indicators for composition	
	ρ	*P*	Weight	ρ	*P*
Day 0				.740	.01
Norovirus	.740	.01	0.348		
Noro	.735	.02	0.346		
Vomiting and bleeding	.650	.04	0.306		
Day 1				.924	<.001
Noro	.950	<.001	0.507		
Norovirus	.924	<.001	0.493		
Day 2				.945	<.001
Norovirus	.945	<.001	0.237		
Noro	.932	<.001	0.234		
Vomiting and diarrhea	.715	.02	0.180		
Nausea and vomiting	.701	.02	0.176		
Viral diarrhea in infants	.688	.03	0.173		
Day 3				.707	.02
Norovirus	.707	.02	0.266		
Why feel headache and nausea	.665	.04	0.250		
Noro	.648	.04	0.243		
Why feel dizziness and nausea	.642	.045	0.241		
Day 4					
Why feel headache and nausea	.673	.03	1	.673	.03

Figure 3. Fluctuant trend of potential case number and mean composite Baidu Index at the time lag of 1 day during February 12 to 21, 2014.

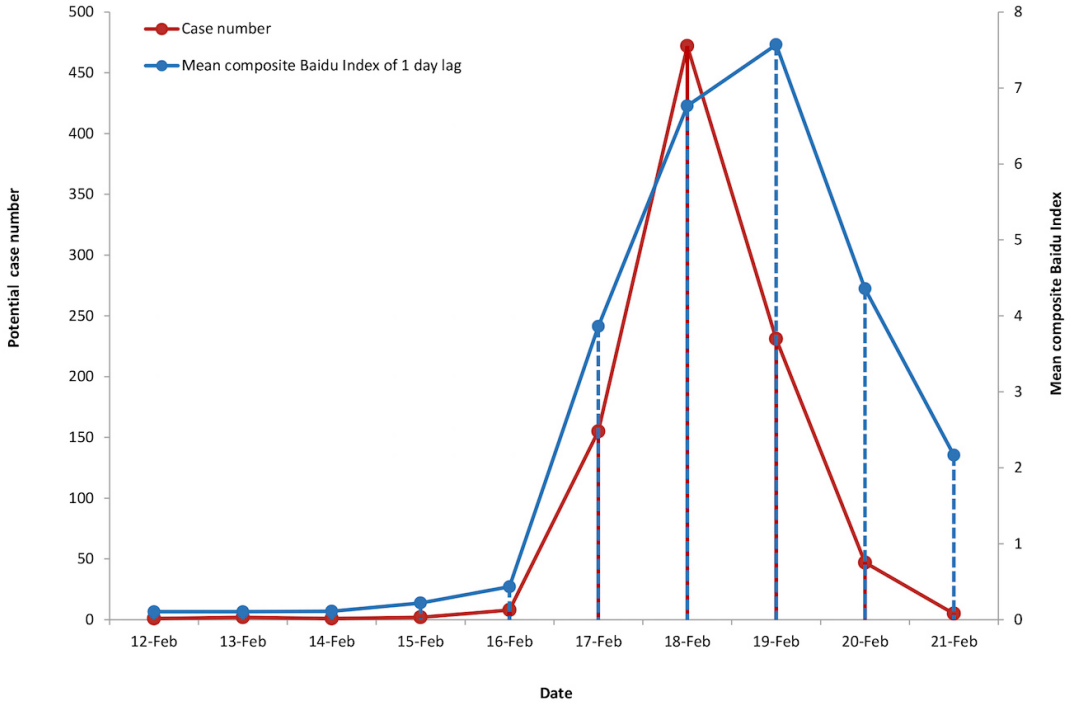


Table 2. Details of model screening for five potential candidate models.

Time lag and model	*F* (df1,df2)	*P*	R^2	Coefficient							
				b_0	*P*	b_1	*P*	b_2	*P*	b_3	*P*
Day 1											
LRM[a]	18.292 (1,8)	.003	.696	–22.213	.59	44.610	.003				
QCRM[b]	9.254 (2,7)	.01	.726	–4.24	.93	10.957	.79	4.827	.41		
CCRM[c]	7.901 (3,6)	.02	.798	30.292	.56	–151.010	.24	66.333	.17	–5.496	.19
ECM[d]	59.664 (1,8)	<.001	.882	1.809	.03	0.764	<.001				
GCM[e]	59.664 (1,8)	<.001	.882	0.593	.15	0.764	<.001				
Day 2											
LRM	19.215 (1,8)	.002	.706	–35.840	.40	95.996	.002				
QCRM	14.546 (2,7)	.003	.806	7.440	.86	–30.034	.68	36.277	.10		
CCRM	9.450 (3,6)	.01	.825	–11.621	.82	103.430	.58	–67.747	.62	19.205	.45
ECM	41.870 (1,8)	<.001	.840	1.535	.06	1.593	<.001				
GCM	41.870 (1,8)	<.001	.840	0.428	.38	1.593	<.001				

[a]LRM: linear regression model.

[b]QCRM: quadratic curve regression model.

[c]CCRM: cubic curve regression model.

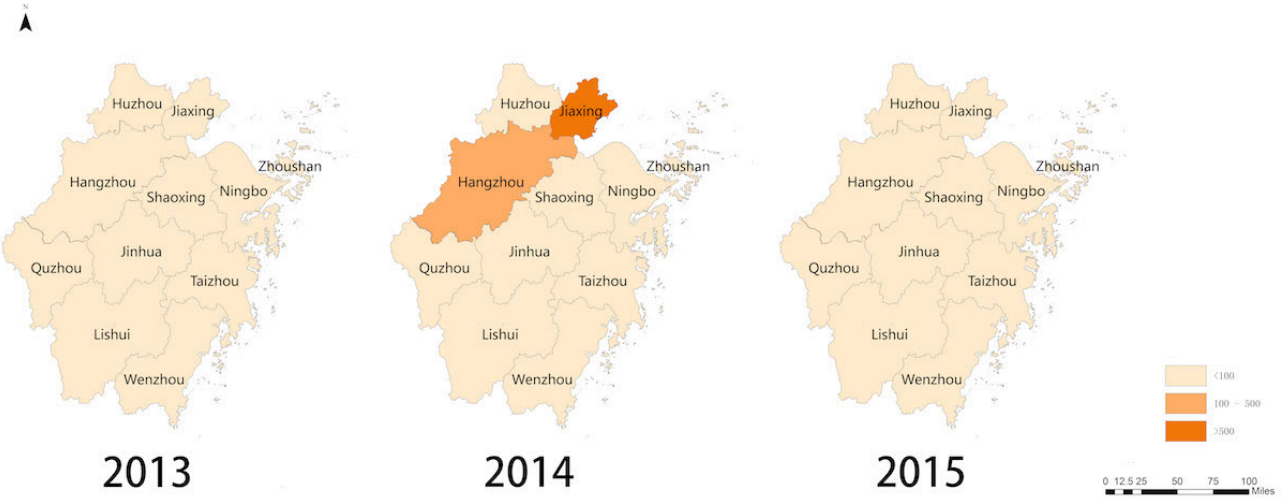
[d]ECM: exponential curve model.

[e]GCM: growth curve model.

Table 3. The overall coincidence rate (OCR) value of the exponential curve model (ECM) and growth curve model (GCM) with different time lags.

Indicators	ECM with 1 day time lag	GCM with 2 days time lag
Total predicted case number	1010	610
Case number during the study period	924	924
OCR	90.69%	66.00%

Figure 4. Predicted norovirus infections in Zhejiang Province from February 12 to 21 in each year from 2013 to 2015.



Spatial Presentation for Predicted Norovirus

Based on the preceding optimal model, potential norovirus cases of 11 municipalities in Zhejiang Province during the study time were evaluated and displayed on the map. The number of possible cases in the same period in 2013 and 2015 was predicted (Figure 4). From the displayed map, Jiaxing Municipality in 2014 showed the peak of the norovirus infection than other areas in the same period. Moreover, there might have been potential norovirus epidemics in other municipalities, such as Hangzhou.

Discussion

Principal Findings

Studies demonstrated that norovirus, a common pathogen of acute gastroenteritis, caused several serious outbreaks in China especially in Zhejiang Province in the last decade, implying imminent demand for effective control and prevention of norovirus epidemics [31-34]. With an independent reporting module yet to be constructed in the Chinese National Notifiable Infectious Disease Reporting System, only norovirus outbreaks were recorded by the Public Health Emergency Management Information System. To some extent, such a circumstance limited the detection of norovirus epidemics at the early stage. Fortunately, Internet-based surveillance offers a potential means for monitoring emergent infectious diseases, whose effectiveness and dependability have been explored and examined in some studies [22,27,35].

In this study, we used Internet-based surveillance to identify the association between possible norovirus case number and the fluctuant retrieval index of norovirus-related keywords from the Baidu search engine, and explored an optimal model with specific time lag for the prediction of norovirus epidemics. Possible norovirus-related keywords were first captured at the Keywords Mining website, which involved the technology of text mining and semantic analysis. Then, the Spearman rank correlations between possible norovirus case numbers and the search index of norovirus-related keywords were calculated to determine the inclusion keywords with different time lags. After that, all inclusion keywords for each time lag were combined to obtain the composite Baidu Index and calculate its related Spearman rank correlation coefficients. In this research, the results suggested that the composite Baidu Index of five included keywords was significantly related to this outbreak at the time lag of 2 days with the largest Spearman rank correlation coefficient of ρ=.945 followed by two included keywords at the time lag of 1 day with Spearman rank correlation coefficient of ρ=.924. Combined with the OCR values of different models, ECM was shown to be optimal including two included keywords with a time lag of 1 day. These findings are similar to previous study implying Internet-based surveillance based on some specific keywords might be effective in identifying epidemics [22]. In contrast to a previous study, the optimal mean composite Baidu Index related to norovirus in this study was essentially on a parallel track with the new cases reported in Figure 3 [22]. Given the short disease course, the serious symptom of acute gastroenteritis, and the characteristics of self-healing, the clustered epidemics of norovirus infection in young groups were more likely to be monitored, particularly in schools. Thus, the early warning of norovirus epidemics by Internet surveillance might be limited for epidemics with short incubation periods and rapid disease progression. Moreover, according to ECM with a time lag of 1 day, the optimal model indicated that an increase of one unit in the composite Baidu Index from 1/100 million netizens contributed to the rise of norovirus infections

by 2.15 times during the outbreak, which further supports the quantitative relationship between Internet surveillance and potential norovirus cases.

Previous studies have identified different optimal time lags in the analysis of diverse diseases, suggesting that this might be attributed to diverse study purposes, various incubation periods, and population susceptibility of different diseases [21,22,36]. Also, the ultimate time lag (1 day) selected in this paper was not the absolutely optimal time lag (2 days) in our study, implying more in-depth studies should be performed at more exquisite scope to explore its potential mechanism, which could mine significantly targeted interventions in public health.

Geographic information system technology has been adopted to present predicted norovirus cases in scale of Zhejiang Province (Figure 4). Compared with the same study period in each year from 2013 to 2015, our predicted results also demonstrated that the selected norovirus outbreak that occurred in Jiaxing Municipality was the largest one, which may prove the reliability of our prediction by Internet surveillance to some extent. Interestingly, some potential cases were identified in Hangzhou Municipality. Although no direct evidence was provided, Hangzhou was shown to be a high-risk region of norovirus infection in the available literature, which also certified the efficiency of Internet surveillance [37-39].

Limitations

Some limitations should be mentioned in our study:

1. The representation of the study was limited because the norovirus outbreak in question involved only schools. Therefore, the conclusion extrapolated to the whole population was insufficient.
2. Although the clinical and epidemiological evidence could be obtained in this study, laboratory tests were not performed for all cases. Thus, the accuracy of the prediction might be affected.
3. The R^2 of ECM was .882, whereas the rest (<12%) were not explored. Other external factors, such as environmental factors and economic factors, might influence the eventual results, which were not considered in this study.
4. Despite some technical means employed to search as many related keywords as possible, omission of subordinate keywords might be inevitable.
5. Other models that were not explored in this study might have a better goodness of fit, which could influence the accuracy of this study.

Conclusions

Over the past decades, the development of the Internet and search engines in China have experienced rapid leaps. A majority of the public sought medical information and expressed personal concerns on the Internet, which provided the underlying possibility for disease surveillance through the Internet, particularly in the field of emergent infectious diseases. The role of forecasting and warning against infectious diseases through the Internet has been identified in some available studies, whereas there is still no record reporting acute infectious diseases such as norovirus that have a short incubation period. In this study, we try to explore the significant keywords involving norovirus, construct an effective model, and eventually identify the potential epidemics of norovirus in Zhejiang Province using Internet surveillance. Despite existing limitations in early warning and unavoidable biases, Internet surveillance may be still useful for the monitoring of norovirus epidemics when a monitoring system is unavailable.

Acknowledgments

This study was supported by the Medical Research Program of Zhejiang Province (2016KYB056 and 2016KYB055), cofunded by Zhejiang Province and Ministry of Public Health (2016154578), and the Epidemiology Supporting Grant from Zhejiang Provincial Health and Family Planning Commission.

Authors' Contributions

KL designed the study and drafted the manuscript. SCH screened, analyzed the data, and drafted the manuscript. ZPM revised the manuscript and drafted the manuscript. KL, SCH, AND ZPM contributed equally to the manuscript. CC and JJ also made equal contributions to the manuscript. BC, TJ, GFC, and ZGJ interpreted and revised the manuscript. YDC and ZTW collected and screened the data. HG screened the data and revised the manuscript. CLC and JMJ designed and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The Internet users of different municipalities in Zhejiang Province by the end of 2013.

[PDF File (Adobe PDF File), 96KB - jmir_v19i8e282_app1.pdf]

References

1. Florez ID, Al-Khalifah R, Sierra JM, Granados CM, Yepes-Nuñez JJ, Cuello-Garcia C, et al. The effectiveness and safety of treatments used for acute diarrhea and acute gastroenteritis in children: protocol for a systematic review and network meta-analysis. Syst Rev 2016 Jan 20;5:14 [FREE Full text] [doi: 10.1186/s13643-016-0186-8] [Medline: 26818403]
2. Pieścik-Lech M, Shamir R, Guarino A, Szajewska H. Review article: the management of acute gastroenteritis in children. Aliment Pharmacol Ther 2013 Feb;37(3):289-303 [FREE Full text] [doi: 10.1111/apt.12163] [Medline: 23190209]
3. Belliot G, Lopman BA, Ambert-Balay K, Pothier P. The burden of norovirus gastroenteritis: an important foodborne and healthcare-related infection. Clin Microbiol Infect 2014 Aug;20(8):724-730 [FREE Full text] [doi: 10.1111/1469-0691.12722] [Medline: 24943671]
4. Robilotti E, Deresinski S, Pinsky BA. Norovirus. Clin Microbiol Rev 2015 Jan;28(1):134-164 [FREE Full text] [doi: 10.1128/CMR.00075-14] [Medline: 25567225]
5. Hall AJ, Wikswo ME, Manikonda K, Roberts VA, Yoder JS, Gould LH. Acute gastroenteritis surveillance through the National Outbreak Reporting System, United States. Emerg Infect Dis 2013 Aug;19(8):1305-1309 [FREE Full text] [doi: 10.3201/eid1908.130482] [Medline: 23876187]
6. Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention. Updated norovirus outbreak management and disease prevention guidelines. MMWR Recomm Rep 2011 Mar 04;60(RR-3):1-18 [FREE Full text] [Medline: 21368741]
7. Mattner F, Sohr D, Heim A, Gastmeier P, Vennema H, Koopmans M. Risk groups for clinical complications of norovirus infections: an outbreak investigation. Clin Microbiol Infect 2006 Jan;12(1):69-74 [FREE Full text] [doi: 10.1111/j.1469-0691.2005.01299.x] [Medline: 16460549]
8. Lopman BA, Steele D, Kirkwood CD, Parashar UD. The vast and varied global burden of norovirus: prospects for prevention and control. PLoS Med 2016 Apr;13(4):e1001999 [FREE Full text] [doi: 10.1371/journal.pmed.1001999] [Medline: 27115709]
9. Mulatu G, Zeynudin A, Zemene E, Debalke S, Beyene G. Intestinal parasitic infections among children under five years of age presenting with diarrhoeal diseases to two public health facilities in Hawassa, South Ethiopia. Infect Dis Poverty 2015 Nov 04;4:49 [FREE Full text] [doi: 10.1186/s40249-015-0081-x] [Medline: 26530964]
10. Petrignani M, van Beek J, Borsboom G, Richardus J, Koopmans M. Norovirus introduction routes into nursing homes and risk factors for spread: a systematic review and meta-analysis of observational studies. J Hosp Infect 2015 Mar;89(3):163-178. [doi: 10.1016/j.jhin.2014.11.015] [Medline: 25601744]
11. Glass RI, Parashar UD, Estes MK. Norovirus gastroenteritis. N Engl J Med 2009 Oct 29;361(18):1776-1785. [doi: 10.1056/NEJMra0804575] [Medline: 19864676]
12. Zhou N, Zhang H, Lin X, Hou P, Wang S, Tao Z, et al. A waterborne norovirus gastroenteritis outbreak in a school, eastern China. Epidemiol Infect 2016 Apr;144(6):1212-1219. [doi: 10.1017/S0950268815002526] [Medline: 26482884]
13. Karst SM, Baric RS. What is the reservoir of emergent human norovirus strains? J Virol 2015 Jun;89(11):5756-5759 [FREE Full text] [doi: 10.1128/JVI.03063-14] [Medline: 25787285]
14. Sai L, Sun J, Shao L, Chen S, Liu H, Ma L. Epidemiology and clinical features of rotavirus and norovirus infection among children in Ji'nan, China. Virol J 2013 Oct 08;10:302 [FREE Full text] [doi: 10.1186/1743-422X-10-302] [Medline: 24099150]
15. Xue Y, Pan H, Hu J, Wu H, Li J, Xiao W, et al. Epidemiology of norovirus infections among diarrhea outpatients in a diarrhea surveillance system in Shanghai, China: a cross-sectional study. BMC Infect Dis 2015 Apr 15;15:183 [FREE Full text] [doi: 10.1186/s12879-015-0922-z] [Medline: 25884557]
16. Zhang S, Zhou Y, Xu W, Tian L, Chen J, Chen S, et al. Impact of co-infections with enteric pathogens on children suffering from acute diarrhea in southwest China. Infect Dis Poverty 2016 Jun 27;5(1):64 [FREE Full text] [doi: 10.1186/s40249-016-0157-2] [Medline: 27349521]
17. Wang X, Yong W, Shi L, Qiao M, He M, Zhang H, et al. An outbreak of multiple norovirus strains on a cruise ship in China, 2014. J Appl Microbiol 2016 Jan;120(1):226-233. [doi: 10.1111/jam.12978] [Medline: 26481457]
18. Zhang X, Huang Q, Long Y, Jiang X, Zhang T, Tan M, et al. An outbreak caused by GII.17 norovirus with a wide spectrum of HBGA-associated susceptibility. Sci Rep 2015 Dec 07;5:17687 [FREE Full text] [doi: 10.1038/srep17687] [Medline: 26639056]
19. Chan MC, Lee N, Hung T, Kwok K, Cheung K, Tin EK, et al. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. Nat Commun 2015 Dec 02;6:10061 [FREE Full text] [doi: 10.1038/ncomms10061] [Medline: 26625712]
20. Tao Z, Xu M, Lin X, Wang H, Song L, Wang S, et al. Environmental surveillance of genogroup I and II noroviruses in Shandong Province, China in 2013. Sci Rep 2015 Nov 30;5:17444 [FREE Full text] [doi: 10.1038/srep17444] [Medline: 26616566]
21. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
22. Gu Y, Chen F, Liu T, Lv X, Shao Z, Lin H, et al. Early detection of an epidemic erythromelalgia outbreak using Baidu search data. Sci Rep 2015 Jul 28;5:12649 [FREE Full text] [doi: 10.1038/srep12649] [Medline: 26218589]

23. China Internet Network Information Center. 2015 Jul. The 36th statistical report on Internet development in China URL: http://www.cnnic.cn/gywm/xwzx/rdxw/2015/201507/t20150723_52626.htm [accessed 2017-07-23] [WebCite Cache ID 6sAfX3iw0]
24. China Internet Network Information Center. 2014. 2014 Chinese Internet users search behavior study URL: http://www.cnnic.net.cn/hlwfzyj/hlwmrtj/201410/t20141017_49359.htm [accessed 2017-07-23] [WebCite Cache ID 6sAfikIYf]
25. Chen T, Lin J. [Comparative analysis of temporal-spatial evolution of online public opinion based on search engine attention-cases of Google Trends and Baidu Index]. J Intell 2013;3(32):7-16.
26. Liu B, Wang Z, Qi X, Zhang X, Chen H. Assessing cyber-user awareness of an emerging infectious disease: evidence from human infections with avian influenza A H7N9 in Zhejiang, China. Int J Infect Dis 2015 Nov;40:34-36 [FREE Full text] [doi: 10.1016/j.ijid.2015.09.017] [Medline: 26432409]
27. Xie T, Yang Z, Yang S, Wu N, Li L. Correlation between reported human infection with avian influenza A H7N9 virus and cyber user awareness: what can we learn from digital epidemiology? Int J Infect Dis 2014 May;22:1-3 [FREE Full text] [doi: 10.1016/j.ijid.2013.11.013] [Medline: 24576842]
28. Althouse BM, Ng YY, Cummings DAT. Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis 2011 Aug;5(8):e1258 [FREE Full text] [doi: 10.1371/journal.pntd.0001258] [Medline: 21829744]
29. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One 2013;8(5):e64323 [FREE Full text] [doi: 10.1371/journal.pone.0064323] [Medline: 23750192]
30. Shang X, Fu X, Zhang P, Sheng M, Song J, He F, et al. An outbreak of norovirus-associated acute gastroenteritis associated with contaminated barrelled water in many schools in Zhejiang, China. PLoS One 2017;12(2):e0171307 [FREE Full text] [doi: 10.1371/journal.pone.0171307] [Medline: 28170414]
31. Xue L, Wu Q, Cai W, Zhang J, Guo W. Molecular characterization of new emerging GII.17 norovirus strains from South China. Infect Genet Evol 2016 Jun;40:1-7. [doi: 10.1016/j.meegid.2016.02.026] [Medline: 26923075]
32. Xue L, Dong R, Wu Q, Li Y, Cai W, Kou X, et al. Molecular epidemiology of noroviruses associated with sporadic gastroenteritis in Guangzhou, China, 2013-2015. Arch Virol 2016 May;161(5):1377-1384. [doi: 10.1007/s00705-016-2784-0] [Medline: 26906693]
33. Chen H, Qian F, Xu J, Chan M, Shen Z, Zai S, et al. A novel norovirus GII.17 lineage contributed to adult gastroenteritis in Shanghai, China, during the winter of 2014–2015. Emerg Microbes Infect 2015 Nov;4(11):e67 [FREE Full text] [Medline: 26975060]
34. Han J, Ji L, Shen Y, Wu X, Xu D, Chen L. Emergence and predominance of norovirus GII.17 in Huzhou, China, 2014-2015. Virol J 2015 Sep 11;12:139 [FREE Full text] [doi: 10.1186/s12985-015-0370-9] [Medline: 26362650]
35. 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 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
36. Cho S, Sohn CH, Jo MW, Shin S, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. PLoS One 2013;8(12):e81422 [FREE Full text] [doi: 10.1371/journal.pone.0081422] [Medline: 24339927]
37. Song CL, Liu Y. [Molecular epidemiology of norovirus in clustering of gastroenteritis in China]. Pract Prevent Med 2013 Dec;20:1294-1296.
38. XU XQ, Wang Z, Zhen YM, Liu BY. [Epidemiology of norovirus gastroenteritis outbreaks in Zhejiang Province]. Disease Surveillance 2011 Sep;26(9):710-713.
39. Lu QB, Lin JF, Li FD, Wu HC, Wu C, Wang Z, et al. [Epidemiology characteristics of school norovirus gastroenteritis outbreaks in Zhejiang Province]. Chin J Sch Health 2015;36(10):1537-1539.

Abbreviations

CCRM: cubic curve regression model
ECM: exponential curve model
GCM: growth curve model
LRM: linear regression model
QCRM: quadratic curve regression model
OCR: overall coincidence rate

Edited by G Eysenbach; submitted 12.04.17; peer-reviewed by Z Zi-ke, N Bragazzi; comments to author 09.05.17; revised version received 05.07.17; accepted 10.07.17; published 08.08.17.

Please cite as:
Liu K, Huang S, Miao ZP, Chen B, Jiang T, Cai G, Jiang Z, Chen Y, Wang Z, Gu H, Chai C, Jiang J
Identifying Potential Norovirus Epidemics in China via Internet Surveillance
J Med Internet Res 2017;19(8):e282
URL: http://www.jmir.org/2017/8/e282/
doi:10.2196/jmir.7855
PMID:28790023

Original Paper

A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals

Michael L Birnbaum[1,2,3*], MD; Sindhu Kiranmai Ernala[4*], MS; Asra F Rizvi[1,2*], MA; Munmun De Choudhury[4*], PhD; John M Kane[1,2,3*], MD

[1]The Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, United States
[2]Feinstein Institute of Medical Research, Manhasset, NY, United States
[3]Hofstra Northwell School of Medicine, Hempstead, NY, United States
[4]Georgia Institute of Technology, Atlanta, GA, United States
[*]all authors contributed equally

Corresponding Author:
Michael L Birnbaum, MD
The Zucker Hillside Hospital
Northwell Health
75-59
263rd Street
Glen Oaks, NY, 11004
United States
Phone: 1 718 470 8305
Fax: 1 718 470 1905
Email: Mbirnbaum@northwell.edu

Abstract

Background: Linguistic analysis of publicly available Twitter feeds have achieved success in differentiating individuals who self-disclose online as having schizophrenia from healthy controls. To date, limited efforts have included expert input to evaluate the authenticity of diagnostic self-disclosures.

Objective: This study aims to move from noisy self-reports of schizophrenia on social media to more accurate identification of diagnoses by exploring a human-machine partnered approach, wherein computational linguistic analysis of shared content is combined with clinical appraisals.

Methods: Twitter timeline data, extracted from 671 users with self-disclosed diagnoses of schizophrenia, was appraised for authenticity by expert clinicians. Data from disclosures deemed true were used to build a classifier aiming to distinguish users with schizophrenia from healthy controls. Results from the classifier were compared to expert appraisals on new, unseen Twitter users.

Results: Significant linguistic differences were identified in the schizophrenia group including greater use of interpersonal pronouns (P<.001), decreased emphasis on friendship (P<.001), and greater emphasis on biological processes (P<.001). The resulting classifier distinguished users with disclosures of schizophrenia deemed genuine from control users with a mean accuracy of 88% using linguistic data alone. Compared to clinicians on new, unseen users, the classifier's precision, recall, and accuracy measures were 0.27, 0.77, and 0.59, respectively.

Conclusions: These data reinforce the need for ongoing collaborations integrating expertise from multiple fields to strengthen our ability to accurately identify and effectively engage individuals with mental illness online. These collaborations are crucial to overcome some of mental illnesses' biggest challenges by using digital technology.

(J Med Internet Res 2017;19(8):e289) doi:10.2196/jmir.7956

KEYWORDS

schizophrenia; psychotic disorders; online social networks; machine learning; linguistic analysis; Twitter

Introduction

Social media provides an unprecedented opportunity to transform early psychosis intervention strategies, especially for youth who are both the highest utilizers of social media and at the greatest risk for the emergence of a psychotic disorder. Social media, defined as any form of online communication through which users create virtual communities to exchange information, ideas, messages, pictures, and videos, has forever changed the way youth interact, learn, and communicate. More than 90% of US youth use social media daily [1], placing it ahead of texting, email, and instant messaging, and they disclose considerably more about themselves online than offline [2]. Globally more than 2 billion users engage with social media regularly [3] and Twitter represents one of the most popular platforms with over 300 million monthly users worldwide.

Individuals with mental illness similarly report regularly engaging with social media [4]. Identified benefits include developing a sense of belonging, establishing and maintaining relationships, accessing support, challenging stigma, raising awareness, and sharing experiences [4,5]. Youth with newly diagnosed schizophrenia in particular report frequently utilizing social networking sites throughout the course of illness development and treatment, engaging in social media activity several times daily, and spending several hours per day online [6].

Harvesting social media activity has become an established source for capturing personalized and population data in the forms of explicit commentary, patterns and frequency of use, as well as in the intricacies of language. The massive amount of data available online has been accompanied by major advancements in computational techniques capable of quantifying language and behavior into statistically meaningful measures. There is now clear and convincing evidence that online activity can be used to reliably monitor and predict health-related behaviors [7] ranging from the spread of the influenza virus across the United States to rates of seasonal allergies, HIV infection, cancer, smoking, and obesity [8-10].

The most robust data source available is made up of the words users post online. Prior work in speech and text analysis has identified reliable linguistic markers associated with schizophrenia, including significant differences in word frequency, word categories, and use of self-referential pronouns [11-15]. These same language analytic tools have been successfully implemented to analyze modern social media-based communication [16] and have demonstrated significant linguistic differences in posts written by individuals with schizophrenia compared to individuals with depression, physical illness, and healthy controls [17]. Furthermore, classifiers designed to automatically sort individual cases into diagnostic categories have achieved success in recognizing participants with psychotic disorders from healthy controls based on linguistic differences in writing samples [15] and speech [13,18].

Researchers have begun to build classifiers aiming to identify individuals online who may have schizophrenia *without* a confirmed clinical diagnosis by scanning publicly available Twitter feeds for self-disclosures. Language-based computational models have achieved more than 80% and 90% accuracy [19,20] in correctly identifying users with self-reported schizophrenia from healthy controls. Unfortunately, however, it is challenging to confirm the authenticity of online self-disclosures. Furthermore, prior work has demonstrated that words that might have been automatically identified as self-disclosure such as “psychosis,” schizophrenia,” and “delusion” are often used inappropriately online [21] and may represent a major limitation to these computational models. To date, limited efforts have involved expert input to evaluate the authenticity of diagnostic self-disclosures.

To move from noisy diagnostic inferences to accurate identification, we propose a human-machine partnered approach, wherein linguistic analysis of content shared on social media is combined with clinical appraisals. This project aims to explore the utility of social media as a viable diagnostic tool in identifying individuals with schizophrenia.

Methods

Initial data acquisition involved extracting publicly available Twitter posts from users with self-disclosed diagnoses of schizophrenia. Case-insensitive examples include “I am diagnosed with schizophrenia,” “told me I have schizophrenia,” and “I was diagnosed with schizoaffective disorder” (Textbox 1). Prior work identifying markers of mental illness online used similar filtering techniques based on self-reported diagnoses [22,23]. Data were extracted from Twitter because posts are often publicly accessible and readily available for analysis by researchers. Approval from the institutional review board was not sought because these data were freely available in the public domain and researchers had no interaction with the users.

These search queries resulted in 21,254 posts by 15,504 users between 2012 and 2016. For each user, Twitter timeline data from 2012 to 2016 were collected using a Web-based Twitter crawler called GetOldTweetsAPI [24], which scrapes public Twitter profiles to obtain historical Twitter data in a structured format. The data included tweet text, username, posting time, hashtags, mentions, favorites, geolocation, and tweet ID. A subsample of 671 users from the primary dataset was randomly selected (each user had equal probability of being selected) and provided to two clinicians for appraisal. As a control group, a random sample of Twitter users was collected from individuals without any mentions of “schizophrenia” or “psychosis” in their timeline. Descriptive statistics of the acquired data are shown in Table 1.

Textbox 1. Search queries for Twitter data collection.

- Diagnosed me with (schizophrenia | psychosis)
- Diagnosed schizophrenic
- I am diagnosed with (psychosis | schizophrenia)
- I am schizophrenic
- I have been diagnosed with (psychosis | schizophrenia)
- I have (psychosis | schizoaffective disorder | schizophrenia)
- I think I have schizophrenia
- My schizophrenia
- They told me I have schizophrenia
- I was diagnosed with (psychosis | schizoaffective disorder | schizophrenia)
- Told me I have (psychosis | schizophrenia)

Table 1. Descriptive statistics of acquired Twitter data.

Results	Schizophrenia group (n=146)	Control group (n=146)
Total tweets by unique users, n	1,940,921	791,092
Mean tweets per user, mean (SD)	13,293.93 (18,134.83)	5418.43 (11,403.54)
Median tweets per user, median (IQR)	5542.5 (14,651.8)	1660.0 (4402.3)
Range of tweets per user (min-max)	8-88,169	1-82,985

Clinician Appraisal

To eliminate noisy data (disingenuous, inappropriate statements, jokes, and quotes) and obtain a cleaner sample of schizophrenia disclosures likely to be genuine, a psychiatrist and a graduate-level mental health clinician (authors MB and AR) from Northwell Health's Early Treatment Program, with extensive expertise in early stage schizophrenia, annotated the data. For each user, their disclosure tweet and the 10 consecutive tweets before and after were extracted to assist in making an authenticity determination. Each user was annotated by categorizing them into one of three classes. Class "yes" contained users who appeared to have genuine disclosures. Class "no" contained users who had inauthentic posts, including jokes, quotes, or were from accounts held by health-related blogs. Class "maybe" contained users for whom the experts could not confidently appraise the authenticity of the disclosure (Textbox 2). Each clinician first categorized users separately and subsequently reviewed findings together to achieve consensus. Interrater reliability for classes "yes" and "no" was 0.81 (Cohen kappa). Disagreement arose on ambiguous disclosure statements. Clinicians then utilized additional input from surrounding tweets to make an authenticity determination. These users were most often annotated as "maybe." The annotation task for 671 users resulted in 146 yes, 101 maybe, and 424 no users. These three classes of users shared 1,940,921, 1,501,838, and 8,829,775 tweets, respectively, with a mean (SD) of 13,293.98 (18,134.83), 14,869.68 (19,245.88), and 20,824.94 (45,098.07) tweets per user.

Classification Method

Data Preparation

To distinguish users with disclosures deemed genuine from the regular Twitter stream, the problem was modeled as a machine learning classification task. Users who had been annotated with class yes, formed the positive examples (class 1) for the classifier. A sample of same size collected from the control group formed the negative examples (class 0). Given the ambiguity of the "maybe" class, it was left out of this initial model. The training dataset, constructed by combining both positive and negative examples resulted in 292 users. The classifier was built and evaluated by applying 10-fold cross-validation, an established technique in supervised machine learning [25].

Classification Framework

Using the training datasets described previously, a supervised learning framework was used to build the classifier. The classification framework involved three steps: featurizing training data, feature selection to improve predictive power, and classification algorithm.

Featurizing Training Data

The textual data from Twitter timelines was used to generate features for the classifier. Each tweet in the user's timeline was represented using the following features:

Textbox 2. Examples of tweets annotated as "yes," "no," and "maybe."

Annotated "yes"

- MY MOM TOOK ME TO THE FUCKING DOCTOR AND MY DOCTOR TOLD ME I HAVE SCHIZOPHRENIA
- Finally home, was in a mental hospital for the last eight days:/ I found out I have schizophrenia...
- My parents and sister are the only family that know about my schizophrenia & everyones talking bad about it
- i have schizophrenia im bound to a life in psych wards hearing voices
- Welcome to crazy town. I figure the best way to tell the family I have psychosis is to take a picture of all my meds post it on fb with the tag of its official"
- Today was basically hell. I had to bullshit my way through it pretending like I was fine with my schizophrenia flaring up again. Urgh.
- I'll give you my Risperdal. it's my old med to treat my schizophrenia, I took it once and I slept for 12 hours
- I have schizophrenia/depression. I am trying to become better by exercise and working I have a job xoxo I love Saturday xx
- I watched your video about depression. I have schizophrenia, epilepsy and depression. I am very proactive although. :)
- And it frightens me to say that I know you don't picture me when you imagine a schizophrenic, even although I'm likely the only one you know.

Annotated "no"

- Twitter is basically an acceptable way to talk to yourself w/o being diagnosed schizophrenic
- Decided to practice my speech at the union. To the naked eye I'm sure it just looks like I have schizophrenia
- My schizophrenia article got approved for my #Psychopharmacology presentation! #yass #cantstopwontstop
- Sometimes I wish I have schizophrenia. So I can escape the reality.
- I always talk about myself as if I have schizophrenia. You gonna do this thing Aidan?" "I don't know. I doubt that I'm going to do that""
- Roses are red Violets are blue I am schizophrenic And so am I
- Texas inmate set to die, but lawyers say he's delusional: Diagnosed schizophrenic killed his in-laws
- She loves my schizophrenia, it embraces every side of me.
- Could schizophrenia simply be an extremely spiritually sensitive person, surrounded by crazy-makers? I think so.
- Watching True Life: I Have Schizophrenia Yessss... My kinda topic, future Clinical Psychologist right here!

Annotated "maybe"

- I am thoroughly convinced that my schizophrenia is a better friend than you.
- Yes, I have schizophrenia. No, I am not crazy.
- Seven days, my schizophrenia breaks-my brain waves distorted. theyre going in the trunk to avoid detection"
- is it my schizophrenia? I always knew it was...
- oh no. (To future employers) it's my schizophrenia
- it's me. I'm the inconsistent lady and i have schizophrenia
- ran up with a shovel. wonder if she felt bad afterwards. I would probably be like sorry it was my schizophrenia
- OMG U R SO FUNNY!1!!!!1!!!!!"it's just my schizophrenia
- can't help it my schizophrenia is hard to contain
- must stop listening to the talking cake, must stop listening to the talking cake, where's my schizophrenia medication

n-Gram language model: a language model of 500 top unigrams, bigrams, and trigrams (ie, sequences of one, two, and three words) was generated from the entire timeline data of all users. Each tweet was represented as a feature vector of normalized term frequency-inverse document frequency (tf-idf) frequency counts of the top 500 *n*-grams.

Linguistic inquiry and word count (LIWC): The widely validated LIWC lexicon [26] was employed, which identifies linguistic measures for the following psycholinguistic categories: (1) affective attributes, including positive and negative affect, anger, anxiety, sadness, swearing; (2) cognitive attributes, including both cognition categories comprising of cognitive mechanisms, discrepancies, inhibition, negation, causation, certainty, and tentativeness, and perception categories comprising of see, hear, feel, percept, insight, and relative; and (3) linguistic style attributes, including lexical density (verbs, auxiliary verbs, adverbs, prepositions, conjunctions, articles, inclusive, and exclusive), temporal references (past, present, and future tenses), social/personal concerns (family, friends, social, work, health,

humans, religion, bio, body, money, achievement, home, sexual, and death), and interpersonal awareness and focus (first-person singular, first-person plural, and second-person and third-person pronouns). Each tweet was represented as a vector of normalized LIWC scores for each of the preceding 50 categories.

Thus, the feature space for the classifier was 550; 500 *n*-grams and 50 LIWC categories.

Feature Selection to Improve Predictive Power

As the linguistic attributes of text contain several correlated features, the classification model tends to be unstable. To improve the predictive power of the model, feature scaling and feature selection methods were employed. First, feature scaling was used to standardize the range of features. The LIWC features were within a normalized range of 0 to1; however, the *n*- gram features represented frequency counts that required standardization. The min-max rescaling technique was used to scale the *n*- gram features to the range of 0 to1. This technique scales a feature vector "x" by converting it to the ratio of difference between x and min(x), and difference between max(x) and min(x), where min(x) and max(x) represent the minimum and maximum value of all values in the vector x.

Next, feature selection was used to eliminate noisy features, which identifies the most salient variables used to predict the outcome. Specifically, the filter method was used where features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. Adopting the ANOVA *F* test reduced the feature space from 550 features to *k* –best features (where *k*=350) by removing noisy and redundant features.

Classification Algorithm

Finally, training data represented by the top *k* features was fed into a model to learn the classification task. The model was trained over several algorithms including the Gaussian naïve Bayes, random forest, logistic regression, and support vector machines [25]. Among these, the best performing algorithm on cross-validation was used for analysis.

Results

Linguistic Characteristics

Table 2 represents comparison data between users with schizophrenia disclosures deemed genuine and the control cohort. Significance using the Mann-Whitney *U* test for all 50 LIWC categories are reported as well as the relative difference in means.

Results of Machine Learning Classification

To evaluate the performance of the classification model, a 10-fold cross-validation method was used. During each fold (iteration), the data was split into a 70% training set and 30% validation set. A model was then constructed on the 70% data and tested on the remaining 30%. Among the several classification algorithms that were applied, a random forest performed best with an average receiver operating characteristic (ROC) area under the curve (AUC) score of 0.88. The best performance for the classifier was 0.95 by the same AUC metric (see Table 3). The ROC curve is presented in Figure 1.

Figure 1. Receiver operating characteristic (ROC) curves for the classification task.

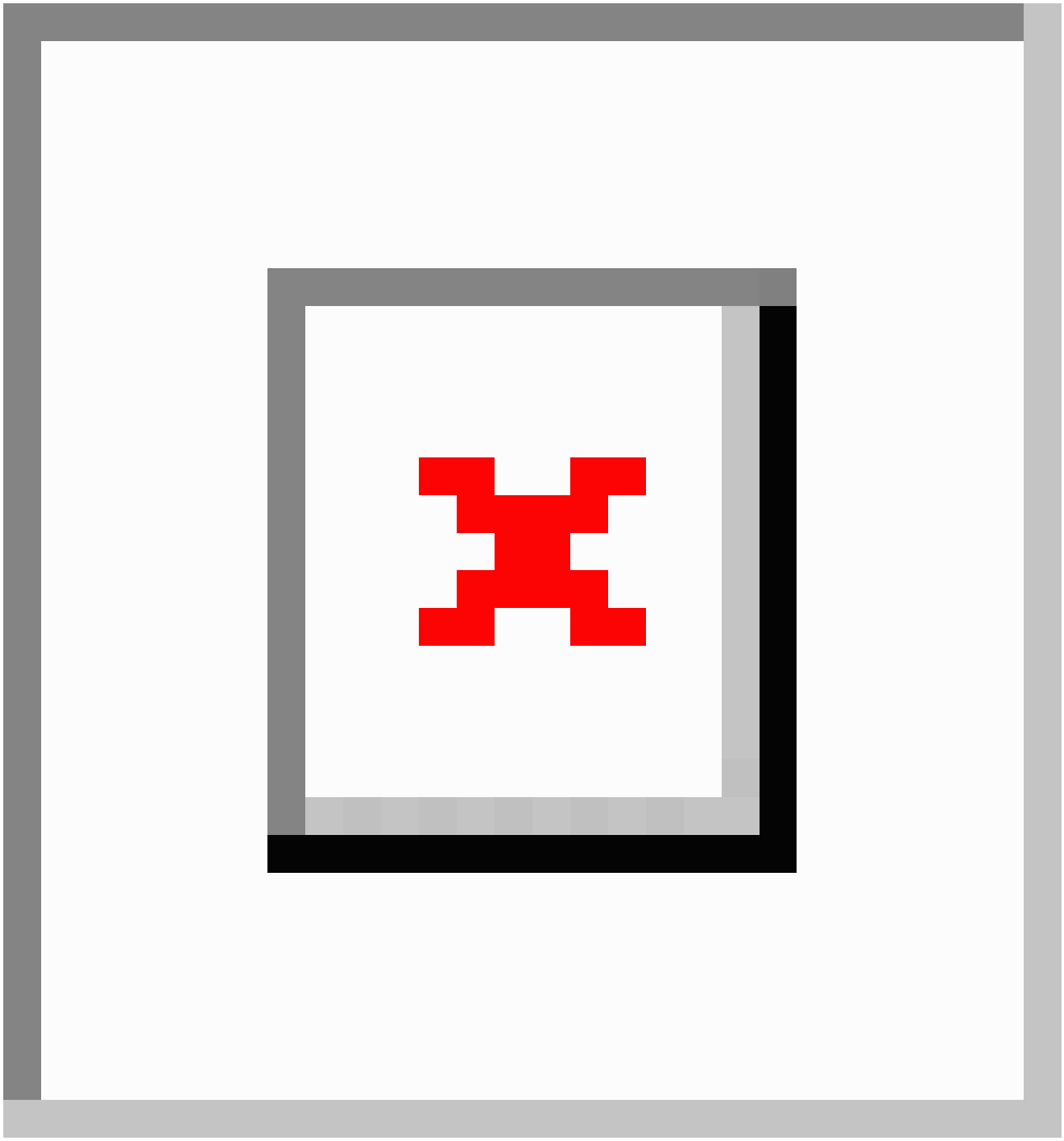


Table 2. Mann-Whitney *U* test results comparing the linguistic differences between users with schizophrenia and the control datasets.

LIWC category	Difference in mean LIWC scores between groups	*U* stat	*P* [a]
Affective attributes			
Positive affect	0.262	8517.5	.002
Negative affect	0.283	7873.5	<.001
Sadness	0.241	5301.5	<.001
Swear	0.164	8557.5	.002
Lexical density			
Auxiliary verbs	0.319	5712.5	<.001
Preposition	0.186	7162.0	<.001
Article	0.426	5812.0	<.001
Inclusive	0.410	8262.5	<.001
Exclusive	0.347	4753.0	<.001
Quantifier	0.079	991.0	<.001
Temporal references			
Past tense	0.194	7809.5	<.001
Present tense	0.304	7501.0	<.001
Future tense	0.185	4130.5	<.001
Interpersonal awareness and focus			
First-person singular	0.024	3387.0	<.001
First-person plural	0.006	8401.5	<.001
Third person	0.243	7329.5	<.001
Indefinite pronoun	0.265	2691.5	<.001
Cognition and perception attributes			
Cognitive mechanisms	0.307	9418.0	.04
Discrepancies	0.220	8975.5	.01
Inhibition	0.257	7738.5	<.001
Negation	0.187	9318.5	.03
Causation	0.353	8023.5	<.001
Certainty	0.110	6101.5	<.001
Tentativeness	0.266	1841.5	<.001
Hear	0.163	1796.5	<.001
Feel	0.270	7555.5	<.001
Perception	0.257	3340.5	<.001
Insight	0.396	7918.5	<.001
Social/Personal concerns			
Friends	–0.068	3269.0	<.001
Work	0.036	5917.5	<.001
Health	1.143	6775.0	<.001
Humans	0.039	2963.5	<.001
Biological Processes	0.427	7587.5	<.001
Body	0.150	8021.5	<.001
Achievement	0.087	6057.5	<.001
Home	0.134	6261.5	<.001

LIWC category	Difference in mean LIWC scores between groups	*U* stat	*P* [a]
Sexual	0.494	8898.5	.007

[a]Based on Bonferroni correction.

Table 3. Classification results to distinguish between schizophrenia users and control users.

Results	Accuracy	Precision	Recall	F1 score	ROC AUC
Best performance	0.90	0.92	0.87	0.90	0.95
Average over 10 folds, mean (SD)	0.81 (0.07)	0.80 (0.09)	0.82 (0.05)	0.80 (0.07)	0.88 (0.04)

Table 4. Confusion matrix showing agreement and disagreement between the machine learning classifier and the experts.

Machine label	Expert annotation	
	Yes	No
Yes	14	37
No	4	45

Verification in Unseen Data

To test the models for predicting new, unseen data, a sample of 100 users was passed through the classifier. The same sample was also provided to clinicians for appraisals. The confusion matrix displaying agreement between the two labels (machine and expert) is presented in Table 4.

By taking the expert annotations as true outcome and the machine labels as predicted outcome, true positive, true negative, false positive, and false negative scores were computed. Precision (positive predictive value) was calculated using true positive/(true positive+false positive) and recall (sensitivity) was calculated using true positive/(true positive+false negative). Accuracy (specificity) was calculated by the proportion of true results (both true positive and true negative) among the total number of cases examined (true positive+true negative)/(true positive+true negative+false positive+false negative). The resulting precision, recall, and accuracy measures were 0.27, 0.77, and 0.59, respectively.

Discussion

Main Findings

These data contribute to a growing body of literature using language to automatically identify individuals online who may be experiencing mental illness, including depression [16,22,27], postpartum mood disorders [28], suicide [29], posttraumatic stress disorder [30], and bipolar disorder [23]. To date, the majority of studies have used a computational approach to flag publicly available social media profiles of users who self-disclose with limited input from mental health clinicians to assess the authenticity of online disclosure. In this study, expert appraisal eliminated more than 70% of Twitter profiles that might have otherwise been recognized by computerized models as belonging to users with schizophrenia. These data reinforce the need for ongoing collaborations integrating expertise from multiple fields to strengthen our ability to accurately identify and effectively engage individuals with mental illness online. These collaborations are crucial to overcome some of mental illnesses' biggest challenges using digital technology.

A major challenge in treating schizophrenia remains the lengthy delay between symptom onset and receiving appropriate care. Results from the Recovery After Initial Schizophrenia Episode-Early Treatment Program (RAISE-ETP) trial [31] suggest that the median duration of untreated psychosis is 74 weeks [32] and support the established hypothesis that lengthy duration of untreated psychosis (DUP) leads to worse outcomes [31,33]. At the same time, there is compelling evidence to suggest that linguistic and behavioral changes manifest on the pages of social media before they are clinically detected, providing the prospect for earlier intervention [22,28,34]. As more and more individuals are regularly engaging with digital resources, researchers must explore novel and effective ways of incorporating technological tools into DUP reduction strategies. Identifying linguistic signals of psychosis online might be an important next step to facilitate timely treatment initiation.

Once identified, social media provides an unparalleled opportunity to explore various engagement strategies. Recently, Birnbaum et al [35] used Google AdWords to explore aspects of digital advertising most effective at engaging individuals online. Digital ads were shown to be a reasonable and cost-effective method to reach individuals searching for behavioral health information. Similar strategies could be employed to engage users via social media platforms identified as potentially experiencing schizophrenia. These strategies would require careful consideration because there is a delicate line between overintrusiveness and concern. More research is needed to better define the trajectory between online activity and making first clinical contact to explore opportunities for digital intervention. Additionally, the ethical and clinical implications of identifying markers of mental illness online require thorough and careful evaluation. Existing ethical principles do not sufficiently guide researchers conducting social media research. Furthermore, new technological approaches to illness identification and symptom tracking will likely result in

a redefinition of existing clinical rules and regulations. Although the potential beneficial impact of social media integration could be transformative, new critical questions regarding clinical expectations and responsibilities will require resolution.

The degree of agreement between the classifier and the experts in this study suggests that the classifier performs well at eliminating inauthentic noisy samples, but was overinclusive in labeling true cases of schizophrenia. For example, although the post "My parents are convinced I have schizophrenia," was labeled by the classifier as a genuine disclosure, clinicians deemed it to be a noisy sample, reflecting a more careful and conservative approach. Therefore, the classifier can theoretically assist in triaging massive amounts of digital data to provide cleaner samples to experts who can then gauge the authenticity of the disclosure.

Comparison With Prior Work

Consistent with prior trials [11-15,18,36], first-person pronouns were found to be significantly increased in the psychosis group, suggesting greater interpersonal focus. Additionally, these data replicate findings that biological processes, including words such as "body" and "health," are more frequently used in psychosis [17], suggesting a greater awareness or focus on health status. Furthermore, the psychosis group was significantly less likely to use words from the "friends" category, possibly associated with social withdrawal. Although language dysfunction, and specifically thought disorder, is an established core symptom of schizophrenia, these data suggest that subtle, more granular changes may additionally be associated with schizophrenia. Furthermore, these data suggest that changes can be detected online, reinforcing exploration of novel Internet-based early identification strategies.

Limitations

Confirming a diagnosis of schizophrenia via Twitter disclosure remains impossible without access to the psychiatric histories of those self-disclosing. Additionally, although some individuals may have psychotic symptoms (in the context of severe depression or mania), they may not meet full diagnostic criteria for schizophrenia. Exploring tweets surrounding the disclosure, taking a deeper look at an individual's profile, and implementing expert consensus certainly improved diagnostic accuracy. Secondly, the research team only had access to publicly available Twitter profiles. It is likely that many individuals who chose to self-disclose online prefer to keep their profiles private and only accessible to select individuals. Many individuals with schizophrenia chose not to self-disclose via social media at all and therefore would not have been identified in this project. To overcome these challenges, we have begun extracting social media data from consenting individuals with known clinical diagnoses of schizophrenia, allowing for exploration of online markers of psychosis from individuals who might not otherwise have publically available data. Additionally, the current classifier was developed using exclusively linguistic variables. Future work must consider incorporating nonlinguistic data including frequency and timing of posts, changes in level of activity, and social engagement online. Finally, these findings may be limited to Twitter users, who may differ from individuals who use other platforms or may use Twitter differently from other sites.

Conclusion

Existing online resources may be capable of sensing changes associated with mental illness offering the prospect for real-time objective identification and monitoring of patients. Ongoing multidisciplinary collaborations are crucial to perfect detection and monitoring capabilities for complex mental illnesses such as schizophrenia. To ensure effective incorporation of digital technology into early psychosis intervention, further research must explore precisely how symptoms of mental illness manifest online through changing patterns of language and activity as well as palatable, respectful, and effective treatment and engagement strategies once an individual is identified online.

Conflicts of Interest

None declared.

References

1. Lenhart A. Pew Research Center. 2015 Apr 09. Teens, social media & technology overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [WebCite Cache ID 6q8zVBDYW]
2. Christofides E, Muise A, Desmarais S. Information disclosure and control on Facebook: are they two sides of the same coin or two different processes? Cyberpsychol Behav 2009 Jun;12(3):341-345. [doi: 10.1089/cpb.2008.0226] [Medline: 19250020]
3. Kemp S. We Are Social. 2017 Jan 24. Digital in 2017: global overview URL: https://wearesocial.com/special-reports/digital-in-2017-global-overview [accessed 2017-06-19] [WebCite Cache ID 6rKzygEBi]
4. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Apr 05;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
5. Highton-Williamson E, Priebe S, Giacco D. Online social networking in people with psychosis: a systematic review. Int J Soc Psychiatry 2015 Feb;61(1):92-101. [doi: 10.1177/0020764014556392] [Medline: 25381145]
6. Birnbaum ML, Rizvi AF, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2015 Mar 23;11(4):290-295. [doi: 10.1111/eip.12237] [Medline: 25808317]

7. Young SD. Behavioral insights on big data: using social media for predicting biomedical outcomes. Trends Microbiol 2014 Nov;22(11):601-602 [FREE Full text] [doi: 10.1016/j.tim.2014.08.004] [Medline: 25438614]
8. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
9. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
10. Kass-Hout TA, Alhinnawi H. Social media in public health. Br Med Bull 2013;108:5-24. [doi: 10.1093/bmb/ldt028] [Medline: 24103335]
11. Buck B, Minor KS, Lysaker PH. Differential lexical correlates of social cognition and metacognition in schizophrenia; a study of spontaneously-generated life narratives. Compr Psychiatry 2015 Apr;58:138-145. [doi: 10.1016/j.comppsych.2014.12.015] [Medline: 25600423]
12. Buck B, Penn DL. Lexical characteristics of emotional narratives in schizophrenia: relationships with symptoms, functioning, and social cognition. J Nerv Ment Dis 2015 Sep;203(9):702-708. [doi: 10.1097/NMD.0000000000000354] [Medline: 26252823]
13. Hong K, Nenkova A, March ME, Parker AP, Verma R, Kohler CG. Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. Psychiatry Res 2015 Jan 30;225(1-2):40-49. [doi: 10.1016/j.psychres.2014.10.002] [Medline: 25480546]
14. Minor KS, Bonfils KA, Luther L, Firmin RL, Kukla M, MacLain VR, et al. Lexical analysis in schizophrenia: how emotion and social word use informs our understanding of clinical presentation. J Psychiatr Res 2015 May;64:74-78. [doi: 10.1016/j.jpsychires.2015.02.024] [Medline: 25777474]
15. Strous RD, Koppel M, Fine J, Nachliel S, Shaked G, Zivotofsky AZ. Automated characterization and identification of schizophrenia in writing. J Nerv Ment Dis 2009 Aug;197(8):585-588. [doi: 10.1097/NMD.0b013e3181b09068] [Medline: 19684495]
16. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS One 2013;8(9):e73791 [FREE Full text] [doi: 10.1371/journal.pone.0073791] [Medline: 24086296]
17. Fineberg SK, Leavitt J, Deutsch-Link S, Dealy S, Landry CD, Pirruccio K, et al. Self-reference in psychosis and depression: a language marker of illness. Psychol Med 2016 Sep;46(12):2605-2615. [doi: 10.1017/S0033291716001215] [Medline: 27353541]
18. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015;1:15030 [FREE Full text] [doi: 10.1038/npjschz.2015.30] [Medline: 27336038]
19. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology; Jun 5, 2015; Denver, CO p. 11.
20. McManus K, Mallory EK, Goldfeder RL, Haynes WA, Tatum JD. Mining Twitter data to improve detection of schizophrenia. AMIA Jt Summits Transl Sci Proc 2015;2015:122-126 [FREE Full text] [Medline: 26306253]
21. Birnbaum ML, Candan K, Libby I, Pascucci O, Kane J. Impact of online resources and social media on help-seeking behaviour in youth with psychotic symptoms. Early Interv Psychiatry 2016 Oct;10(5):397-403. [doi: 10.1111/eip.12179] [Medline: 25200743]
22. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci '13 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris p. 47-56. [doi: 10.1145/2464464.2464480]
23. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: Computational Linguistics and Clinical Psychology Workshop at ACL 2014; Jun 27, 2014; Baltimore, MD p. 27.
24. Henrique J. GitHub. Get old tweets-python computer API URL: https://github.com/Jefferson-Henrique/GetOldTweets-python [accessed 2017-05-01] [WebCite Cache ID 6q8zxN1qp]
25. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.
26. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. The Development and Psychometric Properties of LIWC. Austin, TX: LIWC.net URL: http://www.liwc.net/LIWC2007LanguageManual.pdf [accessed 2017-05-01] [WebCite Cache ID 6q90GAqKV]
27. Nguyen T, Phung D, Dao B, Venkatesh S, Berk M. Affective and content analysis of online depression communities. IEEE Trans Affective Comput 2014 Jul 1;5(3):217-226 [FREE Full text] [doi: 10.1109/TAFFC.2014.2315623]
28. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 27, 2013; Paris. [doi: 10.1145/2470654.2466447]
29. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]

30. Coppersmith G, Harman C, Dredze M. Measuring post traumatic stress disorder in Twitter. In: Proceedings of the International AAAI Conference on Web and Social Media Eighth International AAAI Conference on Weblogs and Social Media. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; Jun 1-4, 2014; Ann Arbor, MI p. 16.
31. Kane JM, Robinson DG, Schooler NR, Mueser KT, Penn DL, Rosenheck RA, et al. Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the NIMH RAISE early treatment program. Am J Psychiatry 2016 Apr 1;173(4):362-372. [doi: 10.1176/appi.ajp.2015.15050632] [Medline: 26481174]
32. Addington J, Heinssen RK, Robinson DG, Schooler NR, Marcy P, Brunette MF, et al. Duration of untreated psychosis in community treatment settings in the United States. Psychiatr Serv 2015 Jul;66(7):753-756. [doi: 10.1176/appi.ps.201400124] [Medline: 25588418]
33. Perkins DO, Gu H, Boteva K, Lieberman JA. Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. Am J Psychiatry 2005 Oct;162(10):1785-1804. [doi: 10.1176/appi.ajp.162.10.1785] [Medline: 16199825]
34. D'Angelo J, Kerr B, Moreno MA. Facebook displays as predictors of binge drinking: from the virtual to the visceral. Bull Sci Technol Soc 2014;34(5-6):159-169 [FREE Full text] [doi: 10.1177/0270467615584044] [Medline: 26412923]
35. Birnbaum ML, Garrett C, Baumel A, Scovel M, Rizvi AF, Muscat W, et al. Using digital media advertising in early psychosis intervention. Psychiatr Serv 2017 Jul 17:appips201600571. [doi: 10.1176/appi.ps.201600571] [Medline: 28712355]
36. Junghaenel DU, Smyth JM, Santner L. Linguistic Dimensions of Psychopathology: A Quantitative Analysis. J Soc Clin Psychol 2008 Jan;27(1):36-55. [doi: 10.1521/jscp.2008.27.1.36]

Abbreviations

AUC: area under the curve
DUP: duration of untreated psychosis
LIWC: language inquiry word count
RAISE-ETP: Recovery After an Initial Schizophrenia Episode-Early Treatment Program
ROC: receiver operating characteristic
tf-idf: term frequency-inverted document frequency

Edited by G Eysenbach; submitted 02.05.17; peer-reviewed by N Berry, TR Soron; comments to author 15.06.17; revised version received 28.06.17; accepted 30.06.17; published 14.08.17.

Please cite as:
Birnbaum ML, Ernala SK, Rizvi AF, De Choudhury M, Kane JM
A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals
J Med Internet Res 2017;19(8):e289
URL: http://www.jmir.org/2017/8/e289/
doi:10.2196/jmir.7956
PMID:28807891

Original Paper

"Am I normal?" The Wishes of Patients With Lymphoma to Compare Their Patient-Reported Outcomes With Those of Their Peers

Simone Oerlemans[1], MSc, PhD; Lindy P Arts[1], MSc; Nicole J Horevoorts[1,2], MA; Lonneke V van de Poll-Franse[1,2,3], MSc, PhD

[1]Department of Research, Netherlands Comprehensive Cancer Organisation, Utrecht, Netherlands

[2]Centre of Research on Psychology in Somatic Diseases (CoRPS), Tilburg University, Tilburg, Netherlands

[3]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, Netherlands

Corresponding Author:
Simone Oerlemans, MSc, PhD
Department of Research
Netherlands Comprehensive Cancer Organisation
Godebaldkwartier 419
PO Box 19079, 3501 DB
Utrecht, 3501 DB
Netherlands
Phone: 31 (0) 88 234 6195
Fax: 31 (0) 88 234 6195
Email: s.oerlemans@iknl.nl

Abstract

Background: Providing feedback to patients on their patient-reported outcomes (PROs) can help patients in monitoring their functioning and symptoms and may help empower them.

Objective: The objective of this study was to investigate whether patients with lymphoma wished to receive PRO feedback, including the option to compare their scores with those of their peers, and how this feedback was evaluated.

Methods: We invited 64 patients participating in a lymphoma cohort who were eligible for a follow-up questionnaire and gave them the option to receive PRO feedback. Patients completed questions about health-related quality of life (HRQoL) and symptoms. PRO feedback was provided via bar charts.

Results: Of the 64 invited patients, 45 participated (response rate 70%) and 36 of those (80%) wished to receive PRO feedback. The vast majority (34/36, 94%) compared their scores with those of a lymphoma reference cohort, and 64% (23/36) compared their score with those of a normative population without cancer. All patients wished to receive feedback on their HRQoL, and 29 (81%) to 33 (92%) wanted feedback on their functioning, fatigue, neuropathy, anxiety, and depressive symptoms. Of the 36 participants wishing to receive PRO feedback, 35 (97%) viewed it as being useful, with reassurance and knowledge about their own functioning in relation to what is "normal" being the most frequently mentioned reasons.

Conclusions: A high number of patients with lymphoma wished to receive PRO feedback. Patients reported the comparison of their scores versus a lymphoma reference cohort as most valuable. Further research should investigate whether PRO feedback could increase empowerment and possibly improve HRQoL.

(J Med Internet Res 2017;19(8):e288) doi:10.2196/jmir.7079

KEYWORDS

lymphoma; health-related quality of life; personalized feedback; self-care; access to information; population-based research

Introduction

Patients with lymphoma are at risk of experiencing adverse physical and psychosocial effects of their cancer and its treatment, such as fatigue, cognitive problems, anxiety, and depression [1-4]. Management of these symptoms is often complex, and patients do not always know if their symptoms are common and are caused by their disease or treatment [5].

Patient-reported outcomes (PROs) provide information about the subjective well-being of patients [6]. PROs are standardized questionnaires that are completed by patients and measure a broad range of health-related constructs, including symptom assessment, and evaluation of function and health-related quality of life (HRQoL) [6,7]. Regular screening of physical and psychosocial symptoms by use of PROs could increase awareness and recognition of symptoms and can contribute to managing them [7-11]. PROs are furthermore useful in identifying issues that are most bothersome to patients [12] and can enable patients and their health professionals to jointly identify goals and priorities for future health and health care [13].

The use of PROs in clinical practice has increased in the past years [14]. Studies have shown that feedback from PROs can lead to improved symptom detection and more dialogue about problems between patients and physicians [7-11,15-19]. However, some studies reported no benefit from PRO feedback in the number of patients referred to psychosocial care or in clinical actions taken [16,18,20,21]. In most of these studies, PRO feedback was provided to a health care provider (eg, a physician or nurse). A limitation of providing feedback to health care providers might be that they may not always see the urgency of a specific problem and forget to discuss it. Some health care providers were found to downgrade or miss symptoms such as fatigue and pain [22-24]. Physicians are furthermore most interested in PRO scores that indicate worsening symptoms, whereas patients prefer to see both worsened and improved scores [25]. The provision of PRO feedback to patients themselves might therefore be another and maybe better solution. Patients can then monitor all symptoms and initiate discussion on symptoms that bother them the most. Patients are moreover best placed to interpret their own subjective PROs within the complex context of their experience [26]. Patients also report that the inclusion of PROs in their clinical follow-up made them feel more in control of their care [27].

Comparison of a patient's outcomes with those of patients with the same age and sex may help to reassure that patient that what he or she is experiencing is "normal" or may empower the patient to take action. The aim of this study was therefore to investigate whether patients with lymphoma wished to receive PRO feedback including the option to compare their scores with those of their peers. We furthermore investigated how patients evaluated the PRO feedback. We hypothesized that around two-thirds of patients would express a wish to receive feedback, as research shows that about 62% of patients with lymphoma want to be fully informed about their illness [28].

Methods

Participants and Setting

This study was part of the Patient Reported Outcomes Following Initial Treatment and Long Term Evaluation of Survivorship (PROFILES) lymphoma registry [29]. This is a longitudinal population-based observational study whereby patients with Hodgkin lymphoma and non-Hodgkin lymphoma as diagnosed by the Netherlands Cancer Registry in 9 hospitals in the Netherlands complete questionnaires either on paper or online for research purposes. The first patients were included in 2009, and each year patients with a new diagnosis between 9 months and 1.5 year after diagnosis are invited for questionnaire completion. Patients with a diagnosis within less than 3 years are invited to complete a questionnaire every 6 months and patients with a diagnosis more than 3 years ago are invited to complete a questionnaire once a year. In January 2016, we invited patients with a diagnosis made less than 3 years previously and who were eligible for a follow-up questionnaire to participate in this study. Patients who participated online were given an option to receive PRO feedback. We obtained ethical approval for this study from a certified medical ethics committee (METC Brabant, the Netherlands; reference number: NL54096.028.15/P1533).

Questionnaire

The questionnaire completed by patients consisted of the following.

We used the Dutch validated version of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) to assess HRQoL [30]. We added the symptom tingling in hands or feet, as it appeared from the literature and focus groups that this might be a prevalent symptom among patients with lymphoma. Answer categories range from 0 (not at all) to 4 (very much). After linear transformation, all scales and single item measures range from 0 to 100 [30].

We used the Hospital Anxiety and Depression Scale (HADS) [31] to measure anxiety and depressive symptoms in separate subscales of 7 items each. Answers range from 0 to 3, and we calculated scores by adding the items, with a higher score meaning more anxiety or depressive symptoms [31].

We also assessed patients' marital status, educational level, and comorbidity in the questionnaire and categorized comorbidity at the time of the survey according to the adapted Self-Administered Comorbidity Questionnaire [32]. We obtained clinical characteristics (ie, sex, age, type of lymphoma, date of diagnosis, stage at diagnosis, and primary treatment) from the Netherlands Cancer Registry.

Procedure

Eligible patients received a letter or email with an invitation to complete the questionnaire. Patients were informed that when they completed the questionnaire online they would have the possibility to receive PRO feedback. After completing the online questionnaire, patients received the following question: "Would you like to receive feedback on your answers to the questionnaire?" If patients answered yes, we asked them on what topics they would like to receive feedback. They could choose from general quality of life, physical functioning, emotional functioning, cognitive functioning, social functioning, fatigue (based on their scores on the EORTC QLQ-C30), tingling hands or feet (based on their score on the question with respect to tingling hands or feet), anxiety or worries, and depressive symptoms (based on their scores on the HADS), or all topics. Subsequently, patients were asked whether they only wanted to see their own scores, and whether they would like to

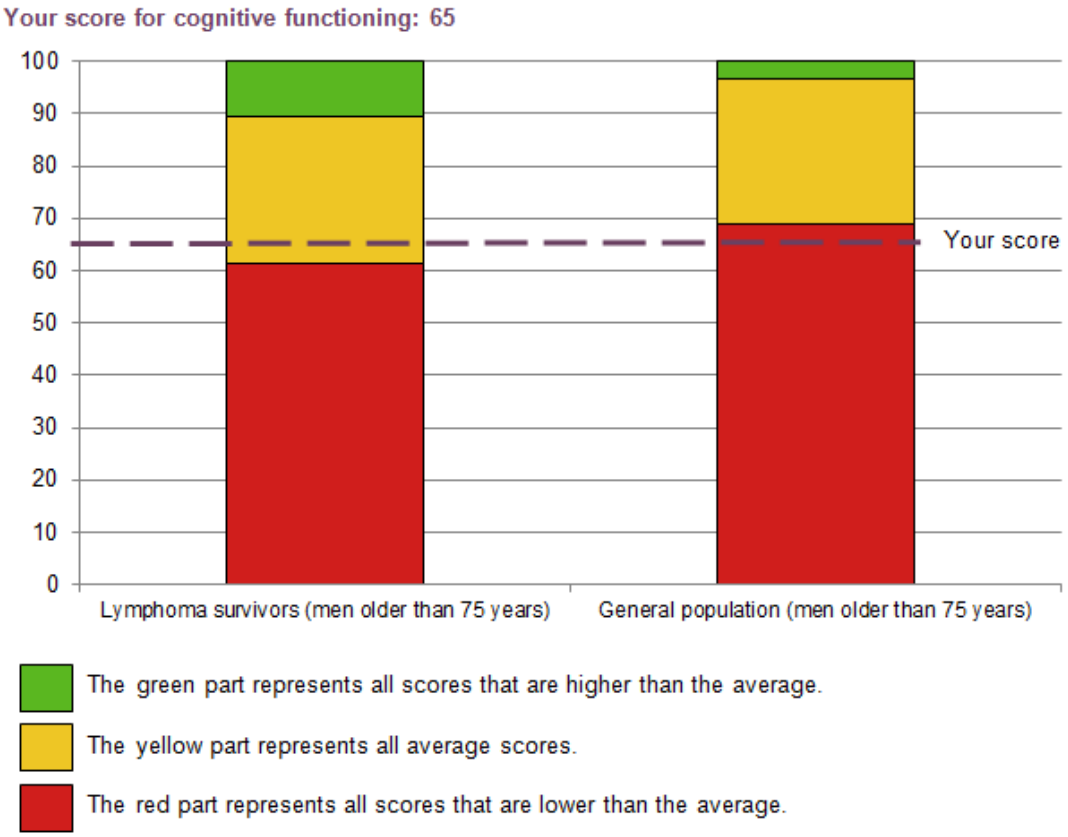
compare their scores with those of other patients with lymphoma or with those of people without cancer, or both. After that, the feedback was generated automatically by computer and was directly shown on the patients' screens. If patients indicated that they did not wish to receive feedback, the feedback was not generated. Patients who viewed their PRO feedback received evaluation questions afterward.

Patient-Reported Outcome Feedback

We based the content and layout of the PRO feedback on examples in the literature [33,34] and on lymphoma patients' preferences reported in an earlier survey on how to provide PRO feedback. In this survey, we presented respondents with 2 examples of PRO feedback: in a bar chart and in a line chart. Respondents had a slight preference for the bar chart. Several examples of PRO feedback presented as bar charts were subsequently evaluated by 12 persons (mean age 55 years; 8/12, 67% female; 5/12, 42% low level of educational attainment). We asked them which colorway they preferred: traffic light colors, pastel colors, or PROFILES house-style colors. Here respondents preferred traffic light colors. Patients furthermore preferred a dotted line over a solid line to indicate "your score" in the bar chart. In this study, we therefore provided the PRO feedback via bar charts in traffic light colors with a dotted line to indicate a patient's score.

If patients wanted to view their own scores, a single bar chart was shown for each PRO feedback topic. If patients had indicated that they wanted to compare their scores with those of a lymphoma reference cohort or a normative population without cancer, both of the same sex and age, either 1 or 2 traffic light-colored bar charts were shown (see Figure 1 for an example). Age was grouped into categories of 10-15 years, ranging from 18-30 years to older than 75 years. The colors of the bar charts were related to clinically relevant mean differences of the evidence-based guidelines of the EORTC QLQ-C30 [35]. A score that differed by less than the minimal medium clinically relevant difference from the mean score was considered average (amber). A score that differed as much as or more than the minimal medium clinically relevant difference from the mean score was considered above average (green) or below average (red). We interpreted anxiety and depressive symptoms according to the published scoring algorithm with 0-7 indicating no or mild symptoms (green), 8-10 indicating moderate symptoms (amber), and ≥11 indicating severe symptoms (red) [31]. We added a detailed description of the meaning of the colors (traffic light model) and how to interpret the scores to assist patients in understanding the graphs (Textbox 1 shows cognitive functioning as an example). Patients with a symptom score in the red part of the bar chart were advised to contact their general practitioner.

Figure 1. The example of cognitive functioning as part of patient-reported outcome feedback provided to participants.



Textbox 1. Description of cognitive functioning (concentration and memory) as an example for interpreting the bar charts.

Cognitive functioning is a component of quality of life. Cognitive functioning, for example, refers to the extent to which one can concentrate or can remember things.

On the cognitive function component, you can score between 0 and 100. The higher the score, and the closer the score is to 100, the higher you will experience your quality of life in this part. Your score is shown in the graph by the purple line.

Your score in comparison with *other lymphoma survivors:*

- Your score falls in the yellow part. This indicates that your score is similar to that of other people with lymphoma with your age and sex.

Your score in comparison with *the general population:*

- Your score falls in the red part. This indicates that your score is lower than the average score of people from the general population with your age and sex.

People with lymphoma score generally lower on cognitive functioning than the general population. Memory and concentration problems are common among people with cancer. Some also experience difficulty working under time pressure or doing different things at the same time. Others must make a greater mental effort to reach the same results as when they were living without cancer. [36]

Lymphoma Reference Cohort and Normative Population

We based the mean scores of the lymphoma reference cohort on data from our previous population-based study on HRQoL among 856 patients with lymphoma [37]. We extracted the mean scores of an age- and sex-matched normative population of 1859 individuals without cancer from a reference cohort from the general Dutch population (CentER panel) [38].

Evaluation Questions

The evaluation questions consisted of 5 open questions with respect to the usefulness, accessibility, clarity, and missing features of the feedback. Patients were furthermore asked whether they would have liked to see different features in the PRO feedback. Based on the average scores on HRQoL and anxiety and depressive symptoms, we evaluated whether both patients with and patients without symptoms wanted to receive PRO feedback.

Statistical Analysis

Analyses were performed using SAS version 9.1 (SAS Institute Inc). $P<.05$ were considered statistically significant. We determined clinically relevant differences using the evidence-based guidelines of the EORTC QLQ-C30 [35].

We used Fischer exact tests or *t* tests to compare differences in sociodemographic and clinical characteristics between respondents and nonrespondents and between patients who wished and those who did not wish to receive PRO feedback.

To evaluate whether scores were on average comparable with those of a lymphoma reference cohort, we compared patients' mean EORTC QLQ-C30 and HADS scores with mean scores of a lymphoma reference group using analysis of covariance with age and sex as covariates. We also compared patients' mean scores, in the same way, with those of a normative population. The numbers of patients scoring in the red, amber, or green part were computed to evaluate whether both patients with and patients without symptoms wished to receive PRO feedback.

Results

Participants

Of the 64 patients who were invited, 45 participated (response rate 70%). Their mean age was 60.7 years and 58% (n=26) were male. Mean time since diagnosis was 2.8 years, and 82% (n=37) had a diagnosis of non-Hodgkin lymphoma. Most patients underwent systemic therapy or radiotherapy, or both. Sociodemographic and clinical characteristics did not statistically differ between respondents and nonrespondents (Table 1).

Evaluation of Patient-Reported Outcome Feedback

A total of 36 (80%) participants wished to receive PRO feedback, with similar percentages for males and females (21/26, 81% vs 15/19, 79%; $P=.29$) and for patients under and above 65 years of age (20/26, 77% vs 16/19, 84%; $P=.25$). Patients who wished to receive PRO feedback had scores on overall HRQoL ($P=.14$) and anxiety ($P=.47$) and depressive symptoms ($P=.25$) similar to those of patients who did not wish to receive feedback.

The vast majority (34/36, 94%) compared their scores with those of the lymphoma reference cohort and 64% (23/36) compared their scores with those of the normative population without cancer, whereas 6% (2/36) viewed only their own scores.

All patients viewed the PRO feedback on their overall HRQoL, and 81% to 92% viewed feedback on their physical, emotional, social, and cognitive functioning, fatigue, tingling in hands or feet, anxiety, and depressive symptoms (Table 2).

Table 1. Sociodemographic and clinical characteristics of respondents and nonrespondents.

Characteristics	Respondents (n=45)	Nonrespondents (n=19)	*P* value, respondents vs nonrespondents
Sex, n (%)			.27
Male	26 (58)	14 (74)	
Female	19 (42)	5 (26)	
Age in years, mean (SD)	60.7 (13.6)	63.8 (14.7)	.28
<65, n (%)	26 (58)	8 (42)	
≥65, n (%)	19 (42)	11 (58)	
Marital status, n (%)			
Partner	34 (76)	N/A[a]	
No partner	11 (24)	N/A	
Educational level attained, n (%)			
Secondary	8 (18)	N/A	
Intermediate vocational	17 (38)	N/A	
High vocational or university	20 (44)	N/A	
Type of lymphoma, n (%)			.26
Hodgkin	8 (18)	1 (5)	
Non-Hodgkin	37 (82)	18 (95)	
Years since diagnosis at time of questionnaire completion, mean (SD)	2.8 (0.8)	2.6 (0.7)	.84
Cancer stage at diagnosis, n (%)			.50
I	8 (22)	4 (29)	
II	10 (28)	2 (14)	
III	5 (14)	4 (29)	
IV	13 (36)	4 (29)	
Primary treatment, n (%)			.11
Radiotherapy only	2 (4)	1 (5)	
Systemic therapy (eg, chemotherapy, immunotherapy)	19 (42)	8 (42)	
Systemic therapy plus radiotherapy	13 (29)	1 (5)	
Active surveillance	11 (25)	9 (47)	
Self-reported comorbidities, mean (SD)	1.3 (1.3)	N/A	
Most frequently reported self-reported comorbidities, n (%)			
Arthritis	10 (22)	N/A	
Heart problems	8 (18)	N/A	
High blood pressure	8 (18)	N/A	

[a]N/A: not available.

Table 2. Overview of patient-reported outcome feedback topics with number and percentage of interested patients.

Topic	n	%
EORTC QLQ-C30[a]		
General HRQoL[b]	36	100
Physical functioning	33	92
Emotional functioning	32	89
Social functioning	33	92
Cognitive functioning	31	86
Fatigue	31	86
Neuropathy	29	81
HADS[c]		
Anxiety	30	83
Depressive symptoms	30	83

[a]EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

[b]HRQoL: health-related quality of life.

[c]HADS: Hospital Anxiety and Depression Scale.

Almost all patients (except 1) viewed the PRO feedback as being useful, with reassurance and knowledge about their own functioning in relation to what is "normal" being the most frequently mentioned reasons. The option to compare their scores with those of a lymphoma reference cohort of the same age and sex was reported as most valuable:

> *This score shows what I actually did expect of my quality of life. The comparison with other patients with lymphoma feels right. I mean, I do not score that different and that again reassures me.* [Female patient with non-Hodgkin lymphoma, 69 years old]
>
> *It is interesting to see how I stand compared to other patients with lymphoma and the general population.* [Male patient with Hodgkin lymphoma, 22 years old]
>
> *The PRO feedback clarifies if symptoms are shared by others or not.* [Female patient with Hodgkin lymphoma, 37 years old]

Some patients reported that the PRO feedback was useful, since it provided new insights for discussion with their physician. No reason was provided by the patient who indicated that the PRO feedback was not useful.

A total of 2 patients reported that the PRO feedback had missing features; 1 patient advised us to provide more information on how to limit symptom burden or improve symptoms; and 1 patient suggested that it would be good to advise patients to go to their general practitioner when experiencing problems:

> *Not everyone has good and regular contact with their doctors, so it would be helpful to advise a patient to contact a doctor when he or she reports problems.* [Female patient with non-Hodgkin lymphoma, 54 years old]

The comment regarding contacting a general practitioner was already covered for the symptoms in the current PRO feedback for patients scoring in the red part of the bar chart, but not for the functioning scales.

With respect to the clarity of the PRO feedback, 1 patient missed the possibility to go back to his answers in the questionnaire to verify that the PRO feedback was correct, because his score on neuropathy was very low according to the PRO feedback, but not in his experience. Furthermore, 1 patient had trouble understanding the colors of the PRO feedback at first, but after looking for a second time it became clear. With respect to things that should be different, some patients indicated that they wished to save their scores for future comparison purposes and to keep track of their scores:

> *Is it possible to download my PRO feedback, so I can compare my scores over time and determine potential deterioration?* [Male patient with non-Hodgkin lymphoma, 84 years old]

Health-Related Quality of Life, Anxiety, and Depression Scores

Mean scores on HRQoL, anxiety, and depressive symptoms of participating patients in this study were not different from the mean scores of the lymphoma reference cohort (Table 3). Compared with the normative population, patients had on average statistically and clinically relevant lower scores on physical, cognitive, and social functioning and higher scores on fatigue (all $P<.05$).

With respect to patients' individual scores on HRQoL, 33% (n=15) of patients reported scores that were lower than the average of the lymphoma reference cohort (red part of bar chart) and 31% (n=14) reported scores higher than the average range of scores (green part of bar chart; Table 4). Compared with the normative population, 33% (n=15) of patients reported scores that were lower than the average and 20% (n=9) reported scores higher than the average of the normative population. The percentages were similar for the other scales (data not shown).

Table 3. EORTC QLQ-C30[a] + tingling hands or feet and HADS[b] mean scores of patients, a lymphoma reference cohort, and a normative population, and clinically important differences between these groups.

Measure	Patients (n=45)	Lymphoma reference cohort (n=876)	Normative population (n=1852)	Patients vs lymphoma cohort		Patients vs normative population	
				P value[c]	Clinical relevance	*P* value[c]	Clinical relevance
EORTC QLQ-C30, mean (SD)							
Physical functioning	83.1 (20)	79.4 (21)	90.5 (15)	.21	No	<.001	Small
Emotional functioning	82.2 (21)	82.8 (21)	87.9 (17)	.86	No	.02	Trivial
Cognitive functioning	80.4 (22)	82.4 (23)	92.5 (14)	.57	No	<.001	Medium
Social functioning	85.9 (25)	84.4 (24)	93.6 (16)	.68	No	.002	Small
Global health status/QoL[d]	73.3 (20)	74.0 (20)	77.6 (17)	.82	No	.10	Small
Fatigue	24.7 (23)	28.9 (27)	17.0 (20)	.30	No	.01	Small
Tingling hands or feet	18.5 (28)	17.0 (29)	N/A[e]	.73	No	N/A	N/A
HADS, mean (SD)							
Anxiety	4.0 (3.8)	4.4 (3.8)	3.6 (3.2)	.51	No	.34	No
Depressive symptoms	3.9 (3.8)	4.7 (3.8)	3.6 (3.2)	.17	No	.54	No

[a]EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

[b]HADS: Hospital Anxiety and Depression Scale.

[c]*P* value is adjusted for age and sex.

[d]QoL: quality of life.

[e]N/A: not available.

Table 4. Number and percentages of patients scoring lower, similar to, or higher than the lymphoma reference cohort and normative population on EORTC QLQ-C30[a] global health status/quality of life.

Relative scores	Compared with lymphoma reference cohort	Compared with normative population
Lower than average (red)	15 (33)	15 (33)
Average (amber)	16 (36)	21 (47)
Higher than average (green)	14 (31)	9 (20)

[a]EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

Discussion

Principal Findings

Of the participating patients with lymphoma, 80% wished to receive PRO feedback, which was higher than the two-thirds of patients that we hypothesized. A similar percentage of men and women and patients younger and older than 65 years wished to receive PRO feedback. They reported the comparison of their scores with those of a lymphoma reference cohort as being very valuable, since it provided information about their functioning in relation to what is "normal."

An advantage of providing PRO feedback to patients themselves is that patients can monitor their symptoms at any specific point in time. Patients are furthermore provided with information that they can use to actively engage with their physician when discussing symptoms [26,27]. However, not all patients will be self-assertive enough to bring up their problems and, in that case, providing feedback to both patients and physicians, as is done in some studies [16-18], might be more effective for discussing problems and taking action with respect to referral to other health care professionals.

Almost all patients indicated that the PRO feedback was useful and reassuring. Even when patients had scores that were below average, they still viewed PRO feedback as useful. The latter pleads for providing PRO feedback as a standard option in care. However, before PRO feedback is provided, patients need to be asked for their preference, as still 20% indicated that they did not want to receive PRO feedback. This is the case for information provision in general, as patients fare psychologically, behaviorally, and physiologically better when the information they receive about their medical condition is tailored to their coping styles, whereby those with a monitoring style tend to do better when given more information, and those with a blunting style do better with less information [39].

Since the feedback was generated automatically after patients completed the questionnaire, implementation in our PROFILES registry is relatively simple. In addition, providing PRO feedback is valuable for other research that is performed with online questionnaires, as well as for patients with other medical

conditions in terms of empowering patients and monitoring their functioning and symptoms.

In this study, we evaluated PRO feedback in a research setting at a fixed time point, but this kind of PRO feedback could also be of merit for patients at any given point in time outside of a research setting. It can, for example, be used as a tool for keeping track of their scores, which may help patients to feel more in control of their cancer and care [27].

Limitations

The sample size was relatively small, although we obtained a response rate of 70%. The PRO feedback was accessible only to patients completing the questionnaire online, which limits the generalizability of the results to the total lymphoma population, as patient characteristics are different for patients who participated online versus patients who participated on paper [40].

Conclusion

Future research should determine whether this kind of feedback could also increase empowerment and possibly improve HRQoL.

Acknowledgments

The PRO feedback was developed with technical assistance for information and communication technology applications by CentERdata, Tilburg University, Tilburg, the Netherlands.

Conflicts of Interest

None declared.

References

1. Behringer K, Goergen H, Müller H, Thielen I, Brillant C, Kreissl S, et al. Cancer-related fatigue in patients with and survivors of Hodgkin lymphoma: the impact on treatment outcome and social reintegration. J Clin Oncol 2016 Dec 20;34(36):4329-4337. [doi: 10.1200/JCO.2016.67.7450] [Medline: 27998235]
2. Daniëls LA, Oerlemans S, Krol ADG, Creutzberg CL, van de Poll-Franse LV. Chronic fatigue in Hodgkin lymphoma survivors and associations with anxiety, depression and comorbidity. Br J Cancer 2014 Feb 18;110(4):868-874 [FREE Full text] [doi: 10.1038/bjc.2013.779] [Medline: 24434433]
3. Oerlemans S, Mols F, Issa DE, Pruijt JHFM, Peters WG, Lybeert M, et al. A high level of fatigue among long-term survivors of non-Hodgkin's lymphoma: results from the longitudinal population-based PROFILES registry in the south of the Netherlands. Haematologica 2013 Mar;98(3):479-486 [FREE Full text] [doi: 10.3324/haematol.2012.064907] [Medline: 22929981]
4. Williams AM, Zent CS, Janelsins MC. What is known and unknown about chemotherapy-related cognitive impairment in patients with haematological malignancies and areas of needed research. Br J Haematol 2016 Sep;174(6):835-846. [doi: 10.1111/bjh.14211] [Medline: 27391367]
5. Weeks JC, Catalano PJ, Cronin A, Finkelman MD, Mack JW, Keating NL, et al. Patients' expectations about effects of chemotherapy for advanced cancer. N Engl J Med 2012 Oct 25;367(17):1616-1625 [FREE Full text] [doi: 10.1056/NEJMoa1204410] [Medline: 23094723]
6. Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health 2003;6(5):522-531 [FREE Full text] [doi: 10.1046/j.1524-4733.2003.65309.x] [Medline: 14627058]
7. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? Qual Life Res 2009 Feb;18(1):115-123. [doi: 10.1007/s11136-008-9430-6] [Medline: 19105048]
8. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
9. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW, PatientViewpoint Scientific Advisory Board. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901 [FREE Full text] [doi: 10.1002/pon.3087] [Medline: 22544513]
10. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008 Mar;17(2):179-193. [doi: 10.1007/s11136-007-9295-0] [Medline: 18175207]
11. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724 [FREE Full text] [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]

12. Snyder CF, Blackford AL, Aaronson NK, Detmar SB, Carducci MA, Brundage MD, et al. Can patient-reported outcome measures identify cancer patients' most bothersome issues? J Clin Oncol 2011 Mar 20;29(9):1216-1220. [doi: 10.1200/JCO.2010.33.2080] [Medline: 21343558]
13. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006 Oct;12(5):559-568. [doi: 10.1111/j.1365-2753.2006.00650.x] [Medline: 16987118]
14. Aaronson NK, Snyder C. Using patient-reported outcomes in clinical practice: proceedings of an International Society of Quality of Life Research conference. Qual Life Res 2008 Dec;17(10):1295. [doi: 10.1007/s11136-008-9422-6] [Medline: 19048409]
15. Brundage MD, Smith KC, Little EA, Bantug ET, Snyder CF. Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. Qual Life Res 2015 Oct;24(10):2457-2472 [FREE Full text] [doi: 10.1007/s11136-015-0974-y] [Medline: 26012839]
16. Carlson LE, Groff SL, Maciejewski O, Bultz BD. Screening for distress in lung and breast cancer outpatients: a randomized controlled trial. J Clin Oncol 2010 Nov 20;28(33):4884-4891. [doi: 10.1200/JCO.2009.27.3698] [Medline: 20940193]
17. Detmar SB, Muller MJ, Schornagel JH, Wever LDV, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. JAMA 2002 Dec 18;288(23):3027-3034. [Medline: 12479768]
18. Hilarius DL, Kloeg PH, Gundy CM, Aaronson NK. Use of health-related quality-of-life assessments in daily clinical oncology nursing practice: a community hospital-based intervention study. Cancer 2008 Aug 01;113(3):628-637 [FREE Full text] [doi: 10.1002/cncr.23623] [Medline: 18543317]
19. Llewellyn AM, Skevington SM. Using guided individualised feedback to review self-reported quality of life in health and its importance. Psychol Health 2015;30(3):301-317. [doi: 10.1080/08870446.2014.972396] [Medline: 25280244]
20. Braeken APBM, Kempen GIJM, Eekers D, van Gils FC, Houben RMA, Lechner L. The usefulness and feasibility of a screening instrument to identify psychosocial problems in patients receiving curative radiotherapy: a process evaluation. BMC Cancer 2011 Nov 08;11:479 [FREE Full text] [doi: 10.1186/1471-2407-11-479] [Medline: 22067707]
21. Rosenbloom SK, Victorson DE, Hahn EA, Peterman AH, Cella D. Assessment is not enough: a randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. Psychooncology 2007 Dec;16(12):1069-1079. [doi: 10.1002/pon.1184] [Medline: 17342789]
22. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009 Dec 2;101(23):1624-1632 [FREE Full text] [doi: 10.1093/jnci/djp386] [Medline: 19920223]
23. Kirchheiner K, Nout R, Lindegaard J, Petrič P, Limbergen EV, Jürgenliemk-Schulz I, et al. Do clinicians and patients agree regarding symptoms? A comparison after definitive radiochemotherapy in 223 uterine cervical cancer patients. Strahlenther Onkol 2012 Oct;188(10):933-939. [doi: 10.1007/s00066-012-0156-6] [Medline: 22895624]
24. Quinten C, Maringwa J, Gotay CC, Martinelli F, Coens C, Reeve BB, et al. Patient self-reports of symptoms and clinician ratings as predictors of overall cancer survival. J Natl Cancer Inst 2011 Dec 21;103(24):1851-1858 [FREE Full text] [doi: 10.1093/jnci/djr485] [Medline: 22157640]
25. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Development of MijnAVL, an interactive portal to empower breast and lung cancer survivors: an iterative, multi-stakeholder approach. JMIR Res Protoc 2015;4(1):e14 [FREE Full text] [doi: 10.2196/resprot.3796] [Medline: 25614924]
26. Llewellyn AM, Skevington SM. Evaluating a new methodology for providing individualized feedback in healthcare on quality of life and its importance, using the WHOQOL-BREF in a community population. Qual Life Res 2016 Mar;25(3):605-614. [doi: 10.1007/s11136-015-1132-2] [Medline: 26370098]
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. Rood JAJ, van Zuuren FJ, Stam F, van der Ploeg T, Eeltink C, Verdonck-de Leeuw IM, et al. Perceived need for information among patients with a haematological malignancy: associations with information satisfaction and treatment decision-making preferences. Hematol Oncol 2015 Jun;33(2):85-98. [doi: 10.1002/hon.2138] [Medline: 24811073]
29. van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, Profiles Registry Group. The Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. Eur J Cancer 2011 Sep;47(14):2188-2194 [FREE Full text] [doi: 10.1016/j.ejca.2011.04.034] [Medline: 21621408]
30. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 3;85(5):365-376. [Medline: 8433390]
31. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]

32. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003 Apr 15;49(2):156-163 [FREE Full text] [doi: 10.1002/art.10993] [Medline: 12687505]
33. Brundage M, Feldman-Stewart D, Leis A, Bezjak A, Degner L, Velji K, et al. Communicating quality of life information to cancer patients: a study of six presentation formats. J Clin Oncol 2005 Oct 1;23(28):6949-6956. [doi: 10.1200/JCO.2005.12.514] [Medline: 16192583]
34. Kuijpers W, Giesinger JM, Zabernigg A, Young T, Friend E, Tomaszewska IM, et al. Patients' and health professionals' understanding of and preferences for graphical presentation styles for individual-level EORTC QLQ-C30 scores. Qual Life Res 2016 Mar;25(3):595-604. [Medline: 26353905]
35. Cocks K, King MT, Velikova G, Martyn SM, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol 2011 Jan 1;29(1):89-96. [doi: 10.1200/JCO.2010.28.0107] [Medline: 21098316]
36. Brouwer T, Schagen S. Problemen met aandacht: geheugenproblemen na kanker. Klankborden 2012;May:4-8 [FREE Full text]
37. Oerlemans S, Husson O, Mols F, Poortmans P, Roerdink H, Daniels LA, et al. Perceived information provision and satisfaction among lymphoma and multiple myeloma survivors--results from a Dutch population-based study. Ann Hematol 2012 Oct;91(10):1587-1595 [FREE Full text] [doi: 10.1007/s00277-012-1495-1] [Medline: 22638756]
38. van de Poll-Franse LV, Mols F, Gundy CM, Creutzberg CL, Nout RA, Verdonck-de Leeuw IM, et al. Normative data for the EORTC QLQ-C30 and EORTC-sexuality items in the general Dutch population. Eur J Cancer 2011 Mar;47(5):667-675. [doi: 10.1016/j.ejca.2010.11.004] [Medline: 21134739]
39. Miller SM. Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease. Implications for cancer screening and management. Cancer 1995 Jul 15;76(2):167-177. [Medline: 8625088]
40. Horevoorts NJ, Vissers PA, Mols F, Thong MS, van de Poll-Franse LV. Response rates for patient-reported outcomes using web-based versus paper questionnaires: comparison of two invitational methods in older colorectal cancer patients. J Med Internet Res 2015;17(5):e111 [FREE Full text] [doi: 10.2196/jmir.3741] [Medline: 25953059]

Abbreviations

EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
HADS: Hospital Anxiety and Depression Scale
HRQoL: health-related quality of life
PRO: patient-reported outcome
PROFILES: Patient Reported Outcomes Following Initial Treatment and Long Term Evaluation of Survivorship

Edited by C Gibbons; submitted 01.12.16; peer-reviewed by A van der Hout, E Tolbert, R Maguire; comments to author 20.01.17; revised version received 30.03.17; accepted 30.06.17; published 15.08.17.

Please cite as:
Oerlemans S, Arts LP, Horevoorts NJ, van de Poll-Franse LV
"Am I normal?" The Wishes of Patients With Lymphoma to Compare Their Patient-Reported Outcomes With Those of Their Peers
J Med Internet Res 2017;19(8):e288
URL: http://www.jmir.org/2017/8/e288/
doi:10.2196/jmir.7079
PMID:28811271

Original Paper

Patient Participation at Health Care Conferences: Engaged Patients Increase Information Flow, Expand Propagation, and Deepen Engagement in the Conversation of Tweets Compared to Physicians or Researchers

Audun Utengen[1*], MBA; Dara Rouholiman[2*], BS; Jamison G Gamble[2*], MPH; Francisco Jose Grajales III[3*], MS, CD, R Kin; Nisha Pradhan[4*], BA; Alicia C Staley[2*], MBA, MS; Liza Bernstein[2*]; Sean D Young[5*], MS, PhD; Kevin A Clauson[6*], PharmD; Larry F Chu[2*], MS, MD

[1]Symplur, Los Angeles, CA, United States

[2]Stanford Medicine X, Stanford University School of Medicine, Stanford, CA, United States

[3]Center for Social Innovation and Impact Investing, Sauder School of Business, University of British Columbia, Vancouver, BC, Canada

[4]Clinical Genetics Service, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[5]University of California Institute for Prediction Technology, Department of Family Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[6]College of Pharmacy, Lipscomb University, Nashville, TN, United States

[*]all authors contributed equally

Corresponding Author:
Larry F Chu, MS, MD
Stanford Medicine X
Stanford University School of Medicine
300 Pasteur Drive
Stanford, CA, 94305
United States
Phone: 1 (650) 723 6632
Fax: 1 (650) 497 9335
Email: lchu@stanford.edu

Abstract

Background: Health care conferences present a unique opportunity to network, spark innovation, and disseminate novel information to a large audience, but the dissemination of information typically stays within very specific networks. Social network analysis can be adopted to understand the flow of information between virtual social communities and the role of patients within the network.

Objective: The purpose of this study is to examine the impact engaged patients bring to health care conference social media information flow and how they expand dissemination and distribution of tweets compared to other health care conference stakeholders such as physicians and researchers.

Methods: From January 2014 through December 2016, 7,644,549 tweets were analyzed from 1672 health care conferences with at least 1000 tweets who had registered in Symplur's Health Care Hashtag Project from 2014 to 2016. The tweet content was analyzed to create a list of the top 100 influencers by mention from each conference, who were then subsequently categorized by stakeholder group. Multivariate linear regression models were created using stepwise function building to identify factors explaining variability as predictor variables for the model in which conference tweets were taken as the dependent variable.

Results: Inclusion of engaged patients in health care conference social media was low compared to that of physicians and has not significantly changed over the last 3 years. When engaged patient voices are included in health care conferences, they greatly increase information flow as measured by total tweet volume (beta=301.6) compared to physicians (beta=137.3, P<.001), expand propagation of information tweeted during a conference as measured by social media impressions created (beta=1,700,000) compared to physicians (beta=270,000, P<.001), and deepen engagement in the tweet conversation as measured by replies to their tweets (beta=24.4) compared to physicians (beta=5.5, P<.001). Social network analysis of hubs and authorities revealed

that patients had statistically significant higher hub scores (mean 8.26×10-4, SD 2.96×10-4) compared to other stakeholder groups' Twitter accounts (mean 7.19×10-4, SD 3.81×10-4; t273.84=4.302, *P*<.001).

Conclusions: Although engaged patients are powerful accelerators of information flow, expanders of tweet propagation, and greatly deepen engagement in conversation of tweets on social media of health care conferences compared to physicians, they represent only 1.4% of the stakeholder mix of the top 100 influencers in the conversation. Health care conferences that fail to engage patients in their proceedings may risk limiting their engagement with the public, disseminating scientific information to a narrow community and slowing flow of information across social media channels.

(J Med Internet Res 2017;19(8):e280) doi:10.2196/jmir.8049

KEYWORDS

social media; patients; physicians; patient participation; congresses as topic; social networking, network analysis

Introduction

Traditionally, health care conferences are attended by experts, industry specialists, or others involved in fields specific to the conference in question. Health care conferences present a unique opportunity to network, spark innovation, and disseminate novel information to a large audience. Twitter is a microblogging and social media site with 313,000,000 monthly users, 82% of which are primarily mobile users. Twitter is gaining in popularity at health care conferences by allowing attendees to interact with one another and with their greater social networks, facilitating the sharing of information and ideas [1-13]. For example, the annual meeting of the American Society of Clinical Oncology saw an increase in tweets from 10,475 in 2012 to 44,034 in 2014 which resulted in 53,001,708 impressions in 2012 (ie, number of times the tweet was seen determined by the total number of followers who could view the tweet) and 154,362,922 impressions in 2014 [11].

We previously reported on the importance of including patients in medical conferences and identified four pillars of patient involvement in academic medical conferences [14]. These four pillars include accommodation (considering the physical needs of patients), codesign (patients codesign conference along with program creator), engagement (including patients in the audience and as presenters), and education and mentorship (guide patients toward conference stakeholder collaboration). By involving patients in health care conferences, a new voice is added to the discussion. Arguably, the ultimate purpose of health care conferences and health care is to improve the lives of patients and their families. By actively including patients in the conversation, patients are able to share their thoughts and express the issues that matter most to them [15]. Inclusion and engagement of patients can help drive information dissemination in health care conferences and widen research agendas to include new patient-centered domains.

Numerous studies have attempted to explain how patients and providers utilize and communicate via social media. In a systematic review of the literature, Smailhodzic et al [16] identified studies that examined patient and provider use of social media and identified six uses of social media for health-related purposes: emotional support, esteem support, information support, network support, emotional expression, and social comparison. Furthermore, the authors identified the primary effects of social media use by patients for health-related reasons, including positive effects (eg, empowerment, enhanced subjective well-being, enhanced psychological well-being, improved self-management and control) and negative effects (eg, diminished subjective well-being, loss of privacy, being targeted for promotion, and addiction to social media). The effects of patient use of social media on provider-patient relationships were also examined and included improved communication, harmonious relationships, and inferior interactions. In general, social media makes it easier to partner with patients and is their preferred method of communication.

Despite these studies, little is known about how stakeholders communicate via social media at health care conferences. To track information dissemination and diffusion during health care conferences, it is important to start by analyzing social networks. Social network analysis can be adapted to understand the flow of information between virtual social communities and to examine how individual user roles affect conversation dynamics [17]. Social networks are comprised of nodes and edges, with nodes representing individual users (represented as a circle) and edges representing connections between individual users (represented by a line). The degree of a node is the total number of edges connected to an individual node [18]. Furthermore, analysis of hubs and authorities within social networks may reveal additional information pertaining to that network. Authorities are defined as reputable sources of information that point to many hubs within a social network. Hubs are not authorities on their own, but point to multiple authorities within a social network. Hence, a good hub points to many good authorities, whereas a good authority is selected by a variety of good hubs. Connection topology of social networks is neither completely random nor completely systematic; it has characteristics of what Watts and Strogatz [19] first called "small-world networks," in which the degree distribution of nodes approximates a power law distribution with pockets of cohesive communities throughout the network [19,20]. Health care conferences often have high community cohesion with quick access to information; however, information often does not disseminate to a broader audience. For a health care conference to disseminate information outside its community to a broader audience, "influential hub" nodes are essential. Influential hub nodes are social network users who, due to their position in the network, have shorter edges that connect them to other nodes/users in different communities [21]. Small network analysis can be utilized to demonstrate that engaged patients act as influential hub nodes during health care conferences and play an essential role in information

dissemination. Engaged patients are broadly defined as any patient who actively participates in their health care through shared decision making, continued mindfulness of personal health needs within the context of their life, proactive seeking of information pertaining to their health, the setting of personal health goals, and the seeking of resources to achieve set goals [22].

In this study, we conducted a retrospective analysis of more than 7.5 million tweets from 1672 health care conferences that occurred from 2014 to 2016, which were registered in the largest online directory of health care conferences using social media [23]. We assessed three primary measures of social media performance concerning information dissemination during live health care conference coverage which included information flow, information propagation, and engagement in conversation in six stakeholder cohorts (patients, physicians and researchers, nonphysician health care professionals [HCPs], journalists, other health care individuals, and pharmaceutical organizations) and assessed performance of these cohorts against one another by these measures (Table 1). Definitions of stakeholder groups are presented in Table 2.

Table 1. Twitter metrics description.

Metric	Description	Purpose
Information flow	Total number of tweets as a performance indicator	Frequency of information disseminated during a health care conference
Engagement in conversation	Number of replies as a quality indicator	Measure of engagement and active conversation
Information propagation	Total number of potential impressions as dissemination network size	Prediction of network size; how many people/groups received your message?

Table 2. Definitions of stakeholder groups.[a]

Stakeholder	Definition
Patient	A person whose primary use of Twitter is to express their point of view as a patient with a specific disease or condition
Physicians and researchers	Those believed to be licensed MDs, DOs, or PhDs who bill directly for services, including residents and persons who work in the field of health-related research and/or academia
Health care professionals (HCPs)	Those believed to be health care professionals (eg, nurses, dietitians, respiratory therapists, nurse practitioners, pharmacists)
Journalists	Person whose profession is journalism or other news-related media
Other health care individual	Person working in the health care industry in a nonclinical role
Pharmaceutical organization	All organizations in the pharmaceutical industry

[a] As defined by Symplur.

Methods

Categorization of Conferences and Stakeholders

Data was collected with the Symplur Signals research platform (Los Angeles, CA, USA) with direct access to the Twitter application program interface (San Francisco, CA, USA) [24]. We analyzed 7,644,549 tweets from 1672 health care conferences registered in Symplur's Health Care Hashtag Project (the world's largest collection of publicly available health care hashtags) from 2014 to 2016 (a total of 5692 conference hashtags), with at least 1000 tweets [23,25]. Metrics used in this study are defined in Table 1. The social network was analyzed to create a list of the top 100 influencers by mention from each conference. Influencers were subsequently identified and categorized by stakeholder group: patients, physicians and researchers, HCPs, journalists, other health care individual or pharmaceutical organizations based on Twitter accounts that publicly self-categorized their biographical description as certain stakeholders (eg, "radiologist," "professor," "nurse"; Table 2).

Multivariate linear regression models were created using stepwise function building to identify factors explaining variability as predictor variables for the model in which conference tweet were taken as the dependent variable. The categorization process involved a multinomial logistic regression multiclass classification model with a manual verification step, by which 156,149 Twitter accounts were categorized.

Statistical Analysis

Comparative Analysis

Statistical analysis was conducted using the open source programing language R, version 3.3.3 (Vienna, Austria, 2017) [26]. A Welch two-sample *t* test was used to compare the relative performance between those conferences with at least one patient among its top 100 influencers by mentions and those conferences without any patients among its influencers, with the number of total tweets from the respective conferences as the information flow indicator.

Predictive Analysis

Multiple regression analysis was constructed to test if the performance metrics significantly ($P<.05$) depended on number of patients among the top influencers of a conference.

Number of Tweets, Replies, and Impressions

Multiple regression analysis was used to test if the health care stakeholders' composition significantly predicted the conferences information flow (ie, number of tweets), engagement in conversations on Twitter (ie, number of replies), and information propagation from conferences on Twitter (ie, impressions). Replies are indicators of engagement in conversation tweets, which represent a quality, back-and-forth conversation and not simply a broadcast tweet or a random retweet. Information propagation was calculated based on total potential impressions by multiplying the number of tweets from each Twitter account with their number of followers then taking the sum of that number for all accounts tweeting during the conference.

Social Network Analysis (Hubs and Authorities)

Stanford Medicine X is an annual health care conference on emerging technology and medicine, focusing on patient-centered innovation and embraces the philosophy of Everyone Included, which places value on the voices of all health care stakeholders [27]. The influential hubs and authorities of the entire social conversation from the 2016 Stanford Medicine X conference was investigated by using the weighted hyperlink-induced topic search (HITS) algorithm. The values of the weighted HITS hub and authority for each Twitter participant as a node in this social network were calculated [28]. HITS was originally introduced by Kleinberg [21] to rate the importance of a node in a complex directed network using authority and hub values, where hub vectors $y=(y_1,...,y_n)^t$ and authority vectors $x=(x_1,...,x_n)^t$ are defined as:

$$x(t+1)=c(t)A^{t}y(t)$$

$$y(t+1)=d(t)Ax(t+1)$$

To shed more light on why patients bring better social performance to conferences, patient nodes were compared to nonpatient nodes using a Welch two-sample *t* test for both the hub scores and authority scores.

Results

Health Care Hashtag Project Conferences

Table 3 displays the total number of conferences in 2014, 2015, and 2016 registered with the Health Care Hashtag Project. During the 3 years included in the study (January 1, 2014 to December 31, 2016) a total of 5692 conferences were identified that utilized the Health Care Hashtag Project. Conferences with at least 1000 tweets were elected for analysis, which yielded 7,644,549 tweets from 1672 conferences, of which 749 had at least one patient in the top 100 influencers by mentions.

Conference Stakeholders

A total of 156,149 Twitter accounts were categorized into 16 stakeholder groups, of which 75,720 belonged to either a patient (n=2355), physician or researcher (n=32,930), HCP (n=10,344), journalist (n=1756), other health care individual (n=26,428), or pharmaceutical organization (n=1907) within the top 100 influencers by mention. Although 16 stakeholder groups were categorized, only six were isolated for analysis. Figure 1 and Table 4 show descriptive statistics of the top 100 influencers by mention within the six primary stakeholder groups analyzed. From 2014 to 2016 each category saw a decrease in the mean number of top 100 influencers by category with the exception of pharmaceutical organizations, which saw a mean increase of 0.15 (Table 4). From 2015 to 2016, the categories of patients, physicians and researchers, and pharmaceutical organizations saw a mean increase of 0.03, 1.43, and 0.11, respectively.

Figure 1. Stakeholder groups among top 100 influencers at health care conferences.

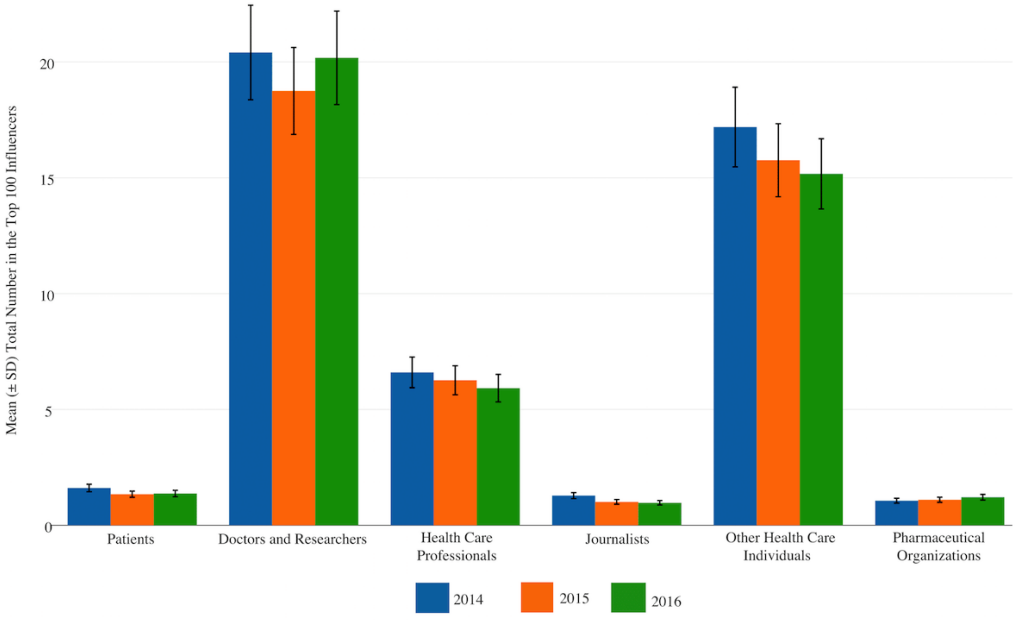


Table 3. Conferences registered with the Health Care Hashtag Project.

Conference metric	Year, n		
	2014	2015	2016
Total conferences	1428	1982	2282
Conferences with >1000 tweets	347	620	705
Total tweets from analyzed conferences	1,543,862	2,710,012	3,390,675

Table 4. Descriptive statistics for stakeholder groups among top 100 influencers at health care conferences.

Stakeholder and year	Mean (SD)	Median (range)
Patients		
2014	1.61 (3.23)	1 (0-39)
2015	1.34 (2.89)	0 (0-28)
2016	1.37 (3.70)	0 (0-51)
Physicians and researchers		
2014	20.41 (17.21)	16 (0-69)
2015	18.75 (16.18)	13 (0-72)
2016	20.18 (16.43)	15 (0-72)
Health care professionals		
2014	6.60 (11.24)	2 (0-68)
2015	6.26 (9.98)	2 (0-65)
2016	5.92 (9.57)	2 (0-61)
Journalists		
2014	1.28 (1.64)	1 (0-39)
2015	1.01 (2.17)	0 (0-28)
2016	0.97 (1.66)	0 (0-51)
Other health care individuals		
2014	17.19 (11.34)	15 (0-68)
2015	15.76 (10.05)	13 (0-51)
2016	15.17 (9.57)	13 (0-51)
Pharmaceutical organizations		
2014	1.06 (2.49)	0 (0-16)
2015	1.10 (2.54)	0 (0-16)
2016	1.21 (2.75)	0 (0-20)

Comparison of Performance Between Conferences

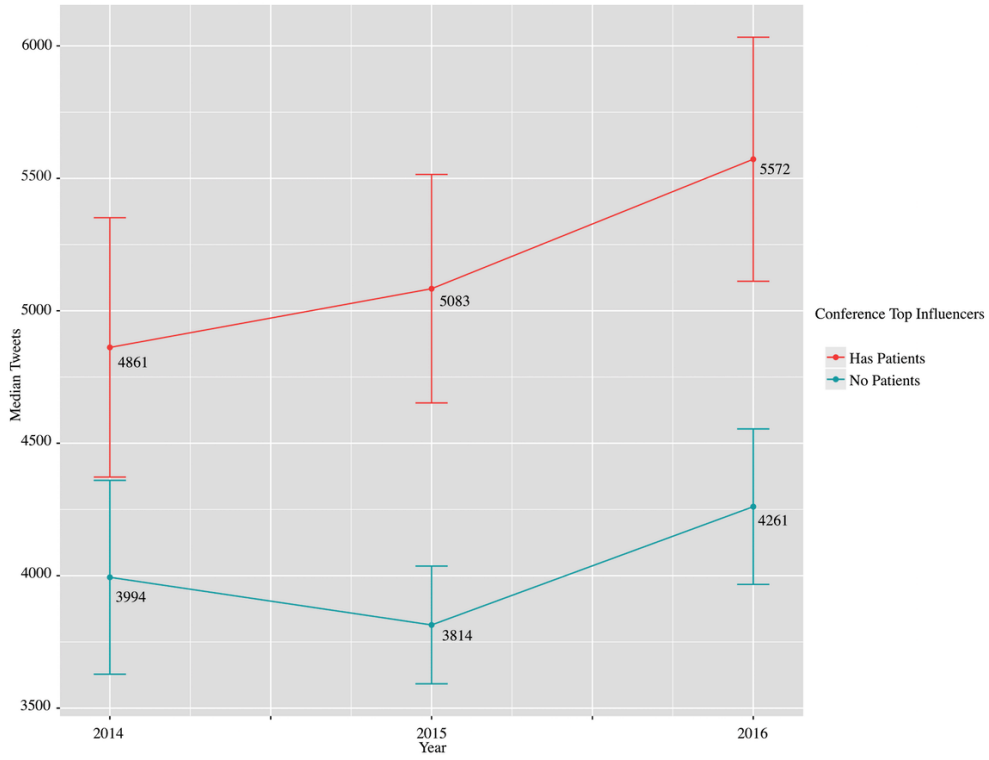
A combined analysis of the 2014, 2015, and 2016 conferences revealed that conferences with patients were found to have statistically significant higher number of tweets (mean 5222, SD 7320) compared to conferences that had no patients (mean 4044, SD 5108: $t_{1292.3}$=3.7271, P<.001). The years 2015 and 2016 had statistically significant differences in the means between conferences with and without patients, although no statistically significant difference was found for 2014 (Figure 2). In 2014, conferences with patients had a mean total number of tweets of 4861 (SD 6604) compared to mean 3994 (SD 4700) for conferences with no patients ($t_{1.42}$=327.05, P=.15). In 2015, conferences with patients had a mean total number of tweets of 5083 (SD 7122) compared to mean 3814 (SD 4144) for conferences with no patients ($t_{2.62}$=411.34, P=.009). In 2016, conferences with patients had a mean total number of tweets of 5572 (SD 7913) compared to mean 4261 (SD 5941) for conferences with no patients in the top 100 influencers ($t_{2.40}$=519.41, P=.01).

Health Care Stakeholder Composition (Information Flow)

Multiple regression analysis was used to test if health care stakeholders' composition significantly predicted the conferences information flow, a performance metric based on the number of tweets. The regression for 2014 to 2016 indicated

that the six predictors explained 21% of the variance (R^2=.21, $F_{6,1665}$=73.35, P<.001). The number of patients among the top influencers significantly predicted better performance for the conferences (beta=309, P<.001), as did the number of physicians and researchers (beta=138, P<.001), HCPs (beta=118, P<.001), journalists (beta=440, P<.001), other health care individuals (beta=130, P<.001), and Twitter accounts representing pharmaceutical organizations (beta=693, P<.001). For every increase of one patient among the top 100 influencers by mention, the conference's predicted number of tweets increased by 309.

Figure 2. Comparison of health care conferences with patients in the top 100 influencers by mention and those without.



Health Care Stakeholder Composition (Engagement in Conversation)

The results of the regression analysis for the combined years of 2014 to 2016 indicated the six predictors explained 24% of the variance (R^2=.24, $F_{6,1665}$=85.51, P<.001). The number of patients among the top influencers significantly predicted better engagement in conversations for the conferences (beta=25, P<.001), as did the number of physicians/researchers (beta=6, P<.001), HCPs (beta=5, P<.001), journalists (beta=12, P<.001), other health care individuals (beta=6, P<.001), and Twitter accounts representing pharmaceutical organizations (beta=8, P<.001).

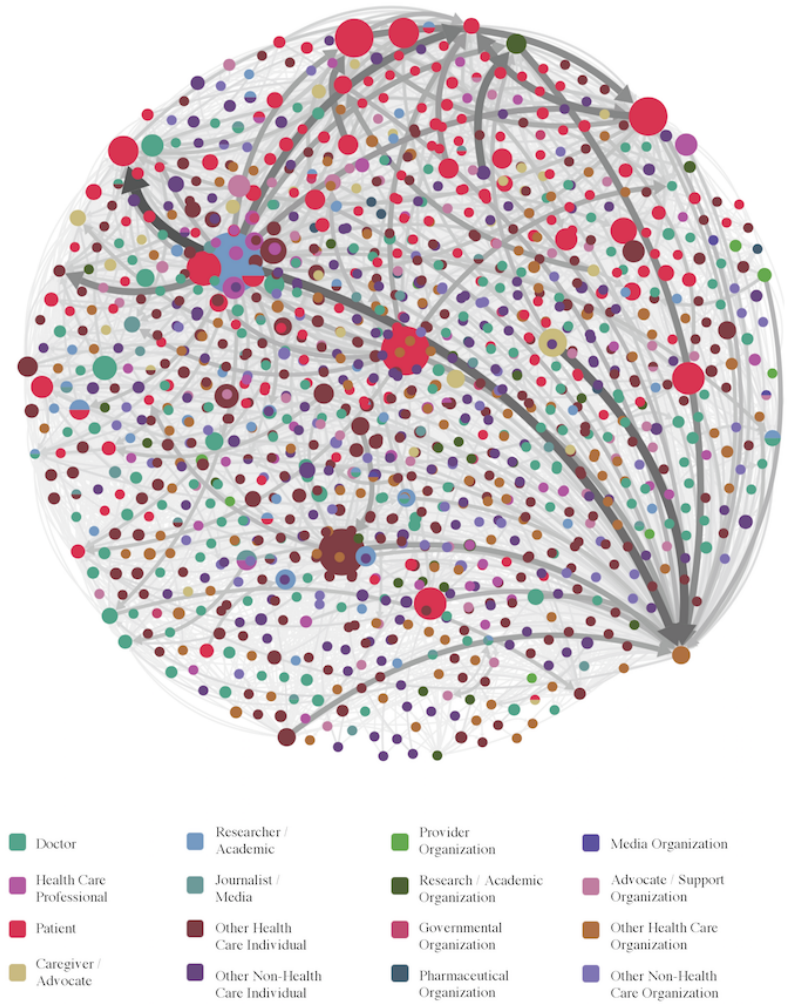
Health Care Stakeholder Composition (Information Propagation)

The results of the regression analysis for the combined years of 2014 to 2016 indicated the six predictors explained 18% of the variance (R^2=.18, $F_{6,1665}$=59.49, P<.001). The number of patients among the top influencers significantly predicted larger audience and wider potential spread of information for the conferences (beta=1,781,222, P<.001), as did the number of physicians/researchers (beta=261,253, P<.001), journalists (beta=2,669,759, P<.001), other health care individuals (beta=261,162, P<.001), and Twitter accounts representing pharmaceutical organizations (beta=2,819,703, P<.001).

Social Network Analysis (Hubs and Authorities)

Social network analysis of hubs and authorities revealed that patients were found having statistically significant higher hub scores (mean 8.26×10^{-4}, SD 2.96×10^{-4}) compared to Twitter accounts not owned by patients (mean 7.19×10^{-4}, SD 3.81×10^{-4}; $t_{273.84}$=4.302, P<.001). There were no statistically significant differences in the authority scores between patient and nonpatient Twitter accounts (Figure 3).

Figure 3. Social network analysis of the 2016 Stanford Medicine X conference based on hubs and authority score.



Discussion

During health care conferences, engaged patients use their position in social networks as an influential hub node to disseminate information to a broader community beyond the conference network. Engaged patients with firsthand experience of a chronic condition contribute their expertise by asking essential questions that lead engaging conversations as measured by increased patient activation levels [29]. The role of the engaged patient goes beyond information diffusion; they add value to the conversations by either questioning the reported information or supporting it based on their personal experiences and expertise. In this study, we demonstrated that engaged patients are more effective than physicians or researchers in three primary measures of social media performance concerning information dissemination during live health care conference coverage. Between the years of 2014 and 2016, 5692 health care conferences with specific Twitter feeds whose hashtags were registered with Symplur's Hashtag project were identified. Each consecutive year saw an increase in registered hashtags. Through analysis of health care conference hashtags, we identified individual conference stakeholders to determine the extent to which patients facilitate the propagation of information flow out of health care conferences via Twitter.

True patient engagement is based on more than just the raw number of tweets that patients contribute but rather on thoughtful tweets, quality replies, meaningful conversations, and exchange of knowledge facilitated by engaged patients and other health care conference stakeholders. Although the total number of tweets at health care conferences increased from 2015 to 2016, our results indicate that conferences that include patients have a significantly higher number of tweets than conferences that do not include patients (beta=309, P<.001). Inclusion of engaged patients in health care conferences increases the estimated number of impressions by beta=1,781,222 (P<.001); however, what is valuable about inclusion of engaged patients is their effect on increase of quality of conversation. Engaged patients increase the number of quality tweets by beta=25 (P<.001), double the impact of any other stakeholder group. Although patients are significant expanders of health care tweet propagation and accelerate information flow out of health care conferences, they make up 1.4% of the stakeholder group in the top 100 influencers of the conversation.

Social network analysis of hubs and authorities provided evidence that patients are functioning as hubs within Twitter at health care conferences to a larger degree than are nonpatients. In the context of this social conversation, authority values are large for nodes with significant incoming mentions and

conversations from large hub nodes, and hub values are large for nodes with significant outgoing mentions and conversations to high-authority nodes. Patients actively engage authorities acting as good hubs and may be seen as a social glue that in itself encourages and creates even more engagement from nonpatients as reflected in our finding for the performance metric number of tweets. By failing to include patients, health care conferences risk attenuated engagement with the public, disseminating scientific information to a narrow social network and reducing the speed of information flow across social media channels, which ultimately deters from the academic missions of health care conferences.

Our results suggest that when health care conferences include patients as conference stakeholders, patients influence the conference Twitter conversation by increasing total number of tweets and increasing spread of conference information across social networks, which will yield better social media performance outcomes for the conference. Strategies such as those outlined by the European Patients' Forum, the patient advocate-originated Cinder Blocks movement, and the Everyone Included initiative facilitate an environment in which patients are trusted, respected, and are appreciated for the expertise they bring to the conversation, openness and experimentation are normal and expected, patients have personal ownership of issues in health, individual patient stories have global impact, and patient voices and choices are incorporated into stakeholder decisions and actions [30-32].

By utilizing patient inclusion frameworks such as those just described, conference developers can build trust and respect with patient populations, create a shared mindset for change, better identify issues that matter most to patients, produce more innovative and creative solutions to health problems, and create a shared, inclusive culture of health. Patient inclusion initiatives have been implemented by the Outcome Measures in Rheumatology (OMERACT) conferences since 2002, where now more than 10% of conference participants are patients [33]. Since including patients as conference stakeholders, OMERACT has identified novel outcome measures important to patients and has incorporated the perspective of patients into the development of novel outcome measures [33-35]. Engaging patients in health care conferences may occur via other methods of multimedia delivery, such as video live streams, bringing the conference directly to patients who may not be able to attend the conference in person but still have thoughts to share.

Patients not only expand information out of health care conferences but also feed knowledge back in by sharing personal experiences and voicing issues that matter to them. This concept is again illustrated through the OMERACT conferences, which after including patients as conference stakeholders, redeveloped their research agenda and developed novel clinical trials based on outcomes that patients identified as the most relevant to their health and quality of life [34]. By including patient voices, conferences such as OMERACT and Stanford Medicine X have increased physicians' knowledge by showing them what it is like to live with a disease in which they are trained to treat [27]. By giving patients a voice at health care conferences, they act as both educators and participants, facilitating discussions and engaging providers by sharing stories and ideas, which acts to widen existing research agendas [36-38].

Some may believe that not all health care-related conferences are pertinent for patient inclusion, such as health care conferences hosted by professional medical societies that are intended to garner continuing medical education units (CMEs) or continuing education units for its members. We believe that patient inclusion in these types of conferences is pertinent, and does benefit both the specific professional medical society hosting the event, and the broader medical community at large. Furthermore, in July of 2016, the Accreditation Council for Continuing Medical Education announced new criteria for accreditation with commendation that incorporates the inclusion and engagement of patients in the planning and delivery of CMEs as planners and faculty in the accredited conference or program [39].

By incorporating patient inclusion frameworks in health care conferences and research, future studies should strive to recognize and include patients as valuable team members who bring novel expertise into the conversation. Patient inclusion frameworks, such as those previously discussed, may be used to facilitate multidisciplinary and interprofessional collaboration in health care and scientific research expanding possible outcome measures. Furthermore, by recognizing and including patients as stakeholders, health care conferences can effectively spread information through social media to new nodes reaching a broad and unique audience. The inclusion of engaged patients leads to higher tweet volume out of health care conferences and facilitates the feeding of knowledge back in via patient expertise, experience, and opinion. By including patient voices, the traditional method of scientific inquiry can be expanded, accelerated, and powered leading to novel research questions and unique patient-centric outcome measures. The simple act of including patients in health care conferences has the potential to revolutionize medical research by shifting focus toward patient-identified issues that may otherwise be overlooked by HCPs and researchers.

There are several limitations of this study. Stakeholder role was determined by self-reported information on Twitter user's personal biographies, which may not be completely accurate. For example, Twitter profiles may underreport patient status whereas others may belong to multiple stakeholder groups, such as an account belonging to an individual who is both a HCP and a patient, giving them a unique perspective. Due to the large amount of data that was collected, it was not possible to analyze tweet content or tweets that were shared or liked by others in the network. Another limitation is that our social network analysis was only based on one health care conference, the 2016 Stanford Medicine X conference, which has upwards of 10% patients as stakeholders.

Future studies should perform network analyses comparing conferences that do and do not include patients among the top 100 stakeholders by mention. Furthermore, future studies should examine *k*-core decomposition, the largest subgraph in which vertices have a minimum of *k* interconnections, of health care conference social networks among conferences that do and do

not engage patients and whether the number of patients included in a conference affect *k* scores [40-42].

Acknowledgments

This work was funded by the Stanford University School of Medicine Department of Anesthesia, Perioperative and Pain Medicine. The views expressed here do not necessarily represent the view of the Memorial Sloan Kettering Cancer Center.

Authors' Contributions

All authors participated in study design, manuscript drafting, data interpretation and final manuscript approval. Statistical analysis was conducted by AU.

Conflicts of Interest

All authors are involved with the organization of the Stanford Medicine X Conference series. AU is the co-founder of Symplur, LLC and holds financial shares in the company

References

1. Twitter. About Twitter URL: https://about.twitter.com/company [WebCite Cache ID 6qUExSGDc]
2. McKendrick DR, Cumming GP, Lee AJ. Increased use of Twitter at a medical conference: a report and a review of the educational opportunities. J Med Internet Res 2012 Dec 11;14(6):e176 [FREE Full text] [doi: 10.2196/jmir.2144] [Medline: 23232765]
3. Awad NI, Cocchio C. Use of Twitter at a major national pharmacy conference. Am J Health Syst Pharm 2015 Jan 1;72(1):65-69. [doi: 10.2146/ajhp140175] [Medline: 25511841]
4. Chung A, Woo H. Twitter in urology and other surgical specialties at global conferences. ANZ J Surg 2016 Apr;86(4):224-227. [doi: 10.1111/ans.13393] [Medline: 26631323]
5. Cochran A, Kao LS, Gusani NJ, Suliburk JW, Nwomeh BC. Use of Twitter to document the 2013 Academic Surgical Congress. J Surg Res 2014 Jul;190(1):36-40. [doi: 10.1016/j.jss.2014.02.029] [Medline: 24655663]
6. Desai T, Shariff A, Shariff A, Kats M, Fang X, Christiano C, et al. Tweeting the meeting: an in-depth analysis of Twitter activity at Kidney Week 2011. PLoS One 2012;7(7):e40253 [FREE Full text] [doi: 10.1371/journal.pone.0040253] [Medline: 22792254]
7. Hawkins CM, Duszak R, Rawson JV. Social media in radiology: early trends in Twitter microblogging at radiology's largest international meeting. J Am Coll Radiol 2014 Apr;11(4):387-390. [doi: 10.1016/j.jacr.2013.07.015] [Medline: 24139963]
8. Mishori R, Levy B, Donvan B. Twitter use at a family medicine conference: analyzing #STFM13. Fam Med 2014 Sep;46(8):608-614 [FREE Full text] [Medline: 25163039]
9. Neill A, Cronin JJ, Brannigan D, O'Sullivan R, Cadogan M. The impact of social media on a major international emergency medicine conference. Emerg Med J 2014 May;31(5):401-404. [doi: 10.1136/emermed-2012-202039] [Medline: 23423992]
10. Nomura JT, Genes N, Bollinger HR, Bollinger M, Reed JF. Twitter use during emergency medicine conferences. Am J Emerg Med 2012 Jun;30(5):819-820. [doi: 10.1016/j.ajem.2012.01.020] [Medline: 22424992]
11. Wilkinson SE, Basto MY, Perovic G, Lawrentschuk N, Murphy DG. The social media revolution is changing the conference experience: analytics and trends from eight international meetings. BJU Int 2015 May;115(5):839-846 [FREE Full text] [doi: 10.1111/bju.12910] [Medline: 25130687]
12. Nason GJ, O'Kelly F, Bouchier-Hayes D, Quinlan DM, Manecksha RP. Twitter expands the reach and engagement of a national scientific meeting: the Irish Society of Urology. Ir J Med Sci 2015 Sep;184(3):685-689. [doi: 10.1007/s11845-015-1277-6] [Medline: 25740094]
13. Winandy M, Kostkova P, de Quincey E, St Louis C, Szomszor M. Follow #eHealth2011: measuring the role and effectiveness of online and social media in increasing the outreach of a scientific conference. J Med Internet Res 2016 Jul 19;18(7):e191 [FREE Full text] [doi: 10.2196/jmir.4480] [Medline: 27436012]
14. Chu LF, Utengen A, Kadry B, Kucharski SE, Campos H, Crockett J, et al. “Nothing about us without us”-patient partnership in medical conferences. BMJ 2016 Sep 14;354:i3883 [FREE Full text] [Medline: 27628427]
15. Wicks P. The BMJ. 2015 May 01. Making sure conferences are “Patients Included” URL: http://blogs.bmj.com/bmj/2015/05/01/paul-wicks-making-sure-conferences-are-patients-included/ [accessed 2017-03-28] [WebCite Cache ID 6qUDclj2n]
16. 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]
17. Gruzd A, Haythornthwaite C. Enabling community through social media. J Med Internet Res 2013 Oct 31;15(10):e248 [FREE Full text] [doi: 10.2196/jmir.2796] [Medline: 24176835]

18. Mishori R, Singh LO, Levy B, Newport C. Mapping physician Twitter networks: describing how they work as a first step in understanding connectivity, information flow, and message diffusion. J Med Internet Res 2014 Apr 14;16(4):e107 [FREE Full text] [doi: 10.2196/jmir.3006] [Medline: 24733146]
19. Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. Nature 1998 Jun 04;393(6684):440-442. [doi: 10.1038/30918]
20. Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. Theory of Comput 2015;11(1):105-147 [FREE Full text] [doi: 10.4086/toc.2015.v011a004]
21. Kleinberg J. Authoritative sources in a hyperlinked environment. J ACM 1999;46(5):604-632. [doi: 10.1145/324133.324140]
22. Zubialde JP, Eubank D, Fink LD. Cultivating engaged patients: a lesson from adult learning. Fam Syst Health 2007;25(4):355-366. [doi: 10.1037/1091-7527.25.4.355]
23. The Healthcare Hashtag Project. 2017. URL: https://www.symplur.com/healthcare-hashtags/ [accessed 2017-03-28] [WebCite Cache ID 6qUDwbVxt]
24. Audun UM. Symplur Signals, a Health Care Social Media Analytics Platform. URL: https://www.symplur.com/signals/ [accessed 2017-05-15] [WebCite Cache ID 6qUFNZ3ga]
25. Pinho-Costa L, Yakubu K, Hoedebecke K, Laranjo L, Reichel CP, Colon-Gonzalez MD, et al. Healthcare hashtag index development: Identifying global impact in social media. J Biomed Inform 2016 Oct;63:390-399. [doi: 10.1016/j.jbi.2016.09.010] [Medline: 27645323]
26. Core Team. R-A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. URL: http://www.r-project.org/ [accessed 2017-03-28] [WebCite Cache ID 6qUFUUAVw]
27. Stanford Medicine X. URL: https://medicinex.stanford.edu/ [accessed 2017-04-04] [WebCite Cache ID 6qUDoIS1x]
28. Tanase R, Radu R. Cornell University Department of Mathematics. 2009. Lecture #4: HITS algorithm - hubs and authorities on the Internet URL: http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture4/lecture4.html [accessed 2017-05-15] [WebCite Cache ID 6qUE883r1]
29. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214 [FREE Full text] [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
30. European Patients' Forum. URL: http://www.eu-patient.eu/ [accessed 2017-04-04] [WebCite Cache ID 6qUEDPfce]
31. Holliday R. Regina Holliday's Medical Advocacy Blog. 2013. Best practices every medical conference should embrace URL: http://reginaholliday.blogspot.com/2013/11/best-practices-every-medical-conference.html [accessed 2017-04-04] [WebCite Cache ID 6qUEIwcnO]
32. Stanford Medicine Everyone Included. Everyone Included in health care URL: http://www.everyoneincluded.org/ [accessed 2017-04-04] [WebCite Cache ID 6qUEe0j5z]
33. de Wit M, Abma T, Koelewijn-Van LM, Collins S, Kirwan J. Facilitating and inhibiting factors for long-term involvement of patients at outcome conferences--lessons learnt from a decade of collaboration in OMERACT: a qualitative study. BMJ Open 2013 Aug 23;3(8):e003311 [FREE Full text] [doi: 10.1136/bmjopen-2013-003311] [Medline: 23975104]
34. de Wit M, Abma T, Koelewijn-van LM, Collins S, Kirwan J. Involving patient research partners has a significant impact on outcomes research: a responsive evaluation of the international OMERACT conferences. BMJ Open 2013 May 09;3(5):1-12 [FREE Full text] [doi: 10.1136/bmjopen-2012-002241] [Medline: 23667160]
35. de Wit MP, Abma TA, Koelewijn-van LM, Collins S, Kirwan J. What has been the effect on trial outcome assessments of a decade of patient participation in OMERACT? J Rheumatol 2014 Jan;41(1):177-184. [doi: 10.3899/jrheum.130816] [Medline: 24128777]
36. Richards T. The BMJ. 2015. Is your conference “Patients Included?” URL: http://blogs.bmj.com/bmj/2015/04/17/tessa-richards-is-your-conference-patients-included/ [accessed 2017-07-14] [WebCite Cache ID 6rxECi01R]
37. Richards T, Coulter A, Wicks P. Time to deliver patient centred care. BMJ 2015 Feb 10;350(feb10 14):h530-h530. [doi: 10.1136/bmj.h530] [Medline: 25670197]
38. Richards T, Montori VM, Godlee F, Lapsley P, Paul D. Let the patient revolution begin. BMJ 2013 May 14;346:f2614. [doi: 10.1136/bmj.f2614] [Medline: 23674136]
39. McMahon G, Reus R, Dorman T. American Council of Continuing Medical Education. 2016. Menu of new criteria for accreditation with commendation URL: http://www.accme.org/sites/default/files/731_20160929_Menu_of_New_Criteria_for_Accreditation_with_Commendation_3.pdf [accessed 2017-07-14] [WebCite Cache ID 6rxEzOXM7]
40. Pei S, Muchnik L, Andrade JS, Zheng Z, Makse HA. Searching for superspreaders of information in real-world social media. Sci Rep 2014 Jul 03;4:5547 [FREE Full text] [doi: 10.1038/srep05547] [Medline: 24989148]
41. Zhang J, Chen D, Dong Q, Zhao Z. Identifying a set of influential spreaders in complex networks. Sci Rep 2016 Jun 14;6:27823 [FREE Full text] [doi: 10.1038/srep27823] [Medline: 27296252]
42. Malliaros FD, Rossi MG, Vazirgiannis M. Locating influential nodes in complex networks. Sci Rep 2016 Jan 18;6:19307 [FREE Full text] [doi: 10.1038/srep19307] [Medline: 26776455]

Abbreviations

CME: continuing medical education
HCP: health care professional
HITS: hyperlink-induced topic search
OMERACT: Outcome Measures in Rheumatology

Edited by G Eysenbach; submitted 15.05.17; peer-reviewed by D Attai, C Bush; comments to author 13.06.17; revised version received 26.06.17; accepted 29.06.17; published 17.08.17.

Please cite as:
Utengen A, Rouholiman D, Gamble JG, Grajales III FJ, Pradhan N, Staley AC, Bernstein L, Young SD, Clauson KA, Chu LF
Patient Participation at Health Care Conferences: Engaged Patients Increase Information Flow, Expand Propagation, and Deepen Engagement in the Conversation of Tweets Compared to Physicians or Researchers
J Med Internet Res 2017;19(8):e280
URL: http://www.jmir.org/2017/8/e280/
doi:10.2196/jmir.8049
PMID:28818821

Original Paper

Validation Relaxation: A Quality Assurance Strategy for Electronic Data Collection

Avi Kenny[1], BA; Nicholas Gordon[1], BS; Thomas Griffiths[1]; John D Kraemer[2], JD, MPH; Mark J Siedner[3], MPH, MD

[1]Last Mile Health, Boston, MA, United States

[2]Georgetown University, Washington, DC, United States

[3]Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Avi Kenny, BA
Last Mile Health
205 Portland St #200
Boston, MA, 02114
United States
Phone: 1 9143163681
Email: akenny@lastmilehealth.org

Abstract

Background: The use of mobile devices for data collection in developing world settings is becoming increasingly common and may offer advantages in data collection quality and efficiency relative to paper-based methods. However, mobile data collection systems can hamper many standard quality assurance techniques due to the lack of a hardcopy backup of data. Consequently, mobile health data collection platforms have the potential to generate datasets that appear valid, but are susceptible to unidentified database design flaws, areas of miscomprehension by enumerators, and data recording errors.

Objective: We describe the design and evaluation of a strategy for estimating data error rates and assessing enumerator performance during electronic data collection, which we term "validation relaxation." Validation relaxation involves the intentional omission of data validation features for select questions to allow for data recording errors to be committed, detected, and monitored.

Methods: We analyzed data collected during a cluster sample population survey in rural Liberia using an electronic data collection system (Open Data Kit). We first developed a classification scheme for types of detectable errors and validation alterations required to detect them. We then implemented the following validation relaxation techniques to enable data error conduct and detection: intentional redundancy, removal of "required" constraint, and illogical response combinations. This allowed for up to 11 identifiable errors to be made per survey. The error rate was defined as the total number of errors committed divided by the number of potential errors. We summarized crude error rates and estimated changes in error rates over time for both individuals and the entire program using logistic regression.

Results: The aggregate error rate was 1.60% (125/7817). Error rates did not differ significantly between enumerators (P=.51), but decreased for the cohort with increasing days of application use, from 2.3% at survey start (95% CI 1.8%-2.8%) to 0.6% at day 45 (95% CI 0.3%-0.9%; OR=0.969; P<.001). The highest error rate (84/618, 13.6%) occurred for an intentional redundancy question for a birthdate field, which was repeated in separate sections of the survey. We found low error rates (0.0% to 3.1%) for all other possible errors.

Conclusions: A strategy of removing validation rules on electronic data capture platforms can be used to create a set of detectable data errors, which can subsequently be used to assess group and individual enumerator error rates, their trends over time, and categories of data collection that require further training or additional quality control measures. This strategy may be particularly useful for identifying individual enumerators or systematic data errors that are responsive to enumerator training and is best applied to questions for which errors cannot be prevented through training or software design alone. Validation relaxation should be considered as a component of a holistic data quality assurance strategy.

(J Med Internet Res 2017;19(8):e297) doi:10.2196/jmir.7813

KEYWORDS

data accuracy; data collection; surveys; survey methodology; research methodology; questionnaire design; mHealth; eHealth

Introduction

A cornerstone of research conduct is the assurance of high-quality data collection. Data quality has been defined as "data that are fit for use by data consumer" [1]. Agmon and Ahituv [2] refer to data quality in terms of "reliability," distinguishing between internal reliability (reliability whose assessment is based on commonly accepted criteria about the characteristics of the data items), relative reliability (reliability of the data in view of the user requirements), and absolute reliability (comparisons between the dataset and reality). Wand and Wang [3] take an ontological approach to identify 4 generic observable data quality issues—loss of information, insufficient (ambiguous) information, meaningless data, and incorrect data. If evidence is generated from underlying data that are of poor quality, incorrect conclusions may be drawn [4,5], leading to both direct and hidden costs [6,7].

The use of mobile phones and tablets for data collection may yield improvements over paper-based methods across a number of data quality dimensions and has been increasingly used in low-income settings [8-14]. Potential advantages of electronic methods over paper-based methods include lower error rates [10,13], reduced likelihood of data loss [8], higher data completeness [9,10,13], reduced time needed for data collection [9,10,13,15], automatic collection of timestamps and geolocation data, and in some cases decreased costs [9,13,16]. Additionally, electronic data collection has been shown to be feasible among users with little to no prior experience with data collection or cell phone use in a number of different settings, provided that they are given some basic training [8,9,12], and has been largely seen as acceptable by managers, users, and data collection subjects [9,12,13,16,17]. Thus, it represents an attractive option for researchers, nongovernmental organizations, governments, and others.

Claims of reduced error rates with mobile data platforms over paper alternatives can be logically attributed to several factors. Programmed skip logic (also called "branching") allows for a question or data element to be displayed or not displayed depending on the user's entry for 1 or more previous data elements, allowing for complex conditional pathways to be automated. This ensures that the proper sequence of questions or data elements are answered, ameliorating the problem of missing data. Real-time validation, notably the use of field constraints, is a restriction of the range or type of possible entries for a data element, limiting entries based on logical rules or previously entered data. This is widely viewed as a strong and appropriate tactic for reducing errors [18] in survey work, as it prevents the entry of logically invalid data. Furthermore, with electronic data collection, there is no manual data entry of paper forms needed, and thus the layer of errors associated with the manual data entry of paper data [19] is completely eliminated.

It has been recognized that data loss is still possible [12] and reductions in data quality have not been seen universally [15]. However, a challenge specific to electronic data collection that has not been explicitly addressed in the data quality literature is "masking" of data recording errors. Masking occurs when an end-user intentionally or unintentionally enters incorrect data that is forced or allowed by the data validation constraints. For example, an insufficiently trained user of a data collection application *without* hard-coded validation rules is likely to enter data that is illogical or internally inconsistent. However, if validation rules *are* applied, the data entered by such a user might still be susceptible to errors, but it will conform to the validation constraints, and thus such errors would not be detectable in the resulting dataset. When such errors could be mitigated by identification, supervision, and retraining, enabling errors to be rapidly identified and addressed is valuable.

In terms of Agmon and Ahituv's dimensions of reliability, electronic data collection has great potential to increase internal reliability, as data constraints can be enforced; however, given the issue of masking, this will not always translate to increased absolute reliability. Similarly, in terms of Wand and Wang's observable data quality issues, the problems of loss of information and meaningless data will be mitigated or eliminated, but this will only partially address the problem of incorrect data. As such, there is an important need to consider alternative methods of data quality oversight for mobile health data collection platforms.

In this paper, we articulate a strategy for assessing the data quality of electronic data collection initiatives by identifying incorrect data, thereby allowing for judgments on absolute reliability. This strategy, which we term *validation relaxation*, involves the intentional omission of validation features for a selection of data elements on which validation would typically be applied in order to allow for the possibility of detectable human errors, along with the creation of a mechanism for monitoring error rates and their trends in real time. Benefits of this approach include identification of instrument comprehension issues, detection of survey or database design errors, and targeted quality improvement efforts for individuals or teams with the highest error rates. We illustrate and evaluate an application of this strategy by describing its use within a cluster sample population survey in rural Liberia.

Methods

Development of the Validation Relaxation Strategy

The validation relaxation strategy was conceived to augment quality assurance of digital survey data collection operations at Last Mile Health, a nongovernmental health care organization operating in rural Liberia. The prior quality assurance approach contained 3 primary components: First, thorough training for survey enumerators, including observed survey practice with frequent instructor feedback and a field-based pilot test. Second, direct observation of a sample of surveys during the data collection period by a field supervisor, along with daily debriefings of field teams to review commonly committed errors. Third, the use of real-time validation and automated skip logic to prevent missing data and avoid illogical or impossible responses.

This approach was based on the Total Data Quality Management methodology, which emphasizes quality checks at multiple time points throughout the data life cycle [20]. The first quality component is a ubiquitous best practice in survey research [21].

The second is straightforward and has been employed in a variety of settings [22]. The third is seen as a major advantage of electronic data collection and has been leveraged extensively, often through the native capabilities of common data collection software packages [23-25] and sometimes through complex software customization that allows for the enforcement of idiosyncratic workflows [9]. However, as mentioned above, this third component can also mask underlying errors and lead to the production data that deceivingly appears clean.

To account for this issue, we created the "validation relaxation" strategy to detect intentional or accidental misuse of electronic data collection applications and avoid collecting poor-quality data. Specifically, we identified select scenarios in which human error can cause data to be collected that is logically valid but factually incorrect with electronic data collection. For example, if enumerators do not comprehend or administer an application correctly, they may intentionally falsify data to conform to data validation structures, an issue that has been previously considered in the context of survey-based research [26]. The validation relaxation strategy was intended to identify such instances by selectively removing form validation to allow for the possibility of unconstrained data entry, therefore making potential misunderstanding or misuse of the application quantitatively detectable, and subsequently monitoring error types and rates. Since only a sample of questions have validation rules removed, the overall *detectable* error rate for a given user may be thought of as a proxy measurement for the overall *undetectable* error rate, although the extent to which these rates correlate within a given set of users may vary between applications. Subsequently, focusing supervision and coaching efforts on the enumerators with the highest error rates may lead to decreases in overall error rates over time. Additionally, if the same survey instrument is used more than once (eg, in a repeated survey series), aggregate error rates can be used as an indicator of overall data quality differences between surveys.

To implement this strategy, we first created the data collection questionnaire and planned a set of validation rules including skip logic and field constraints to be applied, such that logically invalid responses and response patterns were prohibited by the application. We subsequently chose a purposive selection of 11 questions, out of a total of 122 survey questions, for which we removed (or "relaxed") validation rules; this resulted in 11 different possible errors per survey. Questions were selected based on several factors; we were more likely to select questions for which we suspected or found data quality issues in the past (eg, dates), as well as questions that were relatively less important in the context of our ongoing research (to avoid compromising critical data during this evaluation). We also searched opportunistically for questions or sets of questions that allow for a logical rule to be easily validated (eg, the question "Have you ever given birth?" was already asked twice in the questionnaire to facilitate skip logic flow).

Table 1. Classification of detectable errors.

#	Class	Description	Example of error detected
1	Removal of "required" constraint	Removal of a "required question" constraint	User accidentally skips a question on postnatal care that he or she was supposed to complete
2	Illogical response combinations: multiple questions	Inclusion of 2 or more questions for which a certain combination of answers is logically impossible	The first question is "What is your gender?"; user answers "male." The second question is "Have you ever given birth"; user answers "yes."
3	Illogical response combinations: single question	Inclusion of an individual, multiple-response, multiple-choice question for which certain combinations of responses is logically impossible	The question is "Who checked on you during your last pregnancy?" User selects 2 options: "family members" and "I don't know."
4	Intentional redundancy	Repetition of the same question (possibly with slightly different wording or within a different question sequence) more than once in different sections of the questionnaire	At the start of the survey, user answers the question "How many times have you given birth?" with "6." Later in the survey, the user answers a repeated instance of the same question ("How many times have you given birth?") with "5."
5	Manual skip logic	Forcing the user to select the next branch of questions to ask, based on responses to previous questions (instead of automating skip logic)	User answers the question "Have you ever been to a health clinic?" with a "No". User is then prompted with 2 possible options and has to choose one: "Complete clinical questionnaire" or "Skip clinical questionnaire and proceed to child health questionnaire." User selects "Complete clinical questionnaire."
6	Removing minimum or maximum constraints	Removing constraints on the minimum or maximum value that can be entered for a question	User answers "657" to the question "How old are you, in years?"
7	Manual calculation	Prompt the user to enter a value that could be mathematically calculated from previous responses	Survey date is "June 3, 2016." User answers the question "What is your birthday?" with "June 4, 1996." The next question is "What is your age, in years?"; respondent answers "24."
8	Allowing invalid data type	User is allowed to enter a value of an incorrect data type	The question is "How many times have you seen a doctor in the past month?" User answers "sometimes."

We built and thoroughly tested the application, first in the office using a simulated dataset, and then through a field-based pilot test conducted in conditions that approximated the actual conditions in which the application was to be deployed. We created a reporting system to enable active monitoring of errors, disaggregated by the survey date and the enumerator's ID number, which took the form of an automated report within a custom-built Web application written in the PHP (PHP: Hypertext Preprocessor) programming language.

After the implementation of the survey, we created a classification scheme of detectable errors to help facilitate the future selection of questions on which to relax validation. Detectable errors can be categorized based on the types of data elements under examination and the nature of the error that is permitted. This classification is detailed in Table 1.

Data Collection

We assessed the validation relaxation strategy during the implementation of a 2-stage, cross-sectional, cluster sample survey in Rivercess County, Liberia. This was the second survey in a repeated cross-sectional study. Full description of the methods and results from the baseline survey has been described elsewhere [27]. The purpose of the survey was to assess a number of indicators of demographics, maternal health, neonatal health, and child health, as part of ongoing research and evaluation activities of Last Mile Health. The questionnaire was composed of questions adapted from the 2013 Liberia Demographic and Health Survey. Survey data were collected weekly from enumerators in the field by a supervisor and transferred to a secure, cloud-hosted MySQL database.

A total of 7 enumerators were hired to conduct the survey; each received a 5-day training covering the use of the data collection hardware and software, the purpose and meaning of each survey question, field translation in Bassa (the local dialect), and methods to reduce biases. An enumerator served as an alternate and only surveyed 10 women; data from this enumerator were excluded from this analysis.

The platform used was a modified version of Open Data Kit (ODK), an open-source set of tools designed to allow implementers to create information systems in the developing world [23]. Modifications to ODK allowed for data to be transferred wirelessly from one Bluetooth device to another, which was advantageous for prevention of data loss, given that Liberia's poor cellular network coverage meant that users would be out of coverage for many consecutive days. Our modified ODK application was installed on 10 BLU Advance 4.0 Android phones, which were distributed to enumerators and field supervisors. Data collected on the Android devices were stored in XML format, transferred periodically from enumerator phones to supervisor phones via Bluetooth, and ultimately transferred via Bluetooth to a central laptop, where records were uploaded to a custom-built Web application. This application parses the data into JSON (JavaScript Object Notation) format, checks for file integrity, adds several metadata attributes, and sends the resulting dataset into a MySQL database cloud-hosted on a virtual private server.

Enumerators were not informed of the validation relaxation strategy. During the implementation of the survey, we ran the automated error report on a weekly basis, which was used to identify enumerators who were underperforming, as evidenced by high error rates relative to the other enumerators. Each week, the lead field supervisor of the survey examined error rates and focused monitoring and coaching efforts on underperforming enumerators.

Data Analysis

Error rates were summarized using basic descriptive statistics. We then used logistic regression to estimate the association between time (survey day) and the odds of committing an error. No covariates were included in the model. Variance estimation was corrected for the effects of clustering using the clustered sandwich estimator. Next, we collapsed the dataset such that 1 observation represented a single survey day and estimated the daily standard deviation of error rates, and used linear regression to estimate the association between time (survey day) and the standard deviation of error rates between enumerators. Statistical analyses were conducted using Stata Version 14.1 (Statacorp).

Ethics

Ethics approval for the survey was obtained from the institutional review boards of Partners Healthcare, Georgetown University, and the Liberian Institute for Biomedical Research. All respondents gave verbal informed consent.

Results

The survey was conducted between April 12, 2016 and June 7, 2016, and included a sample of 972 women across 1150 households within 86 different communities. Table 2 details the specific errors that were possible within our survey, along with error rates for each. For the calculation of rates, the denominator is equal to the number of times that the requisite question(s) were reached within the application by the enumerator. In other words, the rate is equal to the number of errors divided by the number of opportunities for the error to be made.

The overall error rate was 1.60% (125/7817). This is comparable to error rates in similar settings [28,29]. The most commonly made error was an "intentional redundancy" question in which the respondent was asked twice for the date of birth of her most recently birthed child, with an error rate of 13.6% (84/618). Data for this question were entered through an ODK "date widget" [30], where the enumerator scrolls through month, days, and years to select the correct date. An examination of incorrect dates suggests that the high rate for this particular error may have partially been due to the enumerator accidentally scrolling 1 or 2 ticks past the correct day, month, or year; 30% (25/84) of these errors were 1 tick off and an additional 17% (14/84) were 2 ticks off. However, it is also possible that the respondent's recall is inexact. Four other possible errors had rates between 0.2% and 3.1%, with all other possible errors having rates equal to zero. There was a strong association between the class of the error and the rate at which the error was committed. Specifically, the 4 intentional redundancy errors had the 4 highest error rates.

Table 2. Specific detectable errors implemented in cluster sample survey.

#	Class	Error definition	Number of errors	Error rate, %
1	Intentional redundancy	Gave different answers for the question ("Was your most recent birth in a health facility?") in different sections of the questionnaire	19/618	3.1
2	Intentional redundancy	Gave different answers for the question ("Have you ever given birth?") in different sections of the questionnaire	10/961	1.0
3	Intentional redundancy	Gave different answers for the question ("What was the date of birth of your most recently birthed child?") in different sections of the questionnaire	84/618	13.6
4	Intentional redundancy	Gave different answers for the question ("Is your most recently birthed child still alive?") in different sections of the questionnaire	10/618	1.6
5	Illogical response combinations: single question	Question is "Where you go to get medical advice or treatment?"; answer options included ("refused to respond" OR "unknown") AND ("clinic" OR "drugstore" OR "community health worker" OR "traditional healer" OR "other")	2/895	0.2
6	Illogical response combinations: single question	Question is "What are the signs of someone who can have ebola?"; answer options included ("refused to respond" OR "unknown") AND ("fever" OR "muscle pains" OR "vomiting" OR "sore throat" OR "diarrhea" OR "bleeding" OR "other")	0/895	0.0
7	Removal of "required" constraint	A required question ("Can people get Ebola from touching an Ebola patient?") was skipped	0/895	0.0
8	Removal of "required" constraint	A required question ("Can people get Ebola from the air?") was skipped	0/895	0.0
9	Removal of "required" constraint	A required question ("Can people get Ebola by touching or washing a dead body?") was skipped	0/895	0.0
10	Illogical response combinations: multiple questions	Answers for a multiple-response question ("From whom did the child get treatment [for fever or cough]?") were given; an answer was given to the following question ("From whom did the child get treatment FIRST?") that was not selected in the previous list of responses	0/325	0.0
11	Illogical response combinations: multiple questions	Answers for a multiple-response question ("From whom did the child get treatment [for diarrhea]?") were given; an answer was given to the following question ("From whom did the child get treatment FIRST?") that was not selected in the previous list of responses	0/202	0.0
		Total	125/7817	1.60

Roughly twice per week during the survey implementation, the lead field supervisor reviewed an error report that summarized errors committed so far, disaggregated by the survey date and the enumerator ID number. During the survey data collection period, this report was accessed on 18 different days by the supervisor (with a roughly uniform distribution), based on database usage tracking statistics. The supervisor would then communicate with the 2 other field supervisors, and give the names of the enumerators with high error rates, along with information on which errors were being commonly made. Total enumerator-specific error rates are summarized in Table 3.

Differences between enumerators were not statistically significant for any of the time periods evaluated. An analysis of variance of the data in each of the 4 columns in Table 3 gives the *P* values given in the bottom row of the table.

Table 3. Enumerator-specific error rates.

Enumerator ID#	Error rate, %			Overall error rate, %
	(day 0-14)	(day 15-29)	(day 30-45)	(day 0-45)
2	3.1 (14/458)	0.4 (2/465)	1.2 (5/415)	1.57 (21/1338)
3	2.9 (13/452)	0.8 (3/382)	1.8 (6/334)	1.88 (22/1168)
4	2.8 (17/605)	1.8 (9/506)	1.3 (5/393)	2.06 (31/1504)
5	1.8 (7/386)	1.6 (6/364)	1.0 (3/286)	1.54 (16/1036)
6	2.5 (14/552)	0.9 (5/528)	0.2 (1/436)	1.32 (20/1516)
7	1.4 (7/512)	1.6 (6/380)	0.6 (2/363)	1.20 (15/1255)
	P=.45	*P*=.45	*P*=.42	*P*=.51

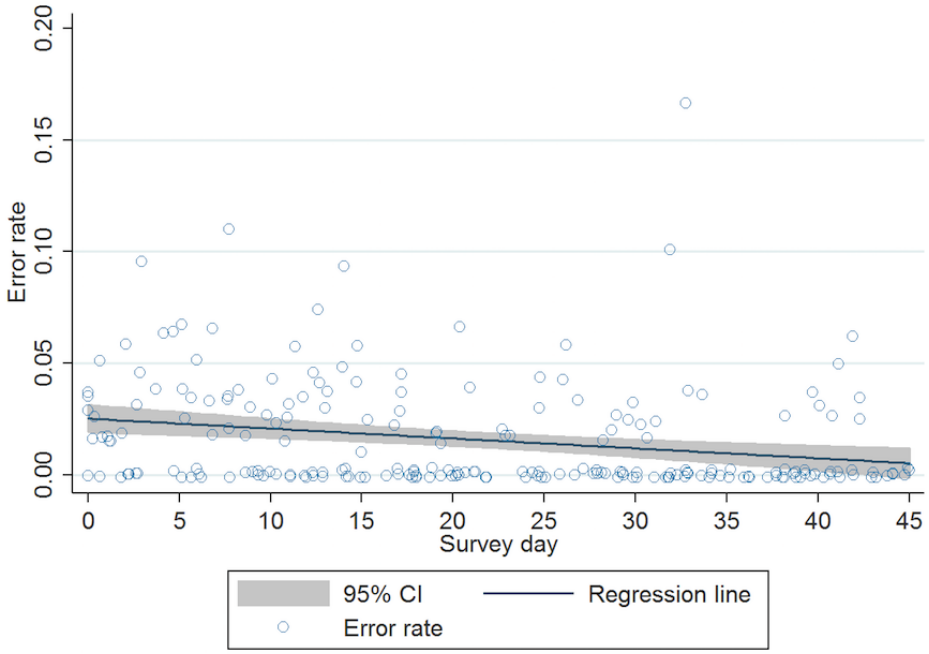
Data quality improved over time. A logistic regression of time on the error variable (a binary variable representing whether an error was committed) was significant (P<.001), with an odds ratio of 0.969 (95% CI 0.955-0.983), representing the change in odds given a 1-day change in time. Thus, the predicted total change in average error rate from the start of the survey (day=0) to the end (day=45) is −1.7%, representing a fourfold decrease in error rate, from 2.3% (95% CI 1.8%-2.8%) to 0.6% (95% CI 0.3%-0.9%) over the observation period. Data are summarized in Table 4.

Data for sensitivity analysis #4 (similar to primary analysis #1, except leveraging aggregated data) are visualized in Figure 1.

Table 4. Change in error rates over time (primary and sensitivity analyses).

Analysis	Type	Number of observations	Odds ratio (OR) or coefficient (beta) (95% CI)	*P* value	Predicted error rate at day=0 (95% CI)	Predicted error rate at day=45 (95% CI)
Primary (#1); all errors included	Logistic regression	9527	OR=0.969 (0.955 to 0.983)	<.001	0.0230 (0.0179 to 0.0281)	0.0056 (0.0027 to 0.0085)
Sensitivity (#2); excludes most common error	Logistic regression	8566	OR = 0.985 (0.964 to 1.007)	.18	0.0064 (0.0041- to 0.0086)	0.0032 (0.0010 to 0.0055)
Sensitivity (#3); includes only 3 most common errors	Logistic regression	2883	OR = 0.965 (0.949 to 0.982)	<.001	0.0710 (0.0512 to 0.0908)	0.0153 (0.0059 to 0.0248)
Sensitivity (#4); includes only 5 most common errors; aggregated data	Logistic regression	4739	OR = 0.968 (0.954 to 0.982)	<.001	0.0461 (0.0356 to 0.0567)	0.0112 (0.0055 to 0.0168)
Sensitivity (#5); all errors included; aggregated data	Linear regression	218	beta = −.000444 (−.000607 to −.000280)	<.001	0.0252 (0.0189 to 0.0315)	0.0052 (−0.0007 to 0.0111)
Sensitivity (#6); excludes most common error; aggregated data	Linear regression	218	beta = −.000051 (−.000171 to −.000070)	.33	0.0069 (0.0048 to 0.0091)	0.0047 (−0.0008 to 0.0101)
Sensitivity (#7); includes only 3 most common errors; aggregated data	Linear regression	218	beta = −.002235 (−.003353 to −.001118)	.004	0.1094 (0.0723 to 0.1465)	0.0088 (−0.0186 to 0.0361)
Sensitivity (#8); includes only 5 most common errors	Linear regression	218	beta = −.000903 (−.001256 to −.000549)	.001	0.0530 (0.0399 to 0.0662)	0.0124 (−0.0032 to 0.0281)

Figure 1. Daily enumerator-specific error rates over time, with fitted regression line (jittered for clarity).



Discussion

Principal Findings

We describe the development and evaluation of *validation relaxation*, a novel strategy that involves the intentional omission of electronic data collection validation features for a selection of data elements to allow for the possibility of detectable human errors, which enables data error rate monitoring and identification of database design and survey comprehension issues. We evaluated this strategy in the field during a population survey in rural Liberia, and found that date question formats were the most problematic, and that error rates were largely consistent between enumerators, and that error rates decreased significantly over time.

This strategy enabled us to learn what types of errors were most commonly occurring and implement training measures to ensure optimal use and comprehension of the data collection platform and survey instrument, respectively. The overall error rate was low at 1.60%, and although error rates did not differ significantly between enumerators, they varied considerably between error types. The highest error rates were found for the "intentional redundancy" errors. There are several possible reasons for this trend. First, 3 of the 4 intentional redundancy questions were grouped in one of the most complicated survey sections in terms of the underlying skip logic. Second, there may have been higher error on the part of the respondents, as they were asked about events that often occurred many years ago. Third, the highest error rate was detected for a date question, and as discussed, the date selector widget was prone to accidental error if the user scrolled too far, resulting in a higher probability that the incorrect value was entered.

Applications

We assessed the validation relaxation strategy during a survey in a low-income setting, but the strategy may also have value across other data collection scenarios including research studies, electronic medical record systems, and mHealth/eHealth initiatives in both developing and high-income settings. It should be considered in addition to other emerging electronic data quality improvement techniques, such as automatic filling of forms [31,32] and dynamic reordering and repeating questions [33], as an additional method to optimize data quality for electronic data collection. Similarly, although we employed validation relaxation to compare error rates between multiple users, it can also be a useful means of assessing trends in data quality. It can also be potentially useful in comparing enumerator or field teams who are individually and simultaneously implementing a data collection instrument. Automated or semiautomated feedback loops can be employed with this strategy to enable real-time detection of errors, which can be used to intervene on faulty survey instruments or to improve enumerator data collection quality [8,34].

Validation relaxation might also allow data managers to detect fraud in data collection applications. Existing approaches to fraud detection focus on conducting repeat interviews for a sample of respondents [35], identifying "at-risk" enumerators [36], examining digit preference (Benford's Law) [37,38], analyzing the statistical qualities of specific variables within the dataset [37,39], leveraging machine learning algorithms to detect anomalies in response distribution [40], and searching for patterns in survey timestamps [8]. The inclusion of intentionally redundant questions, preferably spaced apart within a questionnaire, could lead to patterns of inconsistent response for a single user, which would signal a possible case of falsification.

Finally, although its initial intent was to identify end-user data entry errors, validation relaxation might also help detect errors in application/database design. Often, designers will make assumptions about the potential set of logical response options (eg, an enumerator trying to enter the value "14" on a question that asks for the age of a pregnant woman, where the input range is restricted to 15-49 years). By relaxing validation rules, designers can remove such assumptions regarding valid data ranges, empirically test whether the actual range of collected values falls within the expected range, and subsequently investigate records where values fall outside the expected range.

Limitations

This work was limited to quantitative assessment of the strategy. Future work should include qualitative input from database designers and end-users to further explore the nature of committed errors and enumerator perceptions of the strategy. More data are also needed to better specify the large-scale feasibility and cost of this strategy if applied to large health programs. Moreover, our hypothesis that the *detectable* error rate is a good proxy measurement for the *undetectable* error rate is an assumption that warrants further investigation. Our list of error types and validation domains were self-selected based on our own experience and hypothesis and future iterations of this technique can and should expand upon these to target a more thorough and case-specific error types and validation schemes. Other possible alterations to the strategy to consider for future use include prespecification of a maximum acceptable error rate, use of control charts [41], and use of a formal statistical test to determine whether or not error rates between enumerators or surveys significantly differ.

Conclusions

The validation relaxation strategy can help detect comprehension and platform usability issues for electronic data collection applications, detect end-user and program error rates, and elucidate trends in error rates over time or between user groups. The strategy should be implemented as one component of a holistic data quality approach in the increasingly widespread use of electronic data collection platforms.

Acknowledgments

UBS Optimus Foundation provided programmatic evaluation funds to Last Mile Health for the survey on which this analysis was conducted (no grant identification numbers). The funders had no role in the study design, data collection and analysis, decision

to publish, or preparation of the manuscript. MJS receives research and salary support from the National Institutes of Health (K23 MH099916). JDK receives partial salary support from LMH.

Conflicts of Interest

None declared.

References

1. Wang R, Strong D. Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 1996;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
2. Agmon N, Ahituv N. Assessing data reliability in an information system. J Manag Inf Syst 1987;4(2):34-44. [doi: 10.1080/07421222.1987.11517792]
3. Wand Y, Wang RY. Anchoring data quality dimensions in ontological foundations. Commun ACM 1996;39(11):86-95 [FREE Full text] [doi: 10.1145/240455.240479]
4. Levitt SH, Aeppli DM, Potish RA, Lee CK, Nierengarten ME. Influences on inferences. Effect of errors in data on statistical evaluation. Cancer 1993 Oct 01;72(7):2075-2082 [FREE Full text] [Medline: 8374866]
5. Barchard KA, Pace LA. Preventing human error: the impact of data entry methods on data accuracy and statistical results. Comput Human Behav 2011 Sep;27(5):1834-1839. [doi: 10.1016/j.chb.2011.04.004]
6. Haug A, Zachariassen F, Liempd D V. The costs of poor data quality. J Ind Eng Manag 2011;4(2):168. [doi: 10.3926/jiem.2011.v4n2.p168-193]
7. Eppler M, Helfert M. A classification and analysis of data quality costs. 2004 Presented at: Proceedings of the Ninth International Conference on Information Quality (ICIQ-04); 2004; Cambridge, MA p. 311-325.
8. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decis Mak 2009;9:51 [FREE Full text] [doi: 10.1186/1472-6947-9-51] [Medline: 20030813]
9. Shirima K, Mukasa O, Schellenberg JA, Manzi F, John D, Mushi A, et al. The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania. Emerg Themes Epidemiol 2007;4:5 [FREE Full text] [doi: 10.1186/1742-7622-4-5] [Medline: 17543099]
10. Bernabe-Ortiz A, Curioso WH, Gonzales MA, Evangelista W, Castagnetto JM, Carcamo CP, et al. Handheld computers for self-administered sensitive data collection: a comparative study in Peru. BMC Med Inform Decis Mak 2008 Mar 19;8:11 [FREE Full text] [doi: 10.1186/1472-6947-8-11] [Medline: 18366687]
11. Munro ML, Lori JR, Boyd CJ, Andreatta P. Knowledge and skill retention of a mobile phone data collection protocol in rural Liberia. J Midwifery Womens Health 2014;59(2):176-183 [FREE Full text] [doi: 10.1111/jmwh.12155] [Medline: 24655593]
12. Medhanyie AA, Moser A, Spigt M, Yebyo H, Little A, Dinant G, et al. Mobile health data collection at primary health care in Ethiopia: a feasible challenge. J Clin Epidemiol 2015 Jan;68(1):80-86. [doi: 10.1016/j.jclinepi.2014.09.006] [Medline: 25441699]
13. Thriemer K, Ley B, Ame SM, Puri MK, Hashim R, Chang NY, et al. Replacing paper data collection forms with electronic data entry in the field: findings from a study of community-acquired bloodstream infections in Pemba, Zanzibar. BMC Res Notes 2012 Feb 21;5:113 [FREE Full text] [doi: 10.1186/1756-0500-5-113] [Medline: 22353420]
14. Reitmaier P, Dupret A, Cutting WA. Better health data with a portable microcomputer at the periphery: an anthropometric survey in Cape Verde. Bull World Health Organ 1987;65(5):651-657 [FREE Full text] [Medline: 3322601]
15. Avilés W, Ortega O, Kuan G, Coloma J, Harris E. Quantitative assessment of the benefits of specific information technologies applied to clinical studies in developing countries. Am J Trop Med Hyg 2008 Feb;78(2):311-315 [FREE Full text] [Medline: 18256435]
16. Seebregts CJ, Zwarenstein M, Mathews C, Fairall L, Flisher AJ, Seebregts C, et al. Handheld computers for survey and trial data collection in resource-poor settings: development and evaluation of PDACT, a Palm Pilot interviewing system. Int J Med Inform 2009 Nov;78(11):721-731. [doi: 10.1016/j.ijmedinf.2008.10.006] [Medline: 19157967]
17. Were MC, Kariuki J, Chepng'eno V, Wandabwa M, Ndege S, Braitstein P, et al. Leapfrogging paper-based records using handheld technology: experience from Western Kenya. Stud Health Technol Inform 2010;160(Pt 1):525-529. [Medline: 20841742]
18. Groves RM, Fowler Jr FJ, Couper M, Lepkowski JM, Singer E, Tourangeau R. Survey methodology. Hoboken, NJ: Wiley; 2009.
19. McFadden E. Management of data in clinical trials. Hoboken, NJ: Wiley-Interscience; 2007.
20. Wang R. A product perspective on total data quality management. Commun ACM 1998;41(2):58-65. [doi: 10.1145/269012.269022]
21. United Nations Department of Economic and Social Affairs. Household Sample Surveys in Developing and Transition Countries (Economic & Social Affairs: Studies in Methods, Series F). New York: United Nations Publications; 2005.

22. DH Program. 2017. Demographic Health Survey Supervisor's and Editor's Manual URL: http://www.dhsprogram.com/pubs/pdf/DHSM2/DHS6_Supervisor_Editor_Manual_30Mar2011.pdf [accessed 2017-08-17] [WebCite Cache ID 6smaMkiCP]
23. Hartung C, Lerer A, Anokwa Y, Tseng C, Brunette W, Borriello G. Open Data Kit: Tools to Build Information Services for Developing Regions. 2010 Presented at: Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development - ICTD '10; 2010; London, United Kingdom. [doi: 10.1145/2369220.2369236]
24. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):010. [doi: 10.1016/j.jbi.2008.08.010]
25. Svoronos T, Mjungu D, Dhadialla P, Luk R, Zue C, Jackson J, et al. Researchgate. 2010. CommCare: automated quality improvement to strengthen community-based health URL: https://www.researchgate.net/profile/Theodore_Svoronos/publication/268198360_CommCare_Automated_Quality_Improvement_To_Strengthen_Community-Based_Health/links/5485e4780cf2ef3447892691/CommCare-Automated-Quality-Improvement-To-Strengthen-Community-Based-He [accessed 2017-08-17] [WebCite Cache ID 6smadabSx]
26. Johnson TP, Parker V, Clements C. Detection and prevention of data falsification in survey research. Surv Res 2001;32(3):1-16.
27. Ly J, Sathananthan V, Griffiths T, Kanjee Z, Kenny A, Gordon N, et al. Facility-based delivery during the Ebola virus disease epidemic in rural Liberia: analysis from a cross-sectional, population-based household survey. PLoS Med 2016 Aug;13(8):1-17 [FREE Full text] [doi: 10.1371/journal.pmed.1002096] [Medline: 27482706]
28. Patnaik S, Brunskill E, Thies W. Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice. 2009 Presented at: 2009 International Conference on Information and Communication Technologies and Development, ICTD 2009 - Proceedings; 2009; Doha, Qatar p. 74-84. [doi: 10.1109/ICTD.2009.5426700]
29. Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. PLoS One 2011;6(9):1-11 [FREE Full text] [doi: 10.1371/journal.pone.0025348] [Medline: 21966505]
30. Opendatakit. Open Data Kit: Examples URL: https://opendatakit.org/help/form-design/examples [accessed 2016-09-27] [WebCite Cache ID 6kqJDbUEt]
31. Hermens L, Shlimmer J. A machine-learning apprentice for the completion of repetitive forms. 1993 Presented at: Artificial Intelligence for Applications, 1993. Proceedings., Ninth Conference on; 1-5 March 1993; Orlando, FL p. 268-275. [doi: 10.1109/64.295135]
32. Ali A, Meek C. Predictive models of form filling. Microsoft Res Tech Rep 2009:1-8 [FREE Full text]
33. Chen K, Chen H, Conway N, Dolan H, Hellerstein J, Parikh T. Improving data quality with dynamic forms. 2009 Presented at: 2009 International Conference on Information and Communication Technologies and Development, ICTD 2009 - Proceedings; 2009; Doha, Qatar p. 321-332. [doi: 10.1109/ICTD.2009.5426738]
34. Ricketts D, Newey M, Patterson M, Hitchin D, Fowler S. Markers of data quality in computer audit: the Manchester Orthopaedic Database. Ann R Coll Surg Engl 1993 Nov;75(6):393-396 [FREE Full text] [Medline: 8285541]
35. Biemer P, Stokes L. The optimal design of quality control samples to detect interviewer cheating. Journal of Official Statistics 1989;5(1):23-29.
36. Bredl S, Winker P, Kötschau K. A statistical approach to detect interviewer falsification of survey data. Surv Methodol 2012;38(1):1-10.
37. Al-Marzouki S, Evans S, Marshall T, Roberts I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. Br Med J 2005 Jul 30;331(7511):267-270 [FREE Full text] [doi: 10.1136/bmj.331.7511.267] [Medline: 16052019]
38. Judge G, Schechter L. Detecting problems in survey data using Benford's law. J Hum Resour 2009;44(1):1-24. [doi: 10.1353/jhr.2009.0010]
39. Menold N, Winker P, Storfinger N, Kemper C. A method for ex-post identification of falsifications in survey data. In: Improving Survey Methods. Survey Standardization and Interviewers' Deviations–Impact, Reasons, Detection and Prevention. Oxford: Psychology Press, Taylor & Francis Group; 2015:86-100.
40. Birnbaum B, DeRenzi B, Flaxman A, Lesh N. Automated quality control for mobile data collection. 2012 Presented at: Proceedings of the 2nd ACM Symposium on Computing for Development - ACM DEV '12; 2012; Atlanta URL: http://dl.acm.org/citation.cfm?doid=2160601.2160603
41. Thor J, Lundberg J, Ask J, Olsson J, Carli C, Härenstam KP, et al. Application of statistical process control in healthcare improvement: systematic review. Qual Saf Health Care 2007 Oct;16(5):387-399 [FREE Full text] [doi: 10.1136/qshc.2006.022194] [Medline: 17913782]

Abbreviations

JSON: JavaScript Object Notation

ODK: Open Data Kit
PHP: PHP: hypertext preprocessor

Edited by G Eysenbach; submitted 01.04.17; peer-reviewed by X Wan, M Bujnowska-Fedak; comments to author 14.06.17; revised version received 17.06.17; accepted 19.06.17; published 18.08.17.

Please cite as:
Kenny A, Gordon N, Griffiths T, Kraemer JD, Siedner MJ
Validation Relaxation: A Quality Assurance Strategy for Electronic Data Collection
J Med Internet Res 2017;19(8):e297
URL: http://www.jmir.org/2017/8/e297/
doi:10.2196/jmir.7813
PMID:28821474

Original Paper

Comparing Inpatient Satisfaction Collected via a Web-Based Questionnaire Self-Completion and Through a Telephone Interview: An Ancillary Study of the SENTIPAT Randomized Controlled Trial

Sarah F Feldman[1,2], MSc; Nathanael Lapidus[1,2], MD, PhD; Jacques Cosnes[3], MD, PhD; Emmanuel Tiret[4], MD; Laurent Fonquernie[5], MD; Jean Cabane[6], MD; Olivier Chazouilleres[7], MD; Laure Surgers[5,8], MD; Marc Beaussier[9], MD, PhD; Alain-Jacques Valleron[10], PhD; Fabrice Carrat[1,2], MD, PhD; Gilles Hejblum[1], PhD[‡]

[1]Institut Pierre Louis d'Épidémiologie et de Santé Publique, Unité Mixte de Recherche en Santé 1136, Sorbonne Universités, Université Pierre et Marie Curie, Institut National de la Santé et de la Recherche Médicale, Paris, France

[2]Unité de Santé Publique, Hôpital Saint-Antoine, Assistance Publique–Hôpitaux de Paris, Paris, France

[3]Service de Gastro-Entérologie et Nutrition, Hôpital Saint-Antoine, Assistance Publique–Hôpitaux de Paris, Paris, France

[4]Service de Chirurgie Générale et Digestive, Hôpital Saint-Antoine, Assistance Publique–Hôpitaux de Paris, Paris, France

[5]Service de Maladies Infectieuses et Tropicales, Hôpital Saint-Antoine, Assistance Publique–Hôpitaux de Paris, Paris, France

[6]Service de Médecine Interne, Hôpital Saint-Antoine, Assistance Publique–Hôpitaux de Paris, Paris, France

[7]Service d'Hépato-Gastro-Entérologie, Hôpital Saint-Antoine, Assistance Publique–Hôpitaux de Paris, Paris, France

[8]Centre d'Immunologie et des Maladies Infectieuses, Unité Mixte de Recherche en Santé 1135, Sorbonne Universités, Université Pierre et Marie Curie, Institut National de la Santé et de la Recherche Médicale, Paris, France

[9]Service de Chirurgie Ambulatoire, Hôpital Saint-Antoine, Assistance Publique–Hôpitaux de Paris, Paris, France

[10]Thérapie Génique, Génétique, Epigénétique en Neurologie, Endocrinologie et Développement de l'Enfant, Unité Mixte de Recherche 1169, Université Paris-Sud, Institut National de la Santé et de la Recherche Médicale, Le Kremlin-Bicêtre, France

[‡]on behalf of the SENTIPAT study group

Corresponding Author:
Gilles Hejblum, PhD
Institut Pierre Louis d'Épidémiologie et de Santé Publique, Unité Mixte de Recherche en Santé 1136
Sorbonne Universités, Université Pierre et Marie Curie, Institut National de la Santé et de la Recherche Médicale
27 rue Chaligny
Paris, 75012
France
Phone: 33 149283228
Fax: 33 149283233
Email: gilles.hejblum@inserm.fr

Abstract

Background: Assessing the satisfaction of patients about the health care they have received is relatively common nowadays. In France, the satisfaction questionnaire, I-Satis, is deployed in each institution admitting inpatients. Internet self-completion and telephone interview are the two modes of administration for collecting inpatient satisfaction that have never been compared in a multicenter randomized experiment involving a substantial number of patients.

Objective: The objective of this study was to compare two modes of survey administration for collecting inpatient satisfaction: Internet self-completion and telephone interview.

Methods: In the multicenter SENTIPAT (acronym for the concept of sentinel patients, ie, patients who would voluntarily report their health evolution on a dedicated website) randomized controlled trial, patients who were discharged from the hospital to home and had an Internet connection at home were enrolled between February 2013 and September 2014. They were randomized to either self-complete a set of questionnaires using a dedicated website or to provide answers to the same questionnaires administered during a telephone interview. As recommended by French authorities, the analysis of I-Satis satisfaction questionnaire involved all inpatients with a length of stay (LOS), including at least two nights. Participation rates, questionnaire consistency (measured using Cronbach alpha coefficient), and satisfaction scores were compared in the two groups.

Results: A total of 1680 eligible patients were randomized to the Internet group (n=840) or the telephone group (n=840). The analysis of I-Satis concerned 392 and 389 patients fulfilling the minimum LOS required in the Internet and telephone group, respectively. There were 39.3% (154/392) and 88.4% (344/389) responders in the Internet and telephone group, respectively (P<.001), with similar baseline variables. Internal consistency of the global satisfaction score was higher (P=.03) in the Internet group (Cronbach alpha estimate=.89; 95% CI 0.86-0.91) than in the telephone group (Cronbach alpha estimate=.84; 95% CI 0.79-0.87). The mean global satisfaction score was lower (P=.03) in the Internet group (68.9; 95% CI 66.4-71.4) than in the telephone group (72.1; 95% CI 70.4-74.6), with a corresponding effect size of the difference at −0.253.

Conclusions: The lower response rate issued from Internet administration should be balanced with a likely improved quality in satisfaction estimates, when compared with telephone administration, for which an interviewer effect cannot be excluded.

Trial Registration: Clinicaltrials.gov NCT01769261 ; http://clinicaltrials.gov/ct2/show/NCT01769261 (Archived by WebCite at http://www.webcitation.org/6ZDF5lA41)

(J Med Internet Res 2017;19(8):e293) doi:10.2196/jmir.7061

KEYWORDS

hospital information systems; Internet; patient satisfaction; quality of health care; questionnaires and surveys; patient reported outcome measures; telephone

Introduction

Numerous questionnaires have been developed since the 70s for assessing patient satisfaction with regard to hospital health care delivery. These include the Patient Satisfaction Questionnaire [1], the Client Satisfaction Questionnaire [2], the Service Quality instrument [3-5], the Hospital Consumer Assessment of Health Providers and Systems (HCAHPS) Survey [6,7], the Short Form HK Inpatient Experience Questionnaire [8,9], and the NHS National Adult Inpatient Survey [10]. In France, inpatient hospital satisfaction has been systematically measured since 2015 with the I-Satis questionnaire [11]. These numerous questionnaires attest a worldwide concern for enhancing the central role of the patient in the health care organization. Patient satisfaction assessment is also related to the technical performance and safety of hospital care [12] and is considered as a tool contributing to hospital care evaluation, although controversial [13]. Inpatient hospital satisfaction surveys are usually either self-administrated by pen and paper or conducted by telephone, with telephone interview being a common mode of questionnaire administration. However, the development of Internet has resulted in a widespread use of Web-based questionnaires, with corresponding survey costs lower than those of telephone surveys [14-17]. Moreover, the use of Internet has increased over time, with 78% of people with Internet access at home in France in 2013 [18], thereby suggesting that this mode of administration might result in a satisfactory response rate. Nevertheless, several studies have reported a lower response rate of Internet-based surveys, as compared with other modes of administration [19-21]. On the other hand, Internet self-completion has intrinsic favorable qualities such as the avoidance of any potential bias of responses related to an interviewer effect [22], and patients are more likely to freely express their opinions [23] on websites covering anonymity than through telephone. The way in which the modes of administration of patient satisfaction survey influence response rates and the issued scores remains an important issue. Several teams studied differences between pen-and-paper and Web-based questionnaires in the field of inpatient satisfaction and quality of life [21,24,25]; however, only a few investigated the differences with surveys administered through telephone [20,26], which remains a common mode of administration for inpatient hospital satisfaction surveys [27,28]. In this context, to our knowledge, this study—which is based on the multicenter SENTIPAT (sentinel patients) randomized trial [26,29]—is the first multicenter randomized trial to date comparing inpatient satisfaction collected via the Internet or through a telephone survey. Our objective was to assess whether response rates and satisfaction scores differed between these two modes of investigation of the patients' satisfaction.

Methods

This research was an ancillary study of the multicenter, randomized SENTIPAT trial [29]. We took advantage of the trial to analyze patients' satisfaction with their hospital stay.

General Description of the SENTIPAT Trial

The SENTIPAT multicenter (five adult acute care units in a Parisian teaching hospital participated voluntarily: departments of digestive and general surgery; gastroenterology; hepatology; infectious diseases and tropical medicine; and internal medicine) randomized trial focused on the evolution of patients' health on returning home post hospitalization (follow-up duration of 6 weeks). The general objective was to determine whether the information on patient's health evolution shared by volunteer patients after returning home directly via a dedicated website was comparable with that obtained via telephone interviews. The randomization of 2050 patients (410 from each unit, 205 randomized in the Internet group and 205 randomized in the telephone group) was initially planned. The study was conducted in accordance with French regulation on ethics requirements in biomedical research.

Consecutive inpatients with Internet access at home were eligible for inclusion. Inclusion criteria also required inpatients who were not cognitively impaired and did not have a behavioral disorder, who spoke and wrote French, and were returning home after an acute care hospitalization, regardless of the type of stay—standard hospitalization (scheduled or not) on weekdays only (maximum Monday to Friday or any combination thereof) or outpatient hospitalization (1 day). Inpatients were enrolled

on the day of hospital discharge by a clinical research technician of the trial. At that time, patients were informed about the study. Eligible patients not opposed to participate in the study were randomized into two parallel groups: Internet or telephone follow-up (inherently resulting in an open-label trial) at a ratio of 1:1. On the basis of a centralized randomization that allocated the eligible patient either to the Internet or to the telephone group through a website and using an underlying permutation block randomization stratified by service, the computer-generated list of permutation was established by a statistician independent from the study. At the time of patient inclusion, the technician also collected baseline variables (length of stay [LOS], sex, age, relationship status, level of education, activity, and type of insurance). Patient was then informed and discharged with documents explaining corresponding questionnaire administration.

Characteristics of the Study

Patients

The French authorities have made available the instructions for analyzing I-Satis questionnaire [30], and according to these recommendations, the study was restricted to patients whose LOS included at least two consecutive nights.

Questionnaire Structure

The detailed I-Satis questionnaire used in this study (all questions and corresponding proposed answers) is directly accessible via the Internet [11]. The I-Satis questionnaire comprises 32 items exploring six dimensions: global care (Q1, Q2, Q4, Q13, Q14, and Q15), information to patient (Q16, Q18, Q27, Q28, Q29, and Q30), communication with health care providers (Q3, Q5, Q6, Q17, and Q20), behavior of health care providers (Q7, Q8, Q9, Q10, and Q11), hospital room convenience (Q21, Q22, Q23, and Q24), and hospital catering (Q25 and Q26). The recommendations of the French authorities for I-Satis analysis [30] indicate that 4 questions (Q12, Q19, Q31, and Q32) are not involved in score calculations.

Questionnaire Administration

All patients were informed that their opinions were kept anonymous. For the patients who had been randomized in the telephone group, the I-Satis questionnaire was administered during a telephone interview with a clinical research technician 7 days after discharge (the appointment was scheduled on the day of discharge), with a maximum of three attempts to contact them. For the patients who had been randomized in the Internet group, the same questionnaire was available on the dedicated website on the day of discharge (D0) and was completed directly online by the patient, who had been given oral and written instructions (information sheet) to connect for the first time 7 days post discharge. “Reminders” were sent by email once weekly for 6 weeks after discharge to potential responders of the Internet group, who had not completed the discharge questionnaire yet.

Score Construction

The questionnaires were analyzed according to the French national recommendations of the Direction Générale de l'Offre de Soins [30].

Each item was rated from 0 to 10. Rates 1 to 5 corresponded to increasing ordinal rankings of satisfaction; rates 0, 6, 7, and 8 corresponded to answers of nonrelevancy of the item for the patient (depending on the item: never felt discomfort, no drugs were prescribed, no surgery, etc), and spontaneous answers “I don’t know” and “I don’t wish to answer the question” were rated 9 and 10. Rates 1, 2, 3, 4, and 5 were valued 0, 25, 50, 75, and 100, respectively, in the analysis. Rates 0, 6, 7, 8, 9, and 10 were handled into the analyses as a missing value.

The scores of the dimensions “global care,” “information to patient,” “communication with health care providers,” “behavior of health care providers,” “hospital room convenience,” and “hospital catering” were calculated if at least three, three, three, three, two, and two items comprising the dimension were answered, respectively. The global score was calculated whenever every dimension score was calculated.

Statistical Analysis

The participation rates observed in the Internet and telephone groups were compared using the Fisher exact test, as well as the proportions of nonrelevancy answers observed in these two groups. The delays of questionnaire completion observed in the Internet and telephone groups were compared using Wilcoxon-Mann-Whitney test. Internal consistency of questionnaires was measured by calculating Cronbach alpha [31], taking into account every score that could be calculated according to the abovementioned rules. An alpha coefficient value of greater than .7 was considered as satisfactory. Dimensions’ scores were calculated for each patient as the mean of the corresponding dimensions’ items, and global score was the mean of all answered items of the questionnaire. CIs were obtained by bootstrap. Standardized Cohen *d*-type effect size was measured between the scores of the two groups [32].

Comparisons between the Internet and telephone groups in terms of Cronbach alpha coefficients and in terms of satisfaction scores (including dimensions’ scores) were made using a permutation test [33], with the null hypothesis distribution (distribution of the difference between the two groups under the hypothesis of no difference) generated through 1,000,000 shuffled datasets. A P value of $\leq .05$ defined the significance of comparisons.

Missing data were handled as follows: First, nonresponding patients were excluded from score analyses. Patients for whom less than 16 items were completed were also excluded from score analyses (ie, handled as nonresponders in the analyses). Second, the scores issued from the remaining partially completed questionnaires were calculated as mentioned above (see subsection on Score Construction).

All analyses were made with the R statistical computing freeware version 3.3.0 [34].

Ethic and Legal Approvals

The SENTIPAT study was approved by the Comité de Protection des Personnes Ile de France IX (decision CPP-IDF IX 12-014, June 12, 2012); the Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé (Decision 12.365, June 20, 2012); and the

Commission Nationale de l'Informatique et des Libertés (Decision DR-2012-582, December 12, 2012).

Results

Between February 25, 2013 and September 8, 2014, we managed to enroll in the SENTIPAT study 1680 eligible patients (840 randomized in the Internet group and telephone group each) and not opposed to participating in the trial. Among these, the baseline population of patients fulfilling the minimum LOS of 2 nights required for I-Satis investigation comprised 781 patients, with 392 and 389 patients in the Internet and telephone groups, respectively (Figure 1). Table 1 provides the details of baseline values of the patients who constituted the population investigated in this study. There were no missing data relating to baseline values. Considering all 781 patients, the median LOS was 5 days (interquartile range [IQR]: 2-9); there were as many men as women, participants were aged 19 to 97 years, and median age was 53 years (IQR: 37-64), and 711 patients (91.0%) had a complementary private health insurance in addition to the compulsory health insurance.

There were 154 responders out of the 392 patients in the Internet group (response rate of 39.3%) and 344 responders out of the 389 patients in the telephone group (response rate of 88.4%; $P<.001$), and the corresponding median delays between hospital discharge and questionnaire completion were 6 days (IQR: 3-15.75) and 7 days (IQR: 7-9), respectively ($P=.002$). Missing data in responders concerned 10 patients of the Internet group: answer to question 13 (satisfaction about pain management) was missing in 2 responders, answer to question 20 (satisfaction about the answers of the surgeon about patient's questions on surgery) was missing in 3 responders, and answer to both questions was missing in 5 responders. In addition, there were 13 (8.4%), 95 (61.7%), 43 (27.9%), and 3 (1.9%) responders in the Internet group with 0, 1 to 5, 6 to 10, and more than 10 answers, for which the answer code corresponded to nonrelevancy or refusal (further handled as a missing value in the analyses, see the section on Methods), respectively, whereas the corresponding responders observed in the telephone group were 3 (0.8%), 124 (36.0%), 200 (58.1%), and 15 (4.4%), respectively. Internet responders provided an answer code corresponding to nonrelevancy and refusal less frequently than telephone responders ($P<.001$).

Figure 1. Flow of participants through the study.

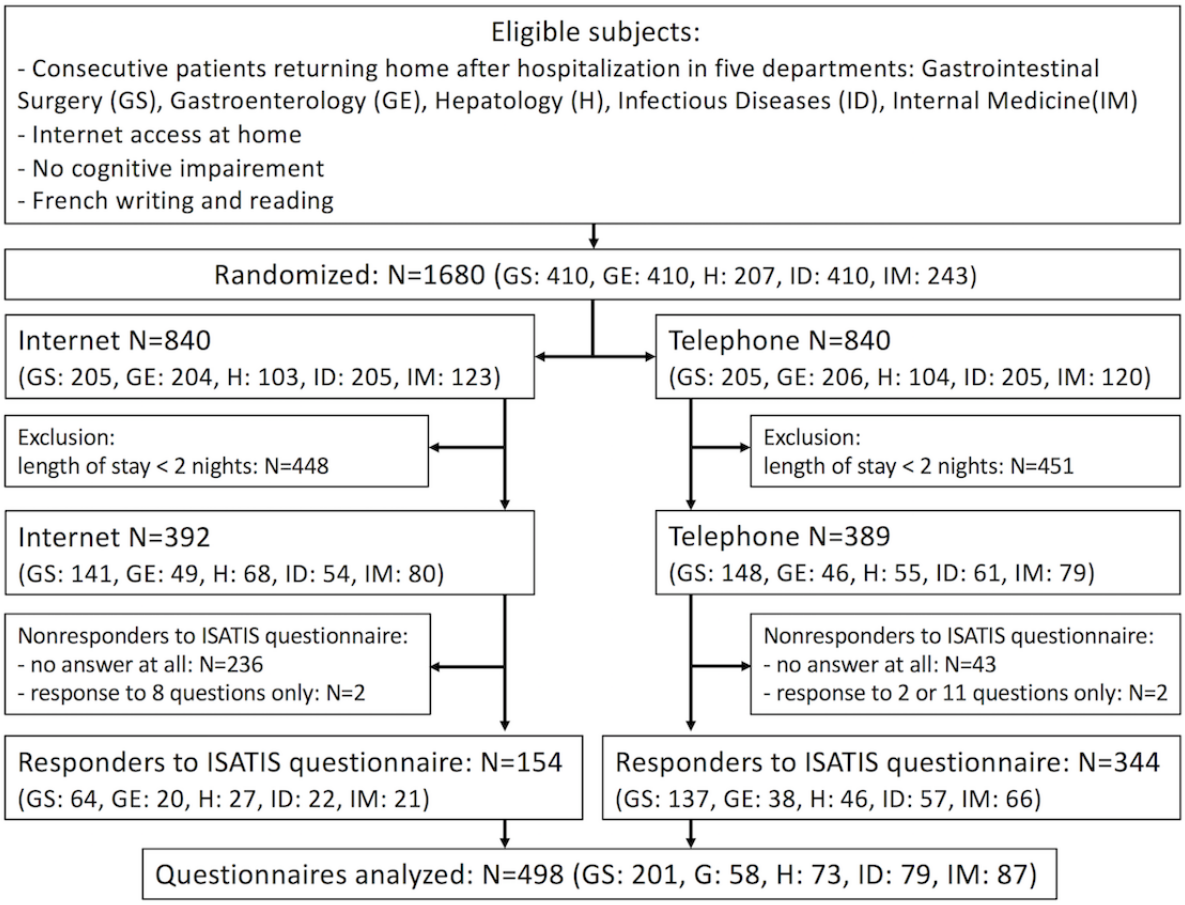


Table 1. Characteristics of study patients.

Variable	Total, N=781	Responders Internet, n=154	Nonresponders Internet, n=238	Responders telephone, n=344	Nonresponders telephone, n=45
LOS[a], in days, median (IQR[b])	5 (2-9)	5 (3-8)	5.5 (3-8)	5 (2-9)	7 (3-10)
Sex, n (%)					
Male	385 (49.4)	71 (46.1)	115 (48.3)	176 (51.2)	23 (51)
Age in years, median (IQR)	53 (37-64)	55 (38-64)	50 (36-64)	53 (38-65)	51 (33-67)
Relationship status, n (%)					
Living alone[c]	342 (43.8)	59 (38.3)	112 (47.1)	146 (42.4)	25 (56)
Living as a couple[d]	439 (56.2)	95 (61.7)	126 (52.9)	198 (57.6)	20 (44)
Level of education, n (%)					
Lower secondary education	88 (11.3)	17 (11.0)	28 (11.8)	35 (10.2)	8 (18)
Upper secondary education	249 (32.0)	46 (29.9)	80 (33.8)	110 (32.0)	13 (30)
Postsecondary nontertiary education or short-cycle tertiary	113 (14.5)	23 (14.9)	32 (13.5)	53 (15.4)	5 (11)
Bachelor's degree or above	329 (42.2)	68 (44.2)	97 (40.9)	146 (42.4)	18 (41)
Activity, n (%)					
Nonworking	321 (41.2)	64 (42.1)	101 (42.4)	139 (40.4)	17 (38)
Employed or student	458 (58.8)	88 (57.9)	137 (57.6)	205 (59.6)	28 (62)
Insurance, n (%)					
Precarious[e]	30 (03.8)	3 (02.0)	13 (05.5)	11 (03.2)	3 (07)
Compulsory health insurance	40 (05.1)	7 (04.6)	15 (06.3)	14 (04.1)	4 (09)
Compulsory health insurance plus complementary private health insurance	711 (91.0)	144 (93.5)	210 (88.2)	319 (92.7)	38 (84)
Delay between hospital discharge and questionnaire completion, in days, median (IQR)	N/A[f]	6 (3-15.75)	N/A	7 (7-9)	N/A

[a]LOS: length of hospital stay.

[b]IQR: interquartile range.

[c]Single, widowed, divorced, separated.

[d]Married, living together under a civil solidarity pact, simply living together without legal ties.

[e]Benefit from state medical help or universal health insurance.

[f]N/A: not applicable.

Table 2 shows the values of Cronbach alpha in the Internet and telephone responders. All estimates, with the exception of those corresponding to room convenience, were >.7. The alpha estimates observed in the Internet group were always greater than those observed in the telephone group, the difference being statistically significant for the two dimensions, global care and room convenience (*P*=.003 and *P*=.03, respectively), and for the global satisfaction (*P*=.03).

Table 3 summarizes the satisfaction scores observed in the Internet and telephone groups. The mean global satisfaction score was 68.89 (95% CI 66.36-71.36) in the Internet group and 72.01 (95% CI 70.36-73.58) in the telephone group. In both groups, the dimension that received the lowest score was hospital catering, with means of 45.77 (95% CI 42.18-49.39) and 45.70 (95% CI 43.32-48.06) in the Internet and telephone group, respectively. Conversely, in both groups, the theme that received the highest score was behavior of health care providers, with means of 87.49 (95% CI 85.05-89.73) and 92.14 (95% CI 90.81-93.39) in the Internet and telephone group, respectively. There were three dimension scores significantly smaller in the Internet group than in the telephone group: information to patients with a mean difference of −5.38 (*P*=.008), communication with health care providers with a mean difference of −7.16 (*P*=.003), and behavior of health care providers with a mean of −4.66 (*P*<.001). The global satisfaction score was significantly smaller in the Internet group with a mean difference of −7.16 (*P*=.003) and behavior of health care providers with a mean of −4.66 (*P*<.001). The global satisfaction score was significantly smaller in the Internet group, with a mean difference of −3.12 (*P*=.03). The absolute values of the effect sizes ranged from 0.003 to 0.352. The satisfaction scores observed in the group of Internet responders according to the delay of questionnaire completion are summarized in Multimedia Appendix 1, and whatever the type of score

considered, the score difference between the two subgroups of Internet responders (questionnaire completed at day 7 after discharge or later versus questionnaire completed earlier) was not significant.

Table 2. Cronbach alpha coefficients.

Dimension of the score (number of items involved)	Internet responders: Cronbach alpha (95% CI); n	Telephone responders: Cronbach alpha (95% CI); n	Comparison between the two groups: *P* value
Global care (6)	.92 (0.87-0.96); 154	.79 (0.73-0.85); 344	.003
Information to patients (6)	.89 (0.81-0.94); 131	.83 (0.76-0.88); 235	.27
Communication with health care providers (5)	.74 (0.65-0.82); 152	.71 (0.64-0.77); 334	.56
Behavior of health care providers (5)	.82 (0.71-0.92); 153	.73 (0.61-0.84); 344	.33
Hospital room convenience (4)	.66 (0.53-0.75); 154	.48 (0.37-0.56); 344	.03
Hospital catering (2)	.86 (0.79-0.91); 136	.76 (0.69-0.82); 285	.07
Global satisfaction score (28)	.89 (0.86-0.91); 116	.84 (0.79-0.87); 189	.03

Table 3. Satisfaction scores.

Dimension of the score	Internet responders: mean score (95% CI); n	Telephone responders: mean score (95% CI); n	Internet-telephone: mean score difference (95% CI), *P* value	Effect size (95% CI)
Global care	70.67 (68.01-73.34); 154	72.15 (70.56-73.75); 344	−1.48 (−4.59 to 1.64), .33	−0.094 (−0.290 to 0.105)
Information to patients	59.62 (56.07-63.23); 131	65.01 (62.75-67.25); 235	−5.38 (−9.53 to −1.19), .009	−0.286 (−0.501 to −0.062)
Communication with health care providers	67.42 (63.28-71.42); 152	74.58 (71.90-77.15); 334	−7.16 (−12.01 to −2.36), .003	−0.287 (−0.477 to −0.096)
Behavior of health care providers	87.49 (85.05-89.73); 153	92.14 (90.81-93.39); 344	−4.66 (−7.40 to −2.02), <.001	−0.352 (−0.554 to −0.154)
Hospital room convenience	61.03 (58.07-63.95); 154	60.97 (59.07-62.88); 344	0.05 (−3.46 to 3.50), .98	0.003 (−0.189 to 0.195)
Hospital catering	45.77 (42.18-49.39); 136	45.70 (43.32-48.06); 285	0.07 (−4.24 to 4.44), .98	0.003 (−0.206 to 0.214]
Global satisfaction score	68.89 (66.36-71.36); 116	72.01 (70.36-73.58); 189	−3.12 (−6.13 to −0.15), .03	−0.253 (−0.490 to −0.014]

Discussion

Principal Findings

The investigation of patient satisfaction after a hospital stay resulted in several differences when comparing the two modes of questionnaire administration: self-reported Internet completion or telephone interview. The comparison between these modes of administration may be discussed according to three topics: response rate, questionnaire reliability, and satisfaction scores. The response rate observed in the group of patients randomized in the Internet group (39.3%) was much lower than that observed in the group of patients randomized in the telephone group (88.4%). Such a difference might have resulted in unbalancing the initial comparability of responders in the two groups even if Table 1 indicates that baseline variables are similar in the responders of the two groups. Unsurprisingly, the observed difference between the two groups in terms of participation rate is in accordance with the previous results issued from the same cohort focusing on patient satisfaction with regard to the hospital discharge process [26]. The difference between the participation rates observed with the two administration modes of the survey might be, at least in part, owing to the fact that it is easier to ignore an email than a phone call scheduled at a date chosen by the patient. The participation rates observed in our study are also similar to those reported by Harewood et al [20] who investigated patient satisfaction with endoscopy and observed a response rate of 34% and 78% in the Internet group and telephone group, respectively. However, comparing the response rates observed in our study with other rates previously reported is probably of limited interest as the study design widely varies from one study to another, as response rates are likely to be highly sensitive to the detailed underlying procedures for selecting participants (eg, face-to-face enrollment vs random selection in an administrative database, issues related to the initial comparability of the participants allocated in the Internet and telephone arms), reaching/soliciting responders (including

reminding procedures for soliciting Web participants to complete the survey or procedure for scheduling the phone calls), and collecting answers (eg, attractiveness of the website and ease in accessing/completing the questionnaire form).

With the exception of the values for the hospital room convenience dimension, which raise concerns, the values of Cronbach alpha were satisfactory for all dimensions investigated and for the global satisfaction score, favoring the conduction of surveys with this questionnaire using either administration mode. Besides, interestingly, considering all six dimensions of the questionnaire, the values of Cronbach alpha were always higher in the Internet group than in the telephone group, with a statistical significance observed for two dimensions (global care and hospital room convenience) and for the global satisfaction score. Here, the adjunction of an interviewer in the telephone group (as compared with self-completion in the Internet group) might be considered as an undesired burden disturbing initial signal.

The observed score differences between the Internet and telephone groups (see Table 3) are contrasted, depending on the dimension investigated. On the one hand, considering hospital room convenience and hospital catering dimensions, both telephone and Internet modes of administration resulted in very similar satisfaction scores, and the difference was only slight when considering global care dimension. On the other hand, in the three dimensions related to interactions with health care providers, that is, information to patients, communication with health care providers, and behavior of health care providers, scores were significantly lower in the Internet group than in the telephone group, although it is worth mentioning that corresponding effect sizes never exceeded 0.35, a value below the medium threshold proposed by Cohen [32]. Such an observation raises a general comment on the surveys conducted in this domain. Those surveys are deployed for investigating patient satisfaction with hospital services, for bringing into light the elements which require improvements and for assessing evolution with time. Additionally, French authorities require the analysis of 120 patients per medical center each year. In such a context, our study's finding that a mean difference of 7 points based on a sample size of 498 responders is modest in terms of effect size, suggests that potential improvements on patient satisfaction are very difficult to evidence. Dynamic trends within a given center from one year to another should be interpreted with great caution and must take into account the underlying variability of the scores, and a similar caution should be required in the interpretation of differences between centers. A potential explanation for the higher scores observed in the telephone group is that a patient might be more reluctant to provide low scores to an interviewer (moreover potentially identified as a member of the hospital staff) than when completing a strict anonymous form via the Internet. Previous studies [35,36] have also mentioned such a social desirability bias [23] as a potential explanation for the higher patient satisfaction scores issued from a questionnaire administered by a telephone interviewer as compared with a self-completed form administered by mail [35] or via the Internet [36]. In addition, the distribution of the delay between hospital discharge and questionnaire completion was more variable in the Internet group. However, as shown in the Multimedia Appendix 1, the scores of the Internet responders did not significantly vary according to the delay of questionnaire completion, indicating that the wider variability in the delay of questionnaire completion observed in the Internet group had a very limited impact (if any) on the differences of scores that were observed between the telephone and Internet modes of administration.

Strengths and Limitations

A strength of the study relies on the fact that it is the first randomized trial reported to date that compared inpatient satisfaction collected either via a telephone interview or via the self-completion of a similar questionnaire on a dedicated website and involved a reasonable sample of inpatients, both in terms of case-mix variability (patients originating from 5 very different hospital wards) and in terms of sample size (498 questionnaires were eventually analyzed). A home access to the Internet and a phone number were two required inclusion criteria for patient eligibility, ensuring the initial comparability of the individuals randomized in the two administration modes of the questionnaire. Moreover, to our knowledge, this study is the first to date that explores, in detail, critical issues relating to the I-Satis questionnaire, which is dedicated to be deployed in all inpatient structures in France; on the one hand, this constitutes an additional strength of the study, while on the other hand, the fact that this questionnaire is yet restricted to France constitutes a limitation of the study.

Conclusions

In conclusion, our study shows that the lower response rate observed with the Internet mode of administration than that observed with the telephone mode of administration must be balanced with other positive features associated with the Internet. Using the latter mode of administration has a potential lower cost than telephone [14-17], and the quality of satisfaction estimates is likely improved because the potential veil of a telephone interviewer is discarded, allowing patients to express more freely. They are more likely to rate their satisfaction about hospital stay with lower scores. This study indicates that some of these score decreases are statistically significant but the corresponding effect sizes are small, indicating that the decreases relate to moderate differences. Communication habits are evolving, and the Web form is progressively adopted as a reference mode for administrating surveys as well as a reference mode for completing questionnaires. With the exception of the higher response rate observed with telephone interview in this study, all other study results support the deployment of Web-based questionnaires for exploring inpatient satisfaction.

Acknowledgments

The authors would like to thank Fabrice Carrat, Bérengère Couturier, Gilles Hejblum, Morgane Le Bail, and Alain-Jacques Valleron from the Scientific Committee of the SENTIPAT study group and express their gratitude to the physicians within the poles and units in which the patients were recruited, including the heads: Marc Beaussier, Jean-Paul Cabane, Olivier Chazouillères, Jacques Cosnes, Jean-Claude Dussaule, Pierre-Marie Girard, Emmanuel Tiret, and Dominique Pateron; additional corresponding physicians: Laurent Beaugerie, Laurence Fardet, Laurent Fonquernie, François Paye, and Laure Surgers; and nursing supervisors: Françoise Cuiller, Catherine Esnouf, Hélène Haure, Valérie Garnier, Josselin Mehal-Birba, Nelly Sallée, and Sylvie Wagener.

The authors are also indebted to the technical team of the study, including the clinical research technicians Élodie Belladame, Azéline Chevance, Magali Girard, and Laurence Nicole, who included patients, collected baseline data, and interviewed the patients followed up by telephone; and the software staff, especially Pauline Raballand, but also Frédéric Chau and Frédéric Fotré, who created and maintained the trial's dedicated website.

We thank all the medical and nursing and administrative staff of the general and digestive surgery (including ambulatory surgery), gastroenterology, hepatology, and infectious diseases and internal medicine departments of Hôpital Saint-Antoine; all the patients who participated in the study; and Isabelle Mounier-Emeury for helpful discussions on the calculations of I-Satis scores.

The Assistance Publique–Hôpitaux de Paris (Département de la Recherche Clinique et du Développement) was the trial sponsor. The SENTIPAT study was funded by grant AOM09213 K081204 from Programme Hospitalier de Recherche Clinique 2009 (Ministère de la Santé). The sponsor and the funders had no role in study design, data collection and analysis, decision to publish, or the preparation of the manuscript.

Authors' Contributions

GH had full access to all of the raw data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. SFF, FC, and GH contributed to the study conception and design, whereas GH contributed to data acquisition. SFF, NL, and GH contributed to the analysis, and SFF, NL, JaC, PMG, ET, JeC, OC, MB, FC, and GH (ie, all authors) contributed to the interpretation of data. SFF and GH worked on the first draft of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Satisfaction scores observed in the Internet responder group according to the delay of questionnaire completion.

[PDF File (Adobe PDF File), 64KB - jmir_v19i8e293_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (v.1.6.1).

[PDF File (Adobe PDF File), 784KB - jmir_v19i8e293_app2.pdf]

References

1. Ware JEJ, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. Eval Program Plann 1983;6(3-4):247-263. [Medline: 10267253]
2. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [Medline: 10245370]
3. Babakus E, Mangold WG. Adapting the SERVQUAL scale to hospital services: an empirical investigation. Health Serv Res 1992 Feb;26(6):767-786 [FREE Full text] [Medline: 1737708]
4. Parasuraman A, Zeithaml VA, Berry LL. Servqual: a multiple-item scale for measuring consumer perceptions of service quality. J Retailing 1998;64(1):12-40.
5. Parasuraman A, Zeithaml VA, Berry LL. A conceptual model of service quality and its implications for future research. J Marketing 1985;49(4):41-50. [doi: 10.2307/1251430]
6. Goldstein E, Farquhar M, Crofton C, Darby C, Garfinkel S. Measuring hospital care from the patients' perspective: an overview of the CAHPS hospital survey development process. Health Serv Res 2005 Dec;40(6 Pt 2):1977-1995 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00477.x] [Medline: 16316434]
7. Weidmer BA, Brach C, Slaughter ME, Hays RD. Development of items to assess patients' health literacy experiences at hospitals for the consumer assessment of healthcare providers and systems (CAHPS) hospital survey. Med Care 2012 Sep;50(9 Suppl 2):S12-S21 [FREE Full text] [doi: 10.1097/MLR.0b013e31826524a0] [Medline: 22895225]

8. Wong EL, Coulter A, Cheung AW, Yam CH, Yeoh EK, Griffiths SM. Patient experiences with public hospital care: first benchmark survey in Hong Kong. Hong Kong Med J 2012 Oct;18(5):371-380 [FREE Full text] [Medline: 23018064]
9. Wong EL, Coulter A, Hewitson P, Cheung AW, Yam CH, Lui SF, et al. Patient experience and satisfaction with inpatient service: development of short form survey instrument measuring the core aspect of inpatient experience. PLoS One 2015;10(4):e0122299 [FREE Full text] [doi: 10.1371/journal.pone.0122299] [Medline: 25860775]
10. DeCourcy A, West E, Barron D. The national adult inpatient survey conducted in the English national health service from 2002 to 2009: how have the data been used and what do we know as a result? BMC Health Serv Res 2012 Mar 21;12:71 [FREE Full text] [doi: 10.1186/1472-6963-12-71] [Medline: 22436670]
11. Legifrance. Arrêté du 22 juin 2012 relatif au modèle de questionnaire à utiliser lors de la mise en œuvre de l'enquête téléphonique de satisfaction des patients hospitalisés (I-SATIS) URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000026148913 [accessed 2016-11-14] [WebCite Cache ID 6m1BIys8u]
12. Isaac T, Zaslavsky AM, Cleary PD, Landon BE. The relationship between patients' perception of care and measures of hospital quality and safety. Health Serv Res 2010 Aug;45(4):1024-1040 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01122.x] [Medline: 20528990]
13. Fenton JJ, Jerant AF, Bertakis KD, Franks P. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. Arch Intern Med 2012 Mar 12;172(5):405-411. [doi: 10.1001/archinternmed.2011.1662] [Medline: 22331982]
14. Roster CA, Rogers RD, Albaum G, Klein D. A comparison of response characteristics from web and telephone surveys. In J Market Res 2004;46(3):359-373.
15. Braunsberger K, Wybenga H, Gates R. A comparison of reliability between telephone and web-based surveys. J Bus Res 2007 Jul;60(7):758-764. [doi: 10.1016/j.jbusres.2007.02.015]
16. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. BMC Med Res Methodol 2012;12:132 [FREE Full text] [doi: 10.1186/1471-2288-12-132] [Medline: 22938205]
17. Yang M, Tan EC. Web-based and telephone surveys to assess public perception toward the national health insurance in taiwan: a comparison of cost and results. Interact J Med Res 2015 Apr 17;4(2):e9 [FREE Full text] [doi: 10.2196/ijmr.4090] [Medline: 25886989]
18. Gombault V. L'internet de plus en plus prisé, l'internaute de plus en plus mobile. INSEE Premiere 2013 Jun;1452:1-4 [FREE Full text]
19. Basnov M, Kongsved SM, Bech P, Hjollund NH. Reliability of short form-36 in an internet- and a pen-and-paper version. Inform Health Soc Care 2009 Jan;34(1):53-58. [doi: 10.1080/17538150902779527] [Medline: 19306199]
20. Harewood GC, Wiersema MJ, de Groen PC. Utility of Web-based assessment of patient satisfaction with endoscopy. Am J Gastroenterol 2003 May;98(5):1016-1021. [doi: 10.1111/j.1572-0241.2003.07400.x] [Medline: 12809822]
21. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. J Med Internet Res 2007;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
22. Dijkstra W. How interviewer variance can bias the results of research on interviewer effects. Qual Quant 1983 Jun;17(3):179-187. [doi: 10.1007/BF00167582]
23. Kreuter F, Presser S, Tourangeau R. Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. Public Opin Q 2009 Jan 26;72(5):847-865. [doi: 10.1093/poq/nfn063]
24. Zuidgeest M, Hendriks M, Koopman L, Spreeuwenberg P, Rademakers J. A comparison of a postal survey and mixed-mode survey using a questionnaire on patients' experiences with breast care. J Med Internet Res 2011;13(3):e68 [FREE Full text] [doi: 10.2196/jmir.1241] [Medline: 21946048]
25. Wu RC, Thorpe K, Ross H, Micevski V, Marquez C, Straus SE. Comparing administration of questionnaires via the internet to pen-and-paper in patients with heart failure: randomized controlled trial. J Med Internet Res 2009;11(1):e3 [FREE Full text] [doi: 10.2196/jmir.1106] [Medline: 19275979]
26. Couturier B, Carrat F, Hejblum G. Comparing patients' opinions on the hospital discharge process collected with a self-reported questionnaire completed via the internet or through a telephone survey: an ancillary study of the SENTIPAT randomized controlled Trial. J Med Internet Res 2015 Jun 24;17(6):e158 [FREE Full text] [doi: 10.2196/jmir.4379] [Medline: 26109261]
27. Centers for Medicare and Medicaid Services. Hcahpsonline. 2016. HCAHPS hospital survey URL: http://www.hcahpsonline.org [accessed 2016-11-14] [WebCite Cache ID 6ToRxrGtn]
28. Kemp KA, Chan N, McCormack B, Douglas-England K. Drivers of inpatient hospital experience using the HCAHPS survey in a Canadian setting. Health Serv Res 2015 Aug;50(4):982-997 [FREE Full text] [doi: 10.1111/1475-6773.12271] [Medline: 25483921]
29. Clinicaltrials.gov. Sentinel patients: value of an information system collecting patient's own report on his healthcare management (SENTIPAT) URL: http://clinicaltrials.gov/ct2/show/NCT01769261 [accessed 2015-06-11] [WebCite Cache ID 6ZDF5lA41]

30. Social-sante.gouv. 2013. Fiche descriptive de l'indicateur de mesure de la satisfaction des patients hospitalisés en médecine, chirurgie ou obstétrique (MCO) URL: http://social-sante.gouv.fr/IMG/pdf/Fiche_de_modalite_de_calcul_I-SATIS.pdf [accessed 2016-11-14] [WebCite Cache ID 6m1A9cGwq]
31. Cortina JM. What is coefficient alpha? an examination of theory and applications. J Appl Psychol 1993;78(1):98-104. [doi: 10.1037//0021-9010.78.1.98]
32. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
33. Good PI. Permutation, Parametric and Bootstrap Tests Of Hypotheses. New York, NY: Springer; 2005:1-314.
34. R Core Team. R-project. Vienna, Austria: R Foundation for Statistical Computing R: a language and environment for statistical computing URL: https://www.r-project.org [accessed 2016-11-14] [WebCite Cache ID 6ks6sL8Lv]
35. Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. Health Serv Res 2009 Apr;44(2 Pt 1):501-518 [FREE Full text] [doi: 10.1111/j.1475-6773.2008.00914.x] [Medline: 19317857]
36. Dillman DA, Phelps G, Tortora R, Swift K, Kohrell J, Berck J, et al. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Soc Sci Res 2009 Mar;38(1):1-18. [doi: 10.1016/j.ssresearch.2008.03.007]

Abbreviations

HCAHPS: Hospital Consumer Assessment of Health Providers and Systems
IQR: interquartile range
LOS: length of stay

Edited by I Weber; submitted 29.11.16; peer-reviewed by K Kemp, C Varsi; comments to author 14.01.17; revised version received 15.02.17; accepted 16.02.17; published 23.08.17.

Please cite as:
Feldman SF, Lapidus N, Cosnes J, Tiret E, Fonquernie L, Cabane J, Chazouilleres O, Surgers L, Beaussier M, Valleron AJ, Carrat F, Hejblum G
Comparing Inpatient Satisfaction Collected via a Web-Based Questionnaire Self-Completion and Through a Telephone Interview: An Ancillary Study of the SENTIPAT Randomized Controlled Trial
J Med Internet Res 2017;19(8):e293
URL: http://www.jmir.org/2017/8/e293/
doi:10.2196/jmir.7061
PMID:28835354

Original Paper

Smartphone Ownership Among US Adult Cigarette Smokers: 2014 Health Information National Trends Survey (HINTS) Data

Jaimee L Heffner[1], PhD; Kristin E Mull[1], MS

Fred Hutchinson Cancer Research Center, Division of Public Health Sciences, Seattle, WA, United States

Corresponding Author:
Jaimee L Heffner, PhD
Fred Hutchinson Cancer Research Center
Division of Public Health Sciences
1100 Fairview Ave N, Mail Stop M3-B232
PO Box 19024
Seattle, WA, 98109
United States
Phone: 1 2066677314
Email: jheffner@fredhutch.org

Abstract

Background: Despite increasing interest in smartphone apps as a platform for delivery of tobacco cessation interventions, no previous studies have evaluated the prevalence and characteristics of smokers who can access smartphone-delivered interventions.

Objective: To guide treatment development in this new platform and to evaluate disparities in access to smartphone-delivered interventions, we examined associations of smartphone ownership with demographics, tobacco use and thoughts about quitting, other health behaviors, physical and mental health, health care access, and Internet and technology utilization using a nationally representative sample of US adult smokers.

Methods: Data were from the National Cancer Institute's 2014 Health Information National Trends Survey 4 (HINTS 4), Cycle 4. This mailed survey targeted noninstitutionalized individuals aged 18 years or older using two-stage stratified random sampling. For this analysis, we restricted the sample to current smokers with complete data on smartphone ownership (n=479).

Results: Nearly two-thirds (weighted percent=63.8%, 248/479) of smokers reported owning a smartphone. Those who were younger (P<.001), employed (P=.002), never married (P=.002), and had higher education (P=.002) and income (P<.001) had the highest rates of ownership. Smartphone owners did not differ from nonowners on frequency of smoking, recent quit attempts, or future plans to quit smoking, although they reported greater belief in the benefits of quitting (P=.04). Despite being equally likely to be overweight or obese, smartphone owners reported greater fruit and vegetable consumption (P=.03) and were more likely to report past-year efforts to increase exercise (P=.001) and to lose weight (P=.02). No differences in health care access and utilization were found. Smartphone owners reported better physical and mental health in several domains and higher access to and utilization of technology and the Internet, including for health reasons.

Conclusions: Smartphone ownership among smokers mirrors many trends in the general population, including the overall rate of ownership and the association with younger age and higher socioeconomic status. Apps for smoking cessation could potentially capitalize on smartphone owners' efforts at multiple health behavior changes and interest in communicating with health care providers via technology. These data also highlight the importance of accessible treatment options for smokers without smartphones in order to reach smokers with the highest physical and mental health burden and prevent worsening of tobacco-related health disparities as interventions move to digital platforms.

(J Med Internet Res 2017;19(8):e305) doi:10.2196/jmir.7953

KEYWORDS

mHealth; mobile health; tobacco; smoking; nicotine use disorder

Introduction

Smartphones are a promising technology platform for the delivery of smoking cessation interventions. Hundreds of cessation apps are currently available, although only a handful have undergone empirical evaluation of efficacy [1-5] and the vast majority have been found lacking in terms of their inclusion of theoretically backed, evidence-based content [6-8]. A major foundational gap in the literature on smartphone apps for smoking cessation is that no previous studies have evaluated, using a nationally representative sample, what proportion of current smokers in the United States own smartphones, or the characteristics of smokers who currently have access to smartphone-delivered interventions versus those who do not. This gap in the literature hinders treatment development in this new platform as well as understanding of disparities in access to smartphone-delivered interventions.

The application of smartphone technologies to tobacco cessation is new; the technology itself is not. The first smartphones were produced in the 1990s, although the mass adoption of smartphones did not occur in the United States until the late 2000s with the release of Apple's first-generation iPhone. In the decade that followed, ownership of smartphones increased sharply over time—it jumped from 35% in 2011 to 64% in 2014 and 77% in 2016 [9]. After this extremely rapid period of growth, smartphones are now approaching market saturation in the United States and other developed markets [10]

As with any technology, adoption within some segments of the population is lagging due to factors ranging from lack of interest in learning to use a new technology to low affordability. It is not clear whether US population trends in adoption of smartphones generalize to the population of smokers because smoking behavior is much more concentrated among individuals with lower education and income [11]. The "digital divide" between those smokers who currently own smartphones versus those who do not provides useful information about the potential reach of smartphone-delivered cessation interventions and informs how they are designed and marketed.

Population-level data for the United States suggest that smartphone ownership is associated with several demographic characteristics, including younger age and higher educational attainment and income [12]. However, demographics provide an incomplete picture of the differences between smokers who own smartphones and those who do not. Understanding at a population level how these two groups differ on health care access, health behaviors, and status, as well as usage of other technologies beyond the smartphone, offers an opportunity to target smartphone-delivered interventions based on users' characteristics and needs, and to identify the needs of smokers who require cessation support through modalities other than the smartphone. Regarding user characteristics and needs, population-level data on smokers who own smartphones could elucidate (1) the prevalence of known barriers to quitting (eg, low motivation, mental health symptoms), (2) the proportion of smartphone-owning smokers whose weight or diet and exercise behaviors place them at even greater risk for poor health outcomes, (3) extent of interest in making changes to these other health behaviors, and (4) potential feasibility of smartphone-delivered treatments that require involvement of the health care system (eg, advising users to seek prescription medications or to discuss withdrawal symptoms with a physician, direct messaging with health care providers).

To guide treatment development on the smartphone platform and to evaluate disparities in access to smartphone-delivered interventions, the aim of this study was to answer three key questions using a large, nationally representative sample:

1. What proportion of current smokers own smartphones?
2. Among smokers, what demographic characteristics are associated with smartphone ownership?
3. How do smokers who own smartphones differ from those who do not on tobacco use and thoughts about quitting, other health behaviors, physical and mental health status, health care access and utilization, and technology and Internet usage?

Methods

Source of Data

Data used in these analyses were from the National Cancer Institute's Health Information National Trends Survey 4 (HINTS 4), Cycle 4. This mailed survey was conducted between August and November 2014. It targeted noninstitutionalized individuals living in the United States aged 18 years or older using two-stage stratified random sampling. The first stage involved selection of a stratified sample of households from a marketing list of nonvacant mailing addresses, and the second stage involved sampling of one adult from each sampled household using the next birthday method. Stratification was designed to overenroll Hispanic and African American participants to increase the precision of estimates for these minority groups. Both English and Spanish versions of the survey were used.

Survey mailings followed a modified Dillman procedure [13]. There were four possible mailings. All selected households received the first survey along with a US $2 incentive, followed by a reminder postcard. Nonrespondents received up to two additional survey mailings. Households flagged as being potentially Spanish speaking received both an English- and Spanish-language version of the survey. The overall weighted response rate was 34.44% (3677/13,996 complete, returned surveys). Additional details regarding the sampling methods and survey procedures can be found in the HINTS 4 database user documentation [14].

Measures

Definitions of Current Smoking and Smartphone Ownership

Participants were classified as current smokers if they reported smoking 100 or more cigarettes in their lifetime and, in response to the item "How often do you now smoke cigarettes?" responded "every day" or "some days." Smartphone ownership was assessed with the item, "Please indicate if you have a smartphone, such as an iPhone, Android, Blackberry, or Windows phone." Of the 3677 respondents, 498 (15.00%) were

classified as current smokers, and 479 (97.4%) of these provided information on smartphone ownership.

Demographics

Demographic survey items included age, gender, current occupational status, household income, marital status, highest grade or level of schooling completed, race, and ethnicity. The US Department of Agriculture's 2013 Rural-Urban Continuum Code was used to classify participants as residing either in metropolitan (codes 1-3) or nonmetropolitan (codes 4-9) areas.

Tobacco Use and Thoughts About Quitting

Respondents were asked whether they stopped smoking for one day or longer in the past year because they were trying to quit and whether they were seriously considering quitting in the next 6 months. They were also asked how much they thought quitting smoking would help reduce the harmful effects of smoking, with response options ranging from "not at all" to "a lot."

Other Health Behaviors

Assessment of current health behaviors included minutes per week of at least moderate exercise (the product of two questions: "In a typical week, how many days do you do any physical activity or exercise of at least moderate intensity, such as brisk walking, bicycling at a regular pace, and swimming at a regular pace" and "On the days that you do any physical activity or exercise of at least moderate intensity, how long do you typically do these activities?"), which was transformed to a binary categorical variable indicating whether or not the respondent met the recommended minimum weekly moderate-intensity exercise duration of 150 minutes [15]. Fruit and vegetable consumption, in total cups per day ("About how many cups of fruit [including 100% pure fruit juice] do you eat or drink each day?" and "About how many cups of vegetables [including 100% pure vegetable juice] do you eat or drink each day?"), were assessed separately as a range, transformed to a continuous variable by taking the midpoint of the range, and combined for analysis. Several items assessed health behavior change efforts in the prior year (ie, "At any time in the past year, have you *intentionally* tried to..."), including attempts to change physical activity and fruit and vegetable consumption and to lose weight.

Physical and Mental Health Status

Body mass index (BMI) was calculated from the respondent's self-reported height and weight and used to classify participants as overweight or obese (BMI ≥25.0 kg/m^2). Physical health status was assessed, in part, by asking whether the respondent had ever been diagnosed with diabetes, hypertension, a heart condition, lung disease, arthritis, and cancer. In terms of mental health, the four-item Patient Health Questionnaire (PHQ-4) [16] was included as a measure of psychological distress. We separated this measure into its component screening tools for depression (PHQ-2) [17] and anxiety (Generalized Anxiety Disorder scale [GAD]-2) [18]. Respondents were also asked whether they had ever been diagnosed with a depressive or anxiety disorder.

Health Care Access and Utilization

Respondents indicated whether or not they had health care coverage as well as how many times they received nonemergent medical care over the past year.

Use of Technology and the Internet

Items included whether or not the respondent used the Internet, device ownership, use of health apps, and which technology-mediated methods they have used to exchange medical information with a health care provider.

Statistical Analyses

Statistical analyses incorporated weights for survey respondents in order to adjust for nonresponse and noncoverage biases, thus ensuring valid inference. Sample weights and replicate weights were calculated as described in the HINTS 4 Cycle 4 Methodology Report [14]. Briefly, full-sample weights were derived by first adjusting for household-level base weights (ie, the reciprocal of the probability of selecting the household for the survey) and household nonresponse, calculating initial weights, and finally calibrating these to population counts. Replicate weights were calculated using the delete one jackknife procedure. We estimated population-level proportions of smartphone ownership and differences between smartphone owners and nonowners on demographics, tobacco use and thoughts about quitting, other health behavior, physical and mental health status, health care access, and use of technology using full-sample weights. Thus, the weighted percentages do not directly correspond with the raw sample proportions. For group comparisons, we implemented chi-square tests of independence and *t* tests using replicate weights. Because a main aim of the study was to better understand characteristics of smokers who own smartphones for intervention planning purposes rather than to identify the set of variables that most efficiently distinguish smartphone-owning smokers from nonowners, we decided against a multivariable approach. However, we accounted for multiple comparisons by adjusting *P* values to control the false discovery rate (FDR) [19].

Results

Proportion of Current Smokers in the United States Who Own Smartphones

Overall, the weighted proportion of current smokers who owned smartphones was 63.8% (248/479). Of those who did not own smartphones (231/479), 53.6% (128/231) owned a basic cell phone. Among smartphone owners, 36.7% (88/245) reported using a health app on their phone or tablet.

Demographic Characteristics of Smokers Associated with Smartphone Ownership

As shown in Table 1, smartphone ownership among smokers differed according to age, education, income, employment, and marital status. There was decreasing prevalence of smartphone ownership from the 18 to 34 years age group (88%, 59/68) through the 65 years and older age group (15%, 17/85; $P<.001$). Increasing levels of education ($P=.002$) and income ($P<.001$) also showed nearly linear increases in smartphone ownership. More than 84% (76/112) of respondents with a college degree

or higher and more than 90% (63/80) of those with household incomes of at least US $75,000 per year owned smartphones. In terms of employment status, employed respondents had the highest rates of ownership (74.1%, 155/224; *P*=.002). Those who were divorced, widowed, or separated (36.3%, 75/171) had lower rates of ownership than those who were married (59.0%, 90/165) or never married (77.3%, 71/122; *P*=.002). Smartphone ownership did not differ by gender, race/ethnicity, or residence in a metropolitan versus nonmetropolitan area.

Table 1. Prevalence of smartphone ownership among smokers by demographic characteristics (n=479).

Demographic characteristic	Prevalence of smartphone ownership, n (%)[a]	*P* [b]
Age (n=457)		<.001
18-34 (n=68)	59 (88)	
35-49 (n=92)	64 (67)	
50-64 (n=212)	114 (47.9)	
≥65 (n=85)	17 (15)	
Gender (n=448)		.93
Female (n=265)	144 (64.6)	
Male (n=183)	88 (64.0)	
Education (n=458)		.002
Less than high school (n=57)	19 (38)	
High school graduate (n=116)	46 (55.1)	
Some college (n=173)	95 (66.8)	
College grad or more (n=112)	76 (84.1)	
Race/ethnicity (n=419)		.93
Non-Hispanic white (n=255)	138 (66.1)	
Non-Hispanic black (n=89)	50 (67)	
Hispanic (n=47)	24 (61)	
Non-Hispanic other (n=28)	16 (78)	
Income (US$) (n=475)		<.001
0-14,999 (n=140)	44 (45.4)	
15,000-34,999 (n=124)	60 (52.9)	
35,000-74,999 (n=131)	80 (64.8)	
75,000-99,000 (n=34)	28 (93)	
≥100,000 (n=46)	35 (90)	
Residence (n=479)		.07
Metro (n=409)	216 (68.1)	
Non-metro (urban or rural) (n=70)	32 (47)	
Employment (n=445)		.002
Employed (n=224)	155 (74.1)	
Unemployed (n=41)	20 (60)	
Retired (n=76)	19 (31)	
Disabled (n=76)	24 (40)	
Other (student, homemaker, other) (n=28)	16 (50)	
Marital status (n=458)		.002
Married/living as married (n=165)	90 (59.0)	
Divorced/widowed/separated (n=171)	75 (36.3)	
Never married (n=122)	71 (77.3)	

[a]Percentage values in table are weighted by the overall sample weight.

[b]*P* values were calculated using chi-square tests for independence on the weighted percentages.

Table 2. Differences between smartphone-owning and nonowning smokers on health behaviors, health status, health care access and utilization, and Internet and technology use.

Variable	Smartphone owners[a] (n=248)	Nonowners[a] (n=231)	*P* [b]
Tobacco use behavior and attitudes, n (%)			
Daily smoker (n=479)	173 (68.5)	192 (80.3)	.20
Made quit attempt in past year (n=476)	159 (59.5)	136 (55.6)	.71
Considering quitting in next 6 months (n=472)	167 (63.4)	136 (57.8)	.51
Believe that quitting reduces harm of smoking “a lot” (n=476)	189 (83.7)	147 (62.0)	.04
Other health behaviors			
Gets recommended ≥150 min/week of moderate exercise, n (%) (n=474)	115 (54.8)	78 (43.6)	.30
Total cups of fruit and vegetable consumption per day, mean (SD) (n=470)	2.7 (0.2)	2.0 (0.2)	.03
Tried to increase exercise in past year, n (%) (n=465)	134 (59.8)	68 (26.2)	.001
Tried to change weight in past year, n (%) (n=466)	138 (55.5)	83 (31.9)	.02
Tried to increase fruit or veg consumption in past year, n (%) (n=459)	123 (42.6)	81 (38.9)	.72
Physical and mental health status, n (%)			
Overweight/obese (BMI ≥25.0) (n=463)	149 (63.3)	135 (54.9)	.47
Positive screen for current depression (n=463)	52 (12.9)	66 (37.5)	.02
Positive screen for current anxiety (n=460)	51 (17.2)	57 (32.7)	.10
Ever diagnosed with depression/anxiety disorder (n=466)	77 (22.0)	97 (45.1)	.04
Ever diagnosed with diabetes (n=466)	32 (7.3)	58 (21.3)	.02
Ever diagnosed with of hypertension (n=468)	96 (24.6)	117 (39.3)	.02
Ever diagnosed with heart condition (n=468)	21 (5.1)	32 (9.6)	.30
Ever diagnosed with lung disease (n=467)	49 (13.4)	56 (22.3)	.23
Ever diagnosed with arthritis (n=466)	65 (17.8)	93 (34.5)	.04
Ever diagnosed with cancer (n=477)	25 (4.5)	26 (7.4)	.30
Health care access and utilization, n (%)			
Have health care coverage (n=478)	192 (78.2)	184 (84.6)	.32
Visited health care provider in past year (excluding ER) (n=469)	187 (69.8)	187 (78.5)	.36
Technology and Internet usage, n (%)			
Use Internet (n=479)	224 (95.1)	127 (52.5)	.001
Have tablet (n=479)	139 (59.0)	47 (24.6)	.001
Exchanged medical information by email with a provider (n=474)	69 (24.7)	22 (6.6)	.03
Exchanged medical information by text with a provider (n=474)	22 (5.7)	15 (4.8)	.72
Exchanged medical information by app with a provider (n=474)	19 (6.6)	7 (3.9)	.51
Exchanged medical information by video conference with a provider (n=474)	2 (0.4)	5 (2.1)	.36
Exchanged medical information by social media with a provider (n=474)	15 (4.4)	8 (4.3)	.98

[a]Percentage values as well as mean and standard deviation in table are weighted by the overall sample weight.

[b]*P* values were calculated using chi-square tests for independence on the weighted percentages.

Differences by Smartphone Ownership in the Characteristics and Behaviors of Smokers

Comparisons of smokers who own smartphones and those who do not are provided in Table 2.

Tobacco Use and Thoughts About Quitting

Smartphone owners did not differ significantly from nonowners on daily versus nondaily smoking, recent quit attempts, or plans to quit in the next 6 months. They did endorse greater belief that quitting smoking reduces smoking-related harm (*P*=.04).

Other Health Behaviors

Smartphone owners reported higher levels of fruit and vegetable consumption (2.7 vs 2.0 cups per day, *P*=.03). A majority had, in the past year, tried to increase exercise (59.8%, 134/240) and lose weight (55.5%, 138/241), which was significantly more common than among those smokers who did not own smartphones (26.2%, 68/225 tried to increase exercise and 31.9%, 83/225 tried to lose weight; *P*=.001 and *P*=.02, respectively). The two groups did not differ on adherence to the recommendation of 150 or more minutes per week of moderate exercise or on attempts to increase fruit and vegetable consumption in the past year.

Physical and Mental Health Status

Despite the aforementioned finding that smartphone owners reported higher levels of fruit and vegetable consumption and were more likely to have engaged in recent efforts to increase exercise and lose weight, smartphone owners were no less likely to be overweight or obese than those smokers who did not own smartphones (63.3%, 149/243 vs 54.9%, 135/220; *P*=.47). Smartphone owners also reported lower rates of some, but not all, physical and mental health problems. They were less likely to screen positive for depression on the PHQ-2 (12.9%, 52/241 vs 37.5%, 66/222; *P*=.02) and less likely to report ever having been diagnosed with a depressive or anxiety disorder (22.0%, 77/242 vs 45.1%, 97/224; *P*=.04), diabetes (7.3%, 32/244 vs 21.3%, 58/222; *P*=.02), hypertension (24.6%, 96/244 vs 39.3%, 117/224; *P*=.02), and arthritis (17.8%, 65/242 vs 34.5%, 93/224; *P*=.04). They did not differ on current anxiety or on lifetime diagnosis of lung disease, heart condition, or cancer.

Health Care Access and Utilization

Smokers with smartphones were no more likely to have health care coverage (78.2%, 192/248) than those without smartphones (84.6%, 184/230; *P*=.32). They were also equally likely to have received nonemergency medical care in the prior year (69.8%, 187/244 vs 78.5%, 187/225; *P*=.36).

Technology and Internet Usage

Nearly all (95.1%, 224/248) smartphone owners reported using the Internet, whereas closer to half (52.5%, 127/231) of those who did not own smartphones reported Internet use (*P*=.001). Smartphone owners were more than twice as likely to own tablet computers (59.0%, 139/248 vs 24.6%, 47/231; *P*=.001). In terms of technology-mediated communication with health care providers, smartphone owners were more likely to have communicated via email (24.7%, 69/247 vs 6.6%, 22/227; *P*=.03) with their provider. There were no differences in communication with a provider by text, app, video conference, or social media, but overall prevalence of use of these methods of communication was very low in both groups (0%-7%).

Discussion

This was the first study to use a nationally representative sample to estimate the prevalence of smartphone ownership among current smokers in the United States and to evaluate the relationship between smartphone ownership and demographics, tobacco use and thoughts about quitting, other health behaviors, physical and mental health, health care access, and Internet and technology utilization. Regarding the prevalence of smartphone ownership among smokers, the 63.8% weighted estimate for this late-2014 HINTS 4 sample of adult current smokers is nearly identical to the 64% rate of smartphone ownership for the US population in late 2014 reported by the Pew Research Center [12]. Where smartphone ownership data specific to smokers are not available, this finding suggests that it is reasonable to assume that population estimates of smartphone ownership among smokers follow the broader population trends. If that is the case, current rates of smartphone ownership among smokers should be approximately 77% based on the 2017 Pew survey results [9].

Considering these ownership data, smartphone apps for smoking cessation have high potential reach. Although the HINTS survey does not assess usage of smoking cessation apps specifically, it does assess use of health apps more broadly. We found that over one-third of smartphone-owning smokers (36.7%) had ever used a health app. The latest Pew survey results on mobile health app use in the general population, which were reported in 2012, found that 19% of adults reported having health apps on their phones [20]. A more recent survey showed much higher rates health app use (58%) [21]. These findings point toward an increase in health app usage over time. Only one prior study has specifically examined cessation app use among smokers, finding that 15% of US adult smokers had ever used a cessation app and 43% were interested in using an app in the future [22]. Taken together with our finding that a high proportion of the 42 million US smokers [23] own smartphones, these data on the usage of and interest in health apps more broadly, and cessation apps more specifically, indicate that smartphone apps are a promising new approach to assist millions of smokers by expanding access to smoking cessation interventions.

As of 2014, smartphone ownership in the broader US population was associated with younger age, higher educational attainment, and higher income, but not with race or ethnicity [12]. Data from this study mirror these general population trends in smartphone ownership, where the younger smokers and those of higher socioeconomic status are more likely to have access to the technology. With the notable exception of the lack of differences by race or ethnicity, this pattern of findings on the demographics of smartphone ownership is reminiscent of the “digital divide” that emerged in the early days of the Internet [24], restricting which smokers could access Web-assisted tobacco treatment [25]. There is reason to believe that this digital health divide will recur with each new technological advancement, perpetuating disparities in treatment access and necessitating consideration of how to best reach those who remain on the other side of the divide.

Traditional modalities such as face-to-face counseling and telephone quitline counseling are effective yet underutilized alternatives [26] to smartphone-delivered interventions. Other technology-driven methods could also be employed to reach smokers who do not own smartphones. For example, both text messaging and Web-based interventions are effective for smoking cessation [27,28]. Within the group of smokers who did not own smartphones, 53.6% owned a basic cellular phone

and 52.5% reported using the Internet, meaning that at least half could access either a text messaging or Web-based program.

Undoubtedly, the demographics of smartphone ownership will continue to change over time and, even now, a substantial proportion of disadvantaged smokers have access to smartphone-delivered treatments. In this HINTS 4 sample, nearly half (45.4%) of smokers in the lowest category of income (US $0-$14,999) reported owning a smartphone. Given the low cost and high accessibility of apps for smokers who own a smartphone, this method of treatment delivery offers many potential benefits for disadvantaged populations of smokers.

Demography is just one facet of understanding the smartphone divide and its implications for treatment development and accessibility. We also evaluated the possibility that smartphone owners differed from nonowners on tobacco use and thoughts about quitting, other health behaviors, physical and mental health, health care access, and technology utilization. The prevalence of daily tobacco use as well as past and planned efforts to quit were similar among smartphone owners and nonowners, although smokers who did not own smartphones expressed less optimism about the health benefits of quitting compared to smokers who did own smartphones. This may be related to the older age and worse mental and physical health reported by this group, including higher rates of depressive symptoms and diagnoses of depression or anxiety disorders, diabetes, hypertension, and arthritis. The clustering of physical and mental health conditions within the group of smokers without smartphones makes this a more challenging population of smokers to treat. As the smartphone-specific digital divide lessens over time, the answer to the question of how effective smartphone apps are for smokers with physical and mental health conditions will become increasingly important. It is also important that tobacco treatment be readily accessible to smokers with physical and mental health conditions through other means, or tobacco-related health disparities will continue to worsen among these vulnerable populations [29].

Surprisingly, we did not observe a divide in health care access between smartphone owners and nonowners. This finding differs from that of an earlier study investigating the digital divide in Web-based tobacco cessation interventions, where those smokers who did not have Internet access were also less likely to have health care access [25]. With the expanded coverage offered by the Affordable Care Act, a majority of smokers should be able to access one or more effective forms of assistance to quit smoking through their health insurance, and those who do not have health care coverage can still access no-cost assistance through a tobacco quitline or websites such as Smokefree.gov. Smoking cessation apps, if proven effective, could expand the safety net of no- or low-cost standalone cessation assistance for the estimated 22% of smartphone-owning smokers who do not have health insurance.

Implications for Future Research on Smartphone App Design

The observed characteristics of smokers who own smartphones raise a number of questions about the optimal design of cessation programs delivered on this platform. First, we found that smartphone owners were more health-conscious with respect to their greater consumption of fruits and vegetables as well as their greater efforts to increase exercise and lose weight in the previous year, suggesting potential synergies between motivation to change smoking, exercise, and nutrition. Although attempting to lose weight while quitting smoking is typically discouraged [26], exercise and proper nutrition can assist smokers in minimizing weight gain after quitting and may be of interest for smokers using smartphone apps to quit. The extent to which smokers would engage with and benefit from a multiple health behavior change app, as opposed to one that focused exclusively on smoking cessation, is an important topic for future research.

Second, smartphone technology offers a medium for synchronous or asynchronous communication with health care providers to support cessation efforts. Email was used by almost one-quarter of smartphone owners as a method of communicating with providers. This is a substantial proportion, particularly given that the option of secure messaging with providers is not available universally. On the other hand, very few participants used apps, text messaging, video conferencing, or social media for this type of communication. It cannot be determined from these data how many smokers would use these methods to communicate with providers specifically about smoking cessation if that option were available, but smokers indicate that they like the idea of having some form of support built into smoking cessation apps [30], and email or other messaging components within an app provides smokers with supportive accountability for behavior change [31]. Smokers' interest in communicating with health care providers through cessation apps should be evaluated further in order to identify the most desirable, effective, and secure methods of communication. Additionally, given the high proportion of smartphone-owning smokers in this study who had health insurance (78%) and who had visited a health care provider in the past year (70%), the extent to which cessation apps could be built for integration into the health care system should be evaluated. Such integration could have a number of benefits, including: (1) provider support for use of the programs, which increases adherence by 10-30% [32]; (2) integration of app data into electronic medical records systems for monitoring of symptoms, patient self-management strategies, and treatment response; and (3) offering apps for behavioral support alongside cessation medications to support quitting, which is consistent with current clinical practice guidelines [26] and has the potential to increase adherence to pharmacotherapy [33].

Limitations

The primary limitation of this analysis stems from the challenge of keeping pace with the speed of technology advancement in population surveys. The HINTS 4, Cycle 4 data were collected in 2014, and changes in smartphone ownership since that time may affect the demographics of smartphone ownership and comparisons between smartphone owners and nonowners. Additionally, the HINTS survey is conducted by the US National Cancer Institute and focuses on US residents; therefore, the results are generalizable only to smokers in the United States. The prevalence and correlates of smartphone ownership among smokers is likely to vary considerably across countries. As described in the Methods section, we accounted for multiple comparisons by controlling the FDR [19], or the expected

proportion of falsely rejected null hypotheses. The FDR is equivalent to the familywise error rate (FWER) when all null hypotheses are true, but it is smaller otherwise. Using the FDR provides a potential gain in power where control of the conservative FWER is unnecessarily stringent [19]. Still, due to the small samples sizes for some comparisons and the *P* value adjustment, results are conservative. Lack of statistical significance should therefore not be interpreted as a conclusive demonstration of no effect and, depending on the context in which they are applied, differences that are not statistically significant may still have practical significance. As such, this research is best characterized as exploratory. Finally, our analyses focus on bivariate correlations between constructs measured at a single time point; thus, we cannot evaluate the effects of time or rule out confounding factors that influence bivariate correlations.

Conclusions

Smartphone ownership among US smokers mirrors many trends in the general population, including the overall rate of ownership and the association with younger age and higher socioeconomic status. Smokers who own smartphones are also healthier, more health-conscious, and are higher users of technology and the Internet than those who do not own smartphones. Design of smartphone-delivered cessation interventions would benefit from additional research on the implications of these user characteristics and behavior, including smartphone owners' interest in multiple health behavior change and interest in communicating with health care providers via technology. These data also highlight the importance of continuing to offer a broad range of intervention strategies that do not require smartphones for access in order to reach the smokers with the highest physical and mental health burden and prevent worsening of tobacco-related health disparities.

Acknowledgments

This work was supported by the Fred Hutchinson Cancer Research Center. The funder had no role in the study design; collection, analysis, and interpretation of the data; the writing of this report; or the decision to submit this manuscript for publication.

Conflicts of Interest

Fred Hutchinson Cancer Research Center holds a patent on the SmartQuit app for smoking cessation.

References

1. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214 [FREE Full text] [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
2. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 1;143:87-94. [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
3. Bricker JB, Copeland W, Mull KE, Zeng EY, Watson NL, Akioka KJ, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2017 Jan 01;170:37-42. [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
4. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
5. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
6. Abroms LC, Lee WJ, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
7. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031. [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
8. Ubhi HK, Michie S, Kotz D, van Schayck OC, Selladurai A, West R. Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features. Transl Behav Med 2016 Sep;6(3):410-417 [FREE Full text] [doi: 10.1007/s13142-015-0352-x] [Medline: 27528530]
9. Pew Research Center. 2017 Jan 12. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2017-05-01] [WebCite Cache ID 6q8kZXCYP]
10. IDC. 2016 Apr 27. Worldwide smartphone growth goes flat in the first quarter as Chinese vendors churn the top 5 vendor list, according to IDC URL: http://www.idc.com/getdoc.jsp?containerId=prUS41216716 [accessed 2017-08-03] [WebCite Cache ID 6sRuLgSzQ]

11. Jamal A, King BA, Neff LJ, Whitmill J, Babb SD, Graffunder CM. Current cigarette smoking among adults - United States, 2005-2015. MMWR Morb Mortal Wkly Rep 2016 Nov 11;65(44):1205-1211 [FREE Full text] [doi: 10.15585/mmwr.mm6544a2] [Medline: 27832052]
12. Smith A. Pew Research Center. 2015 Apr 01. US smartphone use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2017-05-01] [WebCite Cache ID 6YJ3lPbC4]
13. Dillman D, Smyth J, Christian L. Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method. Hoboken, NJ: John Wiley & Sons; 2009.
14. Westat. Health Information National Trends Survey 4 (HINTS 4), Cycle 4 Methodology Report. Bethesda, MD: National Cancer Institute; 2015 Feb. URL: https://hints.cancer.gov/docs/HINTS_4_Cycle_4_Methodology_Report.pdf [accessed 2017-08-03] [WebCite Cache ID 6sRvB1igR]
15. United States Department of Health & Human Services. 2008 Physical Activity Guidelines for Americans: Be Active, Healthy, and Happy. Washington, DC: United States Department of Health and Human Services; 2008.
16. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]
17. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
18. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [Medline: 17339617]
19. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 1995;57:289-300.
20. Fox S, Duggan M. Pew Research Center. 2012 Nov 08. Mobile health 2012 URL: http://pewinternet.org/Reports/2012/Mobile-Health.aspx [WebCite Cache ID 6Ki9MnKwz]
21. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
22. Borrelli B, Bartlett YK, Tooley E, Armitage CJ, Wearden A. Prevalence and frequency of mHealth and eHealth use among US and UK smokers and differences by motivation to quit. J Med Internet Res 2015 Jul 04;17(7):e164 [FREE Full text] [doi: 10.2196/jmir.4420] [Medline: 26149323]
23. United States Department of Health and Human Services. The Health Consequences of Smoking--50 Years of Progress. A Report of the Surgeon General. Atlanta, GA: Office on Smoking and Health, Centers for Disease Control and Prevention, US Department of Health and Human Services; 2014.
24. 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 [FREE Full text] [Medline: 11192412]
25. Stoddard JL, Augustson EM. Smokers who use internet and smokers who don't: data from the Health Information and National Trends Survey (HINTS). Nicotine Tob Res 2006 Dec;8 Suppl 1:S77-S85. [Medline: 17491174]
26. Treating Tobacco Use and Dependence: 2008 Update. Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services, Public Health Service; 2008.
27. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
28. Graham AL, Carpenter KM, Cha S, Cole S, Jacobs MA, Raskob M, et al. Systematic review and meta-analysis of Internet interventions for smoking cessation among adults. Subst Abuse Rehabil 2016;7:55-69 [FREE Full text] [doi: 10.2147/SAR.S101660] [Medline: 27274333]
29. Williams JM, Steinberg ML, Griffiths KG, Cooperman N. Smokers with behavioral health comorbidity should be designated a tobacco use disparity group. Am J Public Health 2013 Sep;103(9):1549-1555 [FREE Full text] [doi: 10.2105/AJPH.2013.301232] [Medline: 23865661]
30. McClure JB, Heffner J, Hohl S, Klasnja P, Catz SL. Design considerations for mHealth programs targeting smokers not yet ready to quit: results of a sequential mixed-methods study. JMIR Mhealth Uhealth 2017 Mar 10;5(3):e31 [FREE Full text] [doi: 10.2196/mhealth.6845] [Medline: 28283465]
31. 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]
32. Aitken M, Lyle J. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2015. URL: http://www.imshealth.com/files/web/IMSH%20Institute/Reports/Patient%20Adoption%20of%20mHealth/IIHI_Paient_Adoption_of_mHealth.pdf [WebCite Cache ID 6t6izD0xo]
33. McClure JB, Anderson ML, Bradley K, An LC, Catz SL. Evaluating an adaptive and interactive mHealth smoking cessation and medication adherence program: a randomized pilot feasibility study. JMIR Mhealth Uhealth 2016 Aug 03;4(3):e94 [FREE Full text] [doi: 10.2196/mhealth.6002] [Medline: 27489247]

Abbreviations

BMI: body mass index
FDR: false discovery rate
FWER: familywise error rate
GAD: generalized anxiety disorder
HINTS: Health Information National Trends Survey
PHQ: Patient Health Questionnaire

Edited by G Eysenbach; submitted 01.05.17; peer-reviewed by A Graham, T Radtke; comments to author 15.06.17; revised version received 28.06.17; accepted 14.07.17; published 31.08.17.

Please cite as:
Heffner JL, Mull KE
Smartphone Ownership Among US Adult Cigarette Smokers: 2014 Health Information National Trends Survey (HINTS) Data
J Med Internet Res 2017;19(8):e305
URL: https://www.jmir.org/2017/8/e305/
doi:10.2196/jmir.7953
PMID:28860108

Original Paper

Parent-Mediated Intervention Training Delivered Remotely for Children With Autism Spectrum Disorder Living Outside of Urban Areas: Systematic Review

Dave Parsons[1*], BSc (OT), MBA; Reinie Cordier[1*], PhD; Sharmila Vaz[1*], BSc (OT), MCISc (OT), PhD; Hoe C Lee[1*], PostGradDip (OT), MPhil, PhD

School of Occupational Therapy and Social Work, Curtin University, Perth, Australia

[*]all authors contributed equally

Corresponding Author:
Dave Parsons, BSc (OT), MBA
School of Occupational Therapy and Social Work
Curtin University
Building 401
Kent St, Bentley
Perth, 6102
Australia
Phone: 61 (08) 92663790
Fax: 61 (08) 9266 3636
Email: dave.parsons@curtin.edu.au

Abstract

Background: Parent training programs for families living outside of urban areas can be used to improve the social behavior and communication skills in children with autism spectrum disorder (ASD). However, no review has been conducted to investigate these programs.

Objective: The aim of this study was to (1) systematically review the existing evidence presented by studies on parent-mediated intervention training, delivered remotely for parents having children with ASD and living outside of urban areas; (2) provide an overview of current parent training interventions used with this population; (3) and provide an overview of the method of delivery of the parent training interventions used with this population.

Methods: Guided by the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement, we conducted a comprehensive review across 5 electronic databases (CINAHL, Embase, ERIC, PsycINFO, and Pubmed) on July 4, 2016, searching for studies investigating parent-mediated intervention training for families living outside of urban centers who have a child diagnosed with ASD. Two independent researchers reviewed the articles for inclusion, and assessment of methodological quality was based on the Kmet appraisal checklist.

Results: Seven studies met the eligibility criteria, including 2 prepost cohort studies, 3 multiple baseline studies, and 2 randomized controlled trials (RCTs). Interventions included mostly self-guided websites: with and without therapist assistance (n=6), with training videos, written training manuals, and videoconferencing. Post intervention, studies reported significant improvements (P<.05) in parent knowledge (n=4), parent intervention fidelity (n=6), and improvements in children's social behavior and communication skills (n=3). A high risk of bias existed within all of the studies because of a range of factors including small sample sizes, limited use of standardized outcome measures, and a lack of control groups to negate confounding factors.

Conclusions: There is preliminary evidence that parent-mediated intervention training delivered remotely may improve parent knowledge, increase parent intervention fidelity, and improve the social behavior and communication skills for children with ASD. A low number of RCTs, difficulty in defining the locality of the population, and a paucity of standardized measures limit the generalization of the findings to the target population. Future studies should investigate the appropriateness and feasibility of the interventions, include RCTs to control for bias, and utilize standard outcome measures.

(J Med Internet Res 2017;19(8):e198) doi:10.2196/jmir.6651

KEYWORDS

Autistic disorder; Internet; parents; rural health services; telemedicine

Introduction

Autism spectrum disorders (ASD) are characterized by deficits in social communication and social behavior, including problems interpreting nonverbal gestures, difficulty developing age-appropriate friendships, adherence to rigid routines, and adapting to environmental change [1,2]. In recent years there has been a marked increase in the prevalence of ASD in children with possible reasons cited including (1) advancement in diagnostic procedures, (2) broadening of the diagnosis criteria, (3) increase awareness of ASD, (4) previous diagnosis, and (5) recognition that ASD is a lifelong condition [3]. The current prevalence rate of ASD ranges from 20 per 10,000 to as high as 110 per 10,000 of the global population [3-10].

The increasing prevalence of ASD exerts major demands on early intervention services and education institutions resulting in calls for innovative service delivery models and methods [11,12]. Limited access to adequate health services and a shortage of adequately trained early intervention health and education professionals are of particular concern in regional and remote areas [13-16]. Families of children with ASD who live in regional and remote areas often experience several barriers to improving the outcomes for their child [17], including (1) increased travel distance to suitably qualified clinicians for effective therapy services, (2) delayed diagnosis due to reduced screening programs, and (3) challenges from the inconsistency of health professionals due to high attrition rates and high workforce transition [13,14,16,18,19]. These challenges highlight the need for innovative and alternative early intervention methods for children with ASD and living outside of urban areas.

Effective early intervention requires skilled health and education professionals and places an increased financial and time burden on families to access services [20,21]. As a result, parents or caregivers may be required to play a larger role in the provision of therapeutic services for their children with ASD [22]. To help overcome these barriers, parents can become active agents in the therapeutic process with the appropriate training and ongoing guidance, thereby delivering these interventions to their children in a more consistent manner [23]. This is particularly pertinent for families living outside of urban areas where there is often a lack of access to suitably trained clinicians.

The rise of technological advances in information communication technology (ICT) has paved the way for alternative modes of delivery for health interventions. Evidence suggests that services provided by health professionals using ICT have high efficacy in areas of health, such as delivering behavioral treatment for people with anxiety and depression [24,25]. Moreover, evidence for using telehealth and ICT for children and adolescents with ASD is emerging, with preliminary findings suggesting that it has potential benefits in the diagnosis and delivery of interventions with this population [11,26-32].

Systematic literature reviews support the use of parent-mediated interventions in children with ASD [23,33,34], as does the use of telehealth in providing education sessions to parents or caregivers who have a child with ASD [26,35]. No systematic review has been published on parent-mediated interventions for families having a child with ASD and living outside of urban areas. This is a unique population, and whereas similarities may exist between this group and the general population, these cannot be generalized due to distinctive characteristics and the barriers these families experience due to remoteness [22].

Evidence suggests that the characteristics of families having a child with ASD and living outside of urban areas are unique; however, categorizing and comparing populations across countries is challenging because of differing definitions and classifications systems. For example, in Canada, all territories outside of an urban area are considered to be rural. Rural areas include those "...having a population of at least 1000 and a density of 400 or more people per square kilometer..." [36]. Similarly, according to the US Census Bureau, rural areas include all population, housing, and territory not contained within an urban area [37]. However, the Australian Bureau of Statistics (ABS) uses a 5-category classification based on the Australian standard geographical classification system (ASGC) [38]. The categories include major cities, inner regional, outer regional, remote, and very remote based on a number of variables including population size and distance by road to service centers.

The purpose of this systematic review was to review the existing evidence for parent-mediated intervention training delivered remotely for parents having a child with ASD and living outside of urban areas. In doing so, this review will (1) provide an overview of the studies involving the use of parent-mediated intervention training delivered remotely to parents who have a child with ASD, (2) provide an overview of current parent training programs used with this population, and (3) provide an overview of the method of delivery of parent training interventions used with this population.

Methods

Protocol and Registration

The systematic review was registered with PROSPERO (registration number CRD42015027300). The preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement guided the methodology and reporting of this systematic review. The statement provides the structure and transparency considered necessary for reporting systematic reviews in areas of health care.

Eligibility Criteria

Participants needed to be parents or caregivers of children diagnosed with ASD. With the recent update to the diagnostic and statistical manual of mental disorders (DSM-V), inclusion criteria were expanded to include participants whose children had a diagnosis of autism, Asperger's syndrome, or pervasive developmental disorder not otherwise specified under criteria of the previous DSM-IV [39]. Studies were included if the children with ASD were aged under 18 years. Given the discrepancies in definition and classification of urban-rural location between countries, for the purpose of this review, we included only those studies in which the population resided

outside of major cities or urban areas and authors explicitly described participants as having limited access to services.

Articles were included if the intervention involved training the parents or caregivers in intervention skills to improve the social behavior and communications skills for their child with ASD using telehealth (remote delivery) methods. Face-to-face training, which required parents to travel to a center for training were excluded. Studies were excluded if training was provided solely to therapy professionals or teachers. Telehealth interventions delivered directly and solely by clinicians were excluded from the review, as one study explicitly addressing this issue already exists [26]. Various modes of delivery were accepted for inclusion, including, DVDs, videoconferencing, and Web-based content, as long as the method of delivery enabled remote delivery. Articles of any methodological design that met the eligibility criteria were included, as long as they were published in English in International Scientific Indexing (ISI) listed scientific journals.

Information Sources

To identify eligible studies, the authors conducted a comprehensive systematic search across 5 electronic databases on July 4, 2016. Databases searched included (1) Cumulative Index to Nursing and Allied Health Literature (CINAHL), (2) Embase, (3) Education Resources Information Center (ERIC), (4) PsycINFO, and (5) Pubmed.

Table 1. Search terms.

Database and Search terms	Limitations	Number of abstracts
Subject Headings:		
CINAHL: (MH "autistic disorder") or (MH "child development disorders, pervasive") or (MH "pervasive developmental disorder—not otherwise specified"), and (MH "rural health centers") or (MH "hospitals, rural") or (MH "rural population") or (MH "rural health services") or (MH "rural areas") or (MH "services for Australian rural and remote allied health") or (MH "rural health") or (MH "rural health personnel") or (MH "telehealth") or (MH "telemedicine") or (MH "videoconferencing") or (MH "teleconferencing")	English language	80
ERIC: SU.EXACT("Asperger syndrome") or SU.EXACT("pervasive developmental disorders") or SU.EXACT("autism"), and SU.EXACT("rural population") or SU.EXACT("rural areas") or SU.EXACT("rural youth") or SU.EXACT("rural environment") or SU.EXACT("rural education") or SU.EXACT("teleconferencing") or SU.EXACT("telecourses") or SU.EXACT("videoconferencing") or SU.EXACT("telecommunications")	English language	29
Embase: autism or Asperger syndrome or "pervasive developmental disorder not otherwise specified," and (rural health care or rural area or urban rural difference or rural population) or (teleconsultation or telediagnosis or telehealth or telemedicine or telemonitoring or teletherapy or videoconferencing or teleconference or health care delivery)	English language	406
PsycINFO: autism or pervasive developmental disorders or Aspergers syndrome, and (exp rural environments or distance education) or (telemedicine or computer mediated communication or telecommunications media)	English language	64
PubMed **:** ("Autistic disorder" [Mesh] or "child development disorders, pervasive" [Mesh]) and ("rural population" [Mesh] or "rural health services" [Mesh] or "rural health" [Mesh] or "remote consultation" [Mesh] or "telemedicine" [Mesh] or "videoconferencing" [Mesh])	English language	45
CINAHL: Autis* or Asperg* or ASD or ("pervasive," "developmental," and "disorder") or PDD, and (rural* or remote* or regional* or telehealth or tele-health or telemedicine or tele-medicine or telerehab* or tele-rehab* or telediagnos* or tele-diagnos* or teletreat* or tele-treat or teletherap* or tele-therap* or telemonitoring or tele-monitoring or teleintervention or tele-intervention or teletreatment or tele-treatment or telepractice or tele-practice or videoconference* or video-conferenc* or teleconference* or tele-conference* or webbased OR web-based or internet-based or ["technology" and "mediated"] or technology-mediated)	English Language Published date: 20140601-20160704	64
Free-text search words		
ERIC:As per CINAHL free text	As per CINAHL free text	45
Embase **:** As per CINAHL free text	As per CINAHL free text	487
PsycINFO **:** As per CINAHL free text	As per CINAHL free text	131
PubMed **:** As per CINAHL free text	As per CINAHL free text	446

Search Strategy

The categories of search terms used were (1) ASD (autism, autism spectrum disorder, pervasive development disorder not otherwise specified, and Asperger's' syndrome) and (2) residing outside of urban areas (rural health, regional health, remote health, telehealth, telemedicine, and videoconferencing) (see Table 1). Search limitations included papers published in English only. A free-text search was completed within the listed databases for literature published from June 16, 2014 to July 4, 2016. The search terms and limitations used for the free-text search are outlined in Table 1. Manual searches of the following

journals were performed: The Journal of Rural Health (United States), Australian Journal of Rural Health, Rural Educator, Journal of Research in Rural Education, Australian and International Journal of Rural Education and Rural Special Education Quarterly . Finally, manual searches were conducted of the reference lists of articles that met the eligibility criteria.

Study Selection

The first author screened titles and abstracts of the entire pool of articles that met the inclusion criteria and removed duplicates. Following the removal of the duplicates, all abstracts were screened independently by 2 authors using the inclusion or exclusion criteria. Full-text articles were sourced for abstracts that met inclusion criteria, and articles that did not meet the inclusion or exclusion criteria were excluded. Agreement between authors was reached on 8 out of the 9 included articles. The remaining disagreement was resolved through discussion and consensus.

Methodological Quality

Methodological quality was assessed using the standard quality assessment criteria as described by Kmet et al [40]. The Kmet appraisal checklist uses a 3-point ordinal system to assess the methodological quality of research papers. The appraisal checklist creates a systematic, quantitative, and reproducible process to assess the methodological quality of a variety of research designs and make comparisons between them. Two authors independently assessed the included articles using the 14-point checklist. Scores were categorized into quality levels: >80% as strong, 70-80% as good, 50-69% adequate, and <50% as limited. The methodological scores are included in Table 2. Disagreements in methodological quality existed between the 2 authors in 2 out of the 9 articles and were resolved through discussion and consensus.

Data Collection

Data were extracted using comprehensive data extraction forms and grouped under the following headings: (1) aims or objectives, (2) study design, (3) level of evidence, (4) participant characteristics (including geographical location and proximity to services), (5) intervention characteristics, (6) outcome measures, (7) discussion, (8) limitations, and (9) implications for future practice. Data extraction was undertaken by the first author. Data extracted was checked by a second author for accuracy. Only minor discrepancies occurred, and these were resolved through consensus. The level of evidence was determined using the hierarchy of evidence as outlined in the National Health and Medical Research Council (NHMRC) guidelines [41]. Additionally, details of the intervention, dosage, method of delivery, and skills or aims being delivered by the researchers were extracted and summarized. Few studies included in the review had large sample sizes, and the lack of control or comparison groups prevented a meta-analysis.

Data Items, Risk of Bias, and Synthesis of Results

Participant characteristics were extracted and are represented in Table 3. Kmet ratings were used to assess the risk of bias of at an individual study level [40]. The extrapolated data from this process are represented in Table 2. Characteristics of the extracted data included (1) aims and objectives, (2) study design, (3) level of evidence, (4) intervention characteristics, (5) outcome measures, (6) results, and (7) methodological quality. Significance of data and calculated effect sizes of the interventions were extracted for synthesis. Effect sizes not reported as Cohen *d* were converted for uniformity as appropriate. The magnitude for Cohen *d* effect sizes was interpreted as small≥0.20, medium≥0.50, or large≥0.80 [42]. None of the researchers authored any of the included published studies; hence, no bias in study selection was introduced in conducting the systematic review.

Results

Study Selection

The PRISMA diagram is presented in Figure 1. Database searches yielded 1797 articles. Four additional articles were identified through manual searches of the included studies' reference lists. From the 2001 articles, 583 duplicates were removed, leaving a total of 1218 abstracts for screening. Following screening of the abstracts, 1202 articles were excluded. The remaining 16 were retrieved for full-text review, and an additional 7 articles were excluded from the study, as participants in four studies were not described as living outside of major cities or urban areas. One study was a summary of a pilot project with no results included, and the 2 remaining studies provided the training to parents in a face-to-face medium. A total of 9 articles met the review eligibility criteria. The articles by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45] were based on the outcomes from one study, and so they were combined for reporting and discussion throughout this paper.

Study Design

The 7 studies included 1 quasi-experimental design by St. Peter et al [46], 1 nonconcurrent multiple-baseline designs by Wacker et al [47], 2 single-subject multiple-baseline design by Vismara et al [12,48], 1 RCT each by Ingersoll et al [44] and Pickard et al [45], and 2 prepost test design studies by Hamad et al [49] and Heitzman-Powell et al [50]. A lack of control groups in 5 of the 7 studies precluded the ability to conduct a meta-analysis of the results. An overview of the included papers is provided in Figure 2.

Level of Evidence

The overall level of the evidence for the studies included in the systematic review was low. The studies by Hamad et al [49], Heitzman-Powell et al [50], Vismara et al [12,48], and Wacker et al [47] demonstrated level IV evidence. The study by St. Peter et al [46] demonstrated level III-1 evidence and the study by Ingersoll et al [44] and Pickard et al [45] was classified as level II [43-45] evidence as per the NHMRC level of evidence guidelines (see Table 2) [41]. The low level of evidence may indicate that it is difficult to conduct research with this population due to travelling distance to research centers and reachability through recruitment strategies, thus making robust study designs more challenging.

Figure 1. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram.

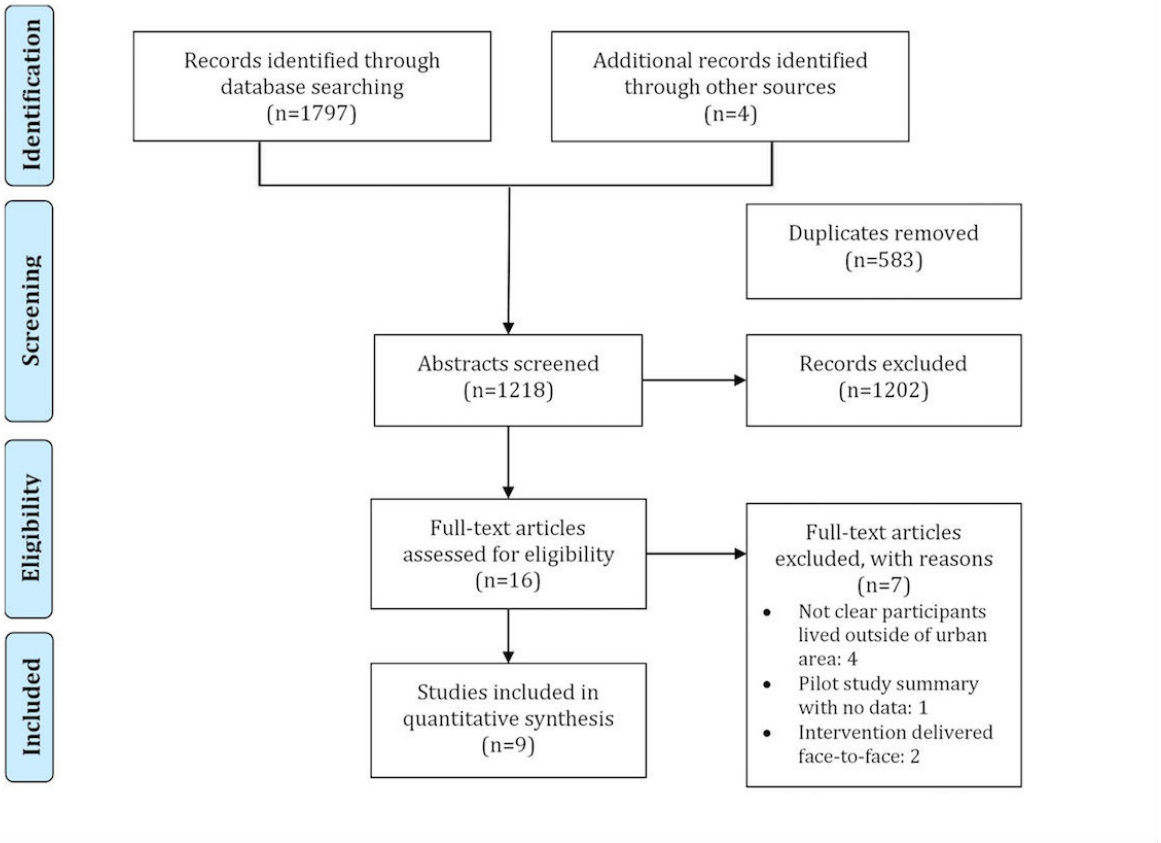


Figure 2. Study Schema.

	Study Design			Participants		Delivery Method		Aims		OM[a]	ES[b]	Interventions – research created or established	
Citation	Pre-Post	Multiple-base design	RCT[c]/ QE[d]	Parents	Child	Online delivered	Face-to-face	Increase parental ASD[e] knowledge	Parent-mediated intervention	Outcome measures created by researchers	Reported Effect Sizes	Researcher Created	Established
Hamad et al. [49]	●			●		●		●	●	●	●	●	
Heitzman-Powell et al. [50]	●			●		●		●	●	●		●	
Ingersoll and Berger [43]			●	●		●		●	●	●		●	
Ingersoll et al. [44]			●	●	●	●		●	●	●	●	●	
Pickard et al. [45]			●	●		●		●	●	●		●	
St. Peter et al. [46]			□[f]	●					●	●		●	●
Vismara et al. [48]		×[g]		●	●	●		●	●	●	●	●	●
Vismara et al. [12]		×		●	●	●		●	●	●	●	●	●
Wacker et al. [47]		†[h]		●	●	●		●	●	●		●	●

[a]OM: outcome measures.
[b]ES: effect sizes.
[c]RCT: randomized controlled trial.
[d]QE: quasi-experimental.
[e]ASD: autism spectrum disorder.
[f]□= quasi-experimental.
[g]×= single subject design.
[h]†= nonconcurrent design.

Table 2. Study characteristics.

Citation and methodology	Aim or objectives	Outcome measures	Results	Methodological quality
Hamad et al [49] Pre- or post-test	Investigate the feasibility of an Internet-based "asynchronous" small-scale three module-Web-based learning course presented in a distance-learning medium.	Parent outcome measures: 25-item Web-based knowledge acquisition measure (test) administered prepost intervention.	Internet-based training curriculum could be effective in training parents about methods and procedures related to behavioral interventions. Pretest scores: mean=68.8, SD=15.6, Posttest scores: mean=82.9, SD=4.9. Large effect size (Cohen *d*=1.21) Paired *t*-tests: mean prepost test scores statistically significant improvement (*P*<.001) for all participants combined (n=51).	Kmet rating: strong (82%) NHMRC[a]level of evidence: level IV
		Child outcome measures: not specified		
Heitzman-Powell et al [50] Pre- or post-test	Evaluate the modified OASIS training intervention for use with parents from a distance.	Parent outcome measures: parent skill assessment in ABA[b]implementation	Implementations of ABA skills (41.23% mean increase)	Kmet rating: good (77%) NHMRC level of evidence: level IV
		Parent knowledge assessment (Web-based) on ASD[c]and ABA principles and procedures	Knowledge assessments (39.15% mean increase)	
		Parent satisfaction with training	High levels of importance and significance of Web-based tutorials (mean scale 1-5:4.62 and 4.71 respectively). High levels of importance and significance of telemedicine coaching sessions (mean scale 1-5:4.62 and 4.8 respectively)	
		Cost savings (driving miles) Child outcome measures	Mean travel savings per family was 2,263 driving miles using telemedicine if compared with face-to-face coaching. Note: Prepost comparison with no statistical analysis for significance.	
		Child outcome measures: not specified		
Ingersoll and Berger [43] Ingersoll et al [44] Pickard et al [45] RCT[d]	Compare parent engagement and effectiveness in self-directed and therapist-assisted versions of a novel telehealth-based parent-mediated intervention for young children with ASD	Parent outcome measures: ImPACT knowledge quiz: 20-item multiple choice quiz taken prepost intervention	Intervention completion was a significant predictor of postintervention knowledge (*P*=.01) in both groups.	Kmet rating: strong (85%) NHMRC level of evidence: level II
		Videotape parent-child interaction for intervention fidelity using the ImPACT intervention fidelity checklist	Intervention completion (*P*=.3) and group assignment (*P*=.45) made significant independent contributions to treatment fidelity. Post intervention fidelity for both groups was significant (*P*=.004) Statistically significant improvement prepost in parent intervention fidelity in both groups (*P*<.01, Large effect size: Cohen *d*=3.21) as well as between groups post intervention (*P*<.01, Large effect size: Cohen *d*=0.3). At follow-up statistically significant (*P*<.001, Large effect size: Cohen *d*=2.92) prepost in both groups but not between groups.	Kmet rating: strong (85%) NHMRC level of evidence: level II
		Parent sense of competence scale	Statistically significant improvement (*P*<.01, Large effect size: Cohen *d*=3.34) prepost intervention in self-efficacy in both groups but not between groups.	

Citation and methodology	Aim or objectives	Outcome measures	Results	Methodological quality
		Parent sense of competence scale	Statistically significant improvement (*P*<.01, Large effect size: Cohen *d*=1.34) prepost intervention in self-efficacy in both groups but not between groups.	
		Family impact questionnaire	Statistically significant improvement (*P*<.05, Large effect size: Cohen *d*=1.03) prepost in parent stress in both groups but not between groups post intervention. Statistically significant improvement (*P*<.05, Large effect size: Cohen *d*=1.47) prepost in positive perception of the child in both groups as well as between groups post intervention (*P*<.05, Large effect size: Cohen *d*=1.16).	
		Parent engagement using website analytics	Therapist-assisted group statistical significantly performed better on parent engagement (number of logins and duration on site) and intervention completion when compared with self-directed groups (*P*<.001 and *P*<.05 respectively)	
		Intervention evaluation survey using 7-point Likert scale measuring treatment appropriateness, website usability, and overall intervention satisfaction.	Participants rated intervention as highly acceptable (mean=6.07, SD=0.79), the website as highly usable (mean=6.36, SD=0.57). Overall satisfaction of intervention was high (mean=6.56, SD=0.71). No statistically significant difference in treatment appropriateness, website usability, and overall intervention satisfaction between groups.	
		49-item 7-point Likert scale quantitative survey administered post intervention examining intervention, appropriateness perceived child social communication gains, burden of the intervention on the family, and frequency of intervention use.	Overall, parent rated intervention favorably with mean scores: 1) Intervention appropriateness 6.59 (SD 0.58), perceived child social communication gains 5.41 (SD 1.24), burden of the intervention on the family 5.72 (SD 1.23), frequency of intervention use 6.36 (SD 0.57) Statistically significant differences between groups (TA[e]vs SD[f]) for intervention appropriateness (*P*=.03, Large effect size: Cohen *d*=0.94) and child social communication gains (*P*=.05, Large effect size: Cohen *d*=0.84). No difference in the burden of intervention on the family and frequency of intervention use domains.	
		Qualitative interviews—semistructured investigated overall perception of intervention and content, perception of feasibility of intervention, experience of support during intervention, and intervention referral preferences.	Qualitative themes: Positive perception of the appropriateness of intervention. The intervention was easy to learn initially but became more challenging as they progressed. The support of a coach would be essential in the later, more complex sections of the intervention. Parents felt more empowered and better able to interact with their child. Perceptions of barriers included time restrictions and technology failure. Parents suggested the intervention should be made available at the time of ASD diagnosis as it may help empower parents at a stressful time.	
		Child outcome measures: language targets	Statistically significant (*P*<.05, Large effect size: Cohen *d*=2.26) prepost improvements in language targets in both groups but not between groups post intervention.	
		MacArthur communicative development inventories: words and gestures	Statistically significant (*P*<.01, Large effect size: Cohen *d*=1.74) prepost improvements in language skills in both groups but not between groups post intervention.	

Citation and methodology	Aim or objectives	Outcome measures	Results	Methodological quality
		Vineland adaptive behavior scales, 2nd edition	Statistically significant (P<.05, Large effect size: Cohen d=1.00) prepost improvements in the communication domain in both groups but not between groups post intervention. No statistically significant differences prepost in the social domains in both groups, however, a statistical difference was observed between groups post intervention (P<.05, Large effect size: Cohen d=0.91)	
St. Peter et al [46] Quasi-randomized	Compare parental adherence during written or asynchronous video teleconsultation designed to teach parents of children with ASD to implement discrete trial instruction.	Parent outcome measures: Parental adherence between the written (control) and video (experimental) groups Child outcome measures: not specified	Adherence in the video group was significantly higher (P<.001) compared with written instructions.	Kmet rating: good (71%) NHMRC level of evidence: level III-1
Vismara et al [48] Single-subject, multiple- baseline design	To assess if a 12-week videoconferencing and DVD learning module (P-ESDM[g]) could improve parents' acquisition of teaching procedures and result in changes in the child's social communicative behavior [51].	Parent outcome measures: Eight item, 5-point response scale evaluating parental satisfaction (feasibility and appropriateness) with the support and ease of the intervention	All parents reported satisfaction with support and ease of the telehealth learning intervention. Six parents identified DVD's as more useful teaching aids compared to handouts. All parents agreed they would recommend an approach to other parents of children with ASD with limited access to community services. Significant increases over time from baseline to follow-up (P<.001)	Kmet rating: good (77%) NHMRC level of evidence: level IV
		P-ESDM fidelity tool—5-point Likert rating tool of 13 parent behavior that define the child-centred, responsive interactive style used in PESDM	Significant increases over time from baseline to follow-up (P<.001)	
		MBRS[h]—A 5-point Likert rating scale measuring the parent's style of interacting to or relating to their child.	Significant increases in parental behavior rating from baseline to follow-up in responsivity (P<.001), affect (P<.001), and achievement orientated behavior (P<.001)	
		Child outcome measures: child social communication behavior—10-min videos transcribed and scored for the production of spontaneous and promoted functional verbal utterances and approximations and imitative play actions on objects and gestures.	Significant overall increases from baseline to follow-up in spontaneous functional verbal utterances (P<.001), prompted words over time (P<.001), and spontaneous imitations (P<.001)	
		CBRS[i][52]—measures engagement and interest in activity as well as joint attention, creativity, and affect demonstrated toward the parent.	Significant increase form baseline to follow-up in child attention (P<.001) and child initiation (P<.001).	
		MacArthur communicative development inventories: words and gestures	Significant increases from baseline to follow-up with vocabulary production (P<.001) and vocabulary comprehension (P<.001).	
		Vineland adaptive behavior scales, 2nd edition	Significant increase from baseline to follow-up on the adaptive behavior composite (P<.05).	

Citation and methodology	Aim or objectives	Outcome measures	Results	Methodological quality
Vismara et al [12] Single-subject, multiple- baseline design	Pilot study of a 12-week telehealth on the Web (videoconferencing and self-guided website) intervention (P-ESDM) and 3-month follow-up to assess: (1) parents' perception of the intervention as a useful learning platform, (2) parents' intervention skills and engagement style improvement, (3) website utility to support the intervention, and (4) improvements in the children's verbal language and joint attention.	Parent outcome measures: Eight item, 5-point response scale evaluating parental satisfaction with the support and ease of the telehealth learning intervention	All parents reported satisfaction with support and ease of the telehealth learning intervention.	Kmet rating: good (77%) NHMRC level of evidence: level IV
		P-ESDM fidelity tool—5-point Likert rating tool of 13 parent behavior that define the child-centred, responsive interactive style used in P-ESDM	Improvement in parent intervention fidelity. Baseline: 0/8 parents meeting criteria for fidelity in tool. Group mean 2.93 (SD 0.6), post intervention: 6/8 parent meeting criteria for fidelity in tool. Group mean 3.69 (SD.51), follow-up: 7/8 parents achieved at least one fidelity score. Group mean 4.15 (SD 0.51)	
		Website use	Average number of logins 30 (SD 18, range 9-60); Average viewing time per day 18 min	
		MBRS [53]—A 5-point Likert rating scale measuring the parent's style of interacting to or relating to their child.	Improvement in parent engagement style. Baseline: low-moderate with MBRS total score mean=2.91, SD=0.68, post intervention: mean=3.50, SD=0.44, follow-up (3 months): moderate to high range with MBRS total score mean=3.87, SD=0.42	
		Child outcome measures: behavior scoring of videotaped probes—functional verbal utterances and nonverbal joint attention initiations without gestures	Increase in the range of vocalizations at all time points Baseline: mean=2.97, SD=1.93, post interventions: mean=3.60, SD=2.51, follow-up: mean=4.14, SD=2.04 Joint attention initiations remained constant between baseline (mean=1.67, SD=1.07) to post intervention (mean=1.67, SD=1.21) but increased at follow-up (mean=2.16, SD=1.34)	
		MacArthur communicative development inventories: words and gestures	Improvements in VP[j]and comprehension, Baseline: VP mean=111.87, SD=156.03, comprehension mean=224.37, SD=133.25, post intervention: VP mean=163.88, SD=156.03, comprehension mean=284.88, SD=141.53, follow-up: VP mean=213.88, SD=155.08, comprehension mean=314.88, SD= 94.16	
Wacker et al [47] Nonconcurrent multiple baseline design	Conduct functional communication training using coaching from trained behavior analysts to parents via telehealth and compare it with completing the same training in-vivo within families' homes.	Parent outcome measures: Parent overall appropriateness—7-point Likert scale	Parents rated training as acceptable (mean=6.47. Comparable with in-vivo training (mean=6.18)	Kmet rating: good (73%) NHMRC level of evidence: level IV
		Costs: mileage and consultant costs	Costs through telehealth were considerably lower that for in-home behavior therapy	

Citation and methodology	Aim or objectives	Outcome measures	Results	Methodological quality
		Child outcome measures: Interobserver agreement on child-targeted problem behavior using interval-by-interval comparisons.	Reduction in child-targeted problem behavior when parents coached via telehealth (mean reduction=93.5%). Comparable with in-vivo training (mean reduction=94.1%).	

[a]NHMRC: National Health and Medical Research Council. Designation of levels of evidence: I—Evidence obtained from a systematic review of all relevant randomized controlled trials, II— evidence obtained from at least one properly designed randomized controlled trial, III-1 —evidence obtained from well-designed pseudo-randomized controlled trials (alternate allocation or some other method), III-2—evidence obtained from comparative studies with concurrent controls and allocation not randomized (cohort studies), case-control studies, or interrupted times series with a control group, III-3—evidence obtained from comparative studies with historical control, two or more single-arm studies, or interrupted time series without a parallel control group, IV—evidence obtained from case series, either post-test or pre-test and post-test.

[b]ABA: applied behavior analysis.

[c]ASD: autism spectrum disorder.

[d]RCT: randomized controlled trials.

[e]TA: therapist-assisted group.

[f]SD: self-directed group.

[g]P-ESDM: parent model—early start Denver model.

[h]MBRS: maternal behavior rating scale.

[i]CBRS: child behavior rating scale.

[j]VP: vocabulary production.

Study Participants

For the purposes of this review, study participants were families having a child with ASD, living outside of urban areas, and having limited access to services as reported by the authors. The inherent difficulty of defining regional and remote localities between different countries made delineating study participants based on geography challenging. None of the studies provided quantitative detail about the participants' proximity and access to services so the interpretation of the findings in relation to this information was impossible. Studies included a total of 197 parents aged between 24 and 69 years involved across the 7 studies. The highest education level achieved by the parents was specified in 5 out of 7 studies. Of the remaining two, one provided a range without specific data and the remaining study did not specify parental level of education. The study populations resided in either the United States, Canada, or Australia. Mothers represented a majority of the parents who received the education and delivered the intervention to the child. Vismara et al [12,48], Wacker et al [47], and the study by Ingersoll et al [44] and Pickard et al [45] measured outcomes for the children receiving the intervention. Refer to Table 3 for detailed demographic information.

Outcomes

The aim of all of the studies was to improve social behavior and communication skills of children with ASD through increasing the knowledge of parents and caregivers by training them in intervention skills (parent-mediated). Outcome measures varied across all of the studies. All 7 studies used measures created by the researchers. Calculated effect sizes were only possible based on the published information in 3 studies included in the review and are reported in Table 2.

Parental satisfaction and perceptions of appropriateness of the intervention were measured by Vismara et al [12,48], Heitzman-Powell et al [50], and the study by Ingersoll et al [44] and Pickard et al [45]. All reported that parents were satisfied with the training they received. When comparing a therapist-assisted and self-guided website versus a self-guided website only, large effect sizes were recorded in parents' perception of the appropriateness of the intervention and child social communication gains (Cohen d=0.94 and 0.84 respectively) in the study by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45].

Parents' self-efficacy was evaluated in the study by Ingersoll et al [44] and Pickard et al [45]. The authors stated that whereas there was a statistically significant (P<.01) improvement and a large effect size (Cohen d=1.34) preintervention to postintervention for both groups, there was no difference between groups. Parents' stress levels were not measured prepost interventions in any of the studies.

Knowledge acquisition by parents was measured by Hamad et al [49], Heitzman-Powell et al [50], and in the study by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45] using quizzes covering the content in the intervention; all studies reported significant increases in knowledge post intervention. Parents' skills in implementing the acquired therapy techniques were investigated by Heitzman-Powell et al [50], St. Peter et al [46], Vismara et al [12,48], Wacker et al [47], and in the study by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45]. All of the studies reported statistically significant improvements in parents' skills in administering skills learnt through the interventions. These findings present evidence that parents who received the appropriate training could gain skills in the delivery of interventions, thus improving the skills in social communication and behavior of their children with ASD.

Vismara et al [12,48] and the study by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45] utilized the MacArthur communicative developmental inventories [54] to measure the child's abilities in vocabulary production and comprehension. In all 3 studies, statistically significant

improvements were reported in the children's vocabulary production and comprehension from baseline to follow-up. Again, this provides preliminary evidence that parents who live in geographically isolated areas are able to learn skills in the provision of therapy and implement it appropriately to help improve the communication skills of their children with ASD.

Improvements in social behavior were measured in 2 studies using the Vineland adaptive behavior scales (2nd edition)[55] with Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45] reporting no significant difference prepost intervention and Vismara et al [48] reporting a significant difference in the social domains. Video-recorded interactions of the children with their parents were used in the studies conducted by Vismara et al [12,48] and Wacker et al [47]. All reported statistically significant improvements prepost intervention in joint attention and affect toward the parents with Wacker et al [47] reporting a reduction in child problem behavior.

Table 3. Participant characteristics.

Study	No. of participants	Geographical location	Demographics: parent	Demographics: child
Hamad et al [49]	51	"Geographically disparate" in the United States	Gender: male n=4, female n=47 Age: not specified Education level: high school n=6, associate degrees n=0, bachelor degrees n=20, master degrees n=12, other n=3 Note: subgroup demographic breakdown not provided	Gender: not specified Age: not specified
Heitzman-Powell et al [50]	7	Remote areas in the United States	Gender: not specified Age: mean age 37.3 (range=32-47) Education level: ranged from graduate degree to high school diploma. Breakdown not specified.	Gender: not specified Age: not specified
Ingersoll and Berger [43] Ingersoll et al [44] Pickard et al [45]	27	70% (19/27) of participants resided in "rural or medically underserved areas"	Gender: male n=1, female n=26 Age: not specified Education: college degree or higher=16, education levels of remaining participants not specified	Gender: male n=19, female n=8 Age: mean chronological age 3 years, 6 months.
St. Peter et al [46]	32	Rural Appalachian counties in West Virginia, Kentucky, Maryland, Virginia, or Pennsylvania, United States	Gender: male n=11, female n=21 Age: mean age of parents=35.87 years (range, 24-69) Education level: 54.15% had received a college degree. Remaining participant education levels not specified.	Not specified
Vismara et al [48]	8	"Very little access to early intervention services" in California, North Carolina, Arkansas, Texas, and Pennsylvania, United States.	Gender: male n=1, female n=7 Age: not specified Education: not specified	Gender: male n=7, female n=1 Age: mean chronological age 2 years, 4 months (standard deviation=7.6 months, range 16-38 months)
Vismara et al [12]	8	"Minimally available intervention services in their community" in the United States and Canada	Gender: male n=1, female n=7 Age: not specified Education level: college n=4, postcollege n=4	Gender: not specified Age: 1 year n=4, 2 years n=2, 3 years n=1
Wacker et al [47]	17	Regional Iowa, United States	Gender: male n=2, female n=16 Age: mean age 33 years Education level: "most" had some level of postsecondary education. Breakdown not specified.	Gender: male n=16, female n=1 Age: 2 years n=3, 3 years n=4, 4 years n=3, 5 years n=5, 6 years n=2 (range 29-80 months)

Table 4. Intervention characteristics.

Study	Intervention description and dosage	Method of delivery to parent	Skills or aims of intervention
Hamad et al [49]	Web-based training intervention in behavioral interventions Dosage: intervention approximately 4-8 h within a 3-week period Three modules	On the Web using Blackboard Vista 4 platform Included: short Web-based lectures, practical exercises, video demonstrations of procedures, study questions, and frequent short Web-based quizzes.	• Positive reinforcement: selection and use of reinforcement. • Relationship building: parent and teaching co-operation. • Prompting and prompt fading.
Heitzman-Powell et al [50]	OASIS training intervention Research-to-practice Applied behavior analysis outreach training model Dosage: Eight modules; timeframe not specified	Training program combines Web-based instructional modules and participation in distance coaching sessions.	• Introduction to ASD[a] and behavioral treatment; • Basic ABA[b] principles and procedures • Use ABA procedures to teach new skills • Use ABA procedure to reduce challenging behavior • Generalize skills to other settings • Collection and analysis of data for data-based intervention decision-making • Working with treatment teams and other provider
Ingersoll and Berger [43] Ingersoll et al [44] Pickard et al [45]	Project ImPACT on the Web—Website-based training for a naturalistic, developmental-behavioral, parent-meditated intervention for children with ASD Dosage: Self-directed—Encourage to complete 1 lesson per week, approximately 80 min for 12 weeks. Therapists assisted—dosage same as self-directed group plus 2 30-min remote coaching sessions per week by trained therapist.	Access to training material was on the Web via personal computer. Included: narrated slideshow with embedded video examples of techniques, written description of lessons, exercises, homework, and reflection questions Training program for the therapist-assisted group was administered by trained therapists using videoconferencing software.	• Promote child social communication within the context of play and daily routines
St. Peter et al [46]	Implementation discrete-trial instructions using a video training materials Dosage: video training was 37 min in duration Written training was a 30-page manual	Written training materials (control) or video training materials (experimental) containing similar content.	• Increase adherence to discrete-trial instruction procedures.
Vismara et al [48]	Parent early start Denver model (P-EDSM) training Dosage: Once-per-week, 1 h parent training sessions for 12 weeks	Telehealth delivery using live, 2-way conferencing with a qualified therapist and the provision of a DVD including all intervention materials with the addition of video recorded examples of the therapist demonstrating skills.	• Increasing child's attention and motivation • Using sensory social routines • Promoting dyadic engagements and joint activity routines • Enhancing nonverbal communication • Building imitation skills • Facilitating joint attention • Promoting sequence relations • Employing promoting, shaping, and fading techniques • Conducting functional assessments of behavior to develop new interventions.

Study	Intervention description and dosage	Method of delivery to parent	Skills or aims of intervention
Vismara et al [12]	Parent early start Denver model (P-EDSM) training Dosage: Once-per-week, 1.5 h parent training sessions for 12 weeks	Telehealth delivery using live, 2-way conferencing with a qualified therapist and a self-guided website.	• Increasing child's attention and motivation • Using sensory social routines • Promoting dyadic engagements and joint activity routines • Enhancing nonverbal communication • Building imitation skills • Facilitating joint attention • Promoting sequence relations • Employing promoting, shaping, and fading techniques • Conducting functional assessments of behavior to develop new interventions.
Wacker et al [47]	Functional communication Dosage: Weekly 60 min sessions until completion of treatment,	Telehealth using PC and video-monitors from behavior consultants	• Child taught to comply with task request and then to mand for a break to play • Child requesting toys after having to wait for increasing period of time • Request attention when adult attention was removed.

[a]ASD: autism spectrum disorder

[b]ABA: applied behavior analysis.

In summary, it appears that interventions targeting parents' knowledge and including fidelity checks have statistically significant improvements with large effect sizes when reported. Additionally, large to small effect sizes were reported in the child's improvement in social behavior and communication skills when reported within the studies.

Interventions

All interventions were developed with consideration of the geographical isolation of participants, with the aim to ease administration of the intervention and increase feasibility of delivery. Parent training interventions investigated in the included articles are summarized in Table 4.

All interventions were developed by the researchers and varied in dosage and method of delivery. This variation makes synthesis of the research challenging and limits the generalizability of these methods of intervention to the targeted population. Dosage for the interventions ranged from an intensive format of 5 h per day for 5 days, to once-a-week over a number of weeks. The most common dosage was once-a-week sessions, with sessions lasting 1-2 h; however, timeframes ranged from 6-12 weeks [12,43-45]. Additionally, the studies by Heitzman-Powell et al [50] and Wacker et al [47] did not have finite periods, and the duration of intervention continued until all training modules were completed at the participants' own pace. The lack of comparison regarding the dosage of education and training provided to the parents prevented the identification of an optimal amount of education and training to achieve the maximum benefit to the children.

The methods of delivery for the parent-mediated interventions were equally as wide-ranging, with Hamad et al [49], Heitzman-Powell et al [50], and the study by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45], requiring parents to access resources on the Web and progress through the content at their own pace. Heitzman-Powell et al [50] coupled the Web-based modules with distant coaching sessions delivered by qualified clinicians. The studies by St. Peter et al [46], Wacker et al [47], and Vismara et al [48] utilized a telehealth delivery model with live, 2-way videoconferencing by qualified clinicians who delivered the intervention in isolation, coupled with a Web-based self-guided website or using teaching materials contained on a DVD. St. Peter et al [46] compared the difference between the effectiveness of delivery methods; training provided using video methods versus training provided via a written manual. Finally, the study by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45] compared 2 groups; one receiving access to a Web-based training program only and the other having access to the same Web-based training program, but with additional weekly therapist-assistance via videoconferencing.

Identifying the superior delivery method of intervention for this population is limited by a lack of between-group comparisons within the included studies. Only the studies by Ingersoll and Berger [43], Ingersoll et al [44,] and Pickard et al [45], and St. Peter et al [46] had comparison groups. Methods with increased user interaction demonstrated some superiority with DVDs having higher adherence to the training program compared with written content. Furthermore, regular therapist-assisted sessions resulted in increased intervention completion, parent appropriateness of intervention, and improvements in parent knowledge and skills.

Overall, these findings suggest that training delivered to parents who live outside of urban areas or with limited access to services can have some effect in improving the social behavior and communication skills in their child with ASD and a large effect on increasing their own knowledge and skills in of ASD interventions. Additionally, no specific content or dosage can be identified as being superior; however, more interactive methods of delivery, such as videos and regular therapist contact for training have been proven to (1) improve adherence, (2)

increase completion rates, and (3) improve fidelity in parent-mediated interventions.

Risk of Bias in Included Studies

The St. Peter et al [46], Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45] studies assigned participants to different intervention groups. The remaining 5 articles have a high risk of selection bias. In the study by St. Peter et al [46], the randomisation process was poorly described with no mention of blinding and allocation procedures by the researchers. The authors reported homogeneity between samples with no significant differences in socioeconomic status, educational level, or previous experience with the intervention between the experimental and control groups and autism severity scores. Therefore, the risk of bias from confounding variables was reduced due to the homogeneity of the 2 groups. Confounding bias was addressed in the study by Ingersoll and Berger [43], Ingersoll et al [44], and Pickard et al [45] by matching participants on their pretreatment expressive language age using a standardized assessment prior to randomisation.

All 7 studies were subject to a high risk of bias due to a lack of blinding. Five of the studies in this review were at a higher risk of confounding bias due to the lack of controls. The small sample sizes of these articles increased the likelihood of type II errors with no article reporting a power calculation relative to the outcome measures.

Discussion

Principal Findings

Findings of this systematic review provide preliminary evidence that parent-mediated intervention training for families living in nonurban areas can assist in improving social behavior and communication skills of children with ASD. Weak study design, lack of standardized outcome measures, lack of measurement outcomes in children with ASD, small participant numbers, high risk of bias, and large variations in interventions limit the generalizability and conclusiveness of the findings to the target population. Despite the limitations, preliminary findings from this review suggest that parent-mediated intervention training delivered remotely could benefit both parents and children with ASD given the barriers they face in accessing traditional services.

The notion that parent-mediated interventions can fully address the gap of limited access to services and be an effective alternative intervention for children with ASD needs further investigation. A systematic review conducted by McConachie and Diggle [23] focused on parent-delivered interventions regardless of geographical location or method of delivery for children with ASD. The authors concluded that whereas these types of interventions can improve the social behavior and communication challenges of children with ASD, the lack of studies with robust study design limits the ability to draw further conclusions and highlighted the need for further research. This paper builds on these findings by reviewing current literature on the effectiveness of parent-mediated intervention training delivered remotely to a nonurban population who face a number of barriers accessing traditional services.

In this review, effect sizes were larger for intervention outcomes that targeted parents' knowledge and intervention fidelity skills, compared with intervention outcomes to improve social behavior and communication skills for their children. Only 2 studies included measures of social behavior and communication skills in the children with ASD despite all the interventions providing training for parents to deliver therapy to address these skills. This finding indicates that parents have the potential to improve their knowledge and intervention fidelity skills and be agents in the delivery of therapeutic interventions, thereby improving the social behavior and communication skills of their children with ASD.

The results of this review indicate that the use of telehealth, Web-based modules, and DVDs all seem to have some effect in educating parents about ASD and increasing the fidelity in the delivery of interventions. A lack of standardized measurements and RCTs limited the comparison of interventions within this review. Interventions that were delivered using videos were more effective and accepted by parents than written information. Additionally, weekly contact with a therapist to answer questions and provide coaching proved to be more effective in the areas of (1) intervention appropriateness, (2) program completion, (3) parent intervention fidelity, (4) parent engagement, and (5) parent's positive perception of their child, when compared to a self-directed program alone. Considering this, the interventions created for families that have limited access to face-to-face therapy could be tailored to meet the needs of the individual parents based on their proximity to services, personal qualities, resources, and preference. Furthermore, interventions clearly benefitted from regular contact with trained professionals throughout the training program.

Defining populations based on their geographical location is challenging due to differing methodologies and definitions adopted by different countries. This disparity in terminology and classification systems makes trying to understand the unique characteristics of families having a child with ASD and living in regional and remote areas difficult due to the wide variability of proximity and access to appropriate services. This is confounded when trying to compare populations from different countries that use vastly different classification systems. The review highlights the importance for researchers to use the relevant geographical classification system in their country to make defining study populations more clearly thereby providing better context for their study.

Finally, evidence is emerging that suggests there is indeed a significant difference in the characteristics and needs of families having a child with ASD residing in urban areas and those residing in rural areas, but further investigation is needed [16-18,22]. Intuitively this discrepancy between the populations makes sense; however, the poor description of participant characteristics, lack of control groups, and absence of comparisons between these 2 groups prevent conclusive findings.

Recommendations for Future Research

Further research into the feasibility, efficacy, and appropriateness of the methods of delivery for this unique population will help inform clinical decisions. This systematic

review provides preliminary evidence on the effectiveness of remotely delivered parent-mediated intervention training. However, more research is needed to determine the most effective balance between parent-mediated intervention and therapist support via Web-based or distance training to provide the best outcome for a child with ASD, while considering the family's proximity to traditional services. Furthermore, investigation into the effectiveness of the parent-mediated intervention training should not only measure parents' knowledge and skill attainment but also the intervention effectiveness in improving social behavior and communication skills of children with ASD.

Future experimental studies on the effectiveness of parent-mediated interventions, including training programs, should include (1) larger sample sizes, (2) RCTs, (3) improved controls for bias, and (4) use of standardized outcome measures. A lack of comparison groups prevented a meta-analysis in this review. Standardized outcome measures should be employed wherever possible, as these were seldom used in the included studies in this review, with nonvalidated measures often created by the researchers to evaluate the effectiveness of their own intervention. This increased the risk of bias in the studies, thus limiting the impact of the studies' findings. Further research could be focused on comparing different parent training interventions, their components, dosage, and the methods of delivery to determine a superior strategy in increasing parent knowledge and intervention fidelity while improving social behavior and communication skills of their children with ASD.

Despite the studies reporting on the parents' perceived appropriateness and overall satisfaction of the intervention, there was limited investigation into the influences of parent engagement in the parent-mediated interventions. Further research in relation to the factors surrounding parent engagement in the intervention could help inform clinicians when devising training interventions related to content, parent commitment, and methods of delivery.

There is emerging evidence that interventions delivered remotely can improve the socioemotional and communication skills of children with ASD and may be an alternative to traditional models of therapy [11,56]. The appropriateness and feasibility for parents to utilize other methods to deliver therapy to their children such as direct one-on-one interventions using telehealth technology or the ever-expanding suite of tablet and other ICT-based interventions remains to be comprehensively investigated. Finally, economic modeling comparing the expense of a variety of methods of delivery and interventions could help inform the most cost-effective and feasible delivery method.

The unique context in which families having children with ASD and living in nonurban settings needs to be further researched. Emerging evidence suggests that the nonurban context is different, yet, the unique enablers and barriers in relation to service delivery that these families experience are yet to be fully understood. Furthermore, there is a need for comparison studies between urban and nonurban populations to better develop effective, appropriate, and feasible interventions to improve the social behavior and communication skills in children with ASD; thus allowing the development of tailor-made interventions for each population.

Limitations

A rigorous process involving (1) the searching of 5 databases, (2) establishing interrater reliability between 2 independent researchers for inclusion or exclusion agreements, (3) standardized data extraction forms, and (4) methodological assessment using the Kmet appraisal checklist was conducted in this study. Despite this, there are some limitations in the review that should be noted. Defining the population was challenging given the poor use of standardized geographical classification systems by authors. Inclusion was based on author report of the participants living in areas described as nonurban, rural, or remote, and as having limited access to services. This could have led to some studies being excluded if this description was not provided by the authors. Additionally, the small number of articles included limits the generalizability of findings to the target population.

Conclusions

Overall, there is preliminary evidence that parent-mediated intervention training delivered remotely can improve parents' knowledge in ASD, parent intervention fidelity, and subsequently improve the social behavior and communication skills of their children with ASD. The studies included in this review had an unclear or high risk of bias due to a lack of control groups and paucity of using standardized outcome measures. Additionally, difficulties in defining the participant characteristics limited the translatability to the target population. Few studies reported on the feasibility and appropriateness of the interventions and the factors of parent engagement in the interventions were evident in most studies. Future research should aim to use RCT designs, incorporate standardized outcome measures, and describe participant characteristics in greater detail. Furthermore, the review highlighted the need to investigate the feasibility and appropriateness of the interventions in addition to the factors influencing parent engagement in the interventions.

Acknowledgments

The authors received no funding to conduct this review or have any financial interest in the interventions included within this review.

Conflicts of Interest

None declared.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). Arlington, VA: American Psychiatric Association; 2013.
2. Happé F, Ronald A. The 'fractionable autism triad': A review of evidence from behavioural, genetic, cognitive and neural research. Neuropsychol Rev 2008 Dec;18(4):287-304. [doi: 10.1007/s11065-008-9076-8] [Medline: 18956240]
3. Matson JL, Kozlowski AM. The increasing prevalence of autism spectrum disorders. Res Autism Spectr Disord 2011 Jan;5(1):418-425. [doi: 10.1016/j.rasd.2010.06.004]
4. Baio J. Prevalence of autism spectrum disorders: Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. MMWR Surveill Summ 2012 Mar 30;61(3):1-19 [FREE Full text] [Medline: 22456193]
5. Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, et al. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). Lancet 2006 Jul 15;368(9531):210-215. [doi: 10.1016/S0140-6736(06)69041-7] [Medline: 16844490]
6. Kim YS, Leventhal BL, Koh Y, Fombonne E, Laska E, Lim E, et al. Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry 2011 Sep;168(9):904-912. [doi: 10.1176/appi.ajp.2011.10101532] [Medline: 21558103]
7. Williams K, MacDermott S, Ridley G, Glasson EJ, Wray JA. The prevalence of autism in Australia. Can it be established from existing data? J Paediatr Child Health 2008 Sep;44(9):504-510. [doi: 10.1111/j.1440-1754.2008.01331.x] [Medline: 18564076]
8. Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, Matthews FE, et al. Prevalence of autism-spectrum conditions: UK school-based population study. Br J Psychiatry 2009 Jun;194(6):500-509 [FREE Full text] [doi: 10.1192/bjp.bp.108.059345] [Medline: 19478287]
9. Hill A, Zuckerman K, Fombonne E. Epidemiology of Autism spectrum disorders. In: Translational approaches to Autism Spectrum Disorder. Switzerland: Springer International Publishing; 2015:13-38.
10. Kogan MD, Blumberg SJ, Schieve LA, Boyle CA, Perrin JM, Ghandour RM, et al. Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. Pediatrics 2009 Nov;124(5):1395-1403. [doi: 10.1542/peds.2009-1522] [Medline: 19805460]
11. Ramdoss S, Machalicek W, Rispoli M, Mulloy A, Lang R, O'Reilly M. Computer-based interventions to improve social and emotional skills in individuals with autism spectrum disorders: A systematic review. Dev Neurorehabil 2012;15(2):119-135. [doi: 10.3109/17518423.2011.651655] [Medline: 22494084]
12. Vismara LA, McCormick C, Young GS, Nadhan A, Monlux K. Preliminary findings of a telehealth approach to parent training in autism. J Autism Dev Disord 2013 Dec;43(12):2953-2969. [doi: 10.1007/s10803-013-1841-8] [Medline: 23677382]
13. Smith KB, Humphreys JS, Wilson MG. Addressing the health disadvantage of rural populations: how does epidemiological evidence inform rural health policies and research? Aust J Rural Health 2008 Apr;16(2):56-66. [doi: 10.1111/j.1440-1584.2008.00953.x] [Medline: 18318846]
14. Struber J. Recruiting and retaining allied health professionals in rural Australia: why is it so difficult. The Internet Journal of Allied Health Sciences and Practice 2004;2(2):1-8.
15. Iacono T, Humphreys J, Davis R, Chandler N. Health care service provision for country people with developmental disability: an Australian perspective. Res Dev Disabil 2004;25(3):265-284. [doi: 10.1016/j.ridd.2003.09.001] [Medline: 15134792]
16. Hutton AM, Caron SL. Experiences of families with children with autism in rural New England. Focus Autism Other Dev Disabl 2005 Jan 01;20(3):180-189. [doi: 10.1177/10883576050200030601]
17. Hoogsteen L, Woodgate RL. Embracing autism in Canadian rural communities. Aust J Rural Health 2013 Jun;21(3):178-182. [doi: 10.1111/ajr.12030] [Medline: 23782286]
18. Murphy MA, Ruble LA. A comparative study of rurality and urbanicity on access to and satisfaction with services for children with autism spectrum disorders. Rural Special Education Quarterly 2012;31(3):3-11.
19. Reese RM, Braun MJ, Hoffmeier S, Stickle L, Rinner L, Smith C, et al. Preliminary evidence for the integrated systems using telemedicine. Telemed J E Health 2015 Jul;21(7):581-587. [doi: 10.1089/tmj.2014.0124] [Medline: 25898157]
20. Horlin C, Falkmer M, Parsons R, Albrecht MA, Falkmer T. The cost of autism spectrum disorders. PLoS One 2014;9(9):e106552 [FREE Full text] [doi: 10.1371/journal.pone.0106552] [Medline: 25191755]
21. 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]
22. Hoogsteen L, Woodgate RL. The lived experience of parenting a child with autism in a rural area: making the invisible, visible. Pediatr Nurs 2013;39(5):233-237. [Medline: 24308088]
23. McConachie H, Diggle T. Parent implemented early intervention for young children with autism spectrum disorder: a systematic review. J Eval Clin Pract 2007 Feb;13(1):120-129. [doi: 10.1111/j.1365-2753.2006.00674.x] [Medline: 17286734]
24. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
25. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]

26. Boisvert M, Lang R, Andrianopoulos M, Boscardin ML. Telepractice in the assessment and treatment of individuals with autism spectrum disorders: a systematic review. Dev Neurorehabil 2010;13(6):423-432. [doi: 10.3109/17518423.2010.499889] [Medline: 20887200]
27. Hepburn SL, Blakeley-Smith A, Wolff B, Reaven JA. Telehealth delivery of cognitive-behavioral intervention to youth with autism spectrum disorder and anxiety: a pilot study. Autism 2016 Feb;20(2):207-218 [FREE Full text] [doi: 10.1177/1362361315575164] [Medline: 25896267]
28. McDuffie A, Oakes A, Machalicek W, Ma M, Bullard L, Nelson S, et al. Early language intervention using distance video-teleconferencing: a pilot study of young boys with fragile X syndrome and their mothers. Am J Speech Lang Pathol 2016 Feb;25(1):46-66. [doi: 10.1044/2015_AJSLP-14-0137] [Medline: 26502382]
29. Ramdoss S, Lang R, Mulloy A, Franco J, O'Reilly M, Didden R, et al. Use of computer-based interventions to teach communication skills to children with autism spectrum disorders: a systematic review. J Behav Educ 2010 Nov 4;20(1):55-76. [doi: 10.1007/s10864-010-9112-7]
30. Terry M. Telemedicine and autism: researchers and clinicians are just starting to consider telemedicine applications for the diagnosis and treatment of autism. Telemed J E Health 2009 Jun;15(5):416-419. [doi: 10.1089/tmj.2009.9965] [Medline: 19548820]
31. Gwynette MF, Morriss D, Warren N, Truelove J, Warthen J, Ross CP, et al. Social skills training for adolescents with autism spectrum disorder using facebook (project rex connect): a survey study. JMIR Ment Health 2017 Jan 23;4(1):e4 [FREE Full text] [doi: 10.2196/mental.6605] [Medline: 28115297]
32. Northrup CM, Lantz J, Hamlin T. Wearable stress sensors for children with autism spectrum disorder with in situ alerts to caregivers via a mobile phone. iproc 2016 Dec 14;2(1):e9. [doi: 10.2196/iproc.6119]
33. Meadan H, Daczewitz ME. Internet-based intervention training for parents of young children with disabilities: a promising service-delivery model. Early Child Dev Care 2014 Apr 22;185(1):155-169. [doi: 10.1080/03004430.2014.908866]
34. Diggle T, McConachie HR, Randle VR. Parent-mediated early intervention for young children with autism spectrum disorder. Cochrane Database Syst Rev 2003(1):CD003496. [doi: 10.1002/14651858.CD003496] [Medline: 12535477]
35. Wainer AL. Increasing the dissemination of an evidence-based ASD intervention via a telehealth parent training program. Michigan: Michigan State University; 2014. URL: https://etd.lib.msu.edu/etd/2934/datastream/OBJ/view [accessed 2015-10-31] [WebCite Cache ID 6qM4tXGzV]
36. Statistics Canada. 2015. From urban areas to population centres URL: http://www.statcan.gc.ca/eng/subjects/standard/sgc/notice/sgc-06 [accessed 2016-03-31] [WebCite Cache ID 6jrwjvcv0]
37. United States Census Bureau. 2015. Urban and rural classification URL: https://www.census.gov/geo/reference/urban-rural.html [accessed 2017-05-10] [WebCite Cache ID 6qM580ccR]
38. Australian Bureau of Statistics. Canberra: Commonwealth of Australia; 2011. Australian standard geographical classification system (ASGC) URL: http://www.abs.gov.au/websitedbs/d3310114.nsf/home/australian+standard+geographical+classification+(asgc) [accessed 2016-03-31] [WebCite Cache ID 6qM5KwvVQ]
39. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV. Washington, DC: American Psychiatric Association; 1994.
40. Kmet LM, Lee RC, Cook LS. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Alberta Heritage Foundation for Medical Research 2004:1-28.
41. National Health Medical Research Council. A guide to the development, implementation and evaluation of clinical practice guidelines. Canberra: Commonwealth of Australia; 1999.
42. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
43. Ingersoll B, Berger NI. Parent engagement with a telehealth-based parent-mediated intervention program for children with autism spectrum disorders: predictors of program use and parent outcomes. J Med Internet Res 2015;17(10):e227 [FREE Full text] [doi: 10.2196/jmir.4913] [Medline: 26443557]
44. Ingersoll B, Wainer AL, Berger NI, Pickard KE, Bonter N. Comparison of a self-directed and therapist-assisted telehealth parent-mediated intervention for children with ASD: a pilot RCT. J Autism Dev Disord 2016 Jul;46(7):2275-2284. [doi: 10.1007/s10803-016-2755-z] [Medline: 26922192]
45. Pickard KE, Wainer AL, Bailey KM, Ingersoll BR. A mixed-method evaluation of the feasibility and acceptability of a telehealth-based parent-mediated intervention for children with autism spectrum disorder. Autism 2016 Oct;20(7):845-855. [doi: 10.1177/1362361315614496] [Medline: 26862084]
46. St. Peter CC, Brunson LY, Cook JE, Subramaniam S, Larson NA, Clingan M, et al. Adherence to discrete-trial instruction procedures by rural parents of children with autism. Behav Intervent 2014 Apr 14;29(3):200-212. [doi: 10.1002/bin.1386]
47. Wacker DP, Lee JF, Dalmau YC, Kopelman TG, Lindgren SD, Kuhle J, et al. Conducting functional communication training via telehealth to reduce the problem behavior of young children with autism. J Dev Phys Disabil 2013 Feb 1;25(1):35-48 [FREE Full text] [doi: 10.1007/s10882-012-9314-0] [Medline: 23543855]
48. Vismara LA, Young GS, Rogers SJ. Telehealth for expanding the reach of early autism training to parents. Autism Res Treat 2012;2012:121878 [FREE Full text] [doi: 10.1155/2012/121878] [Medline: 23227334]
49. Hamad CD, Serna RW, Morrison L, Fleming R. Extending the reach of early intervention training for practitioners: a preliminary investigation of an online curriculum for teaching behavioral intervention knowledge in autism to families and

service providers. Infants Young Child 2010 Jul 1;23(3):195-208 [FREE Full text] [doi: 10.1097/IYC.0b013e3181e32d5e] [Medline: 23504540]
50. Heitzman-Powell LS, Buzhardt J, Rusinko LC, Miller TM. Formative evaluation of an ABA outreach training program for parents of children with autism in remote areas. Focus Autism Other Dev Disabl 2013 Oct 14;29(1):23-38. [doi: 10.1177/1088357613504992]
51. Rogers SJ, Dawson G, Vismara LA. An early start for your child with autism: Using everyday activities to help kids connect, communicate, and learn. New York: Guilford Press; 2012.
52. Mahoney G, Wheeden CA. The effects of teacher style on the interactive engagement of preschool-aged children with disabilities. Journal of Developmental and Learning Disorders 1998;2:293-315.
53. Mahoney G, Powell A, Finger I. The Maternal Behavior Rating Scale. Topics In Earl Child Special Education 1986 Jan 01;6(2):44-56. [doi: 10.1177/027112148600600205]
54. Fenson L, Dale PS, Reznick JS, Thal D, Bates E, Hartung JP, et al. Macarthur Communicative Development Inventories: User's guide and technical manual. San Diego, CA: Singluar Publishing Group, Inc; 1993.
55. Sparrow SS, Cicchetti DV, Balla DA. Vineland Adaptive Behaviour Scales: Survey forms manual. Bloomington, MN: Pearson Assessments; 2005.
56. Ramdoss S, Mulloy A, Lang R, O'Reilly M, Sigafoos J, Lancioni G, et al. Use of computer-based interventions to improve literacy skills in students with autism spectrum disorders: a systematic review. Res Autism Spectr Disord 2011 Oct;5(4):1306-1318. [doi: 10.1016/j.rasd.2011.03.004]

Abbreviations

ASD: Autism spectrum disorder
DSM-V: diagnostic and statistical manual of mental disorders
ICT: information communication technology
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 15.09.16; peer-reviewed by B Ingersoll, B Bie; comments to author 29.12.16; revised version received 14.02.17; accepted 02.03.17; published 14.08.17.

Please cite as:
Parsons D, Cordier R, Vaz S, Lee HC
Parent-Mediated Intervention Training Delivered Remotely for Children With Autism Spectrum Disorder Living Outside of Urban Areas: Systematic Review
J Med Internet Res 2017;19(8):e198
URL: http://www.jmir.org/2017/8/e198/
doi:10.2196/jmir.6651
PMID:28807892

Original Paper

Toward Predicting Social Support Needs in Online Health Social Networks

Min-Je Choi[1], MS; Sung-Hee Kim[2], PhD; Sukwon Lee[3], PhD; Bum Chul Kwon[4], PhD; Ji Soo Yi[5], PhD; Jaegul Choo[1], PhD; Jina Huh[6], PhD

[1]Department of Computer Science and Engineering, Korea University, Seoul, Republic Of Korea

[2]Department of Industrial ICT Engineering, Dong-eui University, Busan, Republic Of Korea

[3]Department of Industrial Engineering, Purdue University, West Lafayette, IN, United States

[4]IBM Thomas J. Watson Research Center, Yorktown Heights, NY, United States

[5]Samsung Electronics Co., Ltd., Seoul, Republic Of Korea

[6]Division of Biomedical Informatics, University of California San Diego, San Diego, CA, United States

Corresponding Author:
Jaegul Choo, PhD
Department of Computer Science and Engineering
Korea University
Room 510A, Science Library Building
145 Anam-ro, Seongbuk-gu
Seoul,
Republic Of Korea
Phone: 82 2 3290 4602
Fax: 82 2 3290 4602
Email: jchoo@korea.ac.kr

Abstract

Background: While online health social networks (OHSNs) serve as an effective platform for patients to fulfill their various social support needs, predicting the needs of users and providing tailored information remains a challenge.

Objective: The objective of this study was to discriminate important features for identifying users' social support needs based on knowledge gathered from survey data. This study also provides guidelines for a technical framework, which can be used to predict users' social support needs based on raw data collected from OHSNs.

Methods: We initially conducted a Web-based survey with 184 OHSN users. From this survey data, we extracted 34 features based on 5 categories: (1) demographics, (2) reading behavior, (3) posting behavior, (4) perceived roles in OHSNs, and (5) values sought in OHSNs. Features from the first 4 categories were used as variables for binary classification. For the prediction outcomes, we used features from the last category: the needs for emotional support, experience-based information, unconventional information, and medical facts. We compared 5 binary classifier algorithms: gradient boosting tree, random forest, decision tree, support vector machines, and logistic regression. We then calculated the scores of the area under the receiver operating characteristic (ROC) curve (AUC) to understand the comparative effectiveness of the used features.

Results: The best performance was AUC scores of 0.89 for predicting users seeking emotional support, 0.86 for experience-based information, 0.80 for unconventional information, and 0.83 for medical facts. With the gradient boosting tree as our best performing model, we analyzed the strength of individual features in predicting one's social support need. Among other discoveries, we found that users seeking emotional support tend to post more in OHSNs compared with others.

Conclusions: We developed an initial framework for automatically predicting social support needs in OHSNs using survey data. Future work should involve nonsurvey data to evaluate the feasibility of the framework. Our study contributes to providing personalized social support in OHSNs.

(J Med Internet Res 2017;19(8):e272) doi:10.2196/jmir.7660

KEYWORDS

online health social network; machine learning; gradient boosting trees; prediction models; social media; online health community

Introduction

The social support model [1,2] received substantial interest in the field of medical informatics. According to the model, social support consists of emotional [3] and informational [2] support; the latter can further be specified into experience-based information [4], unconventional information, and medical facts [5]. Online health social networks (OHSN) users exchange emotional support by encouraging and sympathizing with others. Experience-based information includes a user's experience and feelings about previously tried treatments or diet, or symptoms one had to undergo while suffering from a specific illness [6-8]. Unconventional information, though similar to experience-based information, lacks scientific background and comprises more radical approaches and treatments [9]. Fresh information on upcoming medicine or treatments is also included in this category. Lastly, medical facts refer to traditional medical information, such as experiments and other statistical data as well as published writings on illnesses and treatments, such as a doctor's online blog. Patients have their own social support needs, and correctly understanding social support needs and providing adequate measures has positive effects on patients of both mental and physical conditions [10-12].

This model further developed as the Internet became prevalent to the general public, with OHSNs (eg, PatientsLikeMe [13], WebMD [14], Diabetes Daily [15], and Facebook [9,16]) emerging as a scalable platform for social support and improving health behavior. One underlying reason of its success is the variety of generated contents available. Some OHSN users prefer not to interact with others and just get information, while others prefer to bond with other patients via the Internet [17,18]. Unlike blogs or websites, which are updated by a few moderators with sufficient expertise, content in OHSNs is generated by diverse users with different interests and knowledge levels. Posts and threads are created around interesting topics, which provide better feedback compared with the one-way information delivery inherent in conventional websites. Also, the type of support both requested and generated in OHSNs are not restricted to refined medical facts, but comprises a wide range varying from unscientific, radical experiments to reassurance from peers. Simply put, OHSNs provide excellent grounds for users to fulfill their social support needs.

However, because of this abundance of information, OHSNs also suffer from several problems regarding the sustainability of a community. Most OHSN users visit these social networks in hope of effectively finding information relevant to their immediate needs. Having to instead search through a plethora of text or not being able to receive any replies to questions lead to exhaustion and frustration, making users visit only sporadically and seldom return. This may lead to shortened retention, a common problem with any technology adoption [19]. Monotonous support services to users will increase attrition [20,21] due to the lack of personalization in the content and the support users are receiving, possibly causing user migration [22-24].

A possible solution for OHSNs to both serve as a social support platform as well as maintain users is to provide customized information on an individual basis. This seemingly idealistic task can be materialized by using the vast amount of data provided by OHSNs, such as user-generated postings, user logs (eg, page visit records), and profiles that we can use as potential predictive factors for understanding each user. User characteristics in OHSNs include users' self-reported profiles, visiting frequency, contents of users' posts, the posts users have read, and so on. With an increasing effort from hospital institutions providing OHSNs (eg, Mayo clinic [25], patient-powered research networks [16,26]), medical records data can further be used to predict users' personalized social support needs. By applying state of the art machine learning and analysis techniques on these data sources, we can create a data-driven framework that accurately predicts users' social support types and needs, and then provides useful information or advice based on such prediction results.

However, an unrefined prediction model based on all available data is likely to suffer from high computational cost as well as low accuracy. There is a strong need to identify which features are important for prediction. Also, such raw data is difficult to merge with previously obtained knowledge. In this aspect, surveys are more effective in understanding unknown user behavior as they can be designed based on previous understandings as well as pinpointing possible behavior types.

Thus, in this study, we developed a framework for predicting users' social support needs using a more refined form of data from a carefully designed survey [17]. The survey was based on in-person interviews aimed to identify the most important aspects of OHSN user behaviors. We developed a prediction model using the survey outcomes and evaluated the results to discover which different data types potentially available in actual OHSNs best represent the behavioral aspects of known social needs. Our findings help channel our efforts toward data types critical in generating tailored support for OHSN users.

Methods

Data Collection and Processing

From our prior work, we conducted a Web-based survey about their activities and behaviors in OHSN use [17]. The survey consisted of 21 multiple-choice questions on a 5-point scale ranging from strongly disagree to strongly agree to corresponding statements and 4 open-ended questions. We used the survey results from our previous work [18]. We conducted interviews with OHSN users to identify various characteristics of user needs observed in OHSN and adapted the surveys based on existing validated social support inventories [27,28] and our interview results.

Survey targets were OHSN users interested in chronic diseases, including: HIV, cancer, diabetes, weight management, heart disease, attention deficit hyperactivity disorder, Parkinson's disease, fibromyalgia, depression, and bipolar disorders. Previous results show that patients with chronic diseases increasingly seek more social support in OHSNs [29]. Because our intention was to predict social support needs, and because

this is a first study to predict information needs based on meta information of OHSN use, we decided to scope our work within a previously known group of OHSN users who have increased social support needs—those interested in chronic illness.

We recruited participants from Web-based advertisements including Google and Facebook, and high traffic OHSNs (eg, reddit) suggested from the Google Ranking Algorithm, and we asked the participants to fill out the survey. A total of 184 participants, who have visited any OHSNs at least once in the past and were over 18- years old, responded to the survey.

We encoded the survey responses into a 184 × 38 matrix with 184 respondents, 34 features, and 4 outcome variables (see Multimedia Appendix 1). There were 21 multiple-choice questions with values assigned from 1 to 5, where 1 corresponds to the strongest level of disagreement and 5 to the strongest level of agreement toward the question. Two coders first discussed the coding scheme of grouping similar answers and assigning a categorical value for each group of answers using the first 10 responses of the 4 open-ended questions. Afterward, the coders independently coded all of the rest of the responses, and then compared their responses to reach an agreed result if there were any discrepancies. The resulting agreed results became our coding results

The survey questions included 5 categories: reading, posting, demographic, role, and values sought. The first 4 categories are indicators for collecting features of various user behaviors in OHSNs. These questions contain posting or reading preferences, one's demographic status, self-perceived roles within a community, and so on. On the other hand, values sought questions contain information on the prediction outcome variables. To collect 4 prediction outcome variables on the 4 social support needs, we asked the following question: "The reason for visiting the online health support group is," which had 4 multiple-choice questions asking whether users visit OHSNs to obtain emotional support (ExchangeEmo), experience-based information of others, (HearExp), unconventional information (GetUnusualInfo), or medical facts (SpecificSearch) (See Multimedia Appendix 1 for exact survey questions and corresponding responses).

To develop the 4 binary classifiers predicting social support needs, for each classifier, we assigned TRUE class values to the users who responded 5 (strongly agree) or 4 (agree) to each values sought question. The rest of the users were assigned a NEGATIVE class. For instance, if 1 respondent rated 5 for ExchangeEmo but 1 for HearExp, the user was classified as TRUE for the ExchangeEmo classifier and NEGATIVE for the HearExp classifier. We can assume that this person has a strong need for sharing emotional peer support, but is less inclined toward hearing from the experiences of others. The rationale behind this selection is that lowering the threshold of positive values to 3 dilutes the strength of characteristics inherent in features representing a specific social support need, while further increasing the threshold to separate scores of 4 and 5 increases bias. Also, excluding 3 from classification removes at least 53 of 184 training data samples (28.8%) of the training data (See Multimedia Appendix 1 – SpecificSearch). Therefore, this selection method was adopted to both preserve strong characteristics and data size.

Classification Algorithms

We performed our classification task using a wide variety of machine learning algorithms, which have been heavily applied to binary classification. We selected gradient boosting tree (GBT) [30-32], support vector machines (SVM) [33], decision tree [34], random forest [35], and logistic regression [36] as classifiers to compare the evaluation results. We built 4 models (predicting each of the 4 social support needs) for each classifier.

We used the area under the receiver operating characteristic (ROC) curve (AUC) value as the performance measure [37]. A large AUC value over 0.8 denotes a reasonably good prediction rate [38], while an AUC value of 0.5 is equal to the predictability of a purely random output such as a coin flip. One advantage of adopting the AUC measure is that it is invariant of data imbalance. It measures how well positive data samples are ranked higher than negative samples, and produces reliable results even in positively or negatively skewed datasets. In addition, the AUC measure contains information on all possible precision and recall value pairs as it uses various threshold values of the classifier output about whether a value is positive or negative [39]. For these reasons, we used AUC throughout our experiments.

We included the correlations between the features and each outcome variable to report the direction of association missing in the AUC values (see Multimedia Appendix 3).

Each prediction model followed the following steps, where averaging the outcomes from multiple epochs was performed to compensate for the relatively small sample size:

1. Randomly split data into 70/30, where 70.1% (129/184) of samples are used as a training set and the remaining 29.9% (55/184) as a test set.
   Train on the training set from (1).
   Run predictions on the test set from (1), and compare the predicted scores with the true scores of the outcome variable.
   Calculate AUC score of the predictability shown in (3).
2. Iterate steps (1)~(4) 50 times, with each iteration producing an AUC score.
3. Average the resulting 50 AUC scores.

We then conducted multiple experiments on each outcome variable, using only 1 feature at a time. The AUC values here corresponded to the strength of an individual feature's predictability toward each social support need.

We used 1 feature at a time to measure the prediction power of 1 single variable on the outcome, instead of finding the best subset of features for maximal prediction accuracy. While feature selection methods may improve overall prediction results, this method does not match our aim toward discovering which individual features contain high predictive powers. Furthermore, advanced machine learning algorithms, such as GBT, random forests, and others, can properly use highly important variables while ignoring unimportant variables, given a large number of variables.

All the data training and prediction processes were performed with Matlab R2015a on an Intel Core i7-6700K CPU supported by a Windows 8.1 64-bit environment.

Results

Prediction Model Performance Results

As shown in Table 1, the AUC scores ranged between 0.61 and 0.90 for all support needs. Of all models, GBT consistently produced superior AUC scores compared with others, except for predicting emotional support where it ranked second to SVMs. For this reason, we concluded that GBT could most accurately predict a user's social support need given survey features, and carried out subsequent analyses using this model.

Feature Analysis Results

Figure 1 shows the results of feature analysis. Each social support needed relies on a distinct set of features for prediction. Again, AUC scores were used to indicate the strength of a feature in its predictability of a user's social support need. To provide a better understanding on which features are of actual importance, we devoted the rest of this section to interpreting the results of these AUC scores. We also present the individual scores to provide better information (see Multimedia Appendix 2).

Two features out of the Demographic Information category (PatientOrCare and Satisfaction) proved to be significant in predicting social support needs, especially if the user required experience-based information. The high prediction score for experience-based information in the 'PatientOrCare' feature indicates that users are more likely to visit OHSNs in search of others' experiences when they are patients themselves rather than caregivers. The 'Satisfaction' feature describes that OHSN users seeking for the fourth social support need, medical facts, have a substantially lower satisfaction level upon using OHSNs compared with other users. Meanwhile, whether a user was introduced to an OHSN through recommendation or Web search did not affect the predictability of any of the 4 social support needs (FindSearch/FindRecommend). Features such as users' sex, whether someone nearby or a doctor introduced a user to an OHSN, also did not strongly influence the predictability for any of the 4 social support needs.

Features from the Reading Behavior category showed notable characteristics that differ between social support groups. For example, users who trust others were most likely to seek emotional support (TrustOthers), while those who search for evidence in postings were likely to search for experience-based or fact-based information (NeedEvidence).

The 'Need Evidence' feature was particularly important in that it functioned differently in predicting emotional and informational support. While searching for evidence was not an important feature for predicting emotional support, it largely affected the predictability of the 3 variants of informational support: experience-based information, unconventional information, and medical facts. Users either searching for the experiences of others or medical facts are bound to be skeptical of what was posted and would carefully look for cues of evidence that support the validity of the posted content. However, this feature was less important to those seeking unconventional information, as for such people the type of information they were searching for may often lack concrete scientific evidence but still be worth knowing.

The Posting Behavior category contained the strongest individual features when it came to predicting emotional support. In fact, 8 of 10 features from this category had the strongest predictability when predicting emotional support needs. Users seeking this social support type were most likely to post frequently (PostFreq), ask questions (AskQ), and share their personal stories and emotions with others (SharePersonal/ShareEmo).

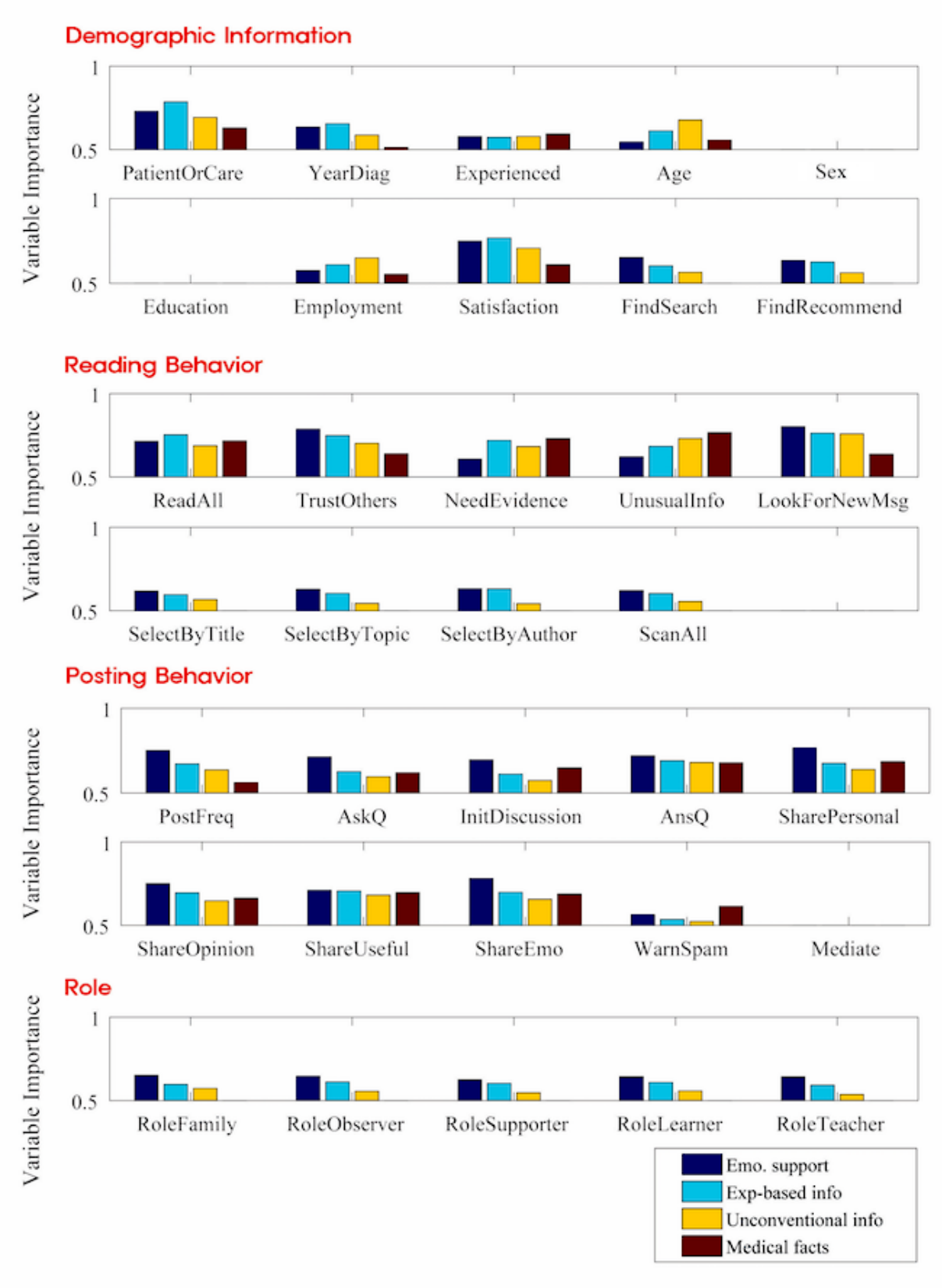
The greatest significance of this category lies in the fact that these features not only help predict which type of social support a user seeks, but even depict how active he/she is as a community member. The number of posts and threads he/she posts rather than reading patterns often determines the activeness of a user within an online community. How frequently one posts (PostFreq), how often one starts up conversations (InitDiscussion), how often one shares opinions with others (ShareOpinion) all serve as measuring sticks. From this perspective, posting behavior can function as an indicator to measure what drives users to actively participate in online communities. Overall, results from this category clearly represent that users of this support need are likely to be the most active users within a community.

The Role category was based on the understanding of users themselves on their self-perceived roles within an OHSN. Contrary to our expectations, features of the Role category did not capture distinguishable characteristics among different social support needs. All role features showed the highest values in predicting users of emotional support.

Table 1. Comparison of different classifier models for obtaining AUC values.

	Classification algorithms				
	Gradient boosting tree	Support vector machine	Decision tree	Random forest	Logistic regression
Emotional support	0.87	0.89	0.78	0.85	0.77
Experience-based information	0.86	0.80	0.76	0.83	0.74
Unconventional information	0.80	0.75	0.69	0.75	0.66
Medical facts	0.83	0.72	0.67	0.83	0.61

Figure 1. Individual feature importance on each social support need.



Discussion

Qualitative Analysis on Prediction Results

Prediction scores, AUC scores in this study, show how our model is capable of accurately predicting the 4 social support needs using features from demographic information, reading behaviors, posting behaviors, and self-perceived roles. Using our model, OHSN administrators can categorize their users based on their social support needs. They can better understand whether patients visit their communities in need of emotional support, experience-based knowledge, or other needs, and provide tailored measures to help users fulfill such desires. Our results showed that the posting activities, more so than their perceived roles or demographic information, had high predictability for social support needs.

Our features can be used to identify the characteristics of active users. Likewise, they can also pinpoint the characteristics for user groups that represent inactive and stationary users. Users with the social support need for medical facts closely fit the description of inactive users, having the lowest prediction scores when using "Satisfaction," "TrustOthers," "LookForNewMsg," and "PostFreq" features. Compared with other support needs, users seeking for medical facts were less likely to read new messages (LookForNewMsg), trust what is posted by others (TrustOthers), and post threads on community boards (PostFreq). In sum, users with the social support need of medical facts were less likely to express themselves in OHSNs compared with other users; thus, partially contributing to the retention or user migration problems mentioned earlier. This knowledge can be used to inform OHSN moderators concerned with the activity level within a community.

Our approach discovering the relationship between OHSN usage patterns and social support types is also shown in work by Wang et al. [40]. Predicted user participation levels using posting behaviors and log data to find that companionship within members lead to low attrition rates [41] found that patients with depression frequently seek interaction with others, which lead to improvements in emotional conditions. Yet these studies have not investigated directly predicting how much each user wants a particular social support need. Predictions using data collected from different time periods will help researchers track the transitioning nature of support needs, such as how users in search of medical facts may gradually shift toward providing and seeking emotional support and experience-based knowledge as they become settled within a community.

Survey Data, User-Generated Data, and User Log Data

Our framework validated through survey data provides a starting point to use user-generated data and log data in OHSNs in predicting social support needs. Even though we used the survey data in this study, we believe that it is worthwhile to compare with other data from different sources. Thus, we compared the advantages and disadvantages of using survey data, user-generated data, and log data in OHSNs for predicting social support needs (see Table 2).

Survey data, as seen from our study, collects direct responses from the OHSN users, such as one's level of agreement or opinion on a particular characteristic. While surveys can cover all necessary features required for prediction directly from the participants, they are costly in data collection. Response rate and completeness are also challenging factors.

User-generated data include posting behaviors, post contents, and their associated log data; from this data, we can generate total word count, sentiment of the post, and posting frequency. These data, in contrast to the survey data, can be easily collected. OHSNs over different resources allow users to continuously interact with each other and produce contents, which directly represent their imminent needs. Such information can be collected over time via text crawling, but requires expertise on natural language processing and data mining to be applied for prediction.

The last data form, user log data, provides click and page view information, search history, and connection time. The vastness of this type of data affords unique opportunities for new discoveries. Predictions based on the objective datasets, such as user logs, will give more realistic results that reflect the real user intentions compared with survey data. The challenge is that such data are usually difficult to obtain because it requires a proprietary access.

Given that user-generated data and user-log data are unstructured and noisy, our work provides guidelines on what features should researchers focus in considering for accurate prediction. For instance, user postings can be analyzed using natural language processing techniques, such as sentiment analysis and topic modeling to see if the writer requires emotional support. Search history logs can provide information on what types of postings the user is interested in. These processes can boost the prediction performance further and provide more insights.

Table 2. Comparison of different data sources for prediction in OHSNs

Data source	Survey data (high-level data)	User-generated data (mid-level data)	User log data (low-level data)
Effort required to collect data			
	Design questionnaires	Perform text mining	Extract data from server database
	Conduct surveys	Apply natural language processing on text	
Data generation rate			
	Slow	Fast	Instantaneous
	Need to conduct new survey to get recent data	Hundreds of posts written by users everyday	New generated with every user action (eg, access time, search history)
Interpretability			
	Very easy to understand	Relatively easy to understand	Difficult to derive meaning from raw data
	Questions directly suited to user's intentions	Requires data processing to extract features from long texts	Requires insight on what features to obtain from given data
Data types			
	Numerical data (eg, scale of 1~10)	Text data (eg, title, user posts, comments)	Periodical data (eg, access time)
	Demographic information (eg, age, sex, region)		Demographic information (eg, user profile information)
	Text data for open-ended questions		Hypertext data (eg, accessed links)
			Text data (eg, keywords typed in for search)
Obtainable characteristics			
	A user's (dis)agreement toward a particular characteristic	Words that represent a user's main interests or concerns	Visiting frequency
	Open-ended answers toward a question	Response to a particular article	Reading preference
			Search preference

Limitations and Future Work

Our sample size was relatively small for prediction and our model was based on the survey data, which is due to the low-response rate from the survey recipients [17,42]. Nonetheless, our high-quality data from survey responses in this study can be used as a good example of users' social needs and other characteristics, which can be predicted using other data sources in future work.

Another potential downfall of surveys is that participants might conceal their true thoughts in fear of being evaluated by others. Although it is challenging to identify the level of honesty in each participant's responses, we can assume the consistent patterns in our prediction results serve as proof that the majority of survey participants completed their survey truthfully.

Our prediction model was less effective in predicting social support needs using results from open-ended questions (FindSearch, FindRecommend, SelectByTitle, SelectByTopic, SelectByAuthor, ScanAll, and questions from the Role category). Not only were individual prediction scores low compared with multiple-answer questions (see Figure 1), but all prediction results showed a regular pattern of decrease as they moved from predicting emotional support to medical facts.

The low-prediction scores are a result of low participation in answering open-ended questions. Unlike multiple-answer questions, the response rates of open-ended questions were in general under 50% (see Multimedia Appendix 1). Although GBT is capable of using features with missing values [31], one cannot expect significant prediction scores when using features with such handicaps.

Future work should also include expanding our data size by collecting features we found useful in this model from various sources, such as user-generated data and log data. A larger dataset means that cross-validation and other techniques can be applied to further increase accuracy. We can also expand the search scope to patients with acute diseases, who tend to show more information-oriented needs compared with those in chronic conditions.

Conclusion

We developed a technical framework to predict the social support needs of OHSN users using users' values and reported

OHSN usage patterns based on survey data. We found dominant features that contributed to successful predictions, not only in predicting a user's desired support needs but also in his/her level of participation. We showed how different granularity of data around OHSN use can be collected and used to make further predictions on OHSN users' social support needs. We also presented strategies for OHSN administrators to identify the characteristics of users and what values they seek.

Our research contributes to not only understanding the different types of OHSN users but also accurately classifying them according to usage patterns. We thus provide a stepping-stone to understanding what features are found to be important in predicting social support needs and what data sources are realistic in being used as a training data for constructing a prediction model. The value of our methodology lies in assisting administrators and moderators by providing them with guidance on what type of support users can benefit most from. We provide a stepping-stone to improving retention in OHSNs. Our study contributes to OHSN as an intervention tool to improve health behavior and social support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey questions and results.

[PDF File (Adobe PDF File), 61KB - jmir_v19i8e272_app1.pdf]

Multimedia Appendix 2

Area under ROC curve (AUC) scores of individual features and target variables.

[PDF File (Adobe PDF File), 18KB - jmir_v19i8e272_app2.pdf]

Multimedia Appendix 3

Correlation scores of individual features and target variables.

[PDF File (Adobe PDF File), 17KB - jmir_v19i8e272_app3.pdf]

References

1. Uchino B. Social Support and Physical Health: Understanding the Health Consequences of Relationships. New Haven: Yale University Press; 2004.
2. Langford CP, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. J Adv Nurs 1997;25:95-100. [Medline: 9004016]
3. Slevin M, Nichols S, Downer S, Wilson P, Lister T, Arnott S, et al. Emotional support for cancer patients: what do patients really want? Br J Cancer 1996;74:1275-1279. [doi: 10.1038/bjc.1996.529]
4. Tilden VP, Weinert C. Social support and the chronically ill individual. Nurs Clin North Am 1987;22:613-620. [Medline: 3649795]
5. Krause N. Journal of Gerontology. 1986. Social Support, Stress, and Well-Being Among Older Adults URL: https://academic.oup.com/geronj/article-abstract/41/4/512/625528/Social-Support-Stress-and-Well-Being-Among-Older?redirectedFrom=fulltext [accessed 2017-07-09] [WebCite Cache ID 6rqFPUFD7]
6. Chelf JH, Agre P, Axelrod A, Cheney L, Cole DD, Conrad K, et al. Cancer-related patient education: an overview of the last decade of evaluation and research. Oncol Nurs Forum 2001;28:1139-1147. [Medline: 11517847]
7. Hwang KO, Trickey AW, Graham AL, Thomas EJ, Street RL, Kraschnewski JL, et al. Acceptability of narratives to promote colorectal cancer screening in an online community. Prev Med 2012;54:405-407 [FREE Full text] [doi: 10.1016/j.ypmed.2012.03.018] [Medline: 22498021]
8. Powell JA, Lowe P, Griffiths FE, Thorogood M. A critical analysis of the literature on the Internet and consumer health information. J Telemed Telecare 2005;11(Suppl 1):41-43. [doi: 10.1258/1357633054461642] [Medline: 16035990]
9. Erfani S, Abedin B, Daneshgar F. A qualitative evaluation of communication in Ovarian Cancer Facebook communities. 2013 Presented at: International Conference on Information Society (i-Society); June 24 - 26, 2013; Toronto, Canada p. 270-272.
10. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010;7:e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
11. Akey JE, Rintamaki LS, Kane TL. Health Belief Model deterrents of social support seeking among people coping with eating disorders. J Affect Disord 2013;145:246-252. [doi: 10.1016/j.jad.2012.04.045] [Medline: 22840616]

12. Kurtz L. Self-Help and Support Groups: A Handbook for Practitioners. Thousand Oaks, CA: Sage; 1997.
13. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008;10:e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
14. Nath C, Huh J, Adupa AK, Jonnalagadda SR. Website sharing in online health communities: a descriptive analysis. J Med Internet Res 2016;18:e11 [FREE Full text] [doi: 10.2196/jmir.5237] [Medline: 26764193]
15. Dobson R, Whittaker R, Murphy R, Khanolkar M, Miller S, Naylor J, et al. The use of mobile health to deliver self-management support to young people with type 1 diabetes: a cross-sectional survey. JMIR Diabetes 2017;2:e4. [doi: 10.2196/diabetes.7221]
16. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME. It's not that I don't have problems, I'm just not putting them on Facebook: Challenges and opportunities in using online social networks for health. In: Proceedings of ACM 2011 Conference on Computer Supported Cooperative Work.: ACM; 2011 Presented at: Proceedings of the ACM Conference on Computer supported Cooperative Work; March 19-23, 2011; Hangzhou, China. [doi: 10.1145/1958824.1958876]
17. Huh J, Kwon BC, Kim S, Lee S, Choo J, Kim J, et al. Personas in online health communities. J Biomed Inform 2016;63:212-225. [doi: 10.1016/j.jbi.2016.08.019] [Medline: 27568913]
18. Huh J, Kwon B, Choo J, Kim S, Yi J. Coddlers, scientists, adventurers, and opportunists: personas to inform online health community development. arXiv:1506.09191 2015:1.
19. Otto P, Simon M. Dynamic perspectives on social characteristics and sustainability in online community networks. Syst Dyn Rev 2008;24:321-347. [doi: 10.1002/sdr.403]
20. McCambridge J, Kalaitzaki E, White IR, Khadjesari Z, Murray E, Linke S, et al. Impact of length or relevance of questionnaires on attrition in online trials: randomized controlled trial. J Med Internet Res 2011;13:e96. [doi: 10.2196/jmir.1733]
21. Poirier J, Cobb NK. Social influence as a driver of engagement in a web-based health intervention. J Med Internet Res 2012;14:e36 [FREE Full text] [doi: 10.2196/jmir.1957] [Medline: 22356829]
22. Xu Y, Yang Y, Cheng Z, Lim J. Retaining and attracting users in social networking services: an empirical investigation of cyber migration. The Journal of Strategic Information Systems 2014;23:239-253. [doi: 10.1016/j.jsis.2014.03.002]
23. Watkins S. What the Migration to Social-Network Sites, Games, and Anytime, Anywhere Media Means for our Future. Boston, MA: Beacon Press; 2009.
24. Kumar S, Zafarani R, Liu H. Understanding User Migration Patterns in Social Media. 2011 Presented at: AAAI'11 Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence; August 8-11, 2011; San Francisco, CA p. 1204-1209.
25. van der Eijk M, Faber MJ, Aarts JWM, Kremer JAM, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013;15:e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
26. Love B, Crook B, Thompson CM, Zaitchik S, Knapp J, Lefebvre L, et al. Exploring psychosocial support online: a content analysis of messages in an adolescent and young adult cancer community. Cyberpsychol Behav Soc Netw 2012;15:555-559. [doi: 10.1089/cyber.2012.0138] [Medline: 22970826]
27. Gottlieb BH, Bergen AE. Social support concepts and measures. J Psychosom Res 2010;69:511-520. [doi: 10.1016/j.jpsychores.2009.10.001] [Medline: 20955871]
28. Ware J, Kosinski M, Dewey J, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. Boston: New England Medical Center; 1993. URL: http://czresearch.com/info/SF36_healthsurvey_ch6.pdf [accessed 2017-07-09] [WebCite Cache ID 6rqG70oGB]
29. Fox S, Duggan M. Health Online 2013. Washington, D.C: Pew Internet & American Life Project; 2013.
30. Becker C, Rigamonti R, Lepetit V, Fua P. Supervised feature learning for curvilinear structure segmentation. Med Image Comput Comput Assist Interv 2013;16(Pt 1):526-533. [Medline: 24505707]
31. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Springer series in statistics. Berlin, Germany: Springer; 2001.
32. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21 [FREE Full text] [doi: 10.3389/fnbot.2013.00021] [Medline: 24409142]
33. Chang C, Lin C. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011 Apr 01;2:1-27. [doi: 10.1145/1961189.1961199]
34. Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. J Chemometrics 2004;18:275-285. [doi: 10.1002/cem.873]
35. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:22.
36. Fan R, Chang K, Hsieh C, Wang X, Lin C. LIBLINEAR: A library for large linear classification. J Mach Learn Res 2008;9:1874. [doi: 10.1145/1390681.1442794]
37. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36. [doi: 10.1148/radiology.143.1.7063747] [Medline: 7063747]
38. Metz C. Basic principles of ROC analysis. Seminars in Nuclear Medicine 1978:298.
39. Fawcett T. ROC graphs: notes and practical considerations for researchers. Machine Learning 2004;31:1-38.

40. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. J Med Internet Res 2017;19:e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]
41. Park A, Conway M. Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community. J Med Internet Res 2017;19:e71 [FREE Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]
42. Leece P, Bhandari M, Sprague S, Swiontkowski MF, Schemitsch EH, Tornetta P, et al. Internet versus mailed questionnaires: a randomized comparison (2). J Med Internet Res 2004;6:e30 [FREE Full text] [doi: 10.2196/jmir.6.3.e30] [Medline: 15471756]

Abbreviations

AUC: area under receiver operating characteristic curve
OHSN: online health social networks
GBT: gradient boosting tree
ROC: receiver operating characteristic
SVM: support vector machines

Edited by G Eysenbach; submitted 09.03.17; peer-reviewed by A Benis, X Wang, J Li; comments to author 30.03.17; revised version received 05.05.17; accepted 23.05.17; published 02.08.17.

Please cite as:
Choi MJ, Kim SH, Lee S, Kwon BC, Yi JS, Choo J, Huh J
Toward Predicting Social Support Needs in Online Health Social Networks
J Med Internet Res 2017;19(8):e272
URL: http://www.jmir.org/2017/8/e272/
doi:10.2196/jmir.7660
PMID:28768609

Original Paper

Engagement as a Driver of Growth of Online Health Forums: Observational Study

Rahul Gopalsamy[1], MSc; Alexander Semenov[2], PhD; Eduardo Pasiliao[3], PhD; Scott McIntosh[4], PhD; Alexander Nikolaev[1], PhD

[1]Social Optimization Laboratory, Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY, United States

[2]SOMEA Group, Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

[3]Munitions Directorate, Air Force Research Laboratory, Eglin, FL, United States

[4]Department of Public Health Sciences, University of Rochester, Rochester, NY, United States

Corresponding Author:
Alexander Nikolaev, PhD
Social Optimization Laboratory
Department of Industrial and Systems Engineering
University at Buffalo
312 Bell Hall
Buffalo, NY, 14221
United States
Phone: 1 716 645 4710
Fax: 1 716 645 3302
Email: anikolae@buffalo.edu

Abstract

Background: The emerging research on nurturing the growth of online communities posits that it is in part attributed to network effects, wherein every increase in the volume of user-generated content increases the value of the community in the eyes of its potential new members. The recently introduced metric *engagement capacity* offers a means of quantitatively assessing the ability of online platform users to engage each other into generating content; meanwhile, the quantity *engagement value* is useful for quantifying communication-based platform use. If the claim that higher engagement leads to accelerated growth holds true for online health forums (OHFs), then engagement tracking should become an important tool in the arsenal of OHF managers. Indeed, it might allow for quantifying the ability of an OHF to exploit network effects, thus predicting the OHF's future success.

Objective: This study aimed to empirically analyze the relationship between internal OHF use (quantified using *engagement measurement)*, and external growth.

Methods: We collected data from 7 OHFs posted between the years 1999 and 2016. Longitudinal analyses were conducted by evaluating engagement in the OHFs over time. We analyzed 2-way causality effects between the engagement value and metrics evaluating OHF growth using Granger causality tests. User activity metrics per week were correlated with engagement metrics, followed by linear regression analyses.

Results: Observational data showed a 1-way causal relationship between the OHF engagement value and reach (P=.02). We detected a 2-way causal relationship between the engagement value and delurking, with further analysis indicating that the engagement value was more likely to cause delurking (P<.001 with lag 2; for the reverse hypothesis, P=.01 with lag 2). Users who engaged each other more were more likely (up to 14 times, depending on how much one user engaged another) to develop personal connections. Finally, we found that the more engaging an OHF user was in a given week, the more likely (up to 2 times, depending on their ability to engage others) they were to remain active in the OHF in the following week.

Conclusions: This study supports the claim that network effects play an important role in accelerating OHF growth, opening the door to exploiting these effects in calculated ways. In such efforts, engagement metrics can be used to monitor the "health" of an OHF and to identify the users most important to its success.

(J Med Internet Res 2017;19(8):e304) doi:10.2196/jmir.7249

KEYWORDS

online health forum; online health community; engagement; engagement capacity; superuser; eHealth; social network analysis

Introduction

Background

Online health forums (OHFs) enable computer-mediated communication on health and health-related issues through the Internet. People use OHFs to seek emotional support, exchange information, ask for help, or simply become part of a community [1]. When the members of an OHF communicate with each other in forum threads, they generate content, which generally benefits other members. For example, members may provide emotional support to each other by sharing personal experiences and stories that may make peers open up in turn. Despite the benefits of OHFs [2], many of them fail due to low user activity. It has been observed that, typically, only 1% of OHF users generate most of the OHF content; meanwhile, 90% of the users are observers, also called lurkers, who rarely leave traces of participation [3].

A growing body of literature addresses the issues of sustaining user activity in online communities. To this end, researchers have used social psychology theories to inform the design of platforms [4,5]. The underlying theme of this line of research is to understand the mechanism of human motivation for the creation of public content. The different branches of this research explore, for example, the uniqueness and role of each user's contributions in accomplishing group goals [6] and user attachment to a group [7] as motivating factors to keep contributing to the group. Another large-scale empirical investigation distilled the identity formation principles in online communities [8]. These studies, among others, concur that psychological factors are important drivers behind the dynamics of growth of online user-generated content.

Early quantitative analyses of health-related online communication employed such aggregate metrics as the number of registered users, number of contributed posts per user, and log-in frequency [9,10]. Individual characteristics, such as age, sex, location, and time of joining the platform, have been used to complement activity-based observations in the search for trends in online user activity. More recently, researchers have employed the methods and tools from social network analysis to quantify pairwise user interaction. The nodes in a social network, constructed based on online platform activity data, represent the users of the platform, while the ties (edges) represent the relationships between the users: for example, 2 users may share a tie if they have often posted messages within the same threads or have labeled each other as friends using the platform's interface. The centrality metrics and the structure of the resulting network reveal most well-connected users and subcommunities, the knowledge of which may be informative about the evolution of the platform's user base [11-13]. Note that all the above-mentioned metrics are good for descriptive purposes; however, they have not been successfully used for predicting OHF future activity.

The Network Effect Hypothesis in Conjunction With Online Health Forum Growth

Positive network effect is a phenomenon initially studied in economics [14]. It explains the mechanism behind the process where the value of goods or services tends to increase as more consumers begin to use them: for example, a sale of each additional unit of goods may increase the value of the goods through positive *network effects*, also referred to as *network externalities*. On an online platform, such effects may occur when its users author new content [15]. This proposition has motivated attempts to quantitatively describe the role that network externalities may play in the growth of online communities. The key premise here is that OHFs grow through their existing users: every contributed post (that is responded to) fosters user "bonding" and enhances the positive network externalities. Therefore, by identifying, encouraging, and perhaps incentivizing the most engaging content contributors, OHF managers could become more successful at keeping their prohealth online forums growing, achieving a greater impact on people's health [16].

Prior Work

This study used a method of measuring engagement in OHF communication based on the recently introduced theoretical work on quantifying online users' ability to engage peers in conversations. To this end, Nikolaev et al [17] recently introduced the terms *engagement value* and *engagement capacity*, with cooperative game theory employed for measuring the latter. In particular, they offered a means to quantify users' ability to engage peers. The details of the engagement quantification are provided in the Methods section, accompanied by an illustrative example. They also presented *reach*, introduced within the RE-AIM (reach, efficacy, adoption, implementation, maintenance) program evaluation framework [18], as the key dimension of impact of an online platform; the main premise of their work is that internal growth of a platform—that is, its use that can be quantified through engagement measurement—is responsible for its reach (external growth). However, so far, the latter claim has been presented only as an application-independent proposition, albeit intuitively justifiable. An empirical validation of this proposition, in particular in application to OHFs, would put the research of engagement on more solid ground, simultaneously establishing its practical value.

Objective

The objective of this study was to analyze the relationship between engagement and OHF growth over time, relying on longitudinal data of real-world OHFs. To this end, we formulated 4 research questions, motivated on the one hand by observations of OHF activity, reported in the existing research literature, and on the other hand by the logic behind the utility of engagement measurement.

An OHF full of conversations is more appealing to external readers, compared with one with little content or lots of messages to which no one has responded. Successful and voluminous conversational engagement can be expected to fuel both the higher intensity of user interaction on an OHF and the word-of-mouth effect, with the latter positively affecting the number of people who become aware of the existence of the OHF [19]. In other words, higher internal engagement is likely to attract more potential new users to it, and also makes it appealing in the eyes of those potential users. This leads us to

the first research question (RQ1): Is there a causal relationship between engagement and the reach of an OHF?

Lurkers make up the majority of the user base of any online community. There are multiple reasons behind this phenomenon [20]. For example, lurkers might not find contents that would prompt them to speak up, breaking the silence of passive reading. Higher volumes of engaging content might drive the lurker to consume more, and eventually find something to respond to. More engaging content may also help new users, hesitant to contribute for an extended period of time, to get comfortable with the OHF norms and style of interaction to become an active contributor. Thus, we formulated the second research question (RQ2): Does delurking tend to occur at a higher rate when engagement increases?

OHFs must facilitate bonding between users in order to sustain voluntary participation. According to the common bond theory, users feel motivated to gel as a group because of bond-based personal attachment [7]. In other words, the more friends a user has, the more likely they are to return to the platform to contribute to it further. Therefore, it is of interest to find out whether engagement measurement can be used to inform (track) the development of virtual bonds in OHFs. To this end, we formulated our third research question (RQ3): Is the engagement value delivered by one forum user to another associated with the development of a personal bond (closer virtual relationship) between them?

One may also expect that those users who are successful in engaging their peers feel encouraged by the fact that they manage to help others and, hence, are more likely to keep contributing. To explore this relationship quantitatively, we formulated the fourth research question (RQ4): Does the success in engaging peers motivate a user to remain active in an OHF?

Methods

Engagement Metrics

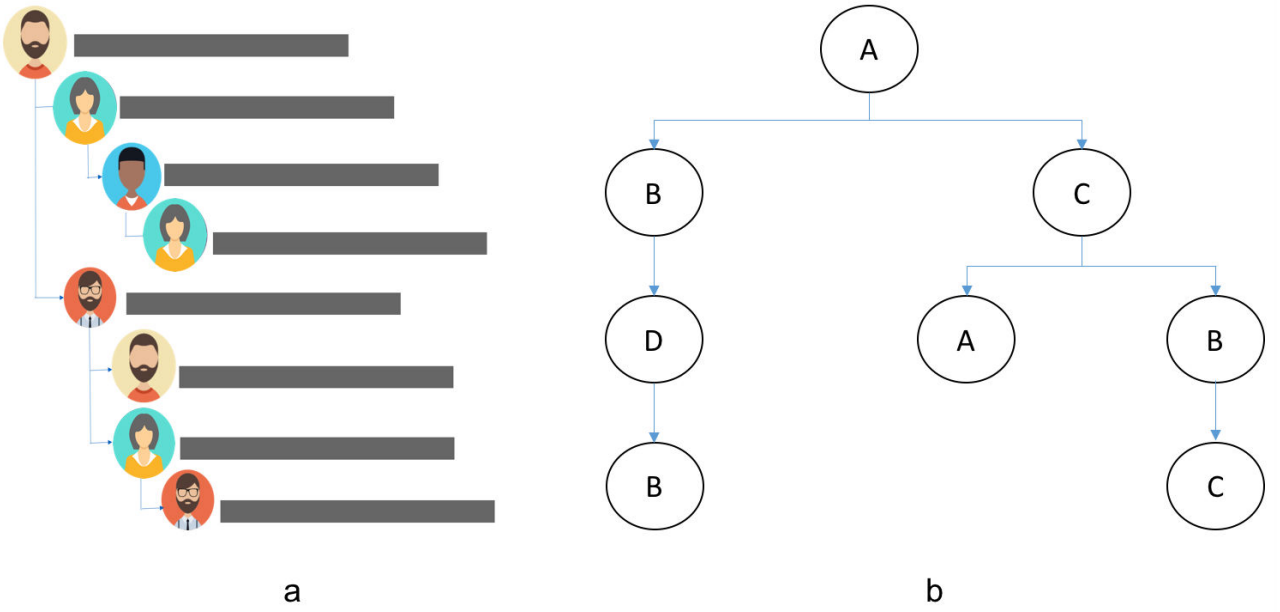
The value of an OHF is generated cooperatively by its users, as they communicate (ie, they author posts). In other words, each user deserves a credit for contributing to the overall value of an OHF (ie, the content it hosts, and consequently, the value it brings to the society). Figure 1 (part a) shows an example of the possible structure of a communication thread in an OHF. This thread can be represented as a network of directed relationships (see Figure 1, part b) reflecting which post "attracted" which. The metric engagement capacity is designed to quantify the credit that each user has earned for engaging their peers, taking into account the flow of communication between them. Note that this credit amount, if computed over a long time (over multiple threads), may also be indicative of the user's ability to engage peers in the future.

The engagement capacity of a user is viewed as their share in the overall forum's ability to attract posts, based on all the threads the user has contributed. It is assumed that engagement capacity can be positive only if a user's content is responded to by other users in the forum. Thus, passive readers and users whose posts do not generate any response do not add to the engagement value of the forum.

To calculate the engagement capacity for each user, all of the engaging *subthreads* in a given OHF must be identified. Every path in the communication network (as in Figure 1, part b) that begins at the root in such a directed graph is called a subthread. Each post, submitted in response to any prior post in a subthread, increments the OHF's overall engagement value by 1 unit. This 1 unit is shared by all the users who participated in the subthread that brought this 1-unit increase (in other words, attracted this new post). For example, the engagement value of 1 generated due to the response of user B to the subthread ABDB is shared between the users A, B, and D. Note that user B in part *self-engages* in this case.

Further, just as OHF users differ by their needs (eg, in seeking information vs emotional support [21,22]), some users may be more successful in engaging a certain set of peers over others. The metric targeted engagement capacity takes this into account and quantifies the total credit allocated to user *i* for successfully engaging user *j*; note that engagement-based relationships are *directed* and nonsymmetric. To put this in a context, for example, for answering RQ3, one can set to check whether targeted engagement can predict the formation of personal connections between users: to this end, the targeted engagement values and the events corresponding to the instances of personal connection formation can be monitored over time.

Figure 1. An online health forum communication thread: an illustrative example. (a) Possible structure of a communication thread, (b) represented as a network of directed relationships.



The Technical Side of Engagement Quantification

To explain the exact formula for the engagement capacity computation, observe how the thread in Figure 1 has been developed. User A is the creator of the thread. First, user A gets 1 unit of engagement value for each of the direct responses of users B and C to this post. Further, when user D responds to user B's post, the generated unit of engagement is shared between users A and B. Note that user C's engagement value is not affected by D's response, as C is not a part of the subthread to which D is responding. The exact share of each subthread contributor in a newly earned unit of engagement value is derived relying on the logic of cooperative game theory [23-25], according to which the OHF users can be viewed as playing games on *k*-coalitions, which are connected ordered sequences of player *appearances* [17].

In the context of a forum of N users with P subthreads, let $\Omega^K(N)$ represent the set of all k-coalitions, in which any player appears at most K times, and let $H(T)$ represent the set of players in coalition T. To denote the position of player i appearing at most K times in a coalition, we use $i(T, k)$, k=1,2,..., K. With the *value function* taken as a total number of post exchanges immediately succeeding such *engaging* subthreads *p is an element of P* that have the same membership, size, and structure as k-coalition *T is an element of* $\Omega^K(N)$, the engagement capacity of user *I is an element of N* is computed by the equation in Figure 2.

The parameter α in this equation controls the distribution of credit for attracting a response among the users in the subthread that generates the response. With α=1, the formula gives equal credit to the authors of all the posts in the subthread. With α<1, the immediate predecessor of a newly submitted response gets more credit for attracting it, with the credit to the earlier predecessors discounted by the factors of α, α^2, α^3, etc, respectively. Note that this property of engagement capacity complies with an observation that the posts, newly attracted to a (long) forum thread, tend to respond to the latest prior contributions in this thread.

All the computations reported in this paper have been performed with α=0.8. With α=0.8, about half the credit for attracting a new post to a very long thread, to be shared by all the contributors, goes to the last 4 contributors, with the contributors further upthread getting less and less credit. A smaller value of α would reward the most recent contributors even more, to the point where the thread originator would get almost no credit. Note, however, that over all the contributions in a given thread, the thread's originator is always guaranteed to earn more credit than any other contributor. Note also that the choice of α does not affect the computational effort required to perform the engagement capacity evaluation.

To illustrate the application of the engagement capacity computation formula, observe that the thread in Figure 1 increments the forum's engagement value by 7 units: 7 of its 8 posts have been submitted in response to some preceding post(s). Figure 3 reports how this value is distributed among the thread contributors.

Note that with each new piece of content (a new post or multiple posts) added to a forum in response to any prior post(s), the engagement analysis does not need to be redone for the past history of the forum. Each response to a prior post adds a value of 1, which is shared only by the users in a particular subthread, and hence, the engagement capacities of only those users need to be updated (incremented by a certain amount, per the engagement capacity computation formula). Hence, the runtime of an efficient engagement quantification algorithm is linear in the number of posts in a forum.

Figure 2. Engagement capacity equation (see text for explanation).

$$\psi_i^{K-\alpha}(N,v) = \sum_{T \in \Omega^K(N), i \in H(T), k=1,\ldots,K} \Delta_v^*(T) \frac{\alpha^{|T|-i(T,k)}}{|T| \sum_{j=0}^{|T|-1} \alpha^j}$$

Figure 3. Engagement capacity values computed for the users in Figure 1.

Users	Engagement Capacity
A	$2 + \frac{\alpha}{1+\alpha} + \frac{2\alpha}{1+\alpha} + \frac{2\alpha^2}{1+\alpha+\alpha^2} = 3.86$
B	$\frac{1}{1+\alpha} + \frac{\alpha}{1+\alpha+\alpha^2} + \frac{1}{1+\alpha+\alpha^2} = 1.29$
C	$\frac{2}{1+\alpha} + \frac{\alpha}{1+\alpha+\alpha^2} = 1.44$
D	$\frac{1}{1+\alpha+\alpha^2} = 0.41$

Data Collection

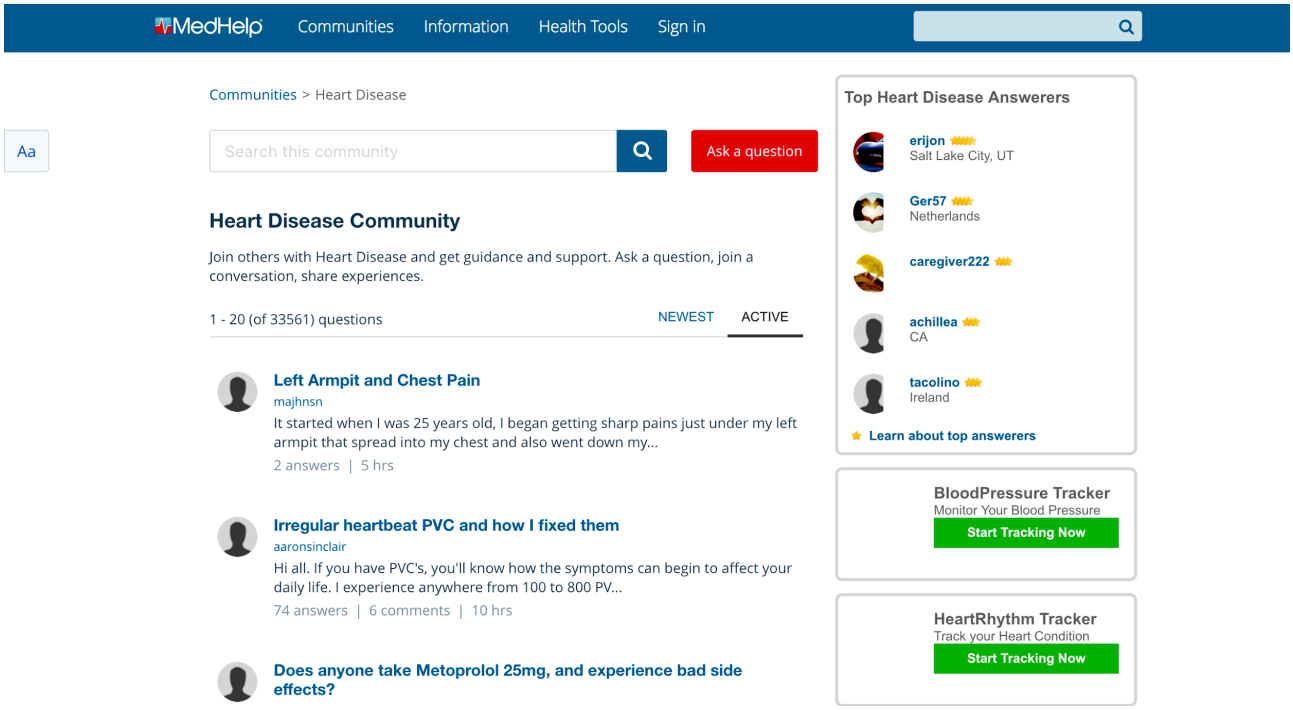
We collected data from a big, active, and freely accessible online prohealth platform, MedHelp.org [26]. Each of its (approximately) 200 predominantly English-speaking OHFs is devoted to a specific health-related topic. The platform users interact through discussion boards, author personal journals, post notes, and post status updates on personal pages.

We collected the data from the Heart Disease, Diabetes, Substance Abuse, Fitness, Depression, Heart Rhythm, and Anxiety OHFs, all active between the years 1999 and 2016. The MedHelp.org data are publicly available [27]. In processing the data, we did not save users' personal profile details, and we anonymized usernames. Figure 4 shows a random screenshot of the MedHelp Heart Disease community forum with snippets of several threads. There were approximately 200,000 threads, although some of the threads did not generate responses (see Table 1). We broke the timeline into fixed intervals for performing longitudinal analyses. Within each time interval, we tracked engagement capacity, targeted engagement capacity between users, number of newly registered users, and delurking and friendship-building events.

Table 1. Statistics (counts) for the observed online health forums (OHFs), rounded to thousands.

OHF community discussion topic	Total messages	Total threads	Threads of >1 posts	Contributing users
Heart Disease	104,000	32,000	23,000	31,000
Diabetes	14,000	4000	3000	5000
Substance Abuse	760,000	81,000	78,000	52,000
Fitness	19,000	5000	4000	9000
Depression	58,000	13,000	12,000	14,000
Heart Rhythm	90,000	18,000	17,000	15,000
Anxiety	166,000	34,000	30,000	33,000

Figure 4. Screenshot of the Heart Disease community forum at MedHelp.org.



Data Analyses

We carried out the following procedures to answer RQ1-RQ4. For RQ1, we calculated the engagement values for each day over 17 years based on the observed communication on the platform (all 7 OHFs lumped together). In this calculation, each new post, generated *in response to* any prior post, contributed an engagement value of 1 to the overall *engagement value*. To quantify the *reach*, we recorded the number of new users joining the platform in each day. Answering RQ2 required the tracking of delurking—the phenomenon (event) where a user breaks the habit of passive reading by adding content to a forum, thereby extending its reach [28]. The OHFs examined in this study did not require their visitors to register for passive reading; therefore, for RQ2, we took delurking to occur whenever a registered user contributed a post after at least a month of inactivity. For RQ3, we assumed a personal virtual connection to have formed (been initiated) when a user posted a note on a peer's profile front page for the first time. We calculated targeted engagement capacity values for user i engaging peer j for all user pairs (i, j). A binary indicator function was used to track the count of personal (virtual) connections resulting from user interaction in an open forum; the function took the value of 1 whenever user j posted a note on i's wall. To answer RQ4, we computed the engagement capacity for each user over fixed time periods (1 week long). A binary indicator function was used to track whether the user was active (ie, contributed to the forum) in the following week; for example, the engagement value for user i was computed for week t and an indicator function was used to record whether user i contributed to the forum through posts or comments in week t +1.

Granger Causality Test

We investigated RQ1 and RQ2 using Granger causality testing, a widely used tool for the analysis of joint temporal dynamics of multiple observed quantities (here, engagement value, new user count, and delurking). Per the theory of Granger causality, one signal (X_1) is said to *Granger-cause* another signal (X_2) if the past value(s) of X_1 contains information that can predict X_2 better than the information contained in the past value(s) of X_2 would do alone [29]. In general, detecting a 1-way Granger causation is desired to definitively establish the nature of a cause-and-effect relationship between 2 temporally varying quantities. The 2-way Granger causality test is typically conducted to first detect any cause-and-effect relationship in both directions. If one observes that X_2 Granger-causes X_1, while X_1 Granger-causes X_2, then further analysis is done to determine the direction in which the causal effect is the strongest. In this study, we took the OHF engagement value as X_1, and new user count and delurking rate as X_2.

Results

Causal Relationship Between Engagement and Reach

The Granger causality test performed between the platform's engagement and new user count indicated that engagement Granger-caused reach (P=.02 with lag 2). The new user count, shown by the orange line in Figure 5, was incremented by 1 if a new user joined the forum; meanwhile, the engagement value of the platform, shown by the blue line, was incremented by 1 if a post was responded to. We calculated both these values for each day between 1999 and 2015. On average, 5076 posts on the platform were responded to daily, with 25,260 being the maximum number of responses. Note that we refer to all 7 OHFs combined as the "platform."

Figure 5. Temporal dynamics of the platform's engagement value and new user counts (with new user counts scaled up by a factor of 10).

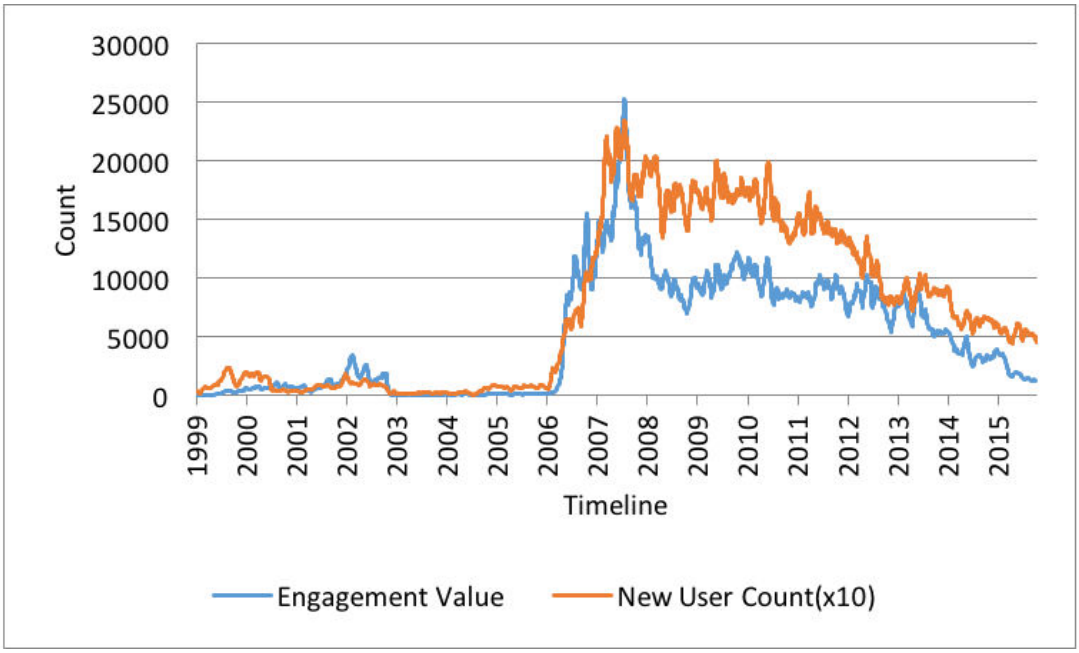


Figure 6. Temporal dynamics of the platform's engagement value and delurking event counts (with delurking event counts scaled up by a factor of 100).

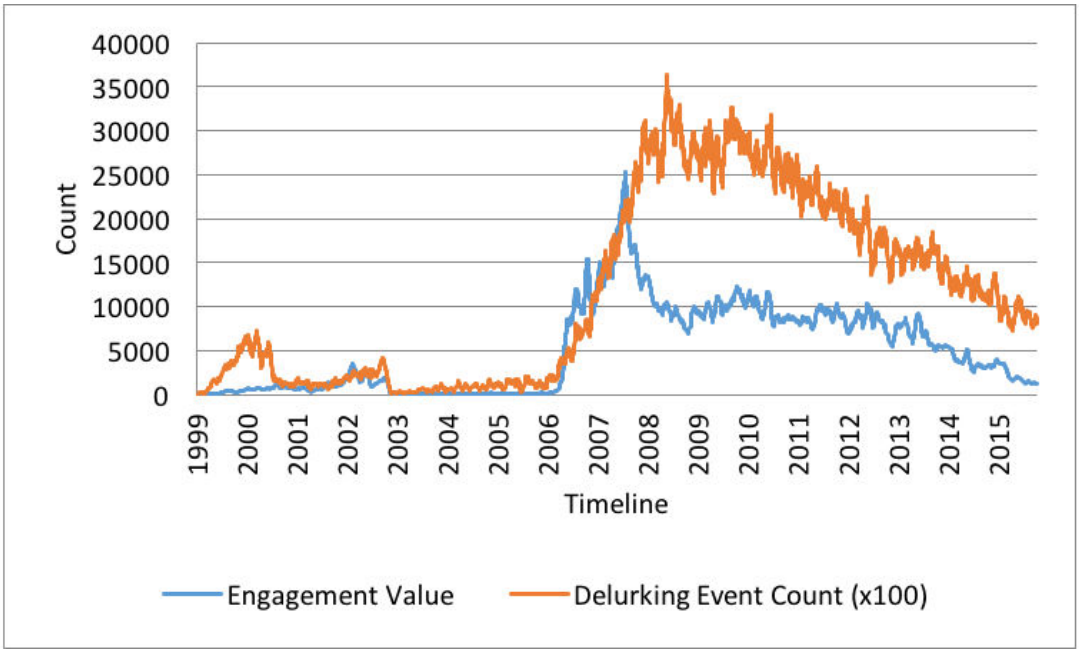
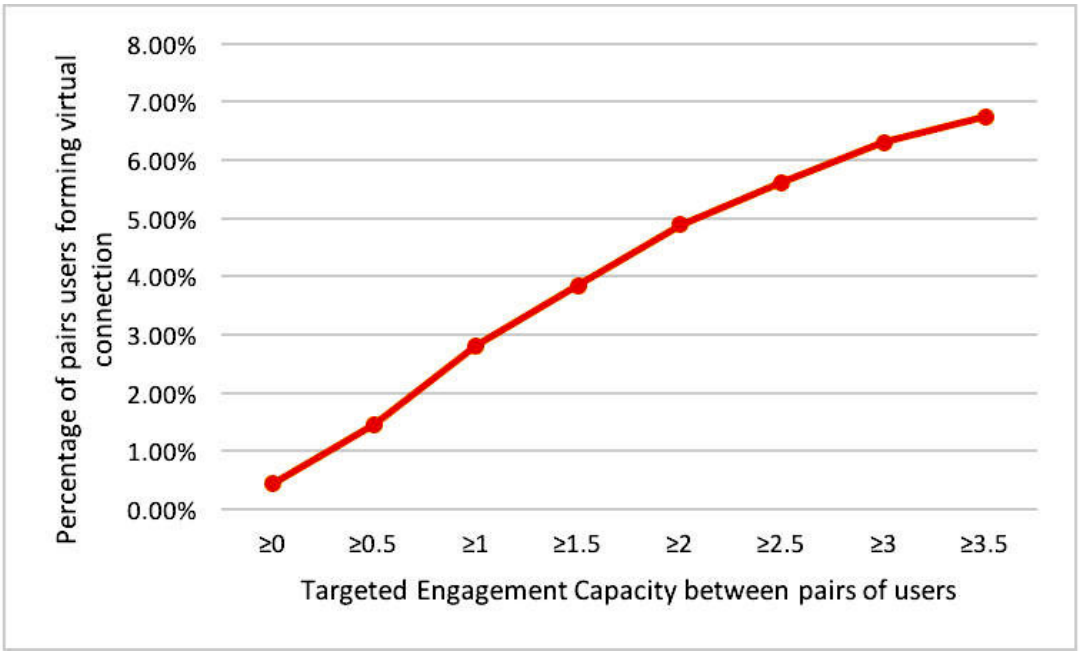


Figure 7. Proportion of total user pairs building a personal connection as a function of targeted engagement capacity between online health forum users.



Causal Relationship Between Engagement and Delurking

Figure 6 shows that the fluctuations in the rate of delurking follow the fluctuations in engagement value. Granger causality analysis revealed a 1-way causation, with the hypothesis of engagement Granger-causing delurking supported by a significance of $P<.001$ with lag 2, and the hypothesis of delurking Granger-causing engagement supported by a significance $P=.01$ with lag 2. We thus concluded that we had a stronger support of the claim that engagement caused delurking (than the other way around). The count of the delurking events, shown by the orange line in Figure 6, was incremented by 1 if a registered user contributed content in 1 of the 7 forums after at least a month of inactivity. On average, 117 people responded to posts in a day after at least a month of inactivity (maximum 364 people per day).

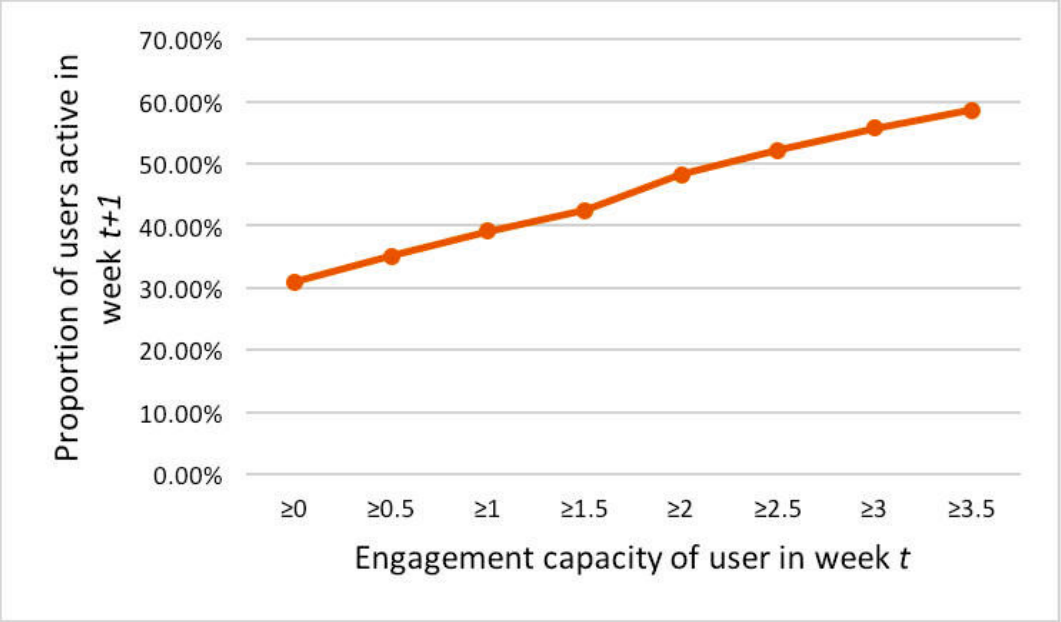
Targeted Engagement Capacity as a Predictor of the Development of Personal Connections Between Users

Figure 7 shows the relationship observed between the propensity of building personal (virtual) connections among OHF user pairs and the targeted engagement capacity delivered by 1 user in a pair to another user in this pair. Any 2 users have zero value if they have not responded to each other's post in any of the 7 forums. Also, note that targeted engagement is directed and nonsymmetric; that is, user *i* engaging user *j* is not the same as *j* engaging *i*. We calculated this metric over 1621,635 pairs and observed a total of 8482 notes in the OHFs. The percentage of pairs of users forming personal virtual connections out of 1,621,635 is shown on the vertical axis. For example, the point (x=1.5, y=3.98%) indicates that, out of 1,621,635 pairs of people accounted for in the targeted engagement calculation, approximately 3.98% ended up developing personal virtual connections if 1 user in the pair managed to engage the other user to earn a targeted engagement value of at least 1.5. Overall, the more frequently user *i* managed to engage user *j* (prompting *j* to respond), the more likely *i* was to receive a posted personal note from *j* for the first time. The mean value of targeted engagement capacity for all the users with established personal connections is 1.6 (SD 4.8). A regression line, fit to the curve in Figure 7, has an adjusted $R^2=.97$.

Engagement Capacity as a Predictor of Future Activity

Figure 8 shows the relationship observed between the propensity of a user to stay active in an OHF and the engagement capacity of this user earned over the past week. We calculated the engagement capacity of each user for each week. For example, the point (x=1, y=39%) in Figure 8 indicates that approximately 39% of users in the platform tended to stay active in the week *t+1* if their posts earned them the engagement capacity of at least 1 in week *t*. Users who received high engagement for their posts were more likely to contribute to the forum. The vertical axis in Figure 8 shows the proportion of users active in week *t* +1 with respect to their engagement capacity in week *t*. We considered a user to be active if they participated by posting content in any of the forum threads in week *t* +1. A regression line, fit to the curve in Figure 8, has an adjusted $R^2=.99$.

Figure 8. Propensity of users to stay active in an online health forum as a function of their engagement capacity earned over time.



Discussion

Principal Findings

The 7 OHFs we observed in this study were diverse, with a varied number of posts and users. Some of the users were simultaneously active in more than 1 OHF. Granger causality testing performed on observational data revealed that engagement Granger-caused reach of the platform. The analysis of delurking showed a 2-way relationship between engagement and delurking: this reveals an interdependence of these 2 variables; further investigation indicated that changes in engagement value dynamics tended to precede those in the delurking count. In summary, engagement measurement appears to provide useful information about the retention rate of users in the platform. We also found engagement capacity to be an informative metric for predicting the propensity of a user to contribute to an OHF. Further, targeted engagement, as the measure of the amount of directed communication between a pair of users, provided information about the development of a virtual connection between them.

Practical Implications

Targeted engagement computation can be useful for predicting the retention rate of users in the platform, which is in line with the common bond theory. Specifically, in the future, this metric can become an integral component of OHF thread recommender systems [16], where suitable threads can be recommended to those users who are more likely to develop personal connections through communication. As noted in the Introduction section, previously proposed metrics of OHF activity, including those that rely on social network analysis tools, have been predominantly used for descriptive purposes. Our results suggest that the engagement capacity and targeted engagement capacity metrics can be used for predictive, and perhaps prescriptive, purposes. Indeed, engagement measurement allows OHF managers to move beyond merely counting contributions, "likes," and time spent by users online, toward identifying such posts and users that are inherently engaging. To inform managerial insights, we foresee outputs based on engagement measurement being displayed in a graphical form, such as plots, also enabling time-dependent tracking. Downward trends in engagement capacity of superusers, engagement value of the platform, etc, can be good indicators for managers to take proactive measures. As part of future developments in this direction, we anticipate an interest from the research community to work toward finding patterns in downward engagement trends and critical points for the platform's growth.

Importantly, *why* or *how* those posts and users manage to be engaging can be studied with the help of text mining techniques [29]. The success of an OHF depends on user participation. To promote the growth of the user network, practitioners are typically interested in identifying those users who contribute most toward the development of the OHF user base as a community [30]. The most important, or influential, users of prohealth platforms are often referred to as core users or superusers [31]. These individuals are integral to OHF growth, as they willingly provide continuous support, advice, and information to other users. The engagement metric offers a new way of understanding what it means for a user to be "important" to an OHF, and the answers to RQ1-RQ4 help pave the way for the development of calculated interventions for strategically growing OHFs, informed by engagement measurement.

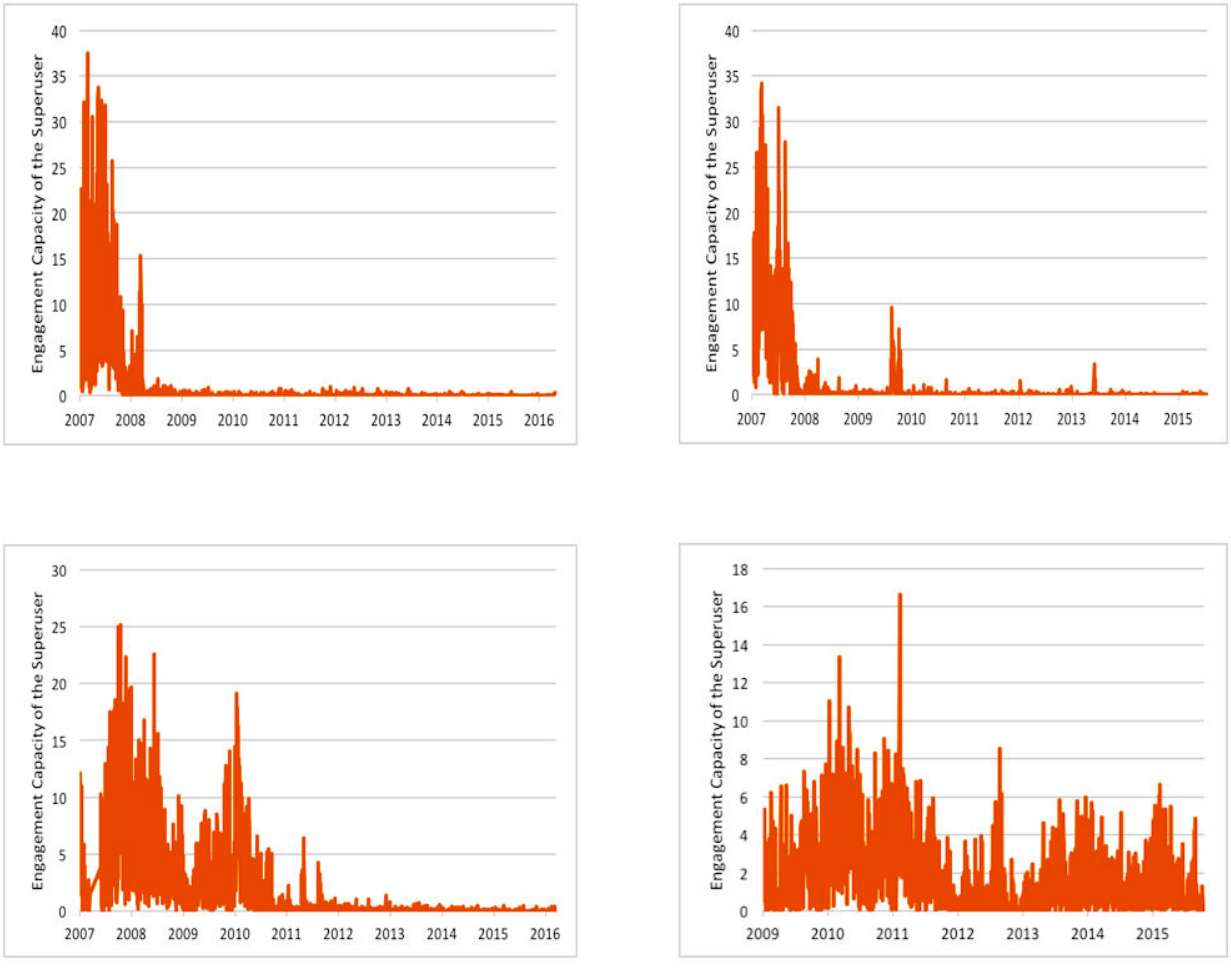
Previously, superusers (and their subtypes) have been defined in the literature [32-34]. These definitions, developed qualitatively, quantitatively, and graphically, have studied a variety of observed parameters, including participation style, social support type, thread initiation, participation inequality, and user life cycles. This study contributes to an enhancement of those existing definitions, using engagement metrics. Superusers, recognized through the lens of the engagement metric, can now be viewed as users having an innate ability to engage others. Indeed, while previous works have considered frequency of contributions and similar frequency-based measurements as the main metrics for identifying superusers,

our work allows for use of a metric such as engagement per post, emphasizing the skill of being engaging as a key quality for a superuser to have.

Displaying the calculated user engagement capacities, earned over time, to users can potentially increase their future contributions via gamification [35]. Further, highly engaging users can be asked or incentivized to spend more time on an OHF, as they now can be recognized as the individuals capable of helping the OHF succeed. Designing suitable incentives to encourage participation of such users will keep the platform more engaging by attracting more posts. Figure 9 shows the engagement capacity, per day, for 4 of the studied OHFs' top 10 most engaging users. Interestingly, the plotted patterns reveal that certain of the most engaging users may have been highly active for prolonged time periods, while others may have been highly active only for shorter time periods. OHF managers can take proactive measures to motivate superusers when there is a downward trend in their engagement capacity. For example, the superuser in Figure 9 (top left) can be seen to have a sudden engagement drop in mid-2008. This could have been avoided by developing suitable interventions to help the superuser stay committed to the platform, or to make sure that other superusers "pick up the slack" in time.

Moreover, further research into the possible exact definitions of engagement-based superusers and quantification of a downward trend in their contribution is in order. For example, nominating a superuser can become time dependent, based on a user receiving a high engagement score, such as over a month of forum life. Also, normalized engagement capacity (ie, engagement capacity per post) can be used to make engagement measurement less subjective.

Figure 9. Engagement capacity dynamics of 4 of the top 10 most engaging online health forum users.



Limitations

The growth of an online forum is dependent on various factors (eg, organizational, psychological, and sociological). Therefore, the engagement of an OHF should be considered as one of the many factors, and not the only factor, responsible for causing the OHF to grow.

On the technical side, the Granger causality test output is known to depend significantly on the functional form of regression used to test a relationship between the variables [36,37]. Further, one must be aware that unobservable variables may play a role in the results of such test, and hence, must be careful not to overinterpret even statistically significant conclusions.

Also, engagement increases as more users join the platform and respond to contributed content; however, conflicts might emerge when online communities pass a certain threshold. While this issue might not be acute in OHF communication, as OHFs are typically well moderated, the nature of engagement—desirable

or undesirable—has to be taken into account during the generation of insights based on engagement analyses.

Conclusion

Our study shows that engagement is one of the important factors responsible for growth of OHFs; the theory-supported engagement metric provides a framework for systematically assessing a platform's health by quantifying users' innate ability to engage others. The predictive property of this metric can potentially inform the development of incentive schemes such as badges and monetary rewards, thread recommender systems, and identification of superusers. Using metrics such as frequency of contribution might result in a causation or correlation issue in identifying superusers, as it favors existing users as opposed to new users joining the platform. By tracking the engagement per post metric, forum managers can overcome this issue, in that users' innate ability to engage their peers can be identified from the time they start contributing to the platform. Research can now be done toward answering the question "what makes a user engaging?"

Acknowledgments

This work was funded in part by the US National Science Foundation (award no. 1635611), National Cancer Institute of the US National Institutes of Health (grant no. R01CA152093-01 to SM), US Air Force Research Laboratory (AFRL) European Office of Aerospace Research and Development (grant no. FA9550-17-1-0030), and the AFRL Mathematical Modeling and Optimization Institute. AN was also supported in part by the USAF Summer Faculty Fellowship Program.

Conflicts of Interest

None declared.

References

1. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Dec;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
2. 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;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
3. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
4. Kraut R, Resnick P, Kiesler S, Burke M, Chen Y, Kittur N, et al. Building Successful Online Communities: Evidence-Based Social Design. Cambridge, MA: MIT Press; 2012.
5. Preece J. Sociability and usability in online communities: determining and measuring success. Behav Inf Technol 2001 Jan;20(5):347-356. [doi: 10.1080/01449290110084683]
6. Ling K, Beenen G, Ludford P, Wang X, Chang K, Li X, et al. Using social psychology to motivate contributions to online communities. J Comput Mediat Commun 2005 Jul;10(4):00. [doi: 10.1111/j.1083-6101.2005.tb00273.x]
7. Ren Y, Kraut R, Kiesler S. Applying common identity and bond theory to design of online communities. Organ Stud 2016 Jun 30;28(3):377-408. [doi: 10.1177/0170840607076007]
8. Ma M, Agarwal R. Through a glass darkly: information technology design, identity verification, and knowledge contribution in online communities. Inf Syst Res 2007 Mar;18(1):42-67. [doi: 10.1287/isre.1070.0113]
9. Gruzd A, Haythornthwaite C. Enabling community through social media. J Med Internet Res 2013;15(10):e248 [FREE Full text] [doi: 10.2196/jmir.2796] [Medline: 24176835]
10. Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. J Med Internet Res 2015;17(6):e141 [FREE Full text] [doi: 10.2196/jmir.3662] [Medline: 26068087]
11. Papadopoulos S, Kompatsiaris Y, Vakali A, Spyridonos P. Community detection in social media. Data Mining Knowl Discovery 2011 Jun 14;24(3):515-554. [doi: 10.1007/s10618-011-0224-z]
12. Borgatti S. Identifying sets of key players in a social network. Comput Math Organ Theory 2006 Apr;12(1):21-34. [doi: 10.1007/s10588-006-7084-x]
13. Freeman L. Centrality in social networks conceptual clarification. Soc Netw 1978 Jan;1(3):215-239. [doi: 10.1016/0378-8733(78)90021-7]
14. Katz ML, Shapiro C. Network externalities, competition, and compatibility. Am Econ Rev 1985;75(3):424-440.
15. van Mierlo T, Hyatt D, Ching AT. Mapping power law distributions in digital health social networks: methods, interpretations, and practical implications. J Med Internet Res 2015;17(6):e160 [FREE Full text] [doi: 10.2196/jmir.4297] [Medline: 26111790]
16. Stearns M, Nambiar S, Nikolaev A, Semenov A, McIntosh S. Towards evaluating and enhancing the reach of online health forums for smoking cessation. Netw Model Anal Health Inform Bioinform 2014 Dec;3:69 [FREE Full text] [doi: 10.1007/s13721-014-0069-7] [Medline: 26075158]

17. Nikolaev A, Gore S, Khopkar S, Govindaraju V. Engagement capacity and engaging team formation for reach maximization of online social media platforms. 2016 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco, CA, USA p. 225-234.
18. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]
19. Ray S, Kim S, Morris J. The central role of engagement in online communities. Inf Syst Res 2014 Sep;25(3):528-546. [doi: 10.1287/isre.2014.0525]
20. Nonnecke B, Preece J. Why lurkers lurk. 2001 Presented at: 7th Americas Conference on Information Systems; Aug 3-5, 2001; Boston, MA, USA p. 294.
21. Wu B, Jiang S, Chen H. Effects of individuals' motivations on communications in online health forums. Soc Behav Pers 2016 Mar 23;44(2):299-312. [doi: 10.2224/sbp.2016.44.2.299]
22. Lin K, Lu H. Why people use social networking sites: an empirical study integrating network externalities and motivation theory. Comput Hum Behav 2011 May;27(3):1152-1161. [doi: 10.1016/j.chb.2010.12.009]
23. Shapley LS. A value for n-person games. In: Kuhn HW, editor. Contributions to the Theory of Games, volume I. (AM-24) (Annals of Mathematics Studies). Princeton, NJ: Princeton University Press; 1950.
24. Sanchez E, Bergantiños G. On values for generalized characteristic functions. Operations Res Spektrum 1997 Sep;19(3):229-234. [doi: 10.1007/BF01545593]
25. Nowak A, Radzik T. The Shapley value for n-person games in generalized characteristic function form. Games Econ Behav 1994 Jan;6(1):150-161. [doi: 10.1006/game.1994.1008]
26. Chuang KY, Yang CC. Interaction patterns of nurturant support exchanged in online health social networking. J Med Internet Res 2012;14(3):e54 [FREE Full text] [doi: 10.2196/jmir.1824] [Medline: 22555303]
27. MedHelp. Support communities. San Francisco, CA: MedHelp International; 2017. URL: http://www.medhelp.org/forums/list [accessed 2017-08-28] [WebCite Cache ID 6t3aCrFFq]
28. Sprague RH. De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. Los Alamitos, CA: IEEE Computer Society Press; 2004 Presented at: 37th Annual Hawaii International Conference on System Sciences; Jan 5-8, 2004; Big Island, Hawaii.
29. 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]
30. van Mierlo T, Li X, Hyatt D, Ching AT. Demographic and indication-specific characteristics have limited association with social network engagement: evidence from 24,954 members of four health care support groups. J Med Internet Res 2017 Feb 17;19(2):e40 [FREE Full text] [doi: 10.2196/jmir.6330] [Medline: 28213340]
31. 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;17(12):e271 [FREE Full text] [doi: 10.2196/jmir.4705] [Medline: 26627369]
32. Carron-Arthur B, Reynolds J, Bennett K, Bennett A, Cunningham JA, Griffiths KM. Community structure of a mental health internet support group: modularity in user thread participation. JMIR Ment Health 2016;3(2):e20 [FREE Full text] [doi: 10.2196/mental.4961] [Medline: 27242012]
33. van Mierlo T, Voci S, Lee S, Fournier R, Selby P. Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian Cancer Society's smokers' helpline online and StopSmokingCenter.net. J Med Internet Res 2012;14(3):e66 [FREE Full text] [doi: 10.2196/jmir.1854] [Medline: 22732103]
34. van Mierlo T, Hyatt D, Ching A. Employing the Gini coefficient to measure participation inequality in treatment-focused digital health social networks. Netw Model Anal Health Inform Bioinform 2016;5(1):32. [doi: 10.1007/s13721-016-0140-7] [Medline: 27840788]
35. Rashid A, Ling K, Tassone R, Resnick P, Kraut R, Riedl J. Motivating participation by displaying the value of contribution. 2006 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 22-27, 2006; Montreal, QC, Canada p. 955-958.
36. Granger C. Some recent development in a concept of causality. J Econometrics 1988 Sep;39(1-2):199-211. [doi: 10.1016/0304-4076(88)90045-0]
37. Roberts DL, Nord S. Causality tests and functional form sensitivity. Appl Econ 2006 May 24;17(1):135-141. [doi: 10.1080/00036848500000011]

Abbreviations

OHF: online health forum
RQ: research question

Edited by G Eysenbach; submitted 02.01.17; peer-reviewed by T van Mierlo, G Petrič; comments to author 26.02.17; revised version received 23.04.17; accepted 29.06.17; published 29.08.17.

Please cite as:
Gopalsamy R, Semenov A, Pasiliao E, McIntosh S, Nikolaev A
Engagement as a Driver of Growth of Online Health Forums: Observational Study
J Med Internet Res 2017;19(8):e304
URL: http://www.jmir.org/2017/8/e304/
doi:10.2196/jmir.7249
PMID:28851677

Original Paper

Individual Prognosis of Symptom Burden and Functioning in Chronic Diseases: A Generic Method Based on Patient-Reported Outcome (PRO) Measures

Niels Henrik Ingvar Hjollund[1,2], MD, PhD

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

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

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

Abstract

Background: Information to the patient about the long-term prognosis of symptom burden and functioning is an integrated part of clinical practice, but relies mostly on the clinician's personal experience. Relevant prognostic models based on patient-reported outcome (PRO) data with repeated measurements are rarely available.

Objective: The aim was to describe a generic method for individual long-term prognosis of symptom burden and functioning that implied few statistical presumptions, to evaluate an implementation for prognosis of depressive symptoms in stroke patients and to provide open access to a Web-based prototype of this implementation for individual use.

Methods: The method used to describe individual prognosis of a PRO outcome was based on the selection of a specific subcohort of patients who have the same score as the patient in question at the same time (eg, after diagnosis or treatment start), plus or minus one unit of minimal clinically important difference. This subcohort's experienced courses were then used to provide quantitative measures of prognosis over time. A cohort of 1404 stroke patients provided data for a simulation study and a prototype for individual use. Members of the cohort answered questionnaires every 6 months for 3.5 years. Depressive symptoms were assessed by the Hospital Anxiety and Depression Scale (HADS) and a single item from the SF-12 (MH4) health survey. Four approaches were compared in a simulation study in which the prognosis for each member of the cohort was individually assessed.

Results: The mean standard deviations were 40% to 70% higher in simulated scores. Mean errors were close to zero, and mean absolute errors were between 0.46 and 0.66 SD in the four approaches. An approach in which missing HADS scores were estimated from the single-item SF-12 MH4 performed marginally better than methods restricted to questionnaires with a genuine HADS score, which indicates that data collected with shorter questionnaires (eg, in clinical practice) may be used together with longer versions with the full scale, given that the design includes at least two simultaneous measurements of the full scale and the surrogate measure.

Conclusions: This is the first description and implementation of a nonparametric method for individual PRO-based prognosis. Given that relevant PRO data have been collected longitudinally, the method may be applied to other patient groups and to any outcome related to symptom burden and functioning. This initial implementation has been deliberately made simple, and further elaborations as well as the usability and clinical validity of the method will be scrutinized in clinical practice. An implementation of the prototype is available online at www.prognosis.dk.

(J Med Internet Res 2017;19(8):e278) doi:10.2196/jmir.8111

KEYWORDS

chronic disease; cohort studies; depression; longitudinal studies; patient-reported outcome measures; prognosis; recovery of function; repeated measurements; stroke; surveys and questionnaires; symptom assessment

Introduction

Prognosis may be defined as foreseeing, predicting, or estimating future outcome based on the patient's clinical profile [1]. The importance of individual prognosis was emphasized already by Hippocrates [2] at a time when effective medicinal treatments were rarely at hand and the principal role of the physician was to evaluate an illness and predict its likely progression based on information collected in detailed case histories. Prognostic evaluations are still an extremely important and integrated part of clinical practice, although in modern medical research, prognostic research has received less attention compared with therapeutic and etiological research [1,3]. As an example, a prognostic model may be developed to predict the short-term outcome after intracerebral hemorrhage. Early survival is known to be strongly dependent on the Glasgow Coma Scale score on admission [4]. Other factors that are known to predict outcome are the size of the hemorrhage and presence of intraventricular hemorrhage [5]. The prediction of outcome in patients with intracerebral hemorrhage can be used in the emergency department for decision support to differentiate between patients who might benefit from intensive care and those who have such poor prognosis that they will not benefit from intensive care [5]. However, another important use of prognostic knowledge is to inform the stroke patient and relatives [5].

Prognosis Based on Patient-Reported Outcomes

In medicine, prognosis commonly relates to the occurrence of specific binary events such as death, relapse of disease, readmission, or specific complications [3,6], and scientific guidelines for prognosis research also focus on the prediction of binary events [7,8]. However, many outcomes are of continuous nature and highly relevant outcomes such as symptom burden and functioning cannot be assessed from clinical data only, but by application of patient-reported outcome (PRO) measures [9]. Information on the prognosis of disease-specific and general symptoms as well as functioning are often crucial for patients, and sometimes needed in order to make important decisions such as retirement or return to work, or whether to move to another house or flat. Social authorities and pension boards also need information on the prognosis of symptoms and functioning to make decisions that substantially affect the patient's future life circumstances. Therefore, requests for prognosis regarding such outcomes are frequently made and every physician is familiar with answering such requests, formally or informally. However, such answers most often rely entirely on the individual physician's personal clinical experience, attitudes, and beliefs, and only rarely on relevant quantifiable prognostic data, still less on biostatistical models [10]. This is somewhat surprising considering the demands for evidence-basedness. Depression is common after stroke, affecting approximately one-third of stroke survivors at any one time after stroke, compared with 5% to 13% of adults without stroke, with a cumulative incidence of 55% [11]. Therefore, the prognosis of depressive symptoms after stroke was selected for this study.

Methodological Shortcomings

This lack of useful methods for individual PRO-based prognosis may be attributed to two reasons: lack of relevant data and the inadequacy of traditional statistical methods for constructing prognostic models. The very purpose of a statistical model is to reduce the original data to a few parameters (eg, estimated regression coefficients). However, such models may explain only a small percentage of the variation over time and sometimes none at all. In the latter case, we label such a study as "negative," meaning that the group mean's association with time was not statistically significant. This way of thinking reflects our focus on group means, not variations. Characteristically, we label unexplained variation as "error"—a noise that we failed to eliminate. However, individual variation is a natural phenomenon in all aspects of life and we should describe it, not eliminate it. From the patient's perspective, it is not sufficient to be informed about group means. He or she would more likely prefer to be informed as much as possible about what courses similar patients have experienced. Furthermore, the traditional approach has a number of methodological limitations. A statistical model implies a number of presumptions about distribution, which may not be fulfilled. In addition, model building is complicated, especially if there are more than two measurements per patient [6]. However, given that relevant PRO data have been systematically collected, another possibility exists which addresses these obstacles. This paper describes a generic method that utilizes the cohort's experience, implies few statistical presumptions, and is easy to extend to other outcomes and patient groups if relevant PRO data are available.

Objective

The aim of this paper was to describe a simple and generic nonparametric, data-based method for on-the-fly individual prognosis that focuses on group means and variation to evaluate an implementation for prognosis of depressive symptoms in stroke patients, and to provide open access to a prototype for individual use.

Methods

The principle used was for a given patient (named "the recipient"; ie, the patient for whom the prognosis is requested) to select the subcohort of patients (named "donors") who have a score matching the patient's score for the variable in question at the same number of days (named the "index day") after the primary event (eg, diagnosis, treatment start). The criterion for donor match was the recipient patient's value at the index date plus or minus a value corresponding to the minimal clinically important difference (MCID). The MCID may be either anchor based or distribution based [12]. In order to preserve the generic approach, the latter approach was preferred and calculated as

one-half the standard deviation of the distribution of scores in cohort members [12,13]. The match criterion was applied for the cohort member's last measurement before the index day. The subsequent trajectories for each member of the subcohort (donors) were simultaneously displayed and quantitatively described with summary statistics.

Type of Questionnaires

Depressive symptoms were measured using the Hospital Anxiety and Depression Scale (HADS) [14]. This scale was developed to identify states of anxiety and depression among hospital outpatients; to avoid potential confounding by somatic illness, the construct excludes somatic symptoms such as insomnia and loss of energy [14]. The HADS consists of two subscales: an anxiety scale (HADS-A) and a depression scale (HADS-D). Each subscale includes seven items rated on a four-point rating scale (range 0-3), higher scores indicating more symptoms. Symptoms of anxiety and depression are assessed by summing the points within each subscale (0-21). Only the depression scale was used in this study. Cut-off values were used as proposed by Singer et al [15]. A maximum of two missing values were allowed by using the individual subscale means as proposed by Bell et al [16]. Two types of questionnaires were used: full-length and brief. Full-length questionnaires included the HADS [14], Multidimensional Fatigue Inventory (MFI-20) [17], WHO-5 Well-Being Index [18], and the 12-item Short Form Medical Outcomes Study (MOS SF-12) [19]. The brief questionnaire included MOS SF-12 as the only scale.

Estimation of HADS-D Score in Brief Questionnaires

The HADS-D scores for brief questionnaires and for full-length questionnaires with missing HADS-D scores were estimated based on the MOS SF-12 MH4 item: "How much of the time during the past 4 weeks did you feel downhearted and depressed?" with the answer categories "all of the time," "most of the time," "some of the time," "a little of the time," and "none of the time." In one approach, the regression estimates for the model HADS-D=β_0+ β_1 * SF-12 MH4 was calculated based on all patients (common regression). In another approach, regression estimates were calculated separately for each patient with at least three concurrent measurements of HADS-D and MOS SF-12 MH4 (individual regression). Finally, in the last approach, genuine HADS-D scores were used when present, supplemented with scores based on individual regression where genuine scores were missing.

Simulation

Internal simulation was used to compare the four approaches: (1) genuine HADS-D score, (2) common regression-based scores, (3) local regression-based scores, and (4) a combined approach, where genuine scores were used, when available, otherwise local regression-based scores were inserted. Each patient was successively selected as a recipient with an index day defined as the date of the first measurement with a genuine HADS-D score. This day was treated as index date for the simulation of prognosis for that recipient patient. Donors were selected as described previously after deleting the actual recipient from the donor cohort. For each such simulation and for each time category, the numbers of measurements for the recipient as well as the donors were recorded as well as the differences between recipients' actual scores and the mean value of donor scores. The intraindividual variation in donor and recipients was also recorded. In the first three approaches, only questionnaires with a valid HADS-D score were included, whereas the present SF-12 MH4 scores from the same patients were allowed in the last combined approach. Recipient patients were excluded if the subcohort of donor patients contained less than four members. For comparability reasons, the common regression-based and local regression-based simulation approaches were restricted to patients and questionnaires included in the genuine HADS-D approach.

Evaluation

The performance of the four approaches was evaluated by comparing the values of means, standard deviations, intraindividual variation, mean error, and mean absolute error (MAE) between recipients and donors and across simulated approaches. MAE is a measure of forecast error in time series analysis [20], which unlike the mean square error, weights deviations proportionally. The Wall statistic was used to evaluate differences in mean scores between the four approaches and the donor population. The R version 3.2.5 package [21] was used for analyses, and the prototype runs on the server version of R-studio version 1.0.136 [22].

Patient Population and Data Collection

The source population consisted of all patients with first-time stroke admitted to any hospital in the Central Denmark Region between October 1, 2008 and December 31, 2011. Patients were identified from the Danish Stroke Register, a nationwide initiative to monitor and improve the quality of care. Participation is mandatory for all Danish hospital departments treating patients for acute stroke. The register has been found to be valid regarding patient registration [23]. Patients younger than 80 years, alive 90 days after stroke, and living in their own homes before the stroke were included and invited to participate. Patients were identified in the register by their unique civil registration number. Information on gender and age at the time of stroke was obtained from the civil registration number. Information on comorbidity was retrieved from the Region Central Denmark patient registry. Information on address and vital status was collected from the Civil Registration System before approaching each patient [24]. Detailed information on the original cohort can be found elsewhere [25]. Data were collected by the WestChronic PRO system, which allows automated data collection with use of Web- and paper-based questionnaires. The system has in previous studies achieved response rates up to 93% for initial questionnaires and 98% to 99% for subsequent questionnaires [10]. The patients answered the initial questionnaire 3 months after the stroke, subsequently followed with repetitive questionnaires every 6 months until at least 3.5 years had elapsed. The first questionnaire was paper-based, but patients were encouraged to answer subsequent questionnaires online. Nonrespondents at a given time were mailed the brief version at the next scheduled date. The HADS scale was not included in the brief questionnaire, which, apart from the first 4 months of the study, was used as the initial questionnaire and when questionnaires were sent to patients

who did not respond to the latest questionnaire. For each questionnaire, a time variable was assigned counting the number of days from the date of the stroke to the date the questionnaire data were received. For analyses and tabulation, the time was categorized in one of eight time categories with the best fit (3, 6, 12, 18, 24, 30, 36, and 42 months) by a method described elsewhere [26]. In all, 3856 patients fulfilled the inclusion criteria and 3499 were mailed a questionnaire 3 months after their stroke (Figure 1).

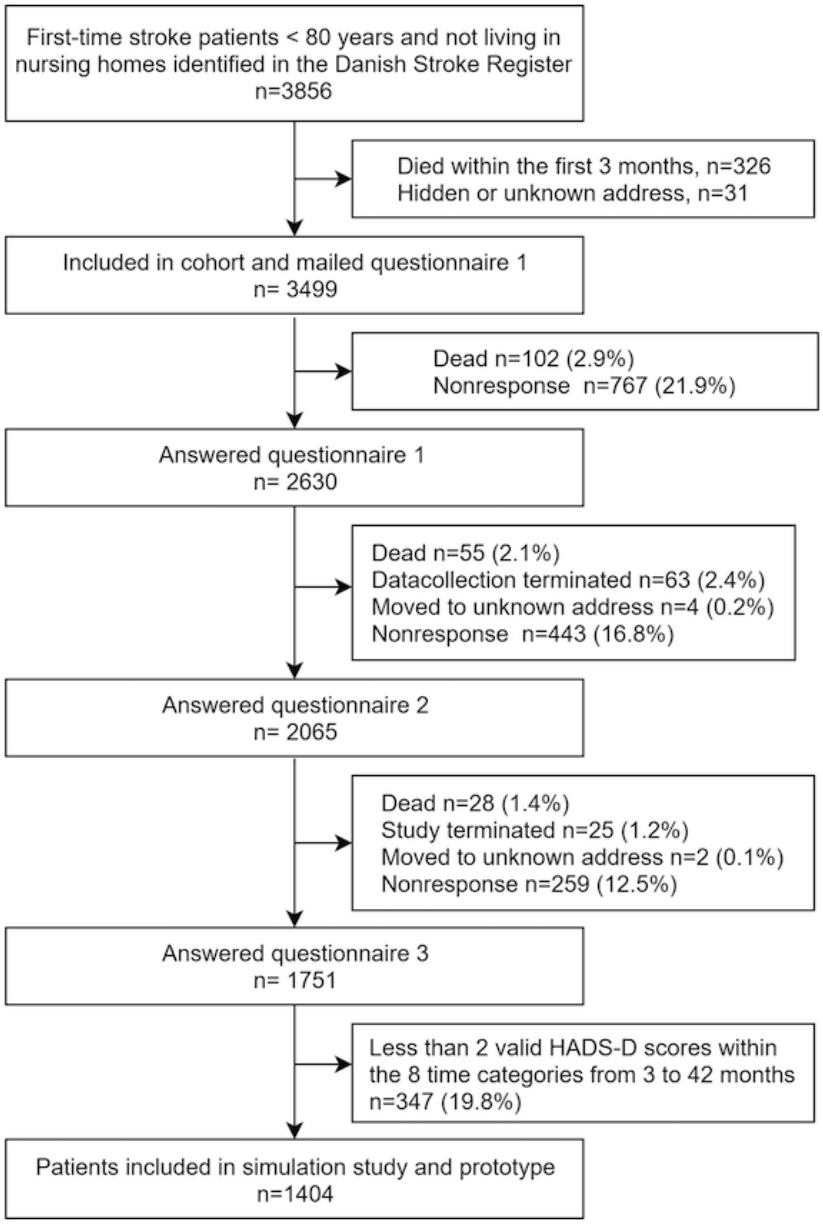
At least three questionnaires with at least two valid HADS-D scores were required for inclusion in the simulation study. A total of 1751 patients answered within at least three time categories, and 1404 patients had at least two valid HADS-D scores and were included in the simulation study (Figure 1). In some cases, a patient had more than one questionnaire in a time category and the second measurement was omitted in the simulation study.

The study was approved by The Danish Data Protection Agency (J.no. 2007-41-0990).

Prototype for Individual Prognosis of Depressive Symptoms

The prototype was based on the same data and method as used in the simulation. However, in the implementation of the prototype, time was measured in days, not in fixed time categories; therefore, all measurements were available. The data used and displayed graphically represent concrete individuals and should not, even theoretically, be identifiable. Before transfer to the prototype server, all data were anonymized and identification numbers replaced with random numbers.

Figure 1. Flowchart of inclusion of stroke patients used in a data-based method for individual prognosis of depression.



Results

Characteristics of the population are given in Table 1. Male patients constituted 63.60% (893/1404) of the population; there were only minor differences in the distribution of characteristics between genders. The mean HADS-D score in the first measurement was 4.48 (SD 3.94). At that time, 208 (14.81%) of the patients had possible signs of depression (score >7-10) and 116 (8.26%) had definite signs of depression (score >10). The 1404 patients contributed with a total of 7273 questionnaires. The median number of questionnaires per patient was five (interdecile range [IDR] 4-7 questionnaires).

A total of 7181 questionnaires could be uniquely classified into one of the predefined time categories (Table 2). In 92 cases, more than one questionnaire was received from the same patient within the same time category. The first one was included in the simulation study, whereas in the prototype, where time is treated as a continuous variable, all questionnaires were eligible. Full-length questionnaires constituted 69.28% (4975/7181) of all questionnaires, whereas a brief questionnaire was received in 30.72% (2206/7181).

The population's distribution of HADS-D scores across time categories is summarized in Table 3. The mean scores varied only slightly over time. The median intraindividual standard deviation was 1.38 (interquartile range [IQR] 0.71-2.12). A detailed analysis of the time trend will be published elsewhere.

Table 1. Characteristics of stroke patients included in simulation study and prototype (N=1404).

Variable	Female (n=511) n (%)	Male (n=893) n (%)	*P* [a]
Age (years)			.06
≤60	173 (33.9)	264 (29.6)	
61-70	164 (32.1)	341 (38.2)	
71-80	174 (34.1)	288 (32.3)	
Comorbidity index[b]			.15
0	284 (55.6)	510 (57.1)	
1	74 (14.5)	155 (17.4)	
2	80 (15.7)	102 (11.4)	
3	22 (4.3)	31 (3.5)	
>3	20 (3.9)	29 (3.2)	
NA	31 (6.1)	66 (7.4)	
Type of stroke			.73
Intracerebral hemorrhage	42 (8.2)	71 (8.0)	
Ischemic	425 (83.2)	729 (81.6)	
Unspecified	37 (7.2)	80 (9.0)	
Missing	7 (1.4)	13 (1.5)	
Year of stroke			.97
2008	39 (7.6)	75 (8.4)	
2009	158 (30.9)	273 (30.6)	
2010	171 (33.5)	299 (33.5)	
2011	143 (28.0)	246 (27.5)	
Type of hospital			
University hospital	199 (38.9)	355 (39.8)	.77
Regional hospital	312 (61.1)	538 (60.2)	
Depression score at entry			
Normal (<7)	374 (73.2)	706 (79.1)	.04
Possible signs of depression (7-10)	89 (17.4)	119 (13.3)	
Definite signs of depression (>10)	48 (9.4)	68 (7.6)	

[a]Data are compared between groups using chi-square test.

[b]Charlson index [27]. Stroke diagnosis not included in calculation.

Table 2. Inclusion and follow-up in stroke patients by time after stroke (N=1404).

	Time after stroke (months)								Total
	3	6	12	18	24	30	36	42	
Inclusion/follow-up, n									
From last month category	NA	1033	1307	1395	1344	1276	1032	586	
Plus entry	1033	274	88	9	0	0	0	0	1404
Minus exit: dead	0	0	0	3	8	4	12	26	53
Minus exit: study termination	0	0	0	8	10	120	132	83	353
Minus exit: attrition	0	0	0	49	50	120	302	231	752
Minus nonresponse this round	0	114	225	229	244	222	4	0	1038
Total received questionnaires	1033	1193	1170	1115	1032	810	582	246	7181
Questionnaire type, n									
Full length	102	827	852	812	823	794	541	224	4975
Brief[a]	931	366	318	303	209	16	41	22	2206
Data collection method, n (%)									
Paper	1031 (99.9)	788 (66.1)	775 (66.2)	735 (65.9)	638 (61.8)	451 (55.7)	398 (68.4)	150 (61.0)	4967 (69.2)
Web	1 (0.1)	405 (34.0)	395 (33.8)	379 (34.1)	394 (38.2)	359 (44.3)	184 (31.6)	96 (39.0)	2214 (30.8)

[a]HADS score estimated from the MOS Short Form 12-item MH4.

Table 3. Hospital Anxiety and Depression Scale depression subscale (HADS-D) scores by time after stroke.

	All	Time after stroke (months)							
		3	6	12	18	24	30	36	42
n	4922	101	818	846	806	820	789	519	223
HADS-D score, mean (SD)	4.5 (3.9)	3.9 (3.6)	4.4 (3.9)	4.7 (3.9)	4.6 (3.8)	4.5 (3.8)	4.4 (3.8)	4.6 (4.2)	4.5 (3.9)
HADS-D score, median (IDR[a])	4 (0-10)	3 (0-9)	3 (0-10)	4 (0-10)	4 (0-10)	4 (0-10)	3 (0-10)	3 (0-11)	3 (1-10)

[a]IDR: interdecile range.

Simulation

In the simulation study, 4922 questionnaires were available, corresponding to the number of questionnaires with a valid HADS-D score, whereas an additional 936 brief questionnaires from the same patients where included in the combined approach (Table 4). At 3 months, when most questionnaires were of the brief type, 543 questionnaires were available for the combined method compared to 101 in the other three approaches (Table 4). With the combined method, 5% of the simulations were based on less than 364 donor questionnaires, whereas for the other three approaches, fewer questionnaires were available (n=116, n=86, and n=117, respectively) (Table 4). The mean scores differed only slightly between the approaches, except for the 3-month value, where the combined approach had a mean score of 5.1 compared to 3.9 in the other approaches (difference 1.20, 95% CI 0.43-1.97) (Table 4).

The results of the simulation obtained in the four approaches are presented in Table 5. The mean values of the standard deviation in simulated scores were 40% to 70% higher than those of the true scores with the largest difference in the approach based on genuine HADS scores (Table 5). In all approaches, the variation in simulated scores was largest in the quintiles with the lowest index score. The mean error was close to zero for all approaches. As expected, the mean errors in relation to time were close to zero. The MAEs were consistently highest for the common regression approach and lowest for the combined approach. Compared to the population standard deviation (SD 3.94; Table 3), the MAEs were 0.61, 0.53, 0.57, and 0.58 SD for the four approaches, respectively.

Prototype for Individual Prognosis

In the online prototype, the recipient patient in question is prompted to complete the HADS questionnaire and enter the date of the stroke (Multimedia Appendix 1). Instantly, the courses of depressive symptoms of each member in a subcohort of donors with matching HADS-D score at the same time after the stroke are presented on the screen (Figure 2), together with descriptive statistics of means and variations over time. Furthermore, sentences with suggested wording of prognosis are displayed (Multimedia Appendix 2).

Table 4. Numbers of simulated patients and donor patient[a] questionnaires in a simulation study in a cohort of 1404 patients by four approaches of simulation.

	Genuine HADS-D score	HADS-D estimated from common regression[b]	HADS-D estimated from individual regression[b]	Combined[c]
Included donor questionnaires	Full length questionnaires	Full length questionnaires	Full length questionnaires	Full length questionnaires supplemented with brief questionnaires
Patients, n				
Simulated	1395	1390	1396	1399
Not simulated[d]	9	14	8	5
Donor questionnaires, n				
Total	1,105,859	1,234,058	1,151,749	1,567,493
Unique	4922	4922	4922	5858
Donors and questionnaires per simulation, n				
Minimum				
Donor patients	4	4	4	4
Questionnaires	11	11	11	15
5th percentile				
Donor patients	29	23	29	63
Questionnaires	116	86	117	364
10th percentile				
Donor patients	46	46	46	91
Questionnaires	186	169	174	516
25th percentile				
Donor patients	127	127	133	154
Questionnaires	492	484	493	830
Median				
Donor patients	199	246	227	211
Questionnaires	801	919	916	1113
Maximum				
Donor patients	423	579	423	488
Questionnaires	1646	2411	1646	2279
HADS-D score, n; mean (SD)				
Overall	4922; 4.5 (3.9)	4922; 4.5 (2.9)	4922; 4.5 (3.8)	5858; 4.6 (3.9)
3 months	101; 3.9 (3.6)	101; 3.9 (3.6)	101; 3.9 (3.6)	543; 5.1 (3.9)[e]
6 months	818; 4.4 (3.9)	818; 4.4 (3.9)	818; 4.4 (3.9)	990; 4.6 (3.9)
12 months	846; 4.7 (3.9)	846; 4.6 (2.9)	846; 4.6 (3.8)	960; 4.6 (3.8)
18 months	806; 4.6 (3.8)	806; 4.5 (2.7)	806; 4.6 (3.7)	897; 4.7 (3.9)
24 months	820; 4.5 (3.8)	820; 4.5 (2.5)	820; 4.5 (3.7)	888; 4.6 (3.8)
30 months	789; 4.4 (3.8)	789; 4.4 (2.4)	789; 4.4 (3.6)	796; 4.4 (3.8)
36 months	519; 4.6 (4.2)	519; 4.7 (2.7)	519; 4.6 (4.1)	551; 4.6 (4.2)
42 months	223; 4.5 (3.9)	223; 4.4 (2.4)	223; 4.6 (3.7)	233; 4.6 (3.9)

[a]Subcohorts of patients with a score matching the score at the index date ±0.5 SD.

[b]Regression based on the SF-12 MH4 item.

[c]Genuine HADS-D score with missing scores estimated by individual regression based on the SF-12 MH4 item.

[d]Less than four donors in the matched subcohort.

[e]*P*=.003.

Table 5. Results from simulation of trajectories of depressive symptoms in a cohort of 1404 patients by four approaches of simulation.

	Genuine HADS-D score	HADS-D estimated from common regression[a]	HADS-D estimated from individual regression[a]	Combined[b]
Standard deviation of scores[c] by quintiles of recipient's[d] initial value, mean values of simulated / true value (% difference)				
1st quintile	1.7/0.5 (240)	2.0/0.5 (300)	1.5/0.5 (200)	1.7/0.5 (240)
2nd quintile	2.0/0.7 (190)	2.1/0.7 (200)	1.8/0.7 (160)	2.0/0.8 (150)
3rd quintile	2.3/1.4 (60)	2.0/1.4 (40)	2.0/1.4 (40)	2.3/1.5 (50)
4th quintile	2.6/1.6 (60)	2.1/1.6 (30)	2.3/1.6 (40)	2.4/1.7 (40)
5th. quintile	2.9/1.7 (70)	2.6/1.7 (50)	2.4/1.7 (40)	2.8/1.7 (60)
Overall	2.4/1.4 (70)	2.1/1.4 (50)	2.0/1.4 (40)	2.3/1.4 (60)
Mean error simulated-true value				
6 months	NA	NA	NA	0.01
12 months	–0.02	0.00	–0.05	0.08
18 months	0.02	–0.16	–0.08	0.13
24 months	–0.02	–0.04	–0.05	0.07
30 months	–0.01	–0.02	0.01	0.08
36 months	–0.07	0.08	0.05	–0.08
42 months	–0.03	0.01	0.08	0.05
Overall	–0.02	–0.03	–0.03	0.05
Mean absolute error				
6 months	NA	NA	NA	1.7
12 months	1.9	2.5	1.9	1.8
18 months	2.1	2.5	2.1	1.8
24 months	2.0	2.5	2.0	1.8
30 months	2.1	2.5	2.0	1.8
36 months	2.4	2.8	2.3	2.0
42 months	2.4	2.7	2.3	2.0
Overall	2.1	2.6	2.1	1.8

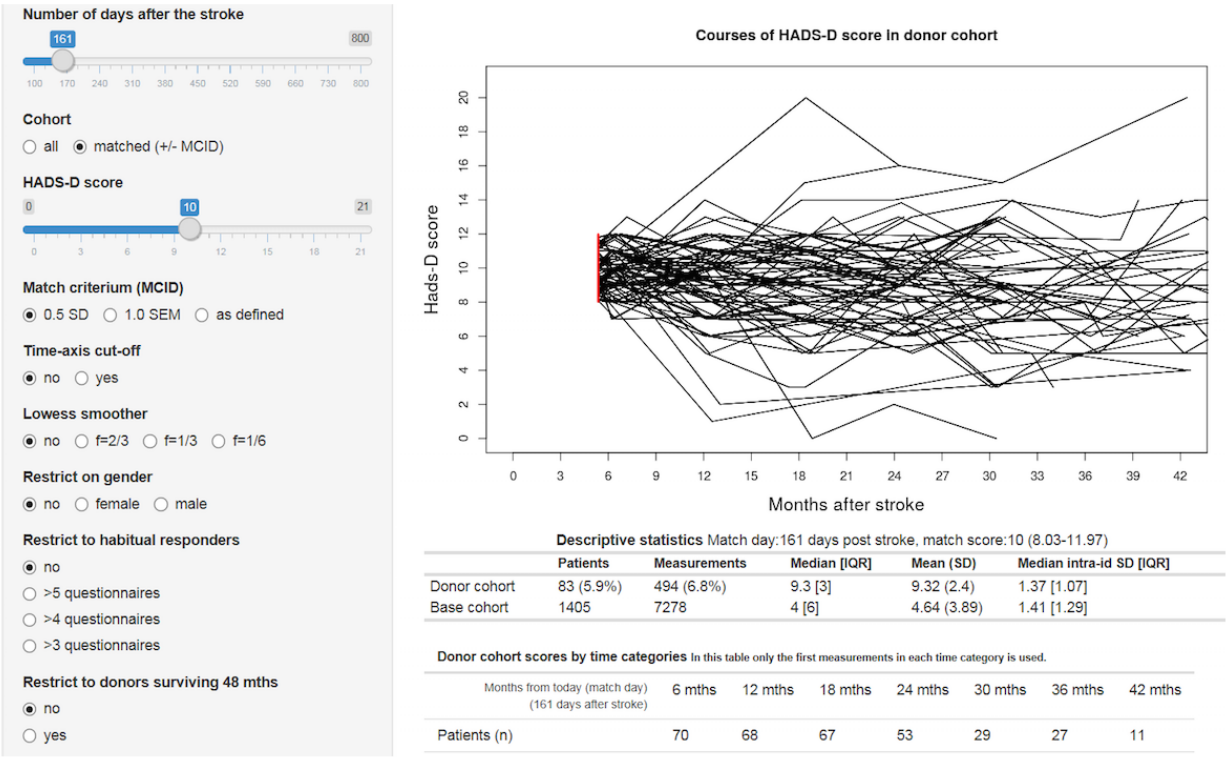
[a]Regression based on the SF-12 MH4 item.

[b]Genuine HADS-D score with missing scores estimated by individual regression based on the SF-12 MH4 item.

[c]In subcohorts of patients with a score matching the score at the index date ±0.5 SD.

[d]The cohort member for which the prognosis is simulated.

Figure 2. Screenshot (extract) from prototype example: individual prognosis of depressive symptoms for a patient who at 23 weeks poststroke had a HADS-D depression score of 10 points.



Discussion

This paper describes a generic method for providing quantitative measures of individual prognosis for PRO-based outcomes. The method is nonparametrical and directly based on original cohort data with repeated PRO assessments.

The mean simulated scores differed only slightly between approaches, except for the 3-month score, where the combined approach had a mean score of 5.1 compared to 3.9 in the other three approaches. Most studies have found that the prevalence of depression is highest in the first period after the stroke [11]. In the first 4 months of data collection, patients were asked to complete the long version of the questionnaire after 3 months, but due to concern about low response rates, we changed the protocol in April 2009 based on the assumption that patients at this early point would more likely complete a briefer version. At 3 months, 931 answered the brief questionnaire, whereas 102 answered the full-length version. The response rate was 69.2% before the change and 81.3% after. Therefore, the lower score in the 101 patients with a genuine HADS-D score at 3 months may be explained by selection bias. The mean 3-month score from the combined approach, which also utilizes data from the brief questionnaires, was more in accordance with other findings, which supports the validity of this approach.

The mean variations in the simulated scores were higher than in the original data (recipients), but did not differ between approaches (Table 5). The variation was, however, only approximately 56% of the overall standard variation (2.2 vs 3.9), which indicates that the method captures additional information. The MAEs were, rather constantly, 2 points, thus in the range of one-half standard deviation among all cohort members. There are two sources for a higher variation in the simulated scores. First, in each simulation, a deviation of up to one-half standard deviation was allowed when selecting donors. Second, the only input used was the actual depression score and the time elapsed since the stroke. In theory, an advanced statistical model would be able to utilize information from more covariates and possibly increase the precision. However, the actual HADS-D score was by far the most important predictor for future HADS scores (data not shown), and other longitudinal studies with similar design have not identified factors of importance for different time trends (ie, interaction on the association between score and time), and the whole model only explained a small percentage of the variation [26]. If previous scores were available for the recipient, not only donor patients with similar actual scores but also those with a similar previous trajectory could be selected. However, given the practical aim of this method, historical scores will typically not be available when the prognosis is requested.

Strengths and Limitations

The internal validity of the method is high, evaluated as the ability to reproduce values from the original cohort. With respect to external validity, this method shares some limitations with model-based methods. The number of severely depressed patients was low, which may be due to underrepresentation of such patients in the original cohort. In etiologic research, selection bias is potential devastating, and it is likely that patients who comply fully with a protocol differ from patients who only answer a few questions or stop answering completely. However, in the setting of data-based individual prognosis, selection and attrition are dynamic, not static, phenomena

because all information regarding future outcomes in the donor cohort is conditioned on surviving (literally and as cohort member) until the index date on which a prognosis is requested. Attrition before that date is therefore merely a question of a reduced number of donors, but given the previously mentioned low predictive values of other covariates, it is less likely to interfere with the distribution of the next measurement in the donor cohort. Nevertheless, the probability of receiving an answer from each donor patient at the next time point will most likely be dependent on the actual health of the patient. Thus, a very good health condition in a patient as well as a very bad one may reduce the probability of answering. This health condition is unknown and unobservable. The combined approach with short questionnaires could be part of a solution, but severely depressed patients will probably still be underrepresented at any point of follow-up. In etiologic research, multiple imputations are often suggested as a solution [28]. However, imputation introduces extra variation, which is a minor problem in etiological research in which the model may only explain a small fraction of the variation, but in the present setting, one of the purposes is to describe the variation itself.

A major strength of the data-based method is its face validity (ie, the intuitive and simple principle easily explainable to clinician and patient), because the prognosis is based on actual scores from actual stroke patients who had reported similar depressive symptoms at the similar point of time after the stroke. This is also the bearing and appealing principle of patient-initiated data capture tools such as PatientsLikeMe [29,30]. However, PRO data collected systematically from a well-defined cohort according to a protocol may be less prone to bias than data collection based on self-selection.

The method is versatile and easy to implement, given that relevant cohort data are available. However, being nonparametric, this method has the disadvantage that subgroups can only be analyzed by stratification, and prediction is only possible for patients who have a combination of covariates that also appear in the source material. However, because only one important predicting variable is involved (the actual score), this is a minor problem, but heterogeneity in the trajectory with respect to a certain covariate cannot be generally ruled out. In the prototype, it is possible to select strata and apply automated in situ nonparametric tests of differences in trajectories between strata of gender, age group, and comorbidity. If other patient cohorts with depression scores are available, it will also be possible to test whether the trajectories in such donor patients differ from those in the actual donor patients and, if not, these patients' trajectories may be included and provide a merged larger cohort for prognosis for severely depressed patients.

The method described here is for application on the individual level and for descriptive use. In analytic epidemiology, when causal factors are searched for (etiology or prediction), parametric or at least semiparametric methods are needed. In case of repetitive data, group-based trajectory modeling may identify latent strata in the longitudinal data [31]. The output of a group-based trajectory model includes estimated probabilities of group membership for each individual [32]. If data on relevant covariates are present, it may, in theory, be possible to also use such models at the individual level to predict future patterns. It would be highly relevant to compare performance of such models with the present model-free method.

Parallel Use of Patient-Reported Outcomes for Multiple Purposes

With the increased application of PRO for clinical use, research, and quality improvement, we will need to address in the near future the problem of the use of multiple different questionnaires for the same patient [10]. An important finding of the simulation study is that shorter questionnaires (eg, for clinical use) may not only coexist with longer questionnaires (for research, quality improvement, as well as use for individual prognosis), but may even provide less-biased longitudinal data given that the design is prepared for individual regression by including at least two simultaneous measurements of the full scale and the surrogate measure.

Access to Prototype

The online prototype of the implementation for depressive symptoms after stroke is available at the website www.prognosis.dk [33].

Conclusion

Internal simulation in a population of stroke patients showed almost similar results in four different approaches, but an approach in which missing scores were calculated based on individual regression coefficients performed best in terms of validity. This is the first description and implementation of a nonparametric cohort-based method for individual prognosis. Further elaborations will be developed and evaluated, and the usability and clinical validity [34] of the method in clinical practice will be scrutinized.

Acknowledgments

Louise Pape Larsen kindly delivered the register-based covariates collected for her PhD thesis on the same cohort. The patients in the stroke cohort, who repeatedly answered the questionnaires, are acknowledged for their effort and adherence. The R development team and the R community are acknowledged for their time and effort in creating and developing R. The original data collection was funded by The Health Foundation and TrygFonden.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Landing page for http://prognosis.dk.

[PNG File, 81KB - jmir_v19i8e278_app1.PNG]

Multimedia Appendix 2

Example of individual prognosis of depressive symptoms after stroke for a patient who 23 weeks after the stroke has a HADS-D score of 10.

[PDF File (Adobe PDF File), 202KB - jmir_v19i8e278_app2.pdf]

References

1. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009 Feb 23;338:b375. [Medline: 19237405]
2. Hippocrates of Cos. The Internet Classics Archive. The Book of Prognostics URL: http://classics.mit.edu/Hippocrates/prognost.html [accessed 2017-05-30] [WebCite Cache ID 6qqbL4aaF]
3. Braitman LE, Davidoff F. Predicting clinical states in individual patients. Ann Intern Med 1996 Sep 01;125(5):406-412. [Medline: 8702092]
4. Lecky F, Woodford M, Edwards A, Bouamra O, Coats T. Trauma scoring systems and databases. Br J Anaesth 2014 Aug;113(2):286-294. [doi: 10.1093/bja/aeu242] [Medline: 25038159]
5. Ariesen MJ, Algra A, van der Worp HB, Rinkel GJ. Applicability and relevance of models that predict short term outcome after intracerebral haemorrhage. J Neurol Neurosurg Psychiatry 2005 Jun;76(6):839-844 [FREE Full text] [doi: 10.1136/jnnp.2004.048223] [Medline: 15897508]
6. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009 May 28;338:b605. [Medline: 19477892]
7. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Cancer 2015 Jan 20;112(2):251-259 [FREE Full text] [doi: 10.1038/bjc.2014.639] [Medline: 25562432]
8. Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG, PROGRESS Group. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. PLoS Med 2014 Jul;11(7):e1001671 [FREE Full text] [doi: 10.1371/journal.pmed.1001671] [Medline: 25003600]
9. Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009 Dec. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf [accessed 2017-07-16] [WebCite Cache ID 6qqbfANOC]
10. Hjollund NH, Larsen LP, Biering K, Johnsen SP, Riiskjær E, Schougaard LM. Use of patient-reported outcome (PRO) measures at group and patient levels: experiences from the generic integrated PRO system, WestChronic. Interact J Med Res 2014 Feb 11;3(1):e5 [FREE Full text] [doi: 10.2196/ijmr.2885] [Medline: 24518281]
11. Towfighi A, Ovbiagele B, El Husseini N, Hackett M, Jorge R, Kissela B, American Heart Association Stroke Council. Poststroke depression: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2017 Feb;48(2):e30-e43. [doi: 10.1161/STR.0000000000000113] [Medline: 27932603]
12. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7(5):541-546. [doi: 10.1016/j.spinee.2007.01.008] [Medline: 17448732]
13. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003 May;41(5):582-592. [doi: 10.1097/01.MLR.0000062554.74615.4C] [Medline: 12719681]
14. Snaith RP. The Hospital Anxiety And Depression Scale. Health Qual Life Outcomes 2003 Aug 01;1:29 [FREE Full text] [doi: 10.1186/1477-7525-1-29] [Medline: 12914662]
15. Singer S, Kuhnt S, Götze H, Hauss J, Hinz A, Liebmann A, et al. Hospital anxiety and depression scale cutoff scores for cancer patients in acute care. Br J Cancer 2009 Mar 24;100(6):908-912 [FREE Full text] [doi: 10.1038/sj.bjc.6604952] [Medline: 19240713]
16. Bell M, Fairclough D, Fiero M, Butow P. Handling missing items in the Hospital Anxiety and Depression Scale (HADS): a simulation study. BMC Res Notes 2016 Oct 22;9(1):479 [FREE Full text] [doi: 10.1186/s13104-016-2284-z] [Medline: 27770833]
17. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995 Apr;39(3):315-325. [Medline: 7636775]
18. Topp C, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]

19. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
20. Hyndman R, Athanasopoulos G. OTexts. 2014. Forecasting: principles and practice URL: https://www.otexts.org/fpp [accessed 2017-07-16] [WebCite Cache ID 6s0CPNUtY]
21. R. The R project for statistical computing URL: https://www.r-project.org/ [accessed 2017-05-30] [WebCite Cache ID 6qqblNJkY]
22. RStudio. Why RStudio? URL: https://www.rstudio.com/about/ [accessed 2017-05-30] [WebCite Cache ID 6qqcL1UZw]
23. Ingeman A, Andersen G, Hundborg H, Johnsen S. Medical complications in patients with stroke: data validity in a stroke registry and a hospital discharge registry. Clin Epidemiol 2010 Aug 09;2:5-13 [FREE Full text] [Medline: 20865097]
24. Pedersen CB. The Danish civil registration system. Scand J Public Health 2011 Jul;39(7 Suppl):22-25. [doi: 10.1177/1403494810387965] [Medline: 21775345]
25. Larsen LP, Johnsen SP, Andersen G, Hjollund NH. Determinants of self-rated health three months after stroke. J Stroke Cerebrovasc Dis 2016 May;25(5):1027-1034. [doi: 10.1016/j.jstrokecerebrovasdis.2015.12.014] [Medline: 26920345]
26. Biering K, Frydenberg M, Hjollund N. Self-reported health following percutaneous coronary intervention: results from a cohort followed for 3 years with multiple measurements. Clin Epidemiol 2014;6:441-449 [FREE Full text] [doi: 10.2147/CLEP.S65476] [Medline: 25506246]
27. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol 2004 Dec;57(12):1288-1294. [doi: 10.1016/j.jclinepi.2004.03.012] [Medline: 15617955]
28. Sterne J, White I, Carlin J, Spratt M, Royston P, Kenward M, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338:b2393 [FREE Full text] [Medline: 19564179]
29. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011 Jan 21;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]
30. Bove R, Secor E, Healy BC, Musallam A, Vaughan T, Glanz BI, et al. Evaluation of an online platform for multiple sclerosis research: patient description, validation of severity scale, and exploration of BMI effects on disease course. PLoS One 2013;8(3):e59707 [FREE Full text] [doi: 10.1371/journal.pone.0059707] [Medline: 23527256]
31. Nagin D, Odgers C. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 2010 Apr 27;6:109-138. [doi: 10.1146/annurev.clinpsy.121208.131413] [Medline: 20192788]
32. Franklin J, Shrank W, Pakes J, Sanfélix-Gimeno G, Matlin O, Brennan T, et al. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. Med Care 2013 Sep;51(9):789-796. [doi: 10.1097/MLR.0b013e3182984c1f] [Medline: 23685406]
33. Individual Prognosis Based on Patient Reported Outcome (PRO) Prototype: Depressive Symptoms After Stroke. URL: http://prognosis.dk/ [accessed 2017-07-17] [WebCite Cache ID 6s1iBG8gt]
34. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000 Feb 29;19(4):453-473. [Medline: 10694730]

Abbreviations

HADS: Hospital Anxiety and Depression Scale
MAE: mean absolute error
MCID: minimal clinically important difference
MOS: Medical Outcomes Study
PRO: patient-reported outcome

Edited by G Eysenbach; submitted 31.05.17; peer-reviewed by J Apolinário-Hagen, B Sapkota; comments to author 21.06.17; revised version received 29.06.17; accepted 29.06.17; published 01.08.17.

Please cite as:
Hjollund NHI
Individual Prognosis of Symptom Burden and Functioning in Chronic Diseases: A Generic Method Based on Patient-Reported Outcome (PRO) Measures
J Med Internet Res 2017;19(8):e278
URL: http://www.jmir.org/2017/8/e278/
doi:10.2196/jmir.8111
PMID:28765099

Original Paper

Psychiatric Consultation at Your Fingertips: Descriptive Analysis of Electronic Consultation From Primary Care to Psychiatry

Margaret Lowenstein[1], MD, MPhil; Olusinmi Bamgbose[2], MD; Nathaniel Gleason[3], MD; Mitchell D Feldman[3], MD, MPhil

[1]University of Pennsylvania Perelman School of Medicine, National Clinician Scholars Program, Philadelphia, PA, United States

[2]Marin General Hospital, Department of Psychiatry, Greenbrae, CA, United States

[3]University of California, San Francisco, Division of General Internal Medicine, San Francisco, CA, United States

Corresponding Author:
Margaret Lowenstein, MD, MPhil
University of Pennsylvania Perelman School of Medicine
National Clinician Scholars Program
Blockley Hall
423 Guardian Drive
Philadelphia, PA,
United States
Phone: 1 215 573 3977
Email: margaw@mail.med.upenn.edu

Abstract

Background: Mental health problems are commonly encountered in primary care, with primary care providers (PCPs) experiencing challenges referring patients to specialty mental health care. Electronic consultation (eConsult) is one model that has been shown to improve timely access to subspecialty care in a number of medical subspecialties. eConsults generally involve a PCP-initiated referral for specialty consultation for a clinical question that is outside their expertise but may not require an in-person evaluation.

Objective: Our aim was to describe the implementation of eConsults for psychiatry in a large academic health system.

Methods: We performed a content analysis of the first 50 eConsults to psychiatry after program implementation. For each question and response, we coded consults as pertaining to diagnosis and/or management as well as categories of medication choice, drug side effects or interactions, and queries about referrals and navigating the health care system. We also performed a chart review to evaluate the timeliness of psychiatrist responses and PCP implementation of recommendations.

Results: Depression was the most common consult template selected by PCPs (20/50, 40%), followed by the generic template (12/50, 24%) and anxiety (8/50, 16%). Most questions (49/50, 98%) pertained primarily to management, particularly for medications. Psychiatrists commented on both diagnosis (28/50, 56%) and management (50/50, 100%), responded in an average of 1.4 days, and recommended in-person consultation for 26% (13/50) of patients. PCPs implemented psychiatrist recommendations 76% (38/50) of the time.

Conclusions: For the majority of patients, psychiatrists provided strategies for ongoing management in primary care without an in-person evaluation, and PCPs implemented most psychiatrist recommendations. eConsults show promise as one means of supporting PCPs to deliver mental health care to patients with common psychiatric disorders.

(J Med Internet Res 2017;19(8):e279) doi:10.2196/jmir.7921

KEYWORDS

mental health; primary care; health care delivery; teleconsultation; telehealth; Internet care delivery

Introduction

Mental and behavioral health problems are among the most common and costly conditions in the United States [1,2]. Primary care providers (PCPs) play an increasingly large role in diagnosis and management, in part, because of limited access to mental health specialty care [3-6]. Shortages of behavioral health providers have been demonstrated in nearly every US county, and in a national survey, two-thirds of PCPs reported they could not obtain high-quality outpatient mental health services for patients [7,8]. Several solutions have been proposed for these gaps in access, many of which focus on developing

more robust systems for mental health care within the primary care setting [9]. These models build on existing evidence that collaborative management between PCPs and behavioral health providers can improve care and access, reduce stigma, and be cost effective [10].

Electronic consultations, or eConsults, are an innovative model of collaborative care developed to improve access to specialty consultation for a variety of physical health conditions. The primary goals of eConsults are to prevent unnecessary referrals for management that could occur in the primary care setting and, in so doing, to decrease wait times for those who require in-person evaluation by the specialist [11]. Improved access to specialists is particularly challenging for systems with provider shortages [12]. Although the format varies, eConsults generally involve a PCP-initiated referral for specialty consultation for a clinical question that is outside their expertise but may not require an in-person evaluation. Most eConsult systems share key features including asynchronous electronic communication between generalist and specialist about clinical questions with documentation within a shared electronic health record (EHR) [12]. Typically, the PCP submits an order for an eConsult with a detailed description of the clinical question, and the consultant responds with recommendations to be implemented by the PCP or, in rare cases, makes the determination that an in-person assessment is needed. Patients are often informed of the eConsult as they would be for a traditional consultation. eConsults are associated with shorter wait times for specialty care, improvement in provider communication and referral quality, and high levels of PCP, specialist, and patient satisfaction [13-20]. eConsults have been successfully implemented in a number of settings, including academic medical centers, the Department of Veterans Affairs, and safety net health systems [21].

eConsults could improve access to mental health care by providing a way for PCPs and psychiatrists to collaborate and expand treatment within the primary care setting. However, acceptability by physicians or patients is not established. In addition, given the nature of behavioral health problems and the importance of the interview in diagnosis, the proportion of eConsult questions that consulting psychiatrists may feel require a face-to-face evaluation is not known. The fact that mental health is often carved out from other medical care creates additional challenges around payment and insurance coverage. Reports on the use of eConsult for mental health are limited but suggest some improvement in PCP perceptions of support for psychiatric diagnosis and treatment as well as access to mental health consultation [22].

Our aim was to test the feasibility of a new psychiatry eConsult program in a large academic health system. We evaluated a consecutive series of the first 50 eConsults to describe the consult questions and responses, assess consult response time and implementation, and determine the rate of conversion to in-person consultation. To our knowledge, this is one of the few studies to describe eConsults for mental health and the first to evaluate consultation content and conversion to in-person referrals.

Methods

Study Setting and eConsult Development

The University of California, San Francisco (UCSF) is a multisite urban academic medical center with 8 primary care practice sites and a diverse payer mix including commercial insurance, Medicare, and Medicaid. PCPs include attending physicians, nurse practitioners, and resident physicians. All clinics use a shared EHR (Epic Systems Corp). For more details on the study site, see Multimedia Appendix 1.

The UCSF eConsult program began in 2012 with 8 internal medicine subspecialties; details of the overall program are described in Multimedia Appendix 2 and elsewhere [23]. eConsults were expanded to include psychiatry in October 2014. Psychiatry eConsult protocols were developed through collaborative efforts of PCPs and psychiatrists. Structured consult templates were created for 3 conditions: depression, anxiety, and bipolar disorder. A generic template was created for questions that fell outside the condition-specific list. Full text of the electronic order templates is available in Multimedia Appendix 3. Templates include prompts for PCPs to state a clear consult question and provide relevant historical information, such as medications used, diagnostic instrument results (such as the Patient Health Questionnaire–9), and concurrent substance use. Relevant laboratory studies also autopopulate the electronic template. The consulting psychiatrists consist of 3 attending physicians who review the consult questions individually on a rotating basis. Consultant psychiatrists are expected to respond within 3 business days, and both question and response text become part of the EHR. If the consultant decides that the clinical question requires in-person evaluation, the consult is converted to a traditional referral for an office visit. eConsult payment is supported by the UCSF health system via the University of California Innovations Fund. Specialists are paid on a time-based work relative value unit (wRVU) system, receiving approximately 0.5 wRVUs for a consult. PCPs also receive 0.5 wRVUs in recognition of the fact that the PCP implements the specialist's recommendation and retains management of the clinical problem. eConsults do not require any insurance authorization, and there is no copayment required from the patient. Further details are available in Multimedia Appendix 2.

eConsult Analysis

We performed a content analysis on a consecutive series of the first 50 eConsults to psychiatry, occurring between October 2014 and August 2015. We determined our initial sample size based on feasibility. We then analyzed the consults sequentially and found no new themes emerging after analysis of approximately the first 25 consults. We opted to present the entirety of the data in this paper as this adds to the richness of the findings.

Question and response texts were extracted from the EHR with personal identifiers removed. Consult type was determined by the template the PCP selected, with the 4 options being depression, anxiety, bipolar disorder, and a generic template as described above. We examined the content of the generic

template and identified a subgroup of consults using this template for questions about psychotic disorders.

We developed a coding framework modeled on prior work classifying eConsult questions and responses and modified by the research team [24-28]. Two independent coders reviewed the text for each eConsult question and response, one internist (ML) and one psychiatrist (OB). First, each question and response was coded as pertaining to one or both of two major domains: diagnosis (identifying the condition) and management (treating the condition). We initially included a third domain for questions pertaining to monitoring, but none of the questions fell into this category. We then defined categories within each domain as below. Each question and response was coded for the presence or absence or each category, and categories were not mutually exclusive.

Within the major domains, we used a hybrid inductive-deductive process to identify categories based on a previously described taxonomy of clinical questions [26] and modified in an iterative fashion to reflect any concepts that emerged during the initial coding process. Categories within the management domain included the following: questions pertaining to medication choice, drug side effects or interactions, navigating the health care system, referral for psychotherapy, and queries as to whether the patient required an in-person evaluation. Specialist responses were categorized in a similar manner. Responses pertaining to diagnosis were coded for the presence of the following categories: interpretation of overall presentation, recommendations for additional testing, and recommendations for obtaining additional history. Within the management recommendations, the following categories were identified: choice of medication, information about side effects, specific medication titration instructions, recommendation for psychotherapy, and recommendation for navigating the health care system.

The two coders analyzed the consult questions and responses independently, and we evaluated agreement using interrater reliability to identify areas of discrepancy and refine the coding structure based on preliminary findings [29,30] **.** Prior to discussion, levels of agreement in the major domains were moderate to high for both PCP consult questions (kappa was 0.4 for diagnosis and 0.66 for management) and specialist responses (kappa was 0.61 for diagnosis and 1 for management). Following the initial independent coding process, the research team met to revise and clarify the coding structure based on these preliminary findings with the final categories derived through consensus process. We then reanalyzed the data using the modified coding structure and any remaining discrepancies were resolved via discussion with the research team.

Additionally, when analyzing questions, we noted whether the PCP had expressed diagnostic uncertainty and whether this was an initial presentation for a patient or the PCP or other provider had tried one or more treatments that failed or resulted in ongoing uncontrolled symptoms. We also noted whether psychiatrist responses included multiple therapeutic options and contingencies or thresholds for starting or adjusting treatment.

Chart Abstraction

Additional information about consults was obtained via chart abstraction, including completion time by the specialist and implementation of consultant recommendations by the PCP. We considered the recommendation to be implemented if the PCP placed an EHR order for one of the recommended medications, tests, or referrals (including in-person psychiatric evaluation). We also considered a recommendation to have been implemented if the PCP documented taking any of the actions recommended by the consultant in the EHR (including documentation of telephone communications, electronic messages to patients, and clinic notes). Simply documenting the recommendation without any action did not count as implementation. We did not categorize whether PCPs implemented all of the recommendations as many responses included multiple therapeutic options for the PCP to choose between. The time frame for documentation was limited to the 6 months following the eConsult, and assessment of PCP implementation of specialist recommendations was limited to eConsults that recommended a change to existing care plan to be carried out within the study window. Both reviewers extracted implementation data for all 50 consults independently with moderate levels of agreement (kappa was 0.55). We used this initial data to clarify our classification system and then reviewed the data again, resolving any remaining disagreements via discussion with the research team.

Statistical Analysis

We used descriptive statistics to report patient characteristics, question and response characteristics, and conversion to in-person consultations. Interrater reliability for categorization of consultative focus and implementation of recommendations was determined using Cohen’s kappa statistic.

Institutional Review

The UCSF Committee on Human Research approved this study.

Results

Demographics

The consult patients ranged from 22 to 80 years of age; 54% (27/50) were female, 62% (31/50) white, and 14% (7/50) were African American (Table 1). Of the 50, 19 had commercial insurance (38%), followed by 16 patients (32%) with Medicaid, 10 (20%) with dual eligibility with Medicare and Medicaid, and 5 (10%) with Medicare. This is compared to the general clinic population which is 59% commercially insured patients, 18% Medicaid, 9% dual-eligible, and 13% Medicare. English was the preferred language of 49 of the 50 patients (98%).

Table 1. Patient demographics for the first 50 psychiatry eConsults.

Patient characteristics		Total
Age, years, median (range)		48.5 (22-80)
Sex, n (%)		
	Female	27 (54)
	Male	23 (46)
Insurance status, n (%)		
	Commercial	19 (38)
	Medicare	5 (10)
	Medicaid	16 (32)
	Medicare-Medicaid	10 (20)
Race, n (%)		
	White	31 (62)
	African American	7 (14)
	Asian-Pacific Islander	6 (12)
	Hispanic	1 (2)
	Other/unknown	5 (10)
Preferred language , n (%)		
	English	49 (95)
	Non-English	1 (2)

Consult Question Types

Depression was the most common consult template selected by PCPs with 40% (20/50) of consults, followed by the generic template with 24% (12/50) of consults, anxiety with 16% (8/50), psychosis with 14% (7/50), and bipolar with 6% (3/50).

For consults that used the generic template, 37% (7/19) of the questions pertained to psychotic disorders. The remaining 12 of the 19 consults placed using the generic template included less common psychiatric disorders, questions about medical comorbidities, access to care, and others. PCPs expressed diagnostic uncertainty in 22% (11/50) of cases, and for 58% (29/50) of patients, PCPs reported that the patient had persistent or worsening symptoms despite having initiated one or more treatments. PCPs noted prior treatment failures for 80% (16/20) of patients with depression, 88% (7/8) of patients with anxiety, 33% (1/3) of patients with bipolar disorder, 43% (3/7) of patients with psychosis, and 20% (2/12) of patients with other diagnoses.

Consult Question and Response Content

Textbox 1 contains a sample eConsult question and response using the depression template.

Textbox 1. This sample consult question was coded as pertaining to management, specifically medication choice and drug side effects or interactions. The question also describes prior treatment failure. The response was coded as pertaining to management, specifically including medication choice, conditions for starting/adjusting treatment, medication titration instructions. The response also included multiple therapeutic options.

Consult Question Text

What might be a good alternative antidepressant for this patient?

She is a 64-year-old woman with a history of chronic pain and depression, who has failed multiple selective serotonin reuptake inhibitors and serotonin norepinephrine reuptake inhibitors in past and was recently started on bupropion, but bupropion has not been effective and patient has also had an episode concerning for possible seizure activity. In addition to depressed mood, significant concerns include fatigue, decreased appetite, and significant weight loss.

Consult Response Text

Based on review of your summary and patient's chart, my recommendations are as follows:

If the patient is more concerned about pain and its treatment, I would initiate duloxetine. Start at 30 mg PO daily, and pending tolerance, increase to 60 mg PO daily. If improving sleep, increasing appetite, and weight gain is more desirable, mirtazapine would be a good idea. Start at 7.5 mg PO daily, and pending tolerance, increase to anywhere from 15 mg to 45 mg. Sedative effects tend to occur at 15 mg or below, less sedation above 15 mg.

Textbox 2. Sample consult question texts by category (categories were not mutually exclusive).

Diagnosis:

[Patient] presented with concern about someone infecting his urine. He has an anxious affect, insomnia, and these new fixed delusions... [I am] concerned for schizophrenia but patient denies he needs psychiatric care or medication. Not a danger to self or others that I can tell...Do you recommend any diagnostic or treatment options for this young man given that he won't see a psychiatrist?

Medication choice:

• Patient with depression and neuropathic pain. Originally on Cymbalta 60 mg daily for neuropathy and depression but depression not well controlled. Started Wellbutrin and has titrated up to 300 mg qAM and 150 mg qPM, but we decreased Cymbalta to 30 mg daily due to increased risk of seizures. Patient has been doing better from a depression perspective but is having more neuropathic pain. Would it be ok to increase Cymbalta to 60 mg daily? Patient without history of seizures or head trauma.

• Patient with anxiety and Generalized Anxiety Disorder–7 Scale score of 12. Functions well during the day but has situational anxiety that is worse, especially flying on a plane. Last primary care provider gave him Ambien, which didn't help. I gave him Ativan 0.5 mg for his last flight. He eventually took 3 of those since it was a turbulent flight. It didn't help. I have given him Ativan 2 mg for his next flight...Is there another med you would recommend for flights or is there an adjunctive med you would recommend?

Side effects or interactions:

• Patient with history of episodic anxiety previously managed with as-needed propranolol and lorazepam while giving talks. Her anxiety is less predictable now but she is giving more talks at work and wanted to start maintenance therapy. She had used fluoxetine 10 mg in past with good effect. However when she took one pill this time she developed abdominal cramping, nausea, leg “tremors,” and increased anxiety...I know some nausea and muscle cramping can occur with starting selective serotonin reuptake inhibitors (SSRI) but the tremors and nervousness raise the question of serotonin syndrome for me. Should we try a different SSRI or is it okay to restart fluoxetine?

• Patient with anxiety successfully treated with Zoloft. However, the patient has severe essential tremor and has noticed worsening of tremor on this medicine. In my read of the other anxiety medications (SSRIs, serotonin norepinephrine reuptake inhibitors), it seems like they could all have this side effect. Would there be any medications for anxiety you might recommend that would be less likely to exacerbate her tremor?

Referral for psychotherapy:

Any ideas for couples’ therapists for this patient and her husband?

Navigating the health care system:

Patient was diagnosed with childhood depression and ADHD...Has some psychotic features such as auditory hallucinations, paranoid thoughts. No history of mania or suicidal ideation. Currently on Lexapro 20 mg daily...Do you have any recommendations as to where he can be seen that can provide him psychiatric services?

Need for psychiatry or other referral:

Patient with history of frontal lobe brain tumor status post resection who presents with mood lability...anxiety and depression. Symptoms may be due to location of tumor. Significantly impacting quality of life...Is this something that would best be managed by psychiatry or neurology?

Textbox 2 includes representative question text for each of the major domains and categories.

Most (49/50, 98%) consult questions pertained to management; only in 4% (2/50) was the PCP seeking consultation on diagnosis. Within the management domain, the majority were queries about medication management, including 74% (37/50) of questions about medication choice and 32% (16/50) of questions about drug side effects or interactions. PCPs asked about navigating the health care system in 14% (7/50) of the questions, specifically for assistance helping patients access specialty mental health treatment. For example, PCPs asked for recommendations for services based on insurance status (particularly for patients with public insurance) or for specific types of treatment such as counseling for posttraumatic stress disorder or diagnostic evaluation for eating disorders. Other questions pertained to recommendations about psychotherapy or in-person evaluation.

Psychiatrist responses focused on both diagnosis and management, with 60% (30/50) of responses addressing the diagnosis and all 50 (100%) making recommendations about management (Table 2). Within the domain of diagnosis, psychiatrists commented on the overall clinical picture, offering an impression, differential diagnosis, or working diagnosis for the clinical scenario in 56% (28/50) of responses. Psychiatrists recommended additional diagnostic testing in 6% (3/50) of cases or obtaining additional history in 14% (7/50) of cases. Recommended diagnostic testing consisted of studies, such as thyroid function testing or brain imaging, to exclude organic causes of symptoms. Recommendations for additional history included further clarification of current or prior symptoms,

responses to prior treatments, family history, and substance use history.

Management responses also focused on medications, with 72% (36/50) offering advice on choice of medication. Responses also included assistance with specific medication titration instructions in 42% (21/50) of cases, information about side effects in 48% (24/50) of cases, and conditions for starting or adjusting treatment in 34% (17/50) of cases. In 50% (25/50) of responses, consultants offered multiple therapeutic options to be guided by PCP choice or clinical conditions. For example, the consultant might recommend several different medication options depending on specific clinical situations, as seen in Textbox 1. Psychiatrists also offered advice on navigating the health care system in 22% (11/50) of responses and recommended psychotherapy as a primary or adjunctive treatment in 24% (12/50) of responses.

Recommendations for In-Person Consultation

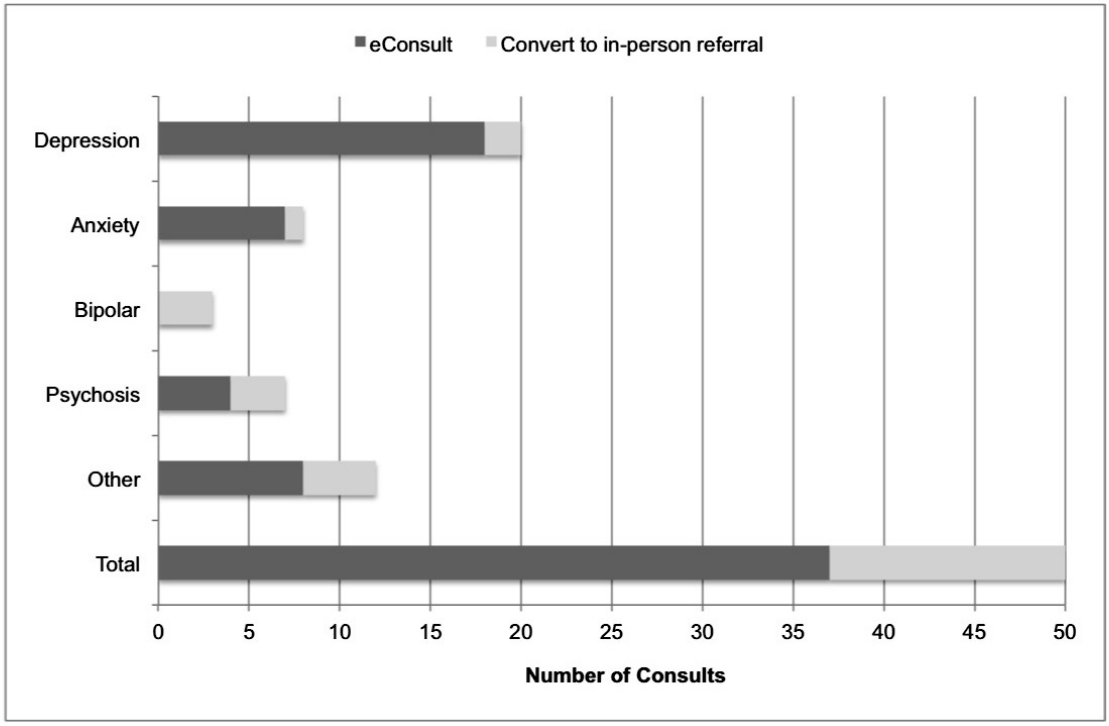
Psychiatry consultants recommended in-person evaluation or treatment in 26% (13/50) of cases and agreed with ongoing management in primary care for the remaining 74% (37/50) of cases (Figure 1). Recommendations also varied by diagnosis, with fewer of the depression and anxiety consults leading to recommendations for in-person evaluation. Psychiatrists recommended in-person evaluation for all 3 patients with bipolar disorder, 3 of 7 of those with psychotic disorders, and 4 of the other 12 consults using the generic template.

Table 2. Breakdown of primary care provider consult question and psychiatrist response text. Question and response texts were coded for 2 domains (diagnosis and management) and for presence and absence of specific categories within these. There could be multiple domains or categories per question or response.

Originator			n (%)
PCP[a] questions (N=50)			
	Diagnostic questions		2 (4)
	Management questions		49 (98)
		Medication choice	37 (74)
		Drug side effects or interactions	16 (32)
		Navigating the health care system	7 (14)
		Psychotherapy recommendations	3 (6)
		Need for in-person referral	3 (6)
Psychiatrist responses (N=50)			
	Diagnostic recommendations		30 (60)
		Interpretation of overall diagnosis	28 (56)
		Additional diagnostic testing	3 (6)
		Obtain additional history	7 (14)
	Management recommendations		50 (100)
		Medication choice	36 (72)
		Conditions for adjusting treatment	17 (34)
		Specific medication titration instructions	21 (42)
		Information about side effects	24 (48)
		Provided multiple therapeutic options	25 (50)
		Recommended psychotherapy	12 (24)
		Navigating health care system	11 (22)

[a]PCP: primary care provider.

Figure 1. Referral for in-person consultation by consult category.



Implementation of Consultant Recommendations

On chart review, PCPs implemented psychiatrist recommendations in 76% (38/50) of consults. The implementation reflected the consultant responses, with the majority of cases consisting of PCPs ordering, modifying, or discontinuing medications. In cases where the recommendations were not implemented, situations varied. For some the patient either improved without intervention or did not follow up. In other cases, PCPs either documented reasons for choosing not to implement recommendations or did not acknowledge the consultant recommendations in the chart. All consults received responses, and the average consultant response time was 1.4 business days, with a median response time of 1 day and range of 0 to 30 days. In comparison, mean and median wait times for an in-person consult in the psychiatry practice during the final 3 months of the study period were 46 days and 38 days, respectively.

Discussion

Principle Findings

This study describes the early experience of a large health practice with an innovative program using eConsults from primary care to psychiatry for management of mental health problems. We found that psychiatry eConsults were a feasible way to support PCPs in managing complex psychiatric issues in the primary care setting, with PCPs using the service and frequently implementing at least one of the recommendations. For the majority of patients, particularly those with depression and anxiety, the consulting psychiatrist supported ongoing management within primary care without requesting an in-person psychiatric evaluation and provided a range of strategies that facilitated ongoing primary care–based treatment. In addition, eConsults provided consultation that circumvented common barriers to mental health treatment, including access to treatment providers and insurance coverage since eConsults were available to patient with all payer types without prior authorization from their insurer.

PCPs mainly used eConsults for guidance about treatment, particularly the management of psychotropic medications. PCPs generally made their own diagnostic assessments and used eConsults when confronted with treatment resistance or treatment failure. They also used eConsults to query about systems of care, either navigating the system more broadly or specific recommendations about psychotherapy or psychiatry. Although they were rarely directly asked, the psychiatric consultants often offered suggestions regarding diagnosis, which generally involved summarizing or clarifying the diagnosis. This may reflect a role for eConsults as both management and educational tools.

Psychiatry eConsults share several key features with eConsults in other specialties. eConsults to psychiatry provided timely recommendations for care, with average response time similar to reported values of less than 3 days [21]. Implementation of consultant recommendations was also similar to the 65% to 85% rate described for medical subspecialties within our own academic health system [24,25]. Question content differed somewhat from medical subspecialties [25], with the majority of psychiatric consultations focusing exclusively on management in contrast with the large percent focused on diagnosis and monitoring for medical subspecialties. This may reflect the different nature of psychiatric diagnosis and treatment, which may require more in-person interaction than laboratory testing and imaging. These differences may explain why in-person evaluations were recommended more frequently when PCPs expressed diagnostic uncertainty.

In our study, some PCPs had queries regarding helping their patients navigate the health care system (14%). In particular, this was common for patients who had public insurance that excluded them from obtaining mental health services in our institution due to a carve-out arrangement. System navigation was not a common theme reported for consults for medical subspecialties [25] and may reflect the challenges of accessing mental health care due to fragmentation of both payment and delivery. Nationally, psychiatrists accept insurance at lower rates than other medical specialties, with particularly low acceptance of Medicaid plans [31]. We were unable to track the rates of psychiatry access for our patient population—largely due to insurance carve-outs that require patients to access psychiatric care outside of our health system—but we did find that our study population had a higher proportion of Medicaid and dual-eligible beneficiaries than the clinic population as a whole. PCPs may be using eConsults for patients who are not able to obtain specialty mental health care due to access barriers or for patient refusal. We also found that only one of the psychiatry eConsults was for a non-English speaking patient, as compared to a higher percentage of non-English speaking patients in the clinic as a whole. The literature demonstrates associations between low English proficiency and disparities in access to mental health services [32,33]. Although our small sample size makes it difficult to draw conclusions about these results, our findings may reflect disparities present for eConsults that are similar to more traditional psychiatric consultation.

Given the challenges of meeting growing mental health care demands, there is a need for novel approaches to integrated behavioral health. There is evidence that collaborative management between PCPs and behavioral health providers can provide high-quality care, improve access, reduce stigma, and be cost effective [10]. Other novel technologies including telepsychiatry and other eMental health interventions have been aimed at expanding and extending the delivery of specialty mental health care [34,35]. Like these technologies, eConsult allows for timely provision of care, particularly for those with barriers to access. However, eConsults differ from most described programs in that they target providers rather than patients, allowing for mental health care managed through the PCP in an integrated care model and via a secure EHR. This may help mitigate concerns raised about other eMental health technologies about security, quality, and a lack of a therapeutic relationship between provider and patient as well as worry that patients may defer needed care. Further, an eConsult model has the potential to take advantage of key strengths of the integrated behavioral health models including high levels of adherence and patient satisfaction and reduced stigma [10,36] because they are also based in the primary care setting.

Limitations

Our analysis has several limitations. First, as a largely qualitative analysis, our sample may not be completely representative of all possible eConsults to psychiatry. Our study was also conducted in a single institution with a robust EHR, and the results may not be generalizable to other institutions with different capabilities. However, we feel our study provides an in-depth look at the nature and scope of PCP questions regarding management of mental health conditions that may be more broadly applicable. Further, our initial interrater reliability was moderate for some portions of our content analysis, although we used this information to identify discrepancies in the coding process, refine our categories, and ultimately create a rigorous coding scheme derived through consensus within the research team. Another important limitation of our analysis is the lack of availability of data for mental health care utilization. Because of separate payment structures for mental health as well as a high percentage of psychiatry referrals outside of our health system, we were unable to track the percentage of eConsults compared to total psychiatry referrals or the impact of eConsult on utilization. We were also unable to evaluate treatment outcomes or compare these with outcomes for traditional care.

Conclusions

Overall, our data suggest that eConsults show promise as one means of supporting primary care providers to deliver mental health care to patients with common psychiatric disorders, particularly those with depression and anxiety. By extending and enhancing traditional psychiatric consultation within the primary care setting, the eConsult model adds to the burgeoning literature on innovative models for integrated mental health. Future efforts should be directed at evaluating the effectiveness of this intervention in terms of clinical and systems outcomes as well as PCP and patient satisfaction to improve overall understanding of best practices for behavioral health integration.

Acknowledgments

The authors would like to acknowledge Jeffrey Tice, MD, James Bourgeois, MD, and Lyudmila Popova, PhD, for comments on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Study setting.

[PDF File (Adobe PDF File), 21KB - jmir_v19i8e279_app1.pdf]

Multimedia Appendix 2

Development and implementation of eConsult.

[PDF File (Adobe PDF File), 21KB - jmir_v19i8e279_app2.pdf]

Multimedia Appendix 3

eConsult templates.

[PDF File (Adobe PDF File), 22KB - jmir_v19i8e279_app3.pdf]

References

1. 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]
2. Roehrig C. Mental disorders top the list of the most costly conditions in the United States: $201 billion. Health Aff (Millwood) 2016 Jun 01;35(6):1130-1135. [doi: 10.1377/hlthaff.2015.1659] [Medline: 27193027]
3. Katon W, Schulberg H. Epidemiology of depression in primary care. Gen Hosp Psychiatry 1992 Jul;14(4):237-247. [Medline: 1505745]
4. Ansseau M, Dierick M, Buntinkx F, Cnockaert P, De SJ, Van den Haute M, et al. High prevalence of mental disorders in primary care. J Affect Disord 2004 Jan;78(1):49-55. [Medline: 14672796]
5. Stafford RS, Ausiello JC, Misra B, Saglam D. National patterns of depression treatment in primary care. Prim Care Companion J Clin Psychiatry 2000 Dec;2(6):211-216 [FREE Full text] [Medline: 15014631]
6. Olfson M, Kroenke K, Wang S, Blanco C. Trends in office-based mental health care provided by psychiatrists and primary care physicians. J Clin Psychiatry 2014 Mar;75(3):247-253. [doi: 10.4088/JCP.13m08834] [Medline: 24717378]
7. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009;28(3):490-501 [FREE Full text] [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
8. 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]
9. Olfson M. The rise of primary care physicians in the provision of US mental health care. J Health Polit Policy Law 2016 Aug;41(4):559-583. [doi: 10.1215/03616878-3620821] [Medline: 27127264]
10. Collins C, Hewson DL, Munger R, Wade T. Evolving models of behavioral health integration in primary care. 2010. URL: https://www.milbank.org/wp-content/uploads/2016/04/EvolvingCare.pdf [accessed 2017-04-26] [WebCite Cache ID 6q0Bsr2xA]
11. Liddy C, Drosinis P, Keely E. Electronic consultation systems: worldwide prevalence and their impact on patient care-a systematic review. Fam Pract 2016 Jun;33(3):274-285. [doi: 10.1093/fampra/cmw024] [Medline: 27075028]
12. Horner K, Wagner E, Tufano J. Electronic consultations between primary and specialty care clinicians: early insights. Issue Brief (Commonw Fund) 2011 Oct;23:1-14. [Medline: 22059281]
13. Angstman KB, Adamson SC, Furst JW, Houston MS, Rohrer JE. Provider satisfaction with virtual specialist consultations in a family medicine department. Health Care Manag (Frederick) 2009;28(1):14-18. [doi: 10.1097/HCM.0b013e318196def8] [Medline: 19225331]
14. Stoves J, Connolly J, Cheung CK, Grange A, Rhodes P, O'Donoghue D, et al. Electronic consultation as an alternative to hospital referral for patients with chronic kidney disease: a novel application for networked electronic health records to improve the accessibility and efficiency of healthcare. Qual Saf Health Care 2010 Oct;19(5):e54. [doi: 10.1136/qshc.2009.038984] [Medline: 20554576]
15. Kim-Hwang JE, Chen AH, Bell DS, Guzman D, Yee HF, Kushel MB. Evaluating electronic referrals for specialty care at a public hospital. J Gen Intern Med 2010 Oct;25(10):1123-1128 [FREE Full text] [doi: 10.1007/s11606-010-1402-1] [Medline: 20512531]
16. 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]
17. 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]
18. Rodriguez KL, Burkitt KH, Bayliss NK, Skoko JE, Switzer GE, Zickmund SL, et al. Veteran, primary care provider, and specialist satisfaction with electronic consultation. JMIR Med Inform 2015 Jan 14;3(1):e5 [FREE Full text] [doi: 10.2196/medinform.3725] [Medline: 25589233]
19. Davis A, Gilchrist V, Grumbach K, James P, Kallenberg R, Shipman SA. Advancing the primary/specialty care interface through eConsults and enhanced referrals. Ann Fam Med 2015;13(4):387-388 [FREE Full text] [doi: 10.1370/afm.1829] [Medline: 26195689]

20. 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 01;36(3):492-499. [doi: 10.1377/hlthaff.2016.1283] [Medline: 28264951]
21. 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]
22. Golberstein E, Kolvenbach S, Carruthers H, Druss B, Goering P. Effects of electronic psychiatric consultations on primary care provider perceptions of mental health care: Survey results from a randomized evaluation. Healthc (Amst) 2017 Feb 02:1 (forthcoming). [doi: 10.1016/j.hjdsi.2017.01.002] [Medline: 28162990]
23. Gleason N, Prasad PA, Ackerman S, Ho C, Monacelli J, Wang M, et al. Adoption and impact of an eConsult system in a fee-for-service setting. Healthc (Amst) 2017 Mar;5(1-2):40-45. [doi: 10.1016/j.hjdsi.2016.05.005] [Medline: 27469441]
24. Cruz ML, Gleason N, Wang M, Wrenn K, Gonzales R. Transforming the endocrine consult: asynchronous provider consultations. Endocr Pract 2015 May;21(5):514-521. [doi: 10.4158/EP14297.OR] [Medline: 25667365]
25. Wrenn K, Catschegn S, Cruz M, Gleason N, Gonzales R. Analysis of an electronic consultation program at an academic medical centre: Primary care provider questions, specialist responses, and primary care provider actions. J Telemed Telecare 2017 Feb;23(2):217-224. [doi: 10.1177/1357633X16633553] [Medline: 26940797]
26. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, et al. A taxonomy of generic clinical questions: classification study. BMJ 2000 Aug 12;321(7258):429-432 [FREE Full text] [Medline: 10938054]
27. Borowitz SM, Wyatt JC. The origin, content, and workload of e-mail consultations. JAMA 1998 Oct 21;280(15):1321-1324. [Medline: 9794310]
28. Donohoe MT, Kravitz RL, Wheeler DB, Chandra R, Chen A, Humphries N. Reasons for outpatient referrals from generalists to specialists. J Gen Intern Med 1999 May;14(5):281-286 [FREE Full text] [Medline: 10337037]
29. Burla L, Knierim B, Barth J, Liewald K, Duetz M, Abel T. From text to codings: intercoder reliability assessment in qualitative content analysis. Nurs Res 2008;57(2):113-117. [doi: 10.1097/01.NNR.0000313482.33917.7d] [Medline: 18347483]
30. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. BMJ 2000 Jan 08;320(7227):114-116 [FREE Full text] [Medline: 10625273]
31. Bishop TF, Press MJ, Keyhani S, Pincus HA. Acceptance of insurance by psychiatrists and the implications for access to mental health care. JAMA Psychiatry 2014 Feb;71(2):176-181 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.2862] [Medline: 24337499]
32. Ohtani A, Suzuki T, Takeuchi H, Uchida H. Language barriers and access to psychiatric care: a systematic review. Psychiatr Serv 2015 Aug 01;66(8):798-805. [doi: 10.1176/appi.ps.201400351] [Medline: 25930043]
33. Cheung S, Mishkin A, Shapiro P. "What we have here is a failure to communicate:" association of preferred language with the rate of psychiatric consultation. Psychosomatics 2017;58(4):421-426. [doi: 10.1016/j.psym.2017.02.006] [Medline: 28413091]
34. 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]
35. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
36. Archer J, Bower P, Gilbody S, Lovell K, Richards D, Gask L, et al. Collaborative care for depression and anxiety problems. Cochrane Database Syst Rev 2012 Oct 17;10:CD006525. [doi: 10.1002/14651858.CD006525.pub2] [Medline: 23076925]

Abbreviations

EHR: electronic health record
PCP: primary care provider
UCSF: University of California, San Francisco
wRVU: work relative value unit

Edited by G Eysenbach; submitted 26.04.17; peer-reviewed by C Liddy, K Wrenn; comments to author 25.05.17; revised version received 12.07.17; accepted 13.07.17; published 04.08.17.

Please cite as:
Lowenstein M, Bamgbose O, Gleason N, Feldman MD
Psychiatric Consultation at Your Fingertips: Descriptive Analysis of Electronic Consultation From Primary Care to Psychiatry
J Med Internet Res 2017;19(8):e279
URL: http://www.jmir.org/2017/8/e279/
doi:10.2196/jmir.7921
PMID:28778852

Original Paper

An Internet-Based Method for Extracting Nursing Home Resident Sedative Medication Data From Pharmacy Packing Systems: Descriptive Evaluation

Tristan Ling[1], BComp (Hons), PhD; Peter Gee[1], BPharm (Hons); Juanita Westbury[2], BPharm, GradDipCommPracPharm, MSc, PhD; Ivan Bindoff[1], BComp (Hons), PhD; Gregory Peterson[1], BPharm (Hons), PhD, MBA, FSHP, FACP, GAICD, AACPA, ARPharmS, MPS

[1]Unit for Medication Outcomes Research and Education, Division of Pharmacy, School of Medicine, University of Tasmania, Hobart, Australia
[2]Wicking Dementia Research and Education Centre, Faculty of Health, University of Tasmania, Hobart, Australia

Corresponding Author:
Tristan Ling, BComp (Hons), PhD
Unit for Medication Outcomes Research and Education
Division of Pharmacy, School of Medicine
University of Tasmania
Private Bag 26
University of Tasmania
Hobart, 7001
Australia
Phone: 61 362267396
Fax: 61 362262870
Email: Tristan.Ling@utas.edu.au

Abstract

Background: Inappropriate use of sedating medication has been reported in nursing homes for several decades. The Reducing Use of Sedatives (RedUSe) project was designed to address this issue through a combination of audit, feedback, staff education, and medication review. The project significantly reduced sedative use in a controlled trial of 25 Tasmanian nursing homes. To expand the project to 150 nursing homes across Australia, an improved and scalable method of data collection was required. This paper describes and evaluates a method for remotely extracting, transforming, and validating electronic resident and medication data from community pharmacies supplying medications to nursing homes.

Objective: The aim of this study was to develop and evaluate an electronic method for extracting and enriching data on psychotropic medication use in nursing homes, on a national scale.

Methods: An application uploaded resident details and medication data from computerized medication packing systems in the pharmacies supplying participating nursing homes. The server converted medication codes used by the packing systems to Australian Medicines Terminology coding and subsequently to Anatomical Therapeutic Chemical (ATC) codes for grouping. Medications of interest, in this case antipsychotics and benzodiazepines, were automatically identified and quantified during the upload. This data was then validated on the Web by project staff and a "champion nurse" at the participating home.

Results: Of participating nursing homes, 94.6% (142/150) had resident and medication records uploaded. Facilitating an upload for one pharmacy took an average of 15 min. A total of 17,722 resident profiles were extracted, representing 95.6% (17,722/18,537) of the homes' residents. For these, 546,535 medication records were extracted, of which, 28,053 were identified as antipsychotics or benzodiazepines. Of these, 8.17% (2291/28,053) were modified during validation and verification stages, and 4.75% (1398/29,451) were added. The champion nurse required a mean of 33 min website interaction to verify data, compared with 60 min for manual data entry.

Conclusions: The results show that the electronic data collection process is accurate: 95.25% (28,053/29,451) of sedative medications being taken by residents were identified and, of those, 91.83% (25,762/28,053) were correct without any manual intervention. The process worked effectively for nearly all homes. Although the pharmacy packing systems contain some invalid patient records, and data is sometimes incorrectly recorded, validation steps can overcome these problems and provide sufficiently accurate data for the purposes of reporting medication use in individual nursing homes.

(J Med Internet Res 2017;19(8):e283) doi:10.2196/jmir.6938

KEYWORDS

electronic health records; information storage and retrieval; inappropriate prescribing; antipsychotic agents; benzodiazepines; nursing homes; systematized nomenclature of medicine; health information systems

Introduction

Reducing Use of Sedatives: "RedUSe"

It has been shown that residents within nursing homes often receive sedating medications contrary to guidelines [1-6]. A multi-strategic, interdisciplinary intervention called Reducing Use of Sedatives (RedUSe) was developed in Tasmania in 2008 aiming to address inappropriate sedative use in nursing homes [7], where the term "sedative" referred to "psycholeptic" medication or antipsychotic or anxiolytic or hypnotic classes (see Table 1 in the Methods section of this paper for the Anatomical Therapeutic Chemical (ATC) codes for included medications). The project approached the problem of inappropriate sedative use by

1. Performing an audit of sedative use across all residents in the nursing home.
2. Presenting audit feedback to nursing home staff, benchmarked against average nursing home sedative use, during an interactive education session.
3. Developing personalized sedative review plans for each resident taking sedative medication, with input from a "champion nurse" and the home's pharmacist, for the attention of the prescriber.

All steps were repeated 3 months after the initial audit, and steps 1 and 2 repeated again at 6 months.

The RedUSe program was tested in a controlled trial in 2009 with 25 homes [7]. Most nursing homes obtain their medications from community pharmacies that utilize commercial computerized medication packing systems to record and pack each resident's medications into separate blister packs or sachets. Audit data was collected by installing software in each supply pharmacy to extract residents' medications from these dispensing and packing databases. The software was compatible with the two most common dispensing and packing systems in Australia, FRED [8] and Webstercare [9]. Data mappings were created between these two systems' antipsychotic and benzodiazepine identifiers to ATC codes, allowing automated production of audit reports charting the prevalence of use of each of these drug groups in each nursing home. This process required an in-person visit to the supply pharmacy and the nursing home for verification against resident medication charts [7].

The RedUSe trial significantly reduced rates of both antipsychotic and benzodiazepine use in intervention homes when compared with control homes [7]. This success led to funding from the Australian government to expand the project on a national scale. However, the remoteness of some nursing homes and the scale of data collection necessitated an improved data extraction method.

Extracting Nursing Home Medication Data

A recent systematic review identified 22 interventions that have been developed to address inappropriate antipsychotic use in nursing homes [4]. This review noted that medication audit initiatives typically collect data by visiting the home and copying resident charts [4,10,11] using cohorts already detailed by other studies [12,13] or by accessing electronic pharmacy records [14,15].

Electronic extraction of medication records from the nursing homes themselves was not possible, as at the time of project implementation many nursing homes were not using electronic medication records. Similarly, the relatively new Australian national electronic health record (EHR) system (formerly the Personally Controlled Electronic Health Record, now My Health) still has a low adoption rate, both from health practitioners and patients [16]. Although prescriptions are being increasingly computerized at the point of prescription, they are not universally so, and accessing the computer systems of each prescriber for each resident in a nursing home would be impractical.

Community pharmacy records are an increasingly viable source of data [17]. Many examples of data collection using pharmacy records exist, with outcomes including an influenza monitoring system [18], validating hospital admission drug charts [19], providing decision support to pharmacists [20], quantifying medicine use [21], and identification of patients for intervention [22,23]. Pharmacies are also a condensed source, as it is generally expected that a nursing home is supplied medications by one or two pharmacies. Additionally, pharmacies often supply multiple nearby nursing homes. Literature has noted that, thus far, pharmacy data collection procedures have typically been small scale, performed by one individual, and lacking a procedure for checking accuracy [4], providing sufficiently accurate data for aggregate analysis but not for personalised intervention.

For larger scale studies, the remoteness of some nursing homes makes in-person visits impractical. An alternative is to recruit a person already employed by each nursing home: within this project a "champion nurse" was designated within each home to promote and lead the project, as published literature has consistently noted this as critical to success [24-28]. However, other project requirements left little capacity for the champion nurse to also facilitate data collection; considering it has been reported that the three most stressful factors in aged care nursing are "not having enough staff," "having too much work to do," and "interruptions to regular work" [29], performing significant data entry in addition to promoting cultural change is likely too large a burden.

Using Extracted Medication Data

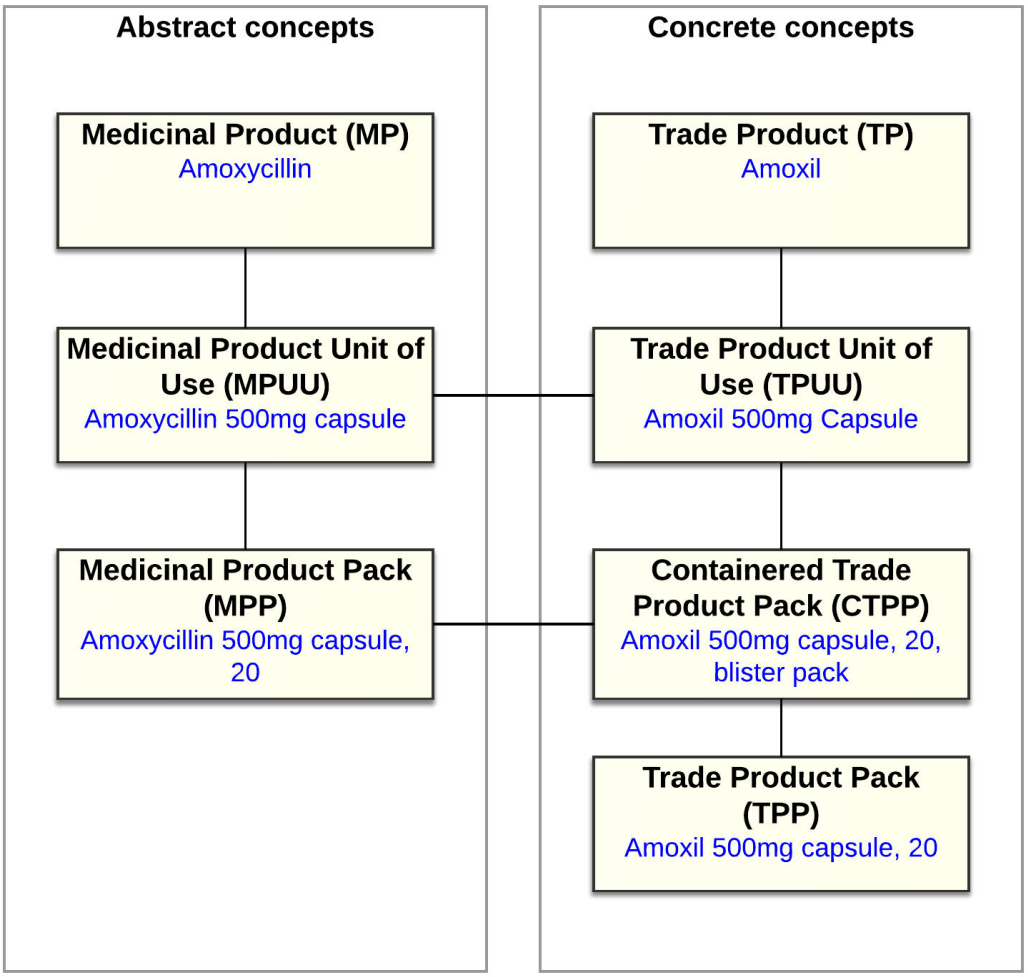
There are a small number of pharmacy medication packing systems in use in Australia, yet each store their data in very different formats. To be able to identify medications of interest and report on them meaningfully, an extract, transform, and load (ETL) process is required that can translate medicines to a common standard for comparison.

In Australia, the Australian Digital Health Agency (ADHA) has created the Australian Medicines Terminology (AMT), an organized collection of codes and descriptions for uniquely identifying originator and generic medication brands used in Australia. The codes are hierarchically linked (see Figure 1) so that for any given medication it is possible to identify the active ingredients, dosage form, strength, trade name, pack type, and pack size [30]. By providing independent, unique codes, the AMT is intended to allow long-term, reliable communication of medication information between different systems. Uptake of the AMT has been slow, particularly in commercial software, but it is being increasingly used in research and analysis [31-33].

The AMT is a subset of the Systemised Nomenclature of Medicine Clinical Terms (SNOMED CT), an international standard for encoding clinical terms [34]. AMT equivalent systems exist outside of Australia, such as the UK NHS Dictionary of Medicines and Devices (dm+d) [35] or the Singapore Drug Dictionary (SDD) [36]. In the United States, drugs can be encoded using the US edition of SNOMED CT's Product and Substance concepts or using the terminology set RxNorm, which has been mapped to SNOMED CT [37].

Given that collecting data on sedative use directly from nursing homes is impractical, pharmacy medication packing systems provide both the next-closest data point, plus the least number of data sources. They also provide the most consistently electronic records, including those kept within nursing homes. An ETL approach to retrieving data from pharmacy systems, combined with a validation and verification process from nursing staff at the point of care, was developed to provide accurate data to facilitate audits of sedative medication use, reporting and education, and personalized medication review plans for nursing home residents on a national scale.

Figure 1. The hierarchical concepts used to describe medicines in the Australian Medicines Terminology.



Methods

Data Extraction

To collect medication data from community pharmacies, a website and client program was developed. Extracted data was securely sent to a webserver, which kept detailed transaction logs of every record added, modified, and deleted. The client program was able to extract data from the FRED, Webstercare, MPS [38], and Minfos [39] medication packing systems. Packing systems were prioritized over dispensing as they provide a simpler measure of quantities supplied. As per the requirements of the broader intervention, data was extracted as a "point-in-time" snapshot, and as with the previous RedUSe study, 3 snapshots were taken at 3-month intervals.

The collection of data is a process run by a staff member of the supplying pharmacy. The staff member

1. Creates an account on the website.
2. Logs into the website from a computer in the pharmacy and downloads and runs client.
3. In the RedUSe client software, follows prompts to identify pharmacy and packing software brand
   a. For Fred and Webstercare, the software locates the packing database; if atypical, it asks staff for location or to select from multiple possibilities.
   b. For MPS and Minfos, the staff member generates a medication report and opens it in the RedUSe client.
4. In the RedUSe client software, links the pharmacy system wing or home group names corresponding to the wing or home names of the RedUSe-participating nursing home(s).

Once the data source was identified in step 3, the software established a connection appropriate to the packing system (eg, opening a database connection or opening the file for reading). Resident and medication data for the chosen homes were then retrieved, with resident names encrypted. Medication data included a packing system medication identifier, medication name, instructions for use, quantity supplied, date started, date ceased, and prescriber. All data was compressed and transmitted to the server using transport layer security (TLS) encryption. The uploaded data from one extraction (referred to as "an upload") contained all residents and medications for all currently participating nursing homes supplied by that pharmacy. The client program was a small (200 kilobyte) download.

Data Enrichment

The medication identifiers used by each of the packing systems were mapped to the appropriate AMT concept. Depending on the software, this was either the Trade Product Unit of Use (TPUU) or the Containered Trade Product Pack (CTPP). Using the AMT, these concepts were then translated to the corresponding Medicinal Product Unit of Use (MPUU). Finally, mappings were created between relevant AMT MPUU concepts and the appropriate ATC codes, providing a link from the medicine identifiers used in each packing program to a consistent naming scheme and drug classification system (see Figure 2 for examples). Where possible, mappings were created automatically based on known codes such as the Australian Register of Therapeutic Goods. Any missing mappings were added by a team of two research pharmacists.

Using the ATC code, medications of interest (for the RedUSe project, sedative medications) were automatically identified and aggregated based on the ATC codes shown in Table 1. Using this data, reports were generated by the server, on request, describing prevalence of antipsychotic and benzodiazepine medication use in each nursing home. Additionally, a measure of sedative load was calculated for each resident: the average daily dose for each medication was determined and compared with the World Health Organization's Defined Daily Dose (DDD) values [40]; to enable reporting at a nursing home level, daily doses were converted to an equivalent strength of diazepam for benzodiazepines and chlorpromazine for antipsychotics. The calculations did not include sedative medications taken pro re nata (PRN, "as required"), rather than taken regularly, due to difficulty quantifying administered PRN doses.

It is noteworthy that this process was implemented using version 2 of the AMT; version 3 was released while this project was ongoing, but there was no capacity to convert while the project was running. As AMT version 2 did not include some newer medications, there were some medications that could not be automatically linked. The impact of this is discussed later.

Table 1. ATC codes used to identify drugs of interest.

ATC Code	Excluding	Name
N05A		Antipsychotics
	N05AN	Excluding Lithium
	N05AB04	Excluding Prochlorperazine
N05BA		Anxiolytics (Benzodiazepine derivatives)
N05CD		Hypnotics and Sedatives (Benzodiazepine derivatives)
N05CF		Benzodiazepine-related drugs
N03AE01		Clonazepam

Figure 2. Example mappings for each packing software brand of a temazepam product.

Packing Brand	Packing ID	AMT CTPP	AMT TPUU	AMT MPUU	ATC
Fred	TEMA5	Temaze 10 mg tablet: uncoated, 25 tablets, bottle	Temaze (temazepam 10 mg) tablet: uncoated	Temazepam 10 mg tablet	N05CD07
Webster	6442	–			
MPS	TEMA5	Temaze 10 mg tablet: uncoated, 25 tablets, bottle			
Minfos	57632	Temaze 10 mg tablet: uncoated, 25 tablets, bottle			

Data Validation and Verification

Once the medication data was uploaded, there was a series of validation and verification steps performed. First, the names of all medication records that were unable to be automatically linked to the AMT were sent to a research pharmacist. If the pharmacist identified any unlinked sedative medications, those mappings were added; if mappings could not be made (eg, the medication did not exist in the AMT), the medication was instead mapped to a chemically equivalent medication to ensure that all sedative use was captured in the resident's DDD. The medication name used in the prescription data was still presented to nursing staff to avoid any confusion.

Following this, the uploaded data was checked by project staff for any obvious errors that would require the upload to be reattempted and for any clear mistakes that could be unambiguously corrected.

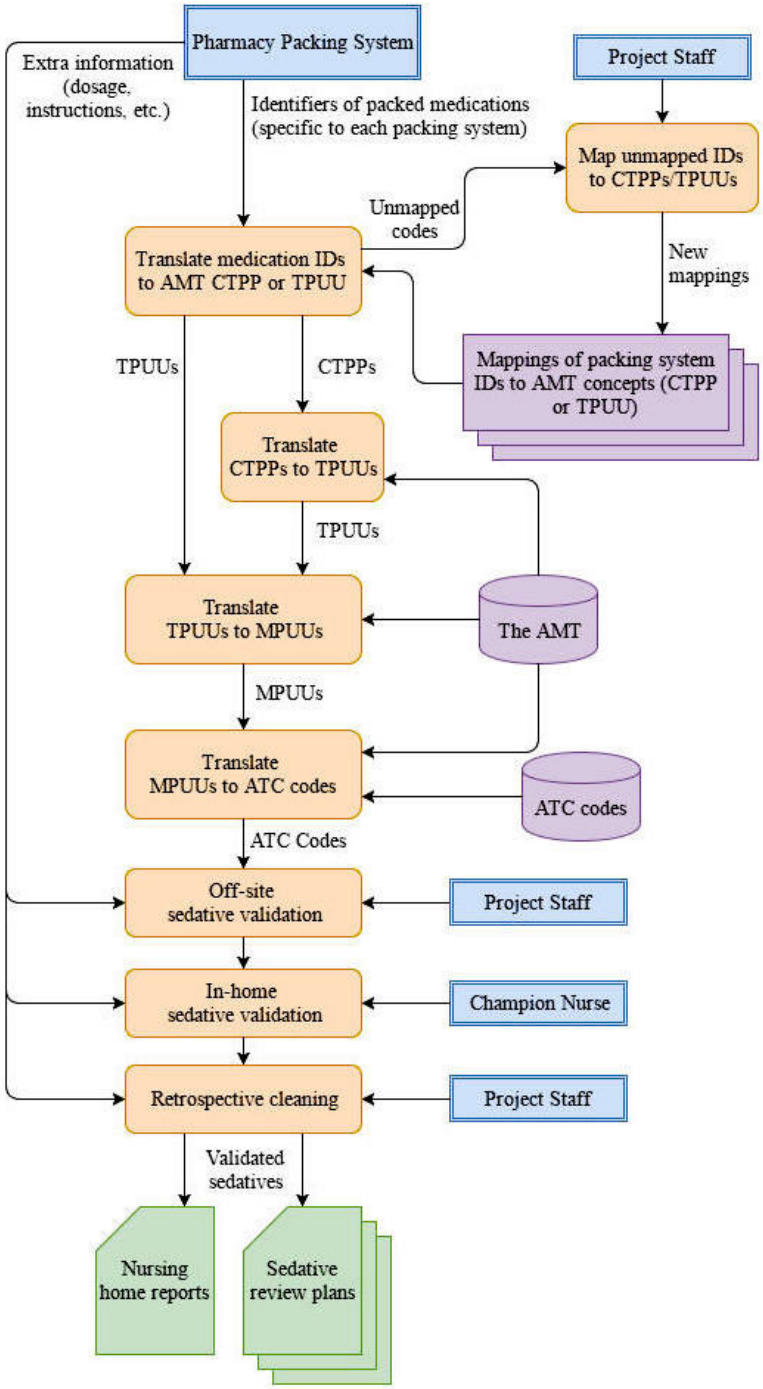
Next, the champion nurse at the nursing home was asked to log into the website and verify that all residents were present and that all sedative medications were present and correct (including name, dose, and instructions). Nurses were not asked to quantify administered doses of PRN medication. They were also asked to add any missing residents and medications, particularly looking for medication dosage forms that may have been packed separately (eg, drops, wafers, or injections). Separately packed medicines were sometimes not uploaded depending on the type of software used by the pharmacy and how the pharmacist recorded the medications. The champion nurse also removed any residents or medications that were incorrect or no longer current; if a resident was identified as no longer residing in the home, their medications were automatically excluded. For these checks, any missing or suspect data (such as doses falling outside of a typical range) were highlighted on the website. Once all residents had their sedative records verified, the option to generate the audit report and sedative review plans became available, with a review plan generated for every resident prescribed one or more sedative medications (the content and process for these review plans are beyond the scope of this paper).

Final data validation was also performed retrospectively by project staff, primarily to merge duplicate or split resident accounts but also to improve the consistency of the data for final reports issued to nursing homes. This addressed issues introduced by errors in the extraction process and any inconsistencies in the previous validation steps between nursing homes. The full ETL and validation process is shown in Figure 3.

Manual Data Entry

The typical approach to conducting medication audits in nursing homes is to visit the home and inspect the resident charts, recording the data by hand. If automated extraction of medication data was not possible, the champion nurse and regional project pharmacist used this as a fallback solution. In this case, the website functionality for data validation and verification could be used as an entry-assistance tool: forms for entering missing residents and medications already existed including auto-completing fields, instruction translation, and error highlighting. Using the same forms also ensured that entered medicines were in the same enriched format as uploaded medicines. To simplify this process, only sedative medication was required (defined according to Table 1 ATC codes), and any data entry snapshots after the first had the capacity to prepopulate from existing data, requiring only the changes since the last snapshot to be entered.

Figure 3. Extract, transform, and load (ETL) and validation process.



Recruitment, Privacy, and Consent

Recruitment of nursing homes aimed to provide a range of sizes, locations, ruralities, and organizations. Two national nursing home groups offered all of their facilities to be involved. Two aged care advocacy bodies assisted in the recruitment of smaller organizations and independent nursing homes by featuring "RedUSe" in industry journals, resulting in over 300 expressions of interest. From this cohort, facilities were selected for invitation, with an effort to meet the described criteria. Invitations continued until consent was gained from the target number of 150 RACFs. Facilities with less than 29 residents or from outer rural or remote locations, and those located in the Northern Territory, were excluded for logistical reasons. Supply pharmacies were approached for consent once the nursing home agreed to participate. Pharmacies were provided detail on the project and the software and were offered a small remuneration for facilitating data collection for a nursing home.

During the project, identifiable patient information was encrypted on the server and only decrypted for authorized staff at that nursing home and the project staff member responsible for contact with that nursing home. Nursing home level aggregations were only available to the management of that

home, the home's pharmacist, the project staff member in contact, and the project lead.

Full ethical approval for the study was obtained from the Human Research Ethics Committee (Tasmania) Network (reference, H0013545).

Results

Extracted Data

Of participating nursing homes, 94.7% (142/150) had resident and medication data successfully uploaded via the automated process, supplied by 81 pharmacies. The eight participating nursing homes that did not have data uploaded were serviced by five pharmacies. One pharmacy owner declined to participate. The other four pharmacies used packing systems not supported by the extraction software: three used a platform that did not allow the software to be run; the other used a new brand of packing software.

Of the 150 nursing homes, 95.3% (143/150) were supplied by a single pharmacy, 4% (6/150) were supplied by two pharmacies, and one home was supplied by three pharmacies. On average, there were 90 residents per nursing home (standard deviation [SD]=37.5).

In total, 311 uploads were completed. Of these, 73.3% (228/311) included one nursing home, 21.9%, (68/311) included two nursing homes, 9 included three nursing homes, 5 included four nursing homes, and 1 included five nursing homes. Facilitating these took the supplying pharmacist a mean of 15 min per upload (standard error=0.67). Data upload size was, on average, 54 kilobytes.

Overall, 17,722 resident profiles were extracted, with 11.07% (1961/17,722) identified as duplicate or secondary records for an existing resident and merged into other records. For homes where data was automatically uploaded, 815 additional residents, or 4.4% of total (815/18,537) were added manually during data validation (shown in Figure 4). For uploaded residents, 546,535 medication records were extracted, of which, 5.13% (28,053/546,535) were identified as antipsychotics or benzodiazepines. There were 2291 antipsychotic and benzodiazepine records (8.16% of 28,053) modified during the validation steps, with a total of 3814 modifications. A further 4.75% (1398/29,451) of total antipsychotic and benzodiazepine medications were added manually (shown in Figure 5). Sedative medications taken PRN were significantly more likely not to be included in packing programs (3.62% [586/16,179] of regular medications were missed, against 6.12% [812/13,272] of PRN medications missed; two-tailed z-score −10.02, $P<.001$).

Figure 4. Residents extracted, removed, and added.

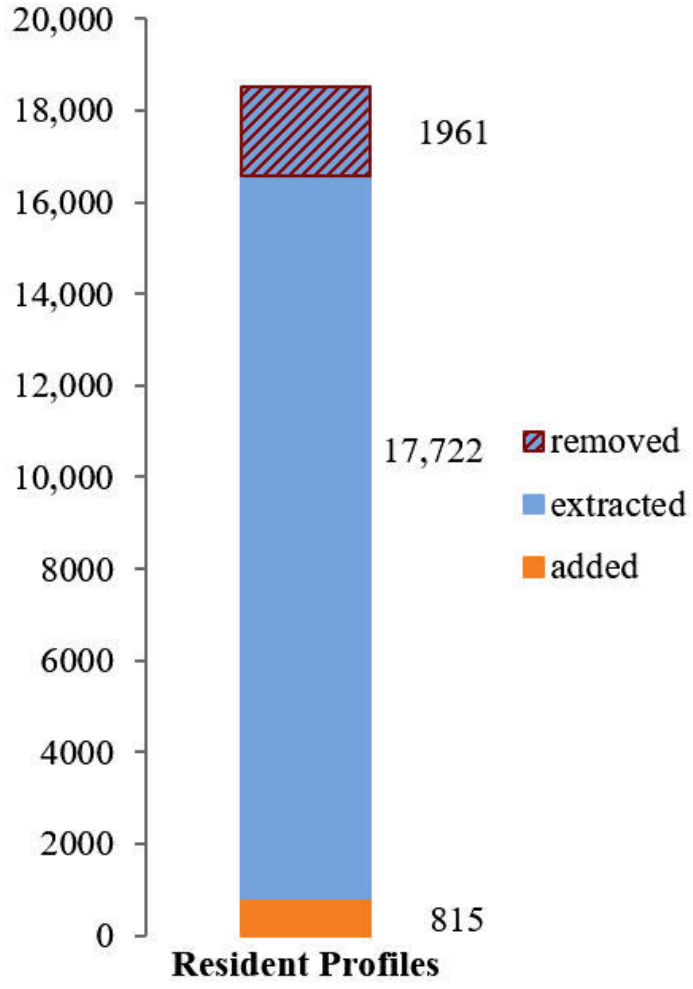
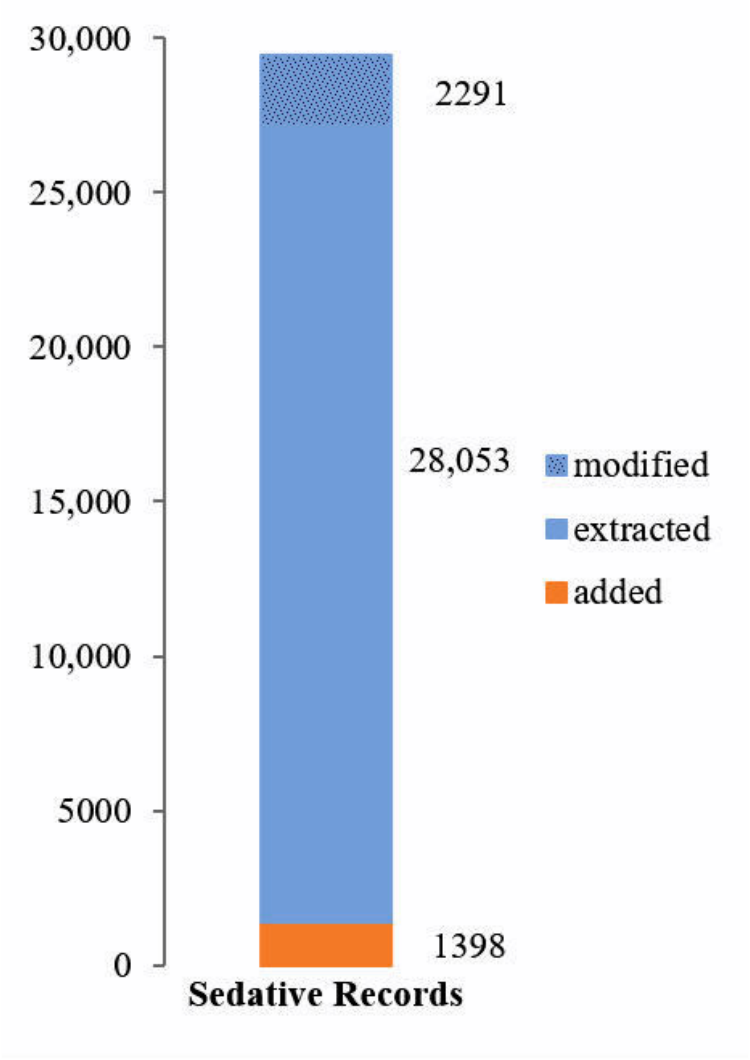


Figure 5. Sedatives extracted, corrected, and added.



Data Errors and Corrections

Validation and Verification

All manual corrections have been broken down into each validation and verification step in Table 2. In phase one, 9.56% (219/2291) of incorrect medications required a manual mapping. Of these, 82.2% (180/219) were recently introduced medications that would have been automatically linked if AMT version 3 was used; the remainder were due to out-of-date pharmacy vendor databases, unusual prescriptions (eg, brandy), or customized medications.

During the third phase validation, 48.69% (1870/3841) of all corrections were made by champion nurses and regional project pharmacists. Off-site validation and retrospective data cleaning, performed by project researchers, accounted for the remaining 44.91% (1725/3841) of modifications made.

Known Errors

Of the errors in extracted medicine data, the most frequent problems were missing dosage quantities. These could all be accounted for by three mistakes in the extraction software. One flaw, for one pharmacy software type, resulted in 325 records missing dosage numbers (24.83% (325/1309) of extracted records for that software in that time). A similar problem with another packing system accounted for another 284 records missing dosage numbers, or 8.73% (284/3254). Once corrected, subsequent dosage extractions from those software types required only 15 modifications, or 1.43% (15/1047) of records from that software in that time. The largest single cause of missing dosages was triggered by a pharmacy packing system changing their data structure during the project. This resulted in another 2108 records missing dosage numbers that needed to be entered based on the extracted instructions (representing 42.95% [2108/4908] of records in that time, up from 3.84% [125/3254] before the change). Altogether, these three missing dosage data issues accounted for 75.58% (2717/3595) of the sedative record corrections after mappings had been made. Removing these from consideration (as they are now corrected issues) alters the accuracy and validation requirements substantially, as shown in the second section of Table 2. Besides dosage quantities, all other extracted medicine attributes had approximately equal frequencies of error, with an average of 2% of uploaded records needing modification.

Table 2. Medication corrections made in each data validation and verification phase.

Phase	Corrections n (%)	Corrections excluding known errors n (%)
1: Identifying unmapped medications	219 (5.74)	219 (19.96)
2: Off-site validation	764 (20.03)	201 (18.32)
3: In-home verification	1870 (49.02)	384 (35)
4: Retrospective cleaning	961 (25.2)	293 (26.71)
	3814	1097

Time Requirements

Website interaction logs provided more detail of the on-site verification of data. By aggregating continuous blocks of interactions, we estimate that the champion nurse required a mean of 33 min of website interaction to check uploaded data for one nursing home; taking into account the size of the homes, this corresponded to 22 seconds per resident (SD=14.8). For homes where an automated upload was not possible, manually entering residents and their medications through the supplied RedUSe website required a mean of 60 min for the first snapshot, 43 min for the second snapshot, and 35 min for the third (shown in Table 3). This equated to 42 seconds per resident at baseline (SD=26.4) and 30 seconds per resident, subsequently (SD=15.2).

Table 3. The number and mean min required to perform data validation, for uploaded versus manually entered data.

Data source	Snapshot	Mean minutes for data entry (Initial validation and on-site verification)	Number of snapshots validated
Manual			
	baseline	60.0	20
	3 months	43.2	13
	6 months	35.2	13
	all manual	48.0	46
Uploaded			
	baseline	38.1	130
	3 months	31.3	137
	6 months	31.3	137
	all uploaded	33.5	404

Discussion

These results demonstrate a successful implementation of an ETL method for retrieving and verifying sedative medication prescribing in nursing homes, on a national scale. This section explores the extent of that success and discusses considerations for future applications of the method.

Upload Coverage

The data extraction method was generally applicable across pharmacies. Apart from one that declined to participate, there were only four pharmacies (5% of the 86 pharmacies invited) that were unable to have medication data uploaded. These were due to two uncommon packing systems, both of which can be included with further development of the extraction software. Likewise, most residents (95.6%, 17,722/18,537) and most of their sedative medications (95.26%, 28,053/29,451) were included in uploads. The residents that were missed were due to individuals being supplied by an alternate pharmacy or due to new residents entering the home between the time the upload occurred and the time the data validation was performed. Similarly, medications could be missed due to this time difference but also due to differences in pharmacy packing protocols: for example, 6.12% (812/13,272) of PRN medications were not included in packing, requiring identification and manual addition during data validation. In addition, certain medication forms presented problems for extraction. For one packing system, injections, wafers, and solutions were always missed; for another, their inclusion was heavily dependent on the particular pharmacy's recording and packing practices, with approximately 1 in 5 pharmacy uploads not including those medications. These highlight the value of the manual validation steps; for example, if timeliness of data is an issue, or if 95% coverage of PRN medications is insufficient, then manual validation is a necessity.

Data Accuracy

Nearly 92% (25,762/28,053) of sedative medications were correct as uploaded. This rate improved to 98.32% (27,583/28,053) if the "known" errors in extraction, including the outdated AMT, were excluded. Given that those errors were corrected, it is tempting to expect that any future use of this

data extraction method would achieve that rate, but this may be unrealistic: the major "known" error was caused by an external update to pharmacy software, and the frequency and timing of those updates is unpredictable.

Other medication data errors were due to individualities in pharmacy packing systems. In the initial validation phase, 82.2% (180/219) of the unknown medication errors were due to using an outdated version of the AMT. The other 17.8% were caused by pharmacy-unique entries: packing systems allow pharmacists to add medications, for new products released before the system vendor can update their database or for pharmacy-prepared medications. Errors in recognized medications were caused by inconsistent entry into the packing systems, such as incomplete records, transposed fields, or typing mistakes. Retrospective corrections were primarily to merge multiple resident profiles: residents were often entered into pharmacy systems more than once, either as an oversight or intentionally to split medication packs. Other retrospective medication data changes were made for internal consistency (eg, correcting dosage quantities to match those given in the instructions).

That 25.2% (961/3814) of the corrections were performed retrospectively might seem to cast doubt on the accuracy of the previous validation steps. It is possible that some mistakes were made but unlikely, given the extensive validation process where the champion nurse, pharmacist, and research team all reviewed the data.

The volume of errors corrected highlights the need for data validation and verification. Adaptability of the extraction software is expected to be crucial for any longitudinal use of this approach as packing systems change between pharmacies and over time. Verifying uploaded data is a crucial step in identifying when those changes occur. Given the accuracy achieved within this project, if due consideration is given to validation, verification, and maintenance, it may be reasonable to aim for an accuracy rate above 95% (before corrections).

Comparison With Manual Entry

Compared with the automated process, the time recorded for manual data entry does not appear too unfavorable, especially when considering the additional 15 min required for the community pharmacist to facilitate an upload; however, this "manual" process was facilitated by a project-developed website with entry-assistance, so it does not fully describe the time benefit of the automated process. The prepopulation of data explains most of the increase in efficiency between the baseline and 3-month snapshots shown in Table 3; the remainder is due to familiarity with the website. It is also notable that although this application was only interested in sedative medications, the automated upload included all medications supplied by the pharmacy, and manual entry only included sedative medications. This translated to an average of 10 more medications per resident (11.84 per uploaded resident against 1.96 per manually added resident).

Applicability of Results

Although this study only validated sedative data, similar accuracy is expected for other medication classes, and the validation cost would be expected to increase proportionally. Maintenance costs should be approximately equivalent; systemic changes typically require the same amount of work regardless of the classes of medication. Managing the burden on nursing staff should be a key consideration, potentially by utilizing more nurses or reducing other project responsibilities.

Although this study included four Australian pharmacy packing systems, the approach is independent of particular systems. The model of community pharmacies supplying medication to nursing homes appears to be the standard in the United States, the United Kingdom, and elsewhere. The complexity of implementing this approach will be dependent on the range of software used by those pharmacies, which will differ by region. Of more significance is the relative uptake of standardized EHRs in the wider health system. If pharmacy software encodes medications to a recognized standard (such as SNOMED CT), then the data transformation cost and maintenance cost of the extraction software is drastically reduced. A standardized, appropriately encoded, widely adopted, and accessible patient EHR would make this data collection approach unnecessary. While many countries are attempting to move toward such a system, the authors are not aware of any that yet meet these criteria.

Conclusions

Extracting pharmacy medication packing records and enriching them through mapping to an independent medication data resource, such as the AMT and ATC codes, was used to collect sedative medication usage data in remote nursing homes on a national scale, with success in 94.6% (142/150) of the homes involved. Without data validation and verification, the accuracy rate for sedative medications could be expected to be above 90% and possibly over 95%. At least 20% of these errors would be expected to be unambiguously correctable with only basic validation. Although this unverified data may be unsuitable for individualized clinical applications, it is useful as a data source for more generalized analyses and avoiding the validation step makes it a much faster process.

With data validation and verification steps included, the method compares favorably to in-person visits and manual resident chart reproduction, taking less time and producing more data. It is also highly scalable. However, to produce reliable clinical data, continual monitoring of data accuracy is essential not only to correct mistakes in individual pharmacy data entry but also to identify variations in pharmacy packing systems. These variations also mean that there needs to be capacity for maintenance to the extraction software.

Acknowledgments

The RedUSe project was funded by the Australian Government under the Dementia and Aged Care Service Fund. We would also like to acknowledge FredHealth, Webstercare, Minfos, and MPS Australia for their support.

Conflicts of Interest

None declared.

References

1. Ballard C, Creese B, Corbett A, Aarsland D. Atypical antipsychotics for the treatment of behavioral and psychological symptoms in dementia, with a particular focus on longer term outcomes and mortality. Expert Opin Drug Saf 2011 Jan;10(1):35-43. [doi: 10.1517/14740338.2010.506711] [Medline: 20684745]
2. Bonner AF, Field TS, Lemay CA, Mazor KM, Andersen DA, Compher CJ, et al. Rationales that providers and family members cited for the use of antipsychotic medications in nursing home residents with dementia. J Am Geriatr Soc 2015 Feb;63(2):302-308. [doi: 10.1111/jgs.13230] [Medline: 25643635]
3. Chen Y, Briesacher BA, Field TS, Tjia J, Lau DT, Gurwitz JH. Unexplained variation across US nursing homes in antipsychotic prescribing rates. Arch Intern Med 2010 Jan 11;170(1):89-95 [FREE Full text] [doi: 10.1001/archinternmed.2009.469] [Medline: 20065204]
4. Thompson-Coon J, Abbott R, Rogers M, Whear R, Pearson S, Lang I, et al. Interventions to reduce inappropriate prescribing of antipsychotic medications in people with dementia resident in care homes: a systematic review. J Am Med Dir Assoc 2014 Oct;15(10):706-718 [FREE Full text] [doi: 10.1016/j.jamda.2014.06.012] [Medline: 25112229]
5. Levinson D. HHS. 2011. Medicare atypical antipsychotic drug claims for elderly nursing home residents URL: https://oig.hhs.gov/oei/reports/oei-07-08-00150.pdf [accessed 2017-07-24] [WebCite Cache ID 6sCAwNvFa]
6. Westbury J, Beld K, Jackson S, Peterson G. Review of psychotropic medication in Tasmanian residential aged care facilities. Australas J Ageing 2010 Jun;29(2):72-76. [doi: 10.1111/j.1741-6612.2010.00409.x] [Medline: 20553537]
7. Westbury J, Jackson S, Gee P, Peterson G. An effective approach to decrease antipsychotic and benzodiazepine use in nursing homes: the RedUSe project. Int Psychogeriatr 2010 Feb;22(1):26-36. [doi: 10.1017/S1041610209991128] [Medline: 19814843]
8. FRED. Fred IT Group URL: http://www.fred.com.au/ [accessed 2016-10-31] [WebCite Cache ID 6lf5zLcfx]
9. Webstercare. 2016. Webstercare MMS URL: http://www.webstercare.com.au/ [accessed 2016-10-31] [WebCite Cache ID 6lgEb3xEo]
10. Dahl LJ, Wright R, Xiao A, Keeven A, Carr DB. Quality improvement in long term care: the psychotropic assessment tool (PAT). J Am Med Dir Assoc 2008 Nov;9(9):676-683 [FREE Full text] [doi: 10.1016/j.jamda.2008.07.002] [Medline: 18992701]
11. Patterson SM, Hughes CM, Crealey G, Cardwell C, Lapane KL. An evaluation of an adapted U.S. model of pharmaceutical care to improve psychoactive prescribing for nursing home residents in northern ireland (fleetwood northern ireland study). J Am Geriatr Soc 2010 Jan;58(1):44-53. [doi: 10.1111/j.1532-5415.2009.02617.x] [Medline: 20002510]
12. Chakraborty A, Linton CR. Antipsychotic prescribing in dementia patients in care homes: proactive in-reach service improved quality of care. Int J Geriatr Psychiatry 2012 Oct;27(10):1097-1098. [doi: 10.1002/gps.2827] [Medline: 22945348]
13. Schultz B, Gambert S. Minimizing the use of psychoactive medications in the institutionalized elderly. Clin Gerontol 1991;11(1):80-82.
14. Hagen BF, Armstrong-Esther C, Quail P, Williams RJ, Norton P, Le NC, et al. Neuroleptic and benzodiazepine use in long-term care in urban and rural Alberta: characteristics and results of an education intervention to ensure appropriate use. Int Psychogeriatr 2005 Dec;17(4):631-652. [doi: 10.1017/S1041610205002188] [Medline: 16246262]
15. Vida S, Monette J, Wilchesky M, Monette M, Friedman R, Nguyen A, et al. A long-term care center interdisciplinary education program for antipsychotic use in dementia: program update five years later. Int Psychogeriatr 2012 Apr;24(4):599-605. [doi: 10.1017/S1041610211002225] [Medline: 22126992]
16. Xu J, Gao X, Sorwar G, Croll P. Implementation of e-health record systems in Australia. The International Technology Management Review 2013;3(2):92-104.
17. Wright D, Twigg M. Community pharmacy: an untapped patient data resource. Integrated Pharmacy Research and Practice 2016;5:19-25.
18. Hsu S, Benton R, Gottumukkala R. Real-time flu monitoring system and decision informatics. In: Proceedings of the 48th Annual Hawaii International Conference on Systems Sciences.: IEEE; 2015 Presented at: HICSS; January 5, 2015; Kauai p. 2794-2803.
19. Tompson AJ, Peterson GM, Jackson SL, Hughes JD, Raymond K. Utilizing community pharmacy dispensing records to disclose errors in hospital admission drug charts. Int J Clin Pharmacol Ther 2012 Sep;50(9):639-646. [doi: 10.5414/CP201720] [Medline: 22762857]
20. Curtain C, Peterson GM, Tenni P, Bindoff IK, Williams M. Outcomes of a decision support prompt in community pharmacy-dispensing software to promote step-down of proton pump inhibitor therapy. Br J Clin Pharmacol 2011 May;71(5):780-784 [FREE Full text] [doi: 10.1111/j.1365-2125.2010.03890.x] [Medline: 21480953]
21. Horsburgh S, Norris P, Becket G, Crampton P, Arroll B, Cumming J, et al. The equity in prescription medicines use study: using community pharmacy databases to study medicines utilisation. J Biomed Inform 2010 Dec;43(6):982-987 [FREE Full text] [doi: 10.1016/j.jbi.2010.08.004] [Medline: 20709187]

22. Gardner SF, Skelton DR, Rollins SD, Hastings JK. Community pharmacy data bases to identify patients at high risk for hypercholesterolemia. Pharmacotherapy 1995;15(3):292-296. [Medline: 7667164]
23. Wright D, Twigg M, Barton G, Thornley T, Kerr C. An evaluation of a multi-site community pharmacy-based chronic obstructive pulmonary disease support service. Int J Pharm Pract 2015 Feb;23(1):36-43. [doi: 10.1111/ijpp.12165] [Medline: 25409898]
24. Damschroder LJ, Banaszak-Holl J, Kowalski CP, Forman J, Saint S, Krein SL. The role of the champion in infection prevention: results from a multisite qualitative study. Qual Saf Health Care 2009 Dec;18(6):434-440. [doi: 10.1136/qshc.2009.034199] [Medline: 19955453]
25. Dunbar L, Brandt T, Wheeler A, Harrison J. Barriers and solutions to implementing metabolic risk assessment in a secondary mental health service. Australas Psychiatry 2010 Aug;18(4):322-325. [doi: 10.3109/10398561003692571] [Medline: 20645897]
26. Fakih MG, George C, Edson BS, Goeschel CA, Saint S. Implementing a national program to reduce catheter-associated urinary tract infection: a quality improvement collaboration of state hospital associations, academic medical centers, professional societies, and governmental agencies. Infect Control Hosp Epidemiol 2013 Oct;34(10):1048-1054. [doi: 10.1086/673149] [Medline: 24018921]
27. Pretorius A, Searle J, Marshall B. Barriers and enablers to emergency department nurses' management of patients' pain. Pain Manag Nurs 2015 Jun;16(3):372-379. [doi: 10.1016/j.pmn.2014.08.015] [Medline: 25440235]
28. Saint S, Kowalski CP, Forman J, Damschroder L, Hofer TP, Kaufman SR, et al. A multicenter qualitative study on preventing hospital-acquired urinary tract infection in US hospitals. Infect Control Hosp Epidemiol 2008 Apr;29(4):333-341. [doi: 10.1086/529589] [Medline: 18462146]
29. Lapane KL, Hughes CM. Considering the employee point of view: perceptions of job satisfaction and stress among nursing staff in nursing homes. J Am Med Dir Assoc 2007 Jan;8(1):8-13. [doi: 10.1016/j.jamda.2006.05.010] [Medline: 17210497]
30. Australian Digital Health Agency. ADHA. 2011. Australian medicines terminology URL: https://www.digitalhealth.gov.au/ [accessed 2017-07-24] [WebCite Cache ID 6sCCbYmYt]
31. Karimi S, Metke-Jimenez A, Nguyen A. Cademiner: a system for mining consumer reports on adverse drug side effects. In: Proceedings of the Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval.: ACM; 2015 Presented at: ESAIR 2015; October 23, 2015; Melbourne, Australia p. 47-50.
32. Leroux H, Lefort L. Using CDISC ODM and the RDF data cube for the semantic enrichment of longitudinal clinical trial data. In: Proceedings of the 5th International Workshop on Semantic Web Applications and Tools for Life Sciences.: CEUR Workshop Proceedings; 2012 Presented at: Semantic Web Applications and Tools for Life Sciences; November 28, 2012; Paris, France.
33. McBride SJ, Lawley MJ, Leroux H, Gibson S. Using Australian medicines terminology (AMT) and SNOMED CT-AU to better support clinical research. Stud Health Technol Inform 2012;178:144-149. [Medline: 22797033]
34. SNOMED. SNOMED clinical terms URL: http://www.snomed.org/snomed-ct [accessed 2017-02-09] [WebCite Cache ID 6o8f03JdU]
35. NHS Business Services Authority. NHSBSA. Dictionary of medicines and devices URL: http://www.nhsbsa.nhs.uk/1121.aspx [accessed 2017-02-09] [WebCite Cache ID 6o8eqrAmE]
36. SNOMED. SNOMED. SNOMED Singapore 2015/2016 update URL: http://www.snomed.org/member/singapore/20152016-update [accessed 2017-02-09] [WebCite Cache ID 6o8ebMlxw]
37. NIH. 2014. RxNorm URL: https://www.nlm.nih.gov/research/umls/rxnorm/ [accessed 2017-06-29] [WebCite Cache ID 6qudnkg4S]
38. Mps-aust. 2016. MPS Australia better medication management URL: http://mps-aust.com.au/ [accessed 2016-11-01] [WebCite Cache ID 6lgEen0Ie]
39. Minfos. 2016. Minfos all-in-one management software for the Australian pharmacy URL: http://www.minfos.com.au/ [accessed 2016-10-31] [WebCite Cache ID 6lgEiIW9W]
40. WHOCC. 2015 Dec 16. WHO collaborating centre for drug statistics methodology URL: http://www.whocc.no/atc_ddd_index/ [accessed 2016-11-01] [WebCite Cache ID 6lgElKh9v]

Abbreviations

ADHA: Australian Digital Health Agency
AMT: Australian Medicines Terminology
ATC: Anatomical Therapeutic Chemical
BPSD: Behavioral and Psychological Symptoms of Dementia
CTPP: Containered Trade Product Pack
DDD: defined daily dose
dm+d: Dictionary of Medicines and Devices
ETL: extract, transform, and load
MPUU: Medicinal Product Unit of Use

PRN: pro re nata (“as needed”)
RedUSe: Reducing the Use of Sedatives project
SDD: Singapore Drug Dictionary
SNOMED CT: Systemized Nomenclature of Medicine Clinical Terms
TLS: transport layer security
TPUU: Trade Product Unit of Use

Edited by G Eysenbach; submitted 02.11.16; peer-reviewed by L Emmerton, S Speedie, T Cornford; comments to author 01.12.16; revised version received 07.04.17; accepted 06.06.17; published 03.08.17.

Please cite as:
Ling T, Gee P, Westbury J, Bindoff I, Peterson G
An Internet-Based Method for Extracting Nursing Home Resident Sedative Medication Data From Pharmacy Packing Systems: Descriptive Evaluation
J Med Internet Res 2017;19(8):e283
URL: http://www.jmir.org/2017/8/e283/
doi:10.2196/jmir.6938
PMID:28778844

Original Paper

Development and Deployment of the OpenMRS-Ebola Electronic Health Record System for an Ebola Treatment Center in Sierra Leone

Shefali Oza[1,2], MSc; Darius Jazayeri[3,4], MEng; Jonathan M Teich[3,5], MD, PhD; Ellen Ball[3,6], MS; Patricia Alexandra Nankubuge[4], MSc; Job Rwebembera[4], BS; Kevin Wing[1,2], PhD; Alieu Amara Sesay[2], BSc; Andrew S Kanter[3,7], MPH, MD; Glauber D Ramos[4], BS; David Walton[4,5], MPH, MD; Rachael Cummings[2,8], MSc; Francesco Checchi[1,2,8], PhD; Hamish S Fraser[3,9], MBChB, MSc

[1]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
[2]Save the Children International, Kerry Town, Sierra Leone
[3]OpenMRS Inc, Indianapolis, IN, United States
[4]ThoughtWorks, Kampala, Uganda
[5]Brigham and Women's Hospital, Boston, MA, United States
[6]Partners In Health, Boston, MA, United States
[7]Columbia University, New York City, NY, United States
[8]Save the Children UK, London, United Kingdom
[9]University of Leeds, Leeds, United Kingdom

Corresponding Author:
Shefali Oza, MSc
Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 020 7636 863
Email: shefali@alum.mit.edu

Abstract

Background: Stringent infection control requirements at Ebola treatment centers (ETCs), which are specialized facilities for isolating and treating Ebola patients, create substantial challenges for recording and reviewing patient information. During the 2014-2016 West African Ebola epidemic, paper-based data collection systems at ETCs compromised the quality, quantity, and confidentiality of patient data. Electronic health record (EHR) systems have the potential to address such problems, with benefits for patient care, surveillance, and research. However, no suitable software was available for deployment when large-scale ETCs opened as the epidemic escalated in 2014.

Objective: We present our work on rapidly developing and deploying OpenMRS-Ebola, an EHR system for the Kerry Town ETC in Sierra Leone. We describe our experience, lessons learned, and recommendations for future health emergencies.

Methods: We used the OpenMRS platform and Agile software development approaches to build OpenMRS-Ebola. Key features of our work included daily communications between the development team and ground-based operations team, iterative processes, and phased development and implementation. We made design decisions based on the restrictions of the ETC environment and regular user feedback. To evaluate the system, we conducted predeployment user questionnaires and compared the EHR records with duplicate paper records.

Results: We successfully built OpenMRS-Ebola, a modular stand-alone EHR system with a tablet-based application for infectious patient wards and a desktop-based application for noninfectious areas. OpenMRS-Ebola supports patient tracking (registration, bed allocation, and discharge); recording of vital signs and symptoms; medication and intravenous fluid ordering and monitoring; laboratory results; clinician notes; and data export. It displays relevant patient information to clinicians in infectious and noninfectious zones. We implemented phase 1 (patient tracking; drug ordering and monitoring) after 2.5 months of full-time development. OpenMRS-Ebola was used for 112 patient registrations, 569 prescription orders, and 971 medication administration

recordings. We were unable to fully implement phases 2 and 3 as the ETC closed because of a decrease in new Ebola cases. The phase 1 evaluation suggested that OpenMRS-Ebola worked well in the context of the rollout, and the user feedback was positive.

Conclusions: To our knowledge, OpenMRS-Ebola is the most comprehensive adaptable clinical EHR built for a low-resource setting health emergency. It is designed to address the main challenges of data collection in highly infectious environments that require robust infection prevention and control measures and it is interoperable with other electronic health systems. Although we built and deployed OpenMRS-Ebola more rapidly than typical software, our work highlights the challenges of having to develop an appropriate system during an emergency rather than being able to rapidly adapt an existing one. Lessons learned from this and previous emergencies should be used to ensure that a set of well-designed, easy-to-use, pretested health software is ready for quick deployment in future.

(J Med Internet Res 2017;19(8):e294) doi:10.2196/jmir.7881

KEYWORDS

Ebola virus disease; electronic health records; eHealth; health information systems; disease outbreaks; disasters; West Africa; Sierra Leone

Introduction

Background

The 2014-2016 West African Ebola epidemic, with more than 28,000 infected and 11,000 deaths, overwhelmed health systems in three of the world's most impoverished countries [1]. Existing health facilities in the affected areas did not have the necessary capacity, staff, and infection control capabilities to adequately cope with this outbreak [2]. Untreated Ebola patients not only have a high mortality rate but also remain a serious infection risk to their communities [3]. In this context, large-scale Ebola treatment centers (ETCs) emerged as emergency health facilities that could be set up quickly to isolate and treat seriously ill patients while providing the rigorous infection control needed to protect staff from the Ebola virus.

ETCs are specialized facilities that must provide efficient care for suspected and confirmed Ebola patients while minimizing the risk of infection to staff and other patients. This presents several challenges [4], including overheating, impaired visibility, and poor dexterity caused by working in highly restrictive personal protective equipment (PPE); limited time for direct patient contact; and the inability to move material—including paper medical records—from highly infectious patient areas (red zone) to low-risk nonpatient areas (green zone).

Data Challenges in Ebola Treatment Centers

A key component of ETC patient management is the collection, transmission, use, and analysis of clinical data. Although the simplicity of paper records makes them a practical and appealing option for health information recording during an emergency, they can also be inflexible and difficult to share. This is particularly true when clinical data are used by multiple teams or by public health authorities and researchers in different locations, as well as where physical and contagion boundaries restrict the use of paper. In the case of Ebola, highly restrictive infection controls in ETCs created problems across a range of standard hospital operations when using paper records such as recording sufficient information legibly on patient charts while wearing PPE, accessing bedside charts, and communicating patient information between the zones.

One of the main data challenges during this outbreak when using paper-based data collection systems involved extracting patient data from the red zone. Methods for doing this imposed severe compromises on the quality, quantity, and confidentiality of patient data that were collected and transferred. Communication methods used by organizations have included shouting or radioing of information, such as prescriptions, to the green zone; photographing red zone whiteboards with identifiable patient information; and time-consuming scanning of often illegible or damaged records over a local wireless network [5,6]. Overall, conventional paper-based data collection and review is difficult in large-scale ETCs and limits the ability to gather urgently needed data for care, surveillance, and research. Due to this, it was recognized that electronic health record (EHR) systems could potentially offer substantial benefits over conventional paper-based records in an ETC. Yet EHRs, which are standard in many health care settings, are still rarely used effectively in health emergencies [7]. This is especially true for emergencies in low-resource settings.

Electronic Patient Records for Ebola

As ETCs began opening in West Africa in 2014, there was no EHR software that could be rapidly deployed in these settings. Most EHR systems are designed for the needs of high-income health facilities that have reliable power, network infrastructure, and technical support. EHR systems for low-resource settings are increasingly common but to date have had limited use in hospital settings, especially for intensive care [8,9]. A district and national health data management system, Health Information System Programme's district health information software (DHIS) 2, was available in West Africa at the time of the Ebola outbreak but is designed mainly for aggregate patient data instead of the individual-level records needed for patient care in an ETC [10]. Other systems, including a number of commercial EHR systems used in Africa, lacked the flexibility and rapid adaptability needed for deployment during this outbreak. Additionally, ETCs have substantially different workflows and information requirements from other acute and intensive care environments. Finally, an EHR must be easy to learn, as well as quick and simple to use in the uncomfortable, time-limited, dexterity- and vision-limited, strictly infection-controlled environment of the ETC.

OpenMRS is an open-source, modular EHR platform [11] for building patient medical record applications. First deployed in Kenya, Rwanda, and South Africa in 2006 [12] to support care of patients with human immunodeficiency virus (HIV) and tuberculosis (TB), it is now used to manage primary care and a range of diseases in more than 60 low- and middle-income countries (LMICs) [13,14]. It did not have a suitable user-friendly tablet interface needed for red zone patient data collection; however, the overall adaptability of OpenMRS, alongside an active community of programmers and users, made it a promising choice as an EHR platform for ETCs.

Study Aims

In this paper, we describe our work on rapidly developing OpenMRS-Ebola, an open-source Ebola EHR system that was implemented in 2015 at Save the Children's Kerry Town ETC in Sierra Leone. We describe our experiences, lessons learned, and recommendations for design and implementation of EHRs in future health emergencies.

Methods

Setting

The 80-bed Kerry Town ETC, based in the Western Area Rural district of Sierra Leone, operated from November 5, 2014 to March 31, 2015. Save the Children International (SCI) ran the site in collaboration with the Sierra Leone and United Kingdom governments. The ETC had wards for suspected and confirmed Ebola patients in the red zone and a range of operational rooms in the green zone, including a pharmacy, clinician station, and offices. The site employed over 100 clinicians, mostly from Sierra Leone and Cuba, with some non-Cuban international staff on rotation. Most patients came from nearby districts. The Kerry Town ETC had reliable power from 2 large generators. For connectivity, the site had a wireless local area network (WLAN) and a Ku-band very small aperture terminal (VSAT) system for Internet. The site also had backup power supply (uninterruptible power supply [UPS]) devices for the server, network routers, and other computing hardware.

EHR Platform

OpenMRS is a highly flexible and configurable EHR platform [11]. It has a core system for managing log-ins, user accounts with security and privileges, data storage, and data retrieval. Plugin software modules extend the basic core system with custom functionality. This provides flexibility for different environments, disease types, and clinical workflows. The OpenMRS data model is built around a concept dictionary that structures and codes nearly all patient data and is mapped to standard medical terminology such as systematized nomenclature of medicine -- clinical terms (SNOMED CT) and International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10; [12]). OpenMRS is also compliant with health-level 7 (HL7) version 2.x and fast healthcare interoperability resources (FHIR) standards for data exchange [12,15]. It is interoperable with other electronic health (eHealth) systems such as DHIS 2, HL7-compliant laboratory information management systems, and several mobile health (mHealth) platforms, including ODK, CommCare, and Sana [16]. Recently, two new user interface (UI) and application tool kits have been developed to better support point-of-care use by clinicians and other staff [17]. Previous OpenMRS-related work in Malawi [18] and Rwanda [19] used touch screen computers. However, these were not designed for modern tablet devices or for the latest versions of the OpenMRS platform and would have required extensive adaptation to meet our red zone design needs.

We chose to use the latest available version of OpenMRS (v1.10.3) because it already had most of the functionality needed for an ETC EHR, it was open source and known to function well in challenging environments, and it had a large global community of developers and implementers.

Strategy for Developing OpenMRS-Ebola

Our strategy for building the OpenMRS-Ebola EHR had four main components: (1) using Agile software methodology; (2) recruiting team members with diverse skills and experience; (3) iterative design based on usability, speed, and clinical needs; and (4) regular communication and feedback between the operations and development teams.

First, we used Agile software development approaches, which emphasize verbal communication, delivering working software early, and responding to changing requirements, to build the system. We designed our own Agile approach because none of the existing concrete agile frameworks such as Scrum matched the needs of this project. Specifically, the operations team lead (product owner) was deeply engaged with the development team, and we reprioritized and revised requirements daily; we did not have traditional sprints, and we performed Agile ceremonies such as retrospectives and showcases at the end of each release. We developed user stories, which we documented and tracked using a JIRA issue tracking system [20]. Some user stories were modified based on direct voice or email communications.

Two main teams worked on developing and deploying OpenMRS-Ebola. The operations team, stationed primarily at the ETC, comprised health information and clinical SCI personnel from the United States, the United Kingdom, and Sierra Leone. This team's main role was to initiate the project, to help guide development based on ground-level needs, and to deploy OpenMRS-Ebola. The operations team lead acted as the product owner for this project. The development team comprised employees from ThoughtWorks (a global technology consultancy and software engineering firm), as well as OpenMRS volunteers and leadership members. The ThoughtWorks team, initially colocated in Uganda, expanded to include 10 staff members around the world. The development team had a range of competencies on developing EHR systems, from programming and project management to medical informatics and UI design.

A core component of this work involved constant communication within and among these teams. We started full-time software development with a 3-day project inception in Uganda with the development team and the lead of the operations team (who dialed in from Sierra Leone). During the project, the operations team lead and the entire development

team had extensive email discussions and daily calls to share brief status updates, which also functioned as daily meetings for the development team. These were followed by a longer daily conversation between the operations team lead and key members of the development team (business analysts, tech lead, and medical informaticist) to (1) showcase work in progress for feedback, (2) reprioritize the next day's work, and (3) agree on feature definitions that would satisfy user requirements with the least development time. The operations team regularly shared existing assessment forms, clinical workflow patterns, drug formulary lists, and other critical information with the development team to inform the design of new functionalities and UIs.

Technical Process of Developing OpenMRS-Ebola

We initiated this project with requirements gathering on the ground by the operations lead and a volunteer effort for development that included analyzing terminology in the paper forms, adding necessary terminology to a shared concept dictionary, and building a proof-of-concept of the inpatient data collection form. Additionally, volunteers organized a hackathon in Brazil in early November that accelerated design work and developed a prototype tablet UI for use in the red zone [21].

During the project inception with ThoughtWorks, we developed a common understanding of the minimum viable product (MVP), intended end product, and foreseeable challenges. The MVP is the smallest work product that would provide value to end users. We focused on deploying the MVP as quickly as possible before moving on to develop other features. We regularly updated the list of features for the MVP and other phases based on changing needs on the ground. Ultimately, OpenMRS-Ebola had three development phases.

We followed a continuous deployment approach. We configured a continuous integration (CI) server to build and test all our code as soon as it was merged, and we used feature toggles to allow partially completed work to be integrated into the master codebase and tested by CI. We automatically deployed all successful builds to an Internet-hosted server for quality assurance and user acceptance testing. We released software updates over the Internet to training and production servers at the ETC as appropriate and as often as daily.

We built OpenMRS-Ebola on top of the OpenMRS Reference Application (version 2.1), an extensible UI with preconfigured functionality atop the OpenMRS platform [22]. We used Java and Groovy to build screens for desktop usage and AngularJS to build screens for tablet usage. We used an international, multilingual, clinical interface terminology developed for low-resource settings produced by the Columbia International eHealth Laboratory (CIEL; [23]) as the concept dictionary behind our EHR. We requested new terminology on a regular basis and incorporated new CIEL releases as often as daily during early development stages. All OpenMRS-Ebola code we wrote is free and fully open source under Mozilla Public License version 2.0 [24].

To increase development speed, we built the software as a browser-based Web application requiring constant WLAN access to the server. The operations team confirmed the suitability of this approach because of reliable power (through generators and UPS devices) and strong wireless signal (tested by walking through patients wards before the ETC opened).

Initial usability testing by the development team was carried out wearing dishwashing gloves and tinted goggles to simulate the ergonomic challenges of using a tablet in the ETC red zone.

Design Decisions

Key design decisions focused on addressing usability, workflow, and communication problems in ETCs. In the green zone, the UI was designed to present detailed information because clinicians can review patient records in the green zone without wearing PPE. In the red zone, the UI was optimized for readability, speed, and ease of use by users wearing PPE. We designed the tablet UI for portrait mode because users found it easier to hold the tablet vertical when using one hand. The tablet functions were based on the initial workflows and forms provided by the clinical team. These functions were optimized for usability and speed so that ordering and administration of medication and intravenous (IV) fluids, as well as entry of patient status, vital signs, and symptoms, could be completed during the short bedside rounds. This led to several design decisions, including high-contrast color schemes, large buttons and text fonts, limiting the amount of information on each screen, building layouts for maximum clarity and rapid entry, limiting the use of complex numeric and text entry, and shortcuts for rapid entry of common items. At various points during development, we obtained feedback from users regarding readability and ease of use and adjusted our designs accordingly.

Hardware Selection

We chose the Sony Xperia Z2 10.1-inch tablet for our EHR because it was waterproof and could be disinfected with chlorine. We charged the tablet using the Sony magnetic charging dock DK39. Clinicians who preferred a stylus instead of their gloved finger used the Boxwave EverTouch Capacitive Stylus with fiber mesh tip that plugged into the tablet headphone jack.

Implementation

OpenMRS-Ebola was deployed in phases at the ETC, with the most essential functions deployed in the MVP (phase 1). Key implementation components involved field testing at the ETC before deployment, training sessions for staff on how to use the hardware and OpenMRS-Ebola software, installing tablets in the red zone patient wards and laptops in the green zone areas, installing OpenMRS-Ebola onto the ETC server, and obtaining user feedback to improve the software.

Evaluation

In February 2015, we asked clinicians to complete a predeployment structured questionnaire about their experiences with the existing paper-based records system and thoughts about an EHR. However, a postdeployment questionnaire was not conducted because of staff departures after the ETC closed. We also obtained informal feedback from users during development, field testing, and deployment.

Patient data were entered in both OpenMRS-Ebola and the existing paper-based record system for more than a month after

the rollout of phase 1 while we made changes based on user feedback. We compared the patient registration and drug-ordering data entered for each patient in both systems to identify differences and to evaluate how the EHR was functioning compared with the standard paper-based system. Although phase 2 was deployed in mid-March, we were unable to do a similar comparison because it was not fully used due to the ETC closing.

Results

Software Product

The final OpenMRS-Ebola EHR includes a browser-based desktop or laptop application for the green zone and a browser-based tablet application for the red zone, each with different UIs but accessing the same data and software infrastructure. Both applications have separate but overlapping modules and functions. The desktop application comprises six key modules for managing patient tracking, entry of some clinical information, and viewing detailed patient summaries (Table 1). The tablet application has five key modules, primarily focused on the following common red zone tasks: drug and IV fluid ordering and administration, entry of patient vital signs and symptoms, and viewing limited patient summaries (Table 1).

A detailed example of a module, tablet-based drug ordering and monitoring, is shown in Figure 1. Additional desktop-based functions for this module include complete drug charting with medication administration time stamps and a list of recently edited prescriptions for pharmacy review (Figure 2). A similar module was designed for IV fluid ordering and monitoring (see Multimedia Appendices 1 and 2).

Table 1. Modules and functionalities of the OpenMRS-Ebola electronic health record (EHR) desktop and tablet applications.

Modules and functions	Description	Application type[a]	Rollout phase[b]
Patient tracking			
Registration	Date, name, demographics, contact information, ID # allocation, quick assessment[c]	Desktop	1
Bed allocation	Selection of ward and bed #	Desktop	1
Discharge	Date and patient outcome	Desktop	1
Clinical input			
Medications	Ordering, medication administration, and discontinuation	Tablet	1
IV[d] fluids	Ordering and administration (with start, hold, restart, and stop functions)	Tablet	2
Patient vital signs	Key vitals, including temperature, pulse, blood pressure, and consciousness level	Tablet	2
Symptoms	Key symptoms, patient status, and observations	Tablet	2
Laboratory tests	Ebola and malaria results by date	Desktop	3
Clinician notes	Time-stamped free text note entry	Desktop	3
Clinical output			
Detailed patient summaries	Full patient details: patient demographics and bed location, vitals, symptoms, medications, full medication administration chart, IV fluids, labs, and clinician notes	Desktop	2
Abridged patient summaries	Patient demographics and bed location, recent vitals and symptoms, active prescriptions, and IV fluids (expandable to full history)	Tablet	2
Additional functionalities			
Active patients	List of active patients by ward with bed #s	Desktop	1
Data editing	Ability to retrospectively edit data as needed	Desktop	1, 2
Data export	Export data from modules to CSV[e] files	Desktop	3

[a]Functionality designed for the tablet application is responsive to different screen layouts and can also be used on the desktop or laptop.

[b]Rollout phases: Phase 1 (deployed in mid-February), phase 2 (deployed in mid-March), and phase 3 (development completed in late March but not deployed because of Ebola treatment center [ETC] closing).

[c]Type of patient (confirmed or suspect), stage of illness.

[d]IV: intravenous.

[e]CSV: comma separated values.

Figure 1. Medication ordering in the tablet application, designed for rapid entry while wearing personal protective equipment (PPE). The patient summary page (a) shows active prescriptions, including buttons to record medication administration and stoppage. Selecting "order medication" from the Actions menu brings up the drug choice page (b), including rapid selection of the 20 most common drugs (accounting for over 90% of orders) and an update-as-you-type search control to access other drugs by name. The next page (c) offers available form and strength options of the selected drug. Dosing instructions are entered (d) based on the Ebola treatment center (ETC) workflow, which had standard rounds per day.

The patient symptom assessment presented particular challenges to make it fast, usable, and meaningful on the tablet. We discussed several tablet application designs with the operations team before reaching a satisfactory single-page button-only form (Figure 3).

An example of the desktop patient summary is shown in Figure 4. This screen includes recent information for all recorded clinical data and options to view the full clinical history during the ETC stay.

Screenshots of the full desktop and tablet Ebola EHR applications, with all modules, are included in Multimedia Appendices 1 and 2.

Figure 2. Desktop-based prescription and medication administration charts.

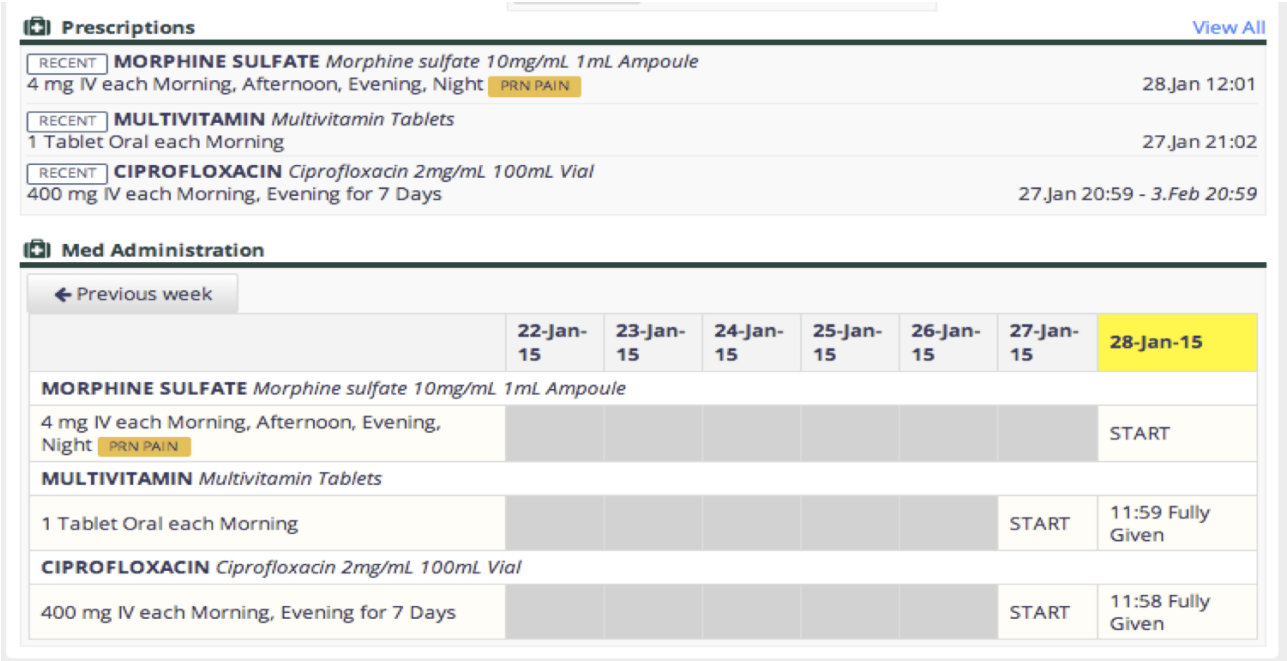


Figure 3. Symptom assessment module in the tablet application.

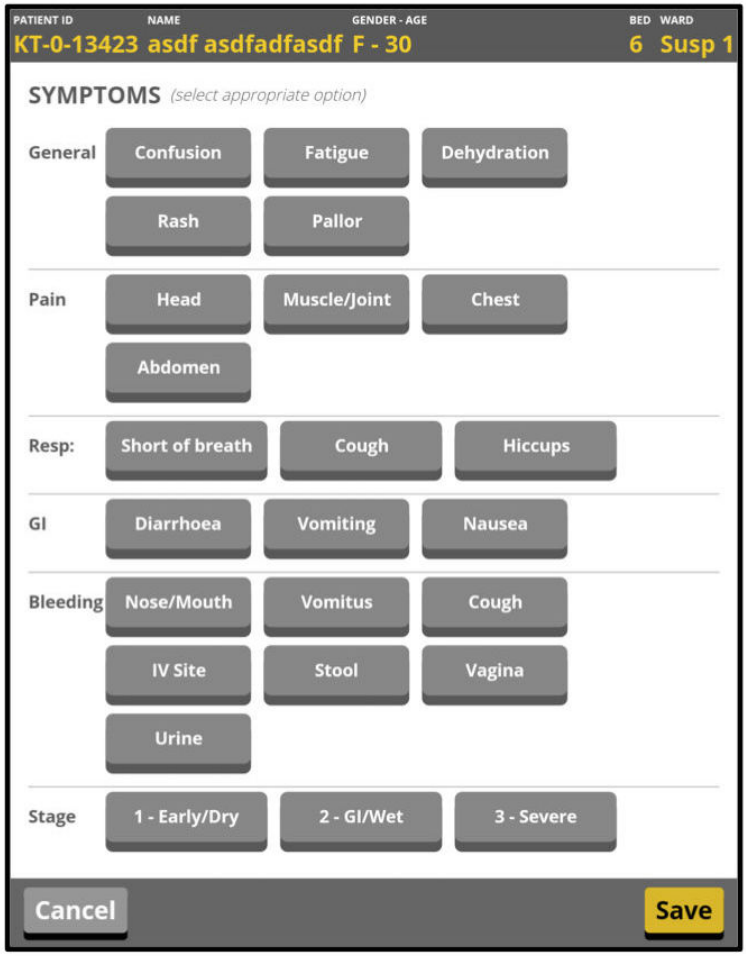
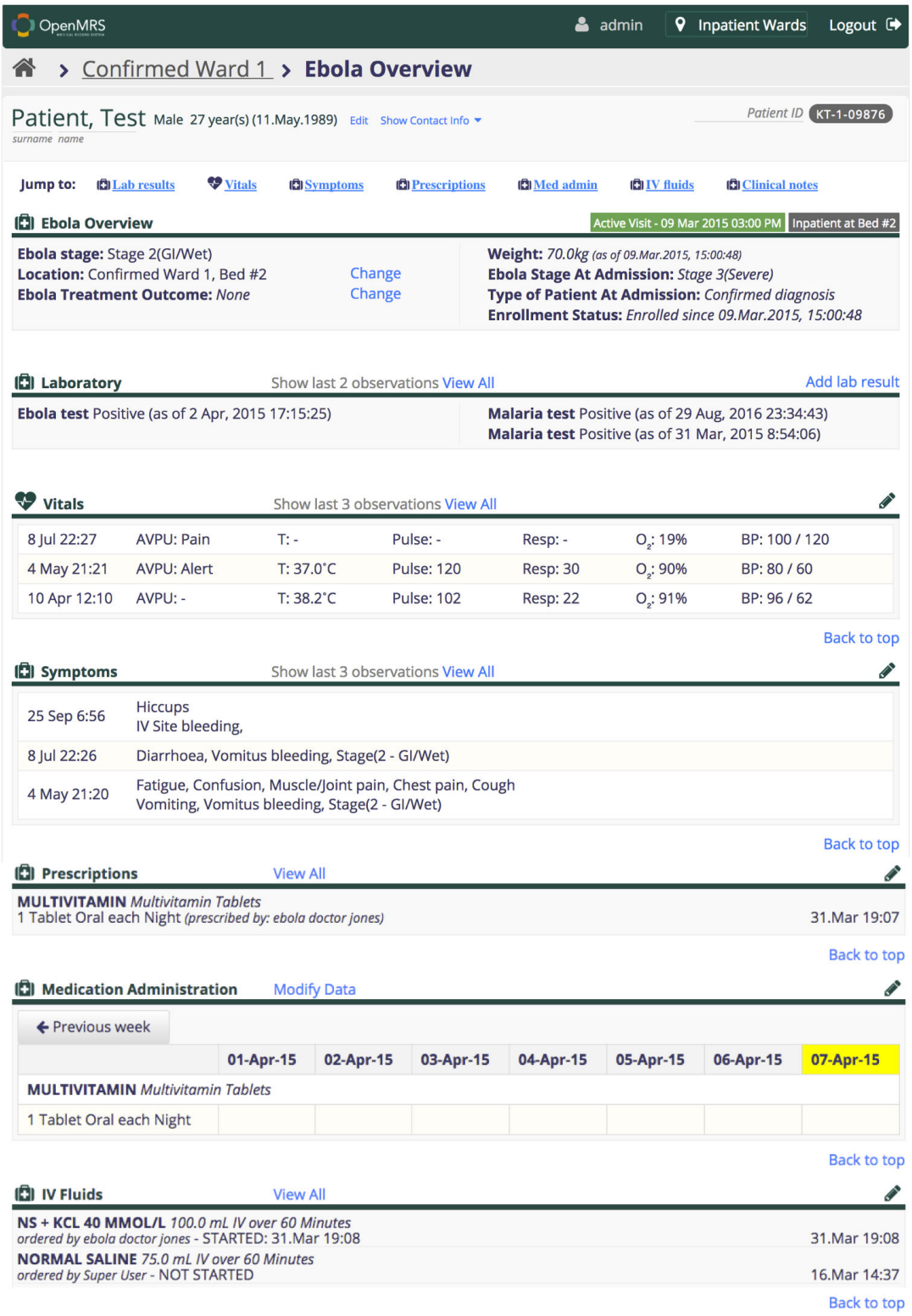


Figure 4. Full desktop summary.



Cost

SCI paid approximately US $50,000 for OpenMRS-Ebola, with about US $38,000 to ThoughtWorks for software development and US $12,000 for hardware. However, the true cost of developing and deploying OpenMRS-Ebola was approximately US $260,000, the majority of which was donated by ThoughtWorks through staff time. The full costs (excluding proof-of-concept and predevelopment volunteer work) were estimated to be US $187,000 for ThoughtWorks staff (based on reported hours and social impact project rates); US $6,000 in ThoughtWorks equipment and travel; US $12,000 by SCI for tablets, charging docks, and styluses; US $50,000 in donated time by 3 key volunteers; and about US $5000 for dual-use hardware already at the ETC (ie, server, 2 laptops, wireless routers, and UPS devices). The cost of redeploying the system

with adaptions in a new emergency would be considerably less with reuse of the workflow, UI designs, and code.

System Deployment and Usage

Volunteer programmers began work on OpenMRS-Ebola in November 2014. ThoughtWorks began full-time development in early December 2014. Phase 1 was completed and deployed in mid-February 2015, with user testing and feedback starting in January while the modules were being finalized. Phase 2 was completed and released in mid-March, and phase 3 development was completed in late March. Phase 3 was not deployed because the Kerry Town ETC closed on short notice at the end of March 2015 after a dramatic decrease in new Ebola cases. Therefore, the phase 2 and 3 modules saw little to no use on actual ETC patients.

From January to March 2015, nearly 100 clinicians were trained to use the system. In total, 112 of 456 Kerry Town ETC patients were registered in OpenMRS-Ebola. All but 2 of those not entered into OpenMRS-Ebola had completed their ETC stay before the system was deployed. The 2 remaining patients were discharged within hours of being admitted. For drug ordering, 569 prescription orders were placed using the system, and 971 medication administrations were recorded.

Evaluation

Sixteen clinicians completed the predeployment questionnaire, including staff from Sierra Leone (n=5), Cuba (n=6), and other countries (n=5). They included the full range of clinical staff, from community health assistants and officers to nurses and doctors. The questionnaires were completed in February 2015, 3 months after the ETC opened. When asked whether they agreed or disagreed that an electronic patient record system could improve patient records at the ETC, 7 strongly agreed, 8 agreed, and 1 was neutral. Some reasons they stated for wanting an EHR over the paper system included likelihood of higher quality data collection in the red zone, not having to rely on memory for transferring data to the green zone, real-time data collection and review, records not being damaged or missing, and legibility. Their concerns about a potential EHR included adoption by users without previous tablet or EHR experience, difficulty if training is inadequate, power outages, equipment breaking, having parallel systems if the EHR could not fully replace the paper system, and cost.

Although we were unable to complete postdeployment user questionnaires, we obtained informal feedback about the EHR during the phase 1 and 2 field testing and deployments. Positive comments included that OpenMRS-Ebola was easy, fast, and intuitive to use; the UI worked well with red zone visibility or dexterity problems; the EHR seemed useful for both red and green zone entry and review; and that the EHR would be favored over the paper system if there was more training and if network connectivity was good. Concerns included needing more training for users unfamiliar with such systems, potential delays in fixing bugs, and having to use both paper and electronic records during the phased rollout.

When comparing the paper and EHR records for the 112 patients registered in OpenMRS-Ebola, ID numbers, age, and sex were incorrectly recorded in the EHR 4, 2, and 2 times, respectively. In comparison, a basic pre-EHR database created for use with the paper records initially had 7, 5, and 3 errors when recording ID numbers, age, and sex, respectively, for these same patients. For the prescription entry, 97.2% (553/569) of the prescriptions in the EHR system correctly matched the paper records. Of the remaining 16 EHR prescriptions, 4 were missing from the paper drug charts but recorded elsewhere in the paper notes, 6 were missing from the paper charts altogether, and 6 were for a patient with missing paper records. A total of 77 prescriptions were recorded in the paper system but not in the EHR. Of these, 67 had specific identifiable reasons: 30 were prescriptions given as needed (pro re nata or PRN) without pharmacy prescription, 24 were illegible or ambiguous, and 13 were because of drugs not listed in the EHR during the first weeks of deployment. The latter were reported as missing and subsequently fixed in the EHR system.

The time taken to perform this evaluation was also informative. The majority of time spent on analysis for this evaluation was because of the time-consuming nature of working with paper records. The code to analyze the EHR data took minutes to write but going through the paper records took days. Additionally, legibility was a major issue for some of the paper records, and missing or damaged pages were a less common but still important problem.

Discussion

Principal Findings

We rapidly developed an open-source Ebola EHR system that was deployed at Save the Children's Kerry Town ETC during the West African Ebola outbreak. OpenMRS-Ebola was designed specifically to address the many challenges in recording patient data within ETCs that arise from severe infection control measures. This EHR supports registration, bed allocation, and discharge of patients; recording of vital signs and symptoms; medication and IV fluid ordering, administration, and monitoring; laboratory results; clinician notes; and data export for analysis. It displays relevant patient information to clinicians in both the infectious and noninfectious zones. To our knowledge, this system is the most comprehensive clinical EHR built for Ebola and is able to function as a stand-alone medical record system in an ETC.

The evaluation suggested that OpenMRS-Ebola worked well in the context of our rollout. There were two main sources of error, with both being ones that would be expected during a rollout. First, some errors were discovered during implementation, such as an initially incomplete drug list that resulted in a few missing prescriptions. Second, during the initial implementation of the EHR, clinicians used it in parallel with the existing paper system. This meant that it was less than fully integrated into the clinical workflow, and some users may have taken the system less seriously than if the parallel paper system was not required. One example of this is the 30 missing PRN drugs (given in the red zone without a pharmacy prescription), for which the medication administration recording workflow differed between the paper and EHR systems. The OpenMRS-Ebola records had fewer age or sex or ID errors than the pre-EHR (ie, single-entered) database for the paper records.

Although those initial database errors were fixed through careful checking of the paper records, this is a time-consuming task. One explanation could be that clinical staff familiar with the patients entered data into the EHR, whereas the database for paper records was completed by data entry clerks who were retrospectively copying from handwritten, possibly illegible, paper charts. This analysis also demonstrates the value of the EHR in terms of how much quicker it is to enter, analyze, report, and evaluate data compared with a paper-based system.

Strengths and Weaknesses of Approach

Although we successfully built the full OpenMRS-Ebola EHR system and implemented its first phase, this system's usefulness was limited because it was deployed around the time when the number of new Ebola cases began to decline in Sierra Leone. The Kerry Town ETC closed soon afterward, just as we were deploying phase 2. This is a major drawback of developing a system during the outbreak. At the time we began building the system, the number of Ebola cases was increasing exponentially, and many believed that the outbreak would last a long time [25]. We therefore had to plan for a wide range of options, including the likely possibility that the Kerry Town ETC would be open for many months. We also hoped to share our EHR with other ETCs because they were all struggling with the same data communication issues between red and green zones and had similar workflows, triaging, and patient care. We believe that OpenMRS-Ebola would have shown greater value as the epidemic continued and may have had wider adoption. Fortunately, the epidemic forecasts when we were developing OpenMRS-Ebola in autumn 2014 were more pessimistic [25,26] than the reality [1], and the epidemic wound down more quickly than most expected. The design experience, added functionalities, and lessons learned from this project should be applicable to future health emergencies, making earlier deployment of an EHR plausible.

We tried to minimize the inherent problem of software development during a health emergency, namely that software requires planning but rapidly evolving situations mean dynamic needs. We first gathered information on the ground to determine how best an Ebola EHR could be designed to address the challenges encountered with the paper-based records system. We started developing OpenMRS-Ebola as the Kerry Town ETC opened, and we recognized that our ideas and designs needed to continuously evolve with changing needs and processes on the ground. To do this efficiently, we selected highly flexible methods, including a modular software platform, Agile software methodology, daily communication between teams, frequent feedback from users on site, regular reevaluation of priorities, and a phased implementation.

We benefited from working with a diverse group, from clinicians and epidemiologists to programmers and UI designers. Our team was located across six continents, which could have produced collaboration problems but instead was beneficial both for perspective and the ability to effectively work around the clock. Several individuals, including the project leads on the operations and development teams, remained on the project from start to finish. This was essential for developing a usable product during rapidly changing conditions but also unusual because of the typical high staff turnover of field staff (eg, operations lead) during such responses.

A key challenge for this work was the dynamic situation on the ground. The clinical workflow at the ETC shifted based on changing clinical protocols, unfamiliarity with treating large numbers of Ebola patients, and frequent turnover of the clinical director and short-term medical teams. For example, drug ordering and monitoring were done at the patient's bedside in the red zone when the ETC opened. Correctly communicating these orders to the green zone was very difficult, especially given the large distances between the red and green zones at this ETC. Since this could have serious safety implications for patients, drug ordering and monitoring were deemed to be a top priority for the EHR and were included in the MVP. However, about 2 months later, the clinical staff changed their workflow to order drugs from the green zone instead, which was nonideal but safer for patients given the lack of instant and accurate communication between zones. The tablet-based drug-ordering module was time-consuming to build but already largely developed at this point. If the software had been ready earlier, the clinicians likely would not have changed their drug ordering workflow, and prescriptions could have been done more safely from the red zone.

One implementation challenge was the limited time to introduce and train staff on the system. We conducted a large showcase, where the development team dialed in to present the product, for all available users before OpenMRS-Ebola was deployed. However, we found that smaller group launches worked better in our setting and should be the norm in health emergencies. For training, we held sessions where clinical staff were able to use and learn the OpenMRS-Ebola interface. We combined our EHR training with more general training on common programs such as Microsoft Word and Excel. This helped increase staff engagement and computing familiarity for those new to such technology. More consistent training and retraining when necessary are critical for the success of new technologies in such situations. This is particularly difficult during an emergency response where hundreds of clinicians, sometimes rotating every few weeks, may provide patient care. The need to adequately train the user must also be balanced with the need to urgently deploy the system. One important consideration is whether all clinical staff should be trained to use the system or if a select set of superusers can mediate all interactions with it. This decision depends at least, in part, on the clinical workflow and skill levels of team members.

Regular communication between the operations team and system users is an important component of user support and buy-in. We had an operations team member speak at morning clinical meetings and discuss overnight software updates with the clinical lead. We also consulted clinicians on how to redesign the clinician station to fit the EHR system and regularly had an operations team member at the station to answer questions about the system as clinicians entered and exited the red zone. We found this difficult to maintain, especially with staff turnover, but it is essential for a smooth and rapid rollout.

We also faced staffing challenges on the ground, including being able to correctly time the deployment of an experienced team

member to lead the EHR implementation, given the difficulty in determining when the MVP would be released. An effective champion on the ground is necessary for a successful implementation, especially during a health emergency where there are few chances for a do-over. Having an organized and properly staffed operations team, both in terms of personnel and skill levels, is critical for a successful rollout. This is especially true during an emergency when there are many demands for health information, including daily reporting to external actors. Deploying our EHR required some effort, at minimum, from an operations lead, a health information manager, 3 to 4 additional health information staff, a key clinician, 3 to 4 additional go-to clinicians, and information technology (IT) support (both on the ground and remote). Without knowledgeable IT staff on the operations side, even a complete EHR may not be deployable on the ground. For our deployment, the SCI IT staff needed to urgently solve network access control and security issues and build a virtual Linux machine on a Windows server. Such requests are not typical IT needs for most emergency responses, so it is essential that proper and accessible IT support is secured before deploying an EHR in such situations. Ideally, local staff can be trained to fill several of the operations positions needed for implementation, which is useful both for the project and overall capacity building. Partly because of the experience received with deploying OpenMRS-Ebola, at least one local health information staff member obtained subsequent employment in EHR-related work at a local hospital after the epidemic.

Finally, a common issue for EHRs—especially in low-resource settings and emergency situations—is lack of evaluation data on performance, usage, and potential evidence of benefits to patient care and health facility management [27,28]. During emergencies, making a case for research or carrying out reliable studies is even harder. Yet, without such data, developing and deploying effective systems will remain an uncertain process and may be hard to justify. We were able to do a partial evaluation because we had a predeployment questionnaire and entered data in both the electronic and paper record systems for about a month as we rolled out phase 1 of the EHR. However, we were unable to perform a more complete evaluation because the ETC closed earlier than expected as the epidemic began winding down. Given the volatile nature of emergency responses, preplanning evaluations and having ideas in place on how to complete the evaluation in changing circumstances is important.

Comparison With Other Work

We conducted a literature review in March 2017 for publications regarding eHealth information systems for supporting the management of Ebola. We found 17 papers after searching PubMed for “Ebola and (“EMR” OR “EHR” OR “Medical Record” OR “mHealth” OR “eHealth” OR “mobile health”).” After reviewing titles and abstracts, we assessed 4 as relevant. These were supplemented with 3 papers recommended by colleagues from the gray literature and from an April 2015 meeting of team leaders from several projects hosted by the International Rescue Committee (IRC). To date, there are only a few publications that describe the development and deployment of EHR systems during the Ebola outbreak and only one that describes an EHR for use in ETCs [5].

No EHRs suitable for ETCs existed when we began developing OpenMRS-Ebola. Around the same time as us, a few other ETC EHR projects were initiated. Médecins Sans Frontières (MSF) partnered with Google Inc to develop and deploy Project Buendia, a basic tablet-based EHR [5]. They also developed robust hardware, including a waterproof tablet enclosure with charging capabilities and a miniature server with built-in low-power backup that allowed offline use with later synchronization of data. VecnaCares partnered with IRC to build an EHR with CliniPAK [29], and Project ELEOS was launched by MSF Belgium to deploy a simple data collection protocol on a personal digital assistant (PDA; [30]).

These other EHR systems, however, recorded a subset of the data that OpenMRS-Ebola collects, and not all were implemented during the outbreak. In particular, they lacked several important functions such as drug and IV fluid ordering, administration, and monitoring. Such data are complex to collect accurately, and errors can put patients at serious risk. Drug order entry systems are considered critical parts of successful EHRs but are rare in low-resource settings.

OpenMRS-Ebola and Project Buendia shared important components. Project Buendia was also built using the OpenMRS platform that was linked through OpenMRS application programming interfaces (APIs) to an Android application. Some clinical vocabulary—a subset of the CIEL dictionary focused on Ebola—was shared with Project Buendia to help speed up development. Project Buendia started in September 2014 and deployed in March 2015, at a cost of about US $1.9 million [5]. Although there was some contact between the OpenMRS-Ebola and Project Buendia teams, the projects diverged early on because of different requirements. In particular, unlike MSF sites, the Kerry Town ETC had unusually reliable power and infrastructure as part of the United Kingdom government's Ebola response. Thus, we were able to save time by developing a browser-based application, whereas Project Buendia needed an application that allowed offline use.

Overall, we believe that our system is the most comprehensive adaptable clinical EHR software developed to date for a health emergency in a low-resource setting. The potential for this system is strengthened by the wealth of software designs and code modules already deployed worldwide by projects using OpenMRS and a growing evidence base of the impact of OpenMRS use on care processes [31,32].

Speeding Up Use of Health Software During Emergencies

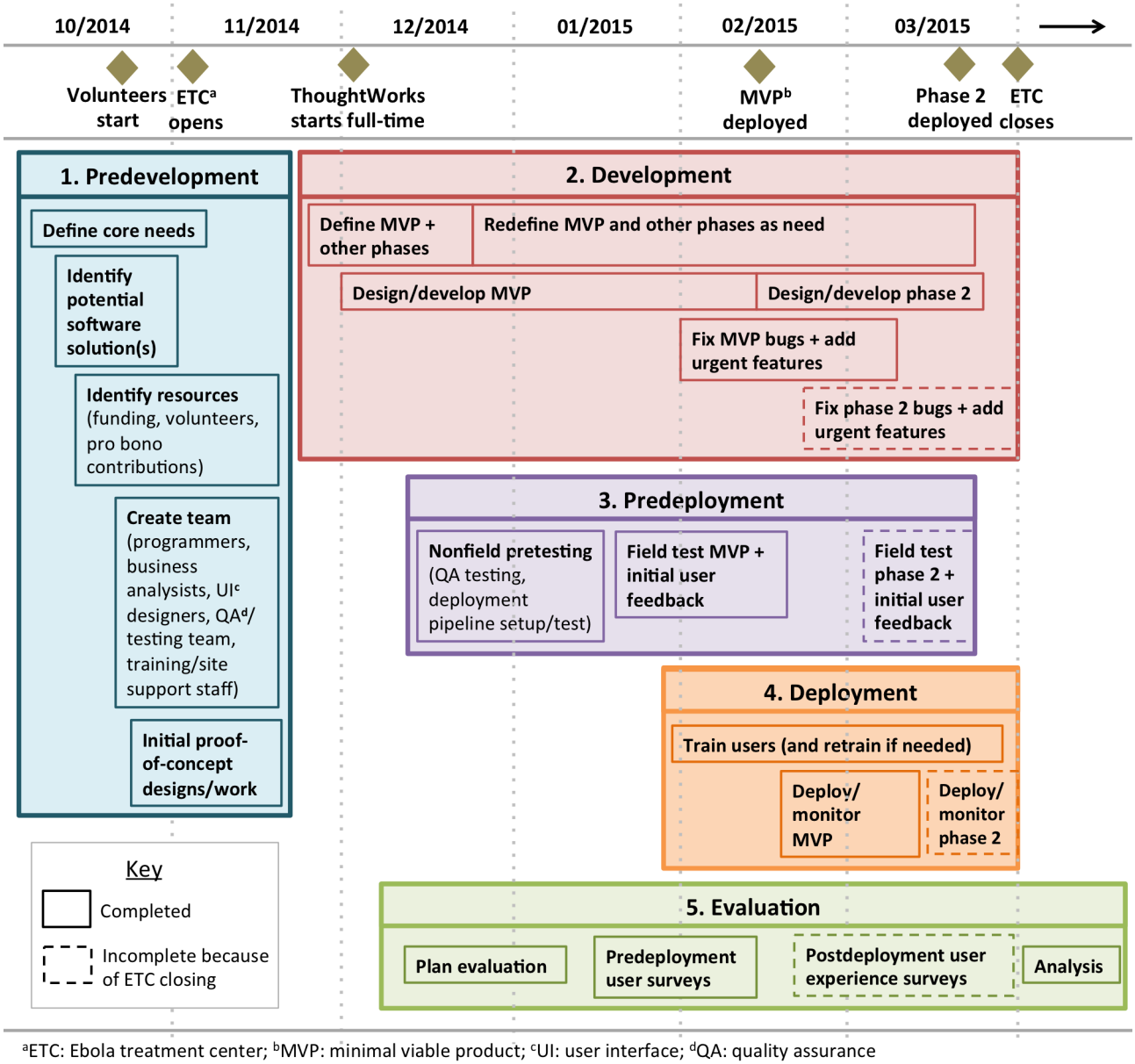
A key contribution from this project has been to advance our understanding of the process required to develop and deploy EHRs for emergencies in low-resource settings, as well as defining the factors slowing this process and how to address them. Figure 5 shows the stages that we recommend for deploying an EHR in such a setting, using the example of OpenMRS-Ebola, and potential areas for improvement.

The biggest delays for deploying an EHR at the Kerry Town ETC occurred for four main reasons. First, the largest delay

occurred because no suitable off-the-shelf EHR existed that could be immediately adapted and put into use at the ETC. Second, the lack of a suitable generic tablet interface for OpenMRS meant that the platform could not be immediately adapted for the most critical need: use in the red zone. Third, although volunteers made major early contributions, it took us a month to realize we needed full-time staff (eg, programmers, UI designers, and business analysts) to complete the EHR as quickly as possible and to contract that full-time development team. Finally, we faced the standard challenge of doing requirements gathering, field testing, and implementation under emergency conditions. This included a late start to field testing and user training because of a long delay in receiving the Sony tablets and other key hardware.

On the basis of our experience with OpenMRS-Ebola, we have listed a set of recommendations for rapidly building, deploying, and evaluating an EHR during a health emergency in Table 2. Even before the health emergency, however, work should be done to design and develop components that are necessary but lacking based on experiences from prior emergencies and anticipated functional needs.

Figure 5. Stages in development and deployment of an electronic health record (EHR) during a health emergency, using OpenMRS-Ebola as an example (note: phase 3 development is not included).



[a]ETC: Ebola treatment center; [b]MVP: minimal viable product; [c]UI: user interface; [d]QA: quality assurance

Table 2. Recommendations for rapidly building, deploying, and evaluating an electronic health record (EHR) during a health emergency.

Process stage	Recommendations
Predevelopment	Hire full-time professional staff to complete the product (ie, do not rely purely on volunteer efforts)
	Select a product owner who is a key stakeholder, deeply knowledgeable of the ground-level needs, and will remain engaged throughout the project
	Ensure that development team is proficient in Agile software approaches
	Set up mechanisms for regular (daily) communication
	Test and select hardware early to ensure (1) suitability based on needs (eg, waterproof, low power, and long battery life) and (2) that software designs are compatible with hardware
	Conduct review of working environment, clinical situation, and needed functionality with health workers
Development	Define the MVP[a] based on both ground-level priorities and time to develop features
	Communicate with ground-level team at least daily, if possible, including demonstrations and review of work in progress
	Conduct operational assessments of hardware and infrastructure needs
	Reprioritize MVP and other phases regularly based on ground-level feedback
Predeployment	Create communication mechanisms for user feedback to reach development team regularly and set up test and training servers to support this
	Ensure that operations team has appropriate staffing and skills required for EHR[b] rollout and troubleshooting
	Prepare training materials in advance and have plans for training and refresher training
	Make sure all hardware is ordered well in advance of user testing and training
	Determine strategy for selecting and training users (eg, all users vs superusers)
	Set up deployment pipeline to load and update software on production server
	Create contingency plans for anticipated problems (eg, locks for hardware, backup paper data collection, and backup power supplies)
Deployment	Ensure that deployment lead is proactive and can create and maintain buy-in from staff
	Conduct regular trainings with user-friendly material (eg, videos and annotated examples) and refresher trainings when needed
	Confirm that communication and feedback channels with development team are functional
Evaluation	Plan evaluation (including templates for pre- and postdeployment user surveys) during early development
	Keep records of informal feedback throughout the process
	Plan for contingencies (eg, obtaining consent and contact information for a Web-based follow-up user survey if emergency ends earlier than anticipated)

[a]MVP: minimal viable product.

[b]EHR: electronic health record.

Future Work

Ideally, the best time to prepare for an emergency response is before the emergency. In this outbreak, the key reason no organization was able to fully implement an EHR before ETCs began closing was because we had to do real software development during the outbreak. Learning from this and previous emergency responses can help identify common themes and modular configurable software that can be made at least partially ready in advance. Although EHR advances by us and others during this outbreak have helped fill many of the gaps we encountered at the start of this work, more development is still needed to extract and extend this work into a comprehensive EHR system suitable for health emergencies. Additional features would include laboratory test ordering, clinical decision support, pharmacy dispensing, integrated data entry quality checks, and customizable automated reports, as well as linkages to external laboratory services, community-level care, and surveillance. Coupling this with hardware advances, such as those started by Project Buendia, can result in a complete product that is ready to go out of the box after minor adaptations. Most importantly, the time to work on further advancing these software and hardware features is now, instead of waiting until the next health emergency hits.

More generally, responding to outbreaks such as this one requires effective and rapidly deployable strategies across a range of health activities, including case detection, triage, contact tracing, public education, and treatment of infected patients. This requires a set of robust technologies that are semantically interoperable and easily adaptable to support high-quality data collection during emergencies. Several open-source applications already exist that fill different niches in the health system, have been implemented at the national

scale, and support interoperability and secure data sharing. These include DHIS 2 at district- and national-level for aggregate data; OpenMRS, Baobab Health (Malawi), iSante (Haiti), and others for health facilities; and several mobile technology platforms in the community such as CommCare and RapidPro. Some of these were deployed in this outbreak and projects such as mHero, a communication tool between health workers and health ministries, combined more than one system [33-36]. However, these systems were largely deployed on an ad-hoc basis and not as part of a network of integrated health information technologies that could efficiently fill complementary data needs. To get to that point, these systems need to be reliably interoperable, easily adaptable, user friendly, and well tested with predesigned and accessible training materials.

Without better coordination and communication—both between and within developer and implementer teams—technology will not be efficiently implemented during an emergency, even if the software is fully ready. A detailed registry of suitable software, ideally managed by the World Health Organization or another international organization, is needed. This would provide a centralized location for governments and emergency response organizations to find and compare IT options, including interoperable ones, at the start of a health emergency instead of doing ad-hoc searches or relying on word-of-mouth contacts. ICT Africa tried but failed to obtain funding for such a registry in 2016 [37]. If we want to be better prepared for future emergencies, we need to fund and maintain such efforts. Similarly, a centralized and easy-to-use site to communicate with other responding organizations could improve coordination during an emergency and promote shared solutions. For example, we connected with other organizations developing EHRs through personal contacts, but a platform to share our work, ideas, and experiences would likely have sped up development and implementation. Such tools may be especially useful for rapid emergency deployment teams, which have increased substantially since the Ebola epidemic. Finally, it is essential that organizations be open and collaborative about their efforts during an emergency, as this can speed up development and implementation. The free and open-source software approach is a powerful framework for collaboration, not only across nonprofits but also with a larger community of developers. At the time of their development, none of the other Ebola EHR projects were open source. As software can be shared at no cost, we think that such cooperation among nonprofits with a common mission should become routine.

Ultimately, though, the best way for relevant technologies to be rapidly deployed in a health emergency is for them to already be integrated within the standard health system. There is understandable reluctance to introduce untried systems and approaches during emergencies, and new technologies come with standard challenges such as additional staff training, setup costs, and potential glitches. For nearly a decade, deploying electronic systems such as EHRs in LMICs in a scalable fashion has been possible [38]. If those systems are developed and deployed using common terminology and coding standards, they can be expanded and adapted for crises while strengthening the core codebase. OpenMRS is now implemented at a national scale in several countries and has been deployed in eastern Sierra Leone after the Ebola outbreak as a standard part of data collection at a large primary care clinic [39]. Systems such as DHIS 2 have also become increasingly popular in Sub-Saharan Africa [40]. Increasing such integration is a needed step toward EHRs becoming commonly used tools in health systems worldwide. There is also an increasing interest in the use of tablets for EHR users in LMICs, and lessons learned from this project should help the design and implementation of such systems.

Due to its fast and easy-to-use interface, OpenMRS-Ebola has potential to help with data collection for a range of other clinical needs (crisis and routine) in LMICs. Its features and adaptability, for example, make it suitable to support data collection in intensive care units and infectious environments. Similarly, the collection of vital signs is a key step in implementing early warning scores for clinical deterioration in hospitals (NEWS Scores). Systems such as OpenMRS-Ebola can help automate that process, similar to innovative systems developed for Oxford hospitals in the United Kingdom [41]. A key priority should be to expand the use of well-functioning technologies into standard health systems. Having the technology already integrated in the health system means having the benefits of an EHR during nonhealth emergency times and an easier transition to high-quality data collection during an emergency response.

Conclusions

The OpenMRS-Ebola EHR is well suited for patient records in an ETC because it allows for instant communication between infectious and noninfectious zones over a local wireless network, access to full clinical histories in both zones, and has a fast, easy UI suited to this difficult environment. Careful user design on a flexible platform can rapidly yield EHRs that are suitable for health emergencies. We were relatively successful in rapid development and deployment, but better preparation could likely have reduced time to implementation of the full system by approximately half (about 2 months). OpenMRS-Ebola can be adapted for future emergencies and is interoperable with other eHealth systems. To make a real impact, however, it must be part of a well-designed and tested set of interoperable electronic systems ready for deployment with appropriate hardware and training materials. Health information is too important a resource in emergency situations to be treated as an afterthought.

Acknowledgments

The authors acknowledge and appreciate the support from the OpenMRS community, ThoughtWorks Inc, Save the Children International, and Save the Children UK. This research was supported by the Save the Children Ebola Emergency Public Appeal. Dr Fraser was a Marie Skłodowska-Curie Fellow, funded from the European Union's Horizon 2020 research and innovation program under grant agreement No. 661289: Global eHealth. The authors are also grateful to the Infectious Disease Epidemiology

department at the London School of Hygiene and Tropical Medicine for publication funding support. The funders had no role in the development or review of this work.

Conflicts of Interest

DJ, PN, JR, GR, and DW are (or were) employees of ThoughtWorks during the course of this project and manuscript preparation. None of the other authors declare any conflicts of interest.

Multimedia Appendix 1

Complete screenshots of the desktop or laptop-based OpenMRS-Ebola application.

[PDF File (Adobe PDF File), 2MB - jmir_v19i8e294_app1.pdf]

Multimedia Appendix 2

Complete screenshots of the tablet-based OpenMRS-Ebola application.

[PDF File (Adobe PDF File), 2MB - jmir_v19i8e294_app2.pdf]

References

1. WHO. 2016. Situation Report: Ebola virus disease: 10 June 2016 URL: http://apps.who.int/iris/bitstream/10665/208883/1/ebolasitrep_10Jun2016_eng.pdf [accessed 2017-04-18] [WebCite Cache ID 6pnxxLNsB]
2. Moon S, Sridhar D, Pate MA, Jha AK, Clinton C, Delaunay S, et al. Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola. Lancet 2015 Nov 28;386(10009):2204-2221. [doi: 10.1016/S0140-6736(15)00946-0] [Medline: 26615326]
3. Whitty CJ, Farrar J, Ferguson N, Edmunds WJ, Piot P, Leach M, et al. Infectious disease: tough choices to reduce Ebola transmission. Nature 2014 Nov 13;515(7526):192-194. [doi: 10.1038/515192a] [Medline: 25391946]
4. Wilson D. CE: Inside an Ebola treatment unit: a nurse's report. Am J Nurs 2015 Dec;115(12):28-38. [doi: 10.1097/01.NAJ.0000475288.30664.70] [Medline: 26559159]
5. Jobanputra K, Greig J, Shankar G, Perakslis E, Kremer R, Achar J, et al. Electronic medical records in humanitarian emergencies - the development of an Ebola clinical information and patient management system. F1000Res Version 3 2016 Jun 23;5:1477 [FREE Full text] [doi: 10.12688/f1000research.8287.3] [Medline: 28357037]
6. Bühler S, Roddy P, Nolte E, Borchert M. Clinical documentation and data transfer from Ebola and Marburg virus disease wards in outbreak settings: health care workers' experiences and preferences. Viruses 2014 Feb 19;6(2):927-937 [FREE Full text] [doi: 10.3390/v6020927] [Medline: 24556792]
7. Abir M, Mostashari F, Atwal P, Lurie N. Electronic health records critical in the aftermath of disasters. Prehosp Disaster Med 2012 Dec;27(6):620-622. [doi: 10.1017/S1049023X12001409] [Medline: 23009722]
8. Clifford GD, Blaya JA, Hall-Clifford R, Fraser HS. Medical information systems: a foundation for healthcare technologies in developing countries. Biomed Eng Online 2008 Jun 11;7:18 [FREE Full text] [doi: 10.1186/1475-925X-7-18] [Medline: 18547411]
9. Fritz F, Tilahun B, Dugas M. Success criteria for electronic medical record implementations in low-resource settings: a systematic review. J Am Med Inform Assoc 2015 Mar;22(2):479-488. [doi: 10.1093/jamia/ocu038] [Medline: 25769683]
10. Nielsen P, Sæbø JI. Three strategies for functional architecting: cases from the health systems of developing countries. Information Technology for Development 2015 Apr 21;22(1):134-151. [doi: 10.1080/02681102.2015.1026304]
11. OpenMRS. OpenMRS website URL: http://openmrs.org/ [accessed 2017-03-29] [WebCite Cache ID 6pKc70Npm]
12. Mamlin BW, Biondich PG, Wolfe BA, Fraser H, Jazayeri D, Allen C, et al. Cooking up an open source EMR for developing countries: OpenMRS - a recipe for successful collaboration. AMIA Annu Symp Proc 2006:529-533 [FREE Full text] [Medline: 17238397]
13. Mohammed-Rajput NA, Smith DC, Mamlin B, Biondich P, Doebbeling BN, Open MRS Collaborative Investigators. OpenMRS, a global medical records system collaborative: factors influencing successful implementation. AMIA Annu Symp Proc 2011;2011:960-968 [FREE Full text] [Medline: 22195155]
14. OpenMRS. 2016 OpenMRS Annual Report URL: http://openmrs.org/wp-content/uploads/2017/03/2016-OpenMRS-Annual-Report.pdf [accessed 2017-03-30] [WebCite Cache ID 6pLm5Nyjf]
15. Kasthurirathne SN, Mamlin B, Grieve G, Biondich P. Towards standardized patient data exchange: integrating a FHIR based API for the open medical record system. Stud Health Technol Inform 2015;216:932. [Medline: 26262234]
16. Greenes R. Clinical decision support: the road to broad adoption. Boston: Academic Press; 2014.
17. Fraser H, Ball E, Jazayeri D. OpenMRS the Open Medical Record System: Software, Community, Philosophy. In: Global Health Informatics: Principles of eHealth and mHealth to Improve Quality of Care. Cambridge, MA: MIT Press; 2017.
18. Douglas GP, Landis-Lewis Z, Hochheiser H. Simplicity and usability: lessons from a touchscreen electronic medical record system in Malawi. Interactions 2011 Nov;18(6):50-53. [doi: 10.1145/2029976.2029990]

19. wiki.openmrs. OpenMRS Rwanda Primary Care documentation URL: https://wiki.openmrs.org/display/docs/Rwanda+Primary+Care [accessed 2017-06-11] [WebCite Cache ID 6r9KGnVGh]
20. issues.openmrs. OpenMRS Ebola Response work tickets URL: https://issues.openmrs.org/browse/ER-96?jql=project%20%3D%20ER%20AND%20status%20%3D%20Accepted%20ORDER%20BY%20fixVersion [accessed 2017-06-11] [WebCite Cache ID 6r9KN8Php]
21. ThoughtWorks. 2015. How ThoughtWorks Brasil Fought Ebola URL: https://www.thoughtworks.com/insights/blog/how-thoughtworks-brasil-helped-fighting-ebola [accessed 2017-03-30] [WebCite Cache ID 6pMByUeeF]
22. wiki.openmrs. OpenMRS 2.1 URL: https://wiki.openmrs.org/display/docs/OpenMRS+2.1 [accessed 2017-03-07] [WebCite Cache ID 6on11eKL1]
23. Kanter AS, Borland R, Barasa M, Iiams-Hauser C, Velez O, Kaonga NN, et al. The importance of using open source technologies and common standards for interoperability within eHealth: perspectives from the Millennium Villages Project. Adv Health Care Manag 2012;12:189-204 [FREE Full text] [Medline: 22894051]
24. GitHub. OpenMRS-Ebola GitHub repository URL: https://github.com/openmrs/openmrs-module-ebolaexample [accessed 2017-03-31] [WebCite Cache ID 6pNp1OUXr]
25. Chan M. Ebola virus disease in West Africa--no early end to the outbreak. N Engl J Med 2014 Sep 25;371(13):1183-1185. [doi: 10.1056/NEJMp1409859] [Medline: 25140856]
26. Meltzer MI, Atkins CY, Santibanez S, Knust B, Petersen BW, Ervin ED, Centers for Disease Control and Prevention (CDC). Estimating the future number of cases in the Ebola epidemic--Liberia and Sierra Leone, 2014-2015. MMWR Suppl 2014 Sep 26;63(3):1-14. [Medline: 25254986]
27. Tomasi E, Facchini LA, Maia MF. Health information technology in primary health care in developing countries: a literature review. Bull World Health Organ 2004 Nov;82(11):867-874 [FREE Full text] [Medline: 15640923]
28. Blaya JA, Fraser HS, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251 [FREE Full text] [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
29. Wang W. JEDI-Development of a joint electronic health data management and decision support interface for an Ebola Treatment Unit in Liberia. 2015 Presented at: APHA Annual Meeting & Expo; 4 November 2015; Chicago, IL URL: https://apha.confex.com/apha/143am/webprogram/Paper328254.html
30. Gallego M, Deneweth W, Adamson B, Vincent-Smith R, Verdonck G. F1000 Research.: MSF Scientific Day; 2015. A barcode/handheld-computer based solution for Ebola Management Centres URL: https://f1000research.com/slides/1000071 [accessed 2017-04-18] [WebCite Cache ID 6pofcl9Je]
31. Amoroso CL, Akimana B, Wise B, Fraser HS. Using electronic medical records for HIV care in rural Rwanda. Stud Health Technol Inform 2010;160(Pt 1):337-341. [Medline: 20841704]
32. Were MC, Nyandiko WM, Huang KT, Slaven JE, Shen C, Tierney WM, et al. Computer-generated reminders and quality of pediatric HIV care in a resource-limited setting. Pediatrics 2013 Mar;131(3):e789-e796 [FREE Full text] [doi: 10.1542/peds.2012-2072] [Medline: 23439898]
33. Sacks JA, Zehe E, Redick C, Bah A, Cowger K, Camara M, et al. Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. Glob Health Sci Pract 2015 Nov 12;3(4):646-659 [FREE Full text] [doi: 10.9745/GHSP-D-15-00207] [Medline: 26681710]
34. Pronyk P, Rogers B, Lee S, Bhatnagar A, Wolman Y, Monasch R, et al. The effect of community-based prevention and care on Ebola transmission in Sierra Leone. Am J Public Health 2016 Apr;106(4):727-732. [doi: 10.2105/AJPH.2015.303020] [Medline: 26890176]
35. Jazayeri D, Oza S, Ramos G, Fraser H, Teich JM, Kanter AS, et al. Design and development of an EMR for Ebola Treatment Centers in Sierra Leone using OpenMRS. Stud Health Technol Inform 2015;216:916. [Medline: 26262218]
36. O'Donovan J, Bersin A. Controlling Ebola through mHealth strategies. Lancet Glob Health 2015 Jan;3(1):e22 [FREE Full text] [doi: 10.1016/S2214-109X(14)70357-2] [Medline: 25539965]
37. kickstarter. All Africa Database of ICT Projects URL: https://www.kickstarter.com/projects/5973466/all-africa-database-of-ict-projects [accessed 2017-03-16] [WebCite Cache ID 6p0nFQCbY]
38. Fraser HS, Biondich P, Moodley D, Choi S, Mamlin BW, Szolovits P. Implementing electronic medical record systems in developing countries. Inform Prim Care 2005;13(2):83-95. [Medline: 15992493]
39. DeSimone D. talk.openmrs. 2016. Recent PIH implementation in Sierra Leone URL: https://talk.openmrs.org/t/recent-pih-implementation-in-sierra-leone/7458 [accessed 2017-03-30] [WebCite Cache ID 6pM6ZXmL7]
40. DHIS2. DHIS 2 in action URL: https://www.dhis2.org/inaction [accessed 2017-03-16] [WebCite Cache ID 6p0s5VIHy]
41. Wong D, Bonnici T, Knight J, Gerry S, Turton J, Watkinson P. A ward-based time study of paper and electronic documentation for recording vital sign observations. J Am Med Inform Assoc 2017 Jul 01:717-721. [doi: 10.1093/jamia/ocw186] [Medline: 28339626]

Abbreviations

API: application programming interface
CIEL: Columbia international eHealth laboratory

CI: continuous integration
CSV: comma separated values
ETC: Ebola treatment center
eHealth: electronic health
EHR: electronic health record
HIV: human immunodeficiency virus
IRC: International Rescue Committee
IT: information technology
IV: intravenous
LMICs: low- and middle-income countries
mHealth: mobile health
MSF: Medècins Sans Frontiéres
MVP: minimum viable product
PPE: personal protective equipment
PRN: pro re nata
SCI: Save the Children International
TB: tuberculosis
UI: user interface
UPS: uninterruptable power supply
VSAT: very small aperture terminal
WLAN: wireless local area network

Edited by G Eysenbach; submitted 18.04.17; peer-reviewed by G Douglas, T Hannan; comments to author 31.05.17; accepted 12.06.17; published 21.08.17.

Please cite as:
Oza S, Jazayeri D, Teich JM, Ball E, Nankubuge PA, Rwebembera J, Wing K, Sesay AA, Kanter AS, Ramos GD, Walton D, Cummings R, Checchi F, Fraser HS
Development and Deployment of the OpenMRS-Ebola Electronic Health Record System for an Ebola Treatment Center in Sierra Leone
J Med Internet Res 2017;19(8):e294
URL: http://www.jmir.org/2017/8/e294/
doi:10.2196/jmir.7881
PMID:28827211

Original Paper

Web-Based Physician Ratings for California Physicians on Probation

Gregory P Murphy[1*], MD; Mohannad A Awad[1,2*], MD; E Charles Osterberg[3*], MD; Thomas W Gaither[4*], BS; Thanabhudee Chumnarnsongkhroh[1*], MD; Samuel L Washington[1*], MD; Benjamin N Breyer[1,5], MD

[1]Zuckerberg San Francisco General Hospital, Department of Urology, University of California, San Francisco, San Francisco, CA, United States

[2]Department of Surgery, King Abdulaziz University, Rabigh, Saudi Arabia

[3]Dell Medical School, Department of Surgery, University of Texas, Austin, TX, United States

[4]School of Medicine, Department of Urology, University of California, San Francisco, San Francisco, CA, United States

[5]Department of Biostatistics and Epidemiology, University of California, San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Benjamin N Breyer, MD
Zuckerberg San Francisco General Hospital
Department of Urology
University of California, San Francisco
Suite 3A
1001 Potrero Ave
San Francisco, CA, 94110
United States
Phone: 1 4152068805
Fax: 1 4152064499
Email: benjamin.breyer@ucsf.edu

Abstract

Background: Web-based physician ratings systems are a popular tool to help patients evaluate physicians. Websites help patients find information regarding physician licensure, office hours, and disciplinary records along with ratings and reviews. Whether higher patient ratings are associated with higher quality of care is unclear.

Objective: The aim of this study was to characterize the impact of physician probation on consumer ratings by comparing website ratings between doctors on probation against matched controls.

Methods: A retrospective review of data from the Medical Board of California for physicians placed on probation from December 1989 to September 2015 was performed. Violations were categorized into nine types. Nonprobation controls were matched by zip code and specialty with probation cases in a 2:1 ratio using the California Department of Consumer Affairs website. Web-based reviews were recorded from vitals.com, healthgrades.com, and ratemds.com (ratings range from 1-5).

Results: A total of 410 physicians were placed on probation for 866 violations. The mean (standard deviation [SD]) number of ratings per doctor was 5.2 (7.8) for cases and 4 (6.3) for controls (P=.003). The mean rating for physicians on probation was 3.7 (1.6) compared with 4.0 (1.0) for controls when all three rating websites were pooled (P<.001). Violations for medical documentation, incompetence, prescription negligence, and fraud were found to have statistically significant lower rating scores. Conversely, scores for professionalism, drugs or alcohol, crime, sexual misconduct, and personal illness were similar between cases and controls. In a univariate analysis, probation was found to be associated with lower rating, odds ratio=1.5 (95% CI 1.0-2.2). This association was not significant in a multivariate model when we included age and gender.

Conclusions: Web-based physician ratings were lower for doctors on probation indicating that patients may perceive a difference. Despite these statistical findings, the absolute difference was quite small. Physician rating websites have utility but are imperfect proxies for competence. Further research on physician Web-based ratings is warranted to understand what they measure and how they are associated with quality.

(J Med Internet Res 2017;19(8):e254) doi:10.2196/jmir.7488

KEYWORDS

online physician ratings; probation; Internet; quality of care

Introduction

Web-based physician ratings systems are becoming an increasingly popular tool to help patients choose their hospitals and providers [1]. Physician rating websites contain information regarding physician licensure, office hours, and disciplinary records in addition to ratings and reviews that are helpful to health care consumers [2]. Studies show mixed results as to whether highly rated hospitals or physicians deliver superior care [3,4]. Okike et al found no correlation between Web-based ratings for cardiac surgeons and 30-day mortality following coronary artery bypass grafting [5]. Regarding hospitals, a review of social media and rating site literature indicated a relationship between improved hospital ratings and better mortality and infection rates [6]. A systematic review by Doyle et al suggested a positive correlation between higher ratings and patient safety and clinical effectiveness [7]. Similarly, Greaves et al demonstrated that Web-based hospital ratings are associated with improved mortality and infection rates [8]. Higher Yelp ratings were also associated with higher Hospital Consumer Assessment of Healthcare Providers and Systems ratings [9]. Regarding social media, Facebook ratings have been correlated to lower hospital readmission rates, whereas Twitter comments were not associated with quality metrics [10,11].

When studying the quality of physician care, many different outcome metrics have been used with differing strengths. These outcomes can be influenced by a multitude of factors, some directly related to physician skill whereas others, such as patient health, are outside of the doctor's control. Physician ratings have been correlated with such diverse outcomes as board certification, education, malpractice claims, mortality, infection, and readmission rates [2-12]. Though it can be difficult to equate physician competence with outcomes, doctors who violate codes of conduct and fail to meet the standard of care are placed on probation by their state medical board after a review of the evidence. It is unknown how patients perceive and rate physicians on probation as it has never been used as a quality metric. We sought to determine whether patients rated probated providers differently than physicians in good standing by comparing Web-based physician ratings from three consumer rating websites. We hypothesize that physicians on probation will have similar ratings to nonprobation controls.

Methods

We retrospectively reviewed publically available data from the Medical Board of California for physicians who were placed on probation from December 1989 to September 2015. Rationales for probation were independently categorized into nine types of infractions by five independent reviewers (MAA, TWG, TC, SLW, and GPM). After reviewing all infractions, we used an inductive approach to create nine probation categories [13]. If there were any questions as to the appropriate categorization by an individual reviewer, this was brought to the group and a consensus decision was made. Nonprobation controls were matched by zip code and specialty with probation cases in a 2:1 ratio using the California Department of Consumer Affairs website. Web-based reviews were recorded from vitals.com, healthgrades.com, and ratemds.com. Ratings on these websites ranged from 1 to 5 and were weighted by the number of ratings so as not to overemphasize a small number of outlier ratings.

Statistical analysis was performed with STATA version 14 (College Station, TX). Parametric and nonparametric statistics were run despite the nonnormal statistical distribution. The results were similar between statistical analyses as the sample size was large [14]. For ease of reporting, parametric tests were reported. Specifically, a student's *t*-test was used to compare mean ratings for violations. A rating below 3 was considered low and logistic univariable and multivariable regression analyses were done to determine predictors for low rating. We determined covariates including age, gender, type of specialty, and type of violation to be included in the multivariable model a priori. All tests were two-sided and *P* values <.05 were considered significant.

Results

A total of 410 physicians (cases) were placed on probation for 866 violations and were matched with 818 controls. The mean (standard deviation [SD]) number of ratings per physician was 5.2 (7.8) for cases and 4.0 (6.3) for controls (*P*=.003). The mean rating for physicians on probation was 3.7 (SD 1.6) compared with 4.0 (SD 1.0) for controls when all three rating websites were pooled (*P*<.001). Figure 1 depicts the overall violations using a violin plot that shows the median, interquartile range, and distribution of the ratings for cases and controls.

Tables 1 and 2 show the differences in average weighted ratings between cases and controls for each website stratified by violation type. Violations for medical documentation, incompetence, prescription negligence, and fraud were found to have statistically significant differences in physician ratings between cases and controls. Conversely, physician ratings for cases and controls were not statistically different when the violation pertained to professionalism, drugs or alcohol, crime, sexual misconduct, and personal illness.

Table 1. Mean rating of probation cases and controls by violation type.

Violation	Total (%)[a]	Vitals cases	Vitals controls	*P*-value	Health Grades cases	Health Grades controls	*P*-value
Medical records, mean (SD[b])	176 (42.9)	3.7 (1.41)	4.1 (0.94)	<.001	3.6 (2.1)	3.8 (1.29)	.02
Professionalism, mean (SD)	166 (40.5)	3.9 (1.35)	4.1 (0.82)	.02	3.9 (1.98)	3.9 (1.09)	.81
Incompetence, mean (SD)	161 (39.3)	3.6 (1.46)	4 (0,93)	<.001	3.6 (2.08)	3.8 (1.21)	.08
Prescription negligence, mean (SD)	106 (25.9)	3.9 (1.6)	4 (1.05)	.31	3.8 (2.11)	3.9 (1.23)	.49
Drug or alcohol addiction, mean (SD)	84 (20.5)	4 (1.57)	4.1 (0.88)	.63	3.8 (2.1)	3.9 (1.11)	.66
Committing a crime, mean (SD)	63 (15.4)	3.8 (1.53)	3.8 (0.83)	.91	3.8 (1.87)	3.9 (0.94)	.45
Fraud, mean (SD)	44 (10.7)	3.9 (1.28)	4.1 (0.82)	.11	3.7 (1.81)	3.9 (1.05)	.32
Sexual misconduct or battery, mean (SD)	42 (10.2)	4 (1.27)	4.1 (0.81)	.28	4 (2.02)	4 (1.11)	.86
Personal illness, mean (SD)	24 (5.9)	3.7 (1.32)	4.2 (0.83)	.02	3.6 (2.3)	3.8 (1.09)	.56
All violations, mean (SD)	866 (0)	3.8 (1.47)	4.1 (0.96)	<.001	3.7 (2.03)	3.8 (1.2)	.01

[a]Some doctors have more than one violation. The percentage is for number of doctors on probation (cases group) having each violation out of the total number of doctors on probation (410).

[b]SD: standard deviation.

Table 2. Mean rating of probation cases and controls by violation type.

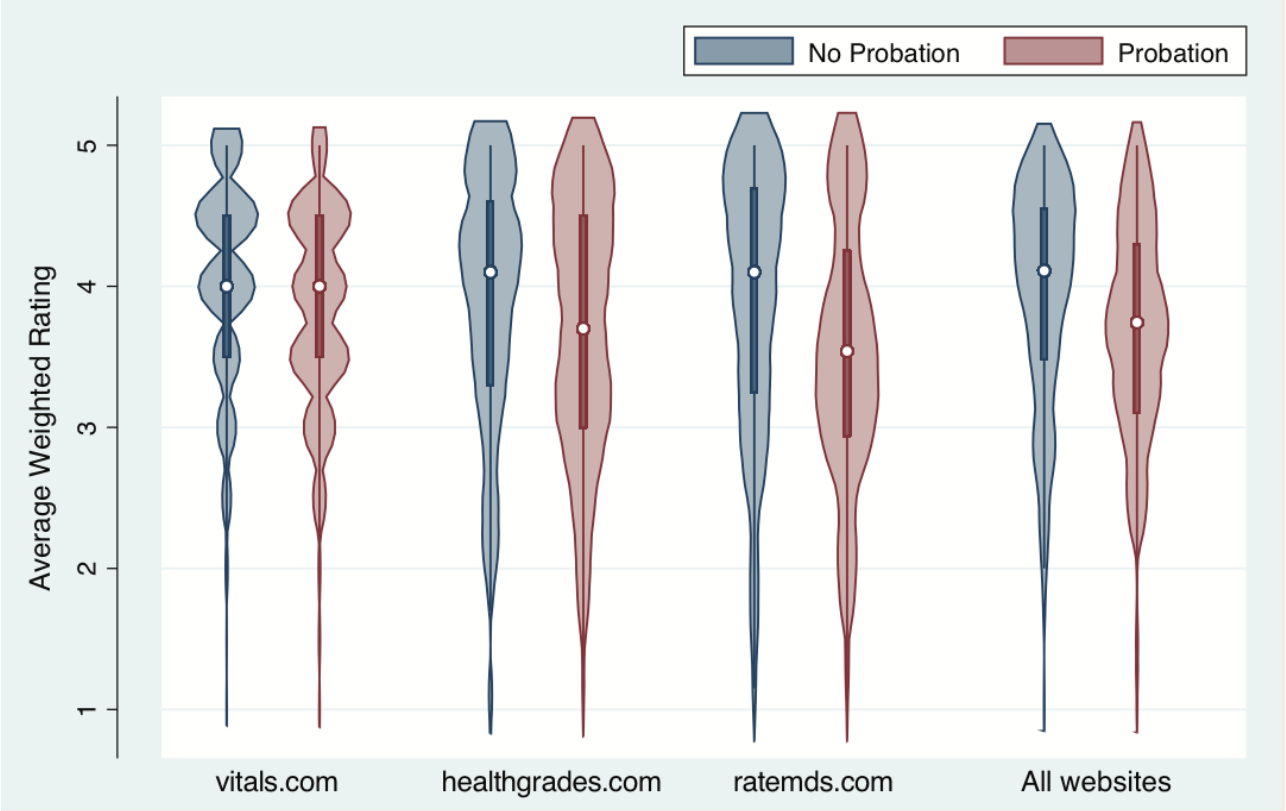
Violation	RateMDs cases	RateMDs controls	*P*-value	Cases in all websites	Controls in all websites	*P*-value
Medical records, mean (SD[a])	3.5 (2.06)	3.8 (1.4)	.04	3.5 (1.58)	3.9 (1.04)	<.001
Professionalism, mean (SD)	3.9 (2)	4 (1.14)	.78	3.9 (1.58)	4 (0.87)	.54
Incompetence, mean (SD)	3.3 (2.21)	3.8 (1.43)	.03	3.5 (1.55)	4 (1.01)	<.001
Prescription negligence, mean (SD)	3.8 (2.12)	4 (1.32)	.41	3.8 (1.59)	4 (1.05)	.04
Drug or alcohol addiction, mean (SD)	3.8 (1.93)	3.9 (1.09)	.86	4 (1.78)	4 (0.93)	.64
Committing a crime, mean (SD)	3.7 (2.1)	3.7 (1.05)	.91	3.9 (1.59)	3.8 (0.83)	.62
Fraud, mean (SD)	3.3 (2)	3.6 (1.23)	.35	3.6 (1.37)	3.9 (0.89)	.03
Sexual misconduct or battery, mean (SD)	3.5 (1.83)	3.8 (1.11)	.26	3.9 (1.53)	4 (0.94)	.59
Personal illness, mean (SD)	3.4 (1.65)	3.4 (1.32)	.95	3.6 (1.33)	3.9 (0.8)	.17
All violations, mean (SD)	3.5 (2)	3.9 (1.31)	.001	3.7 (1.56)	4 (1.01)	<.001

[a]SD: standard deviation.

In the univariable analysis, probation was found to be associated with lower rating, odds ratio=1.5 (95% CI 1.001-2.2). This association was not significant in a multivariable model when we included age and gender, odds ratio=1.4 (95% CI 0.9-2.2). In addition, age, gender, type of specialty, and type of violation all did not predict a low rating.

Healthgrades.com included the probation status of the physician on their website in 328 of 389 (84%) cases in our cohort, yet there was no significant difference in mean ratings for doctors who had their violations published compared with those who did not. A sensitivity analysis was done to determine if timing of probation effected ratings. The majority of our violations are from the recent past. When we excluded violations before 2005, only 17 physicians were excluded and no difference in results was seen.

Figure 1. Violin plot showing ratings of physicians on probation versus controls by website.



Discussion

Principal Findings

Web-based physician ratings from three websites were lower for doctors on probation, indicating that patients perceive a difference in physician performance. Despite this difference being statistically significant, the absolute difference is quite small with overlapping rating distributions. The vast majority of reviews were positive with small numbers of reviews per doctor. This is similar to other studies, making it difficult to draw strong conclusions about Web-based rating utility [6,15].

When categorizing the violations into subtypes, patients rate physicians lower when the probation infringements correspond to infractions for medical documentation, incompetence, prescription negligence, and fraud. In contrast, doctors on probation for professionalism, drugs or alcohol, sexual abuse, or personal illness received statistically similar ratings to controls. This finding may reflect what websites are measuring: patient perception and consumer experience, which is influenced mostly by bedside manner, wait times, and staffing issues [16]. Physicians with violations for medical documentation, for example, may run higher volume clinics, which negatively influences the patient experience. A German study showed higher physician ratings correlated to lower number of patients in the practice more than quality metrics [17].

Comparison With Prior Work

The relationship between physician skills, Web-based ratings and outcome metrics is complex. Polled physicians are concerned about the usefulness of Web-based ratings and how they could negatively impact their practice [18]. Comparing websites scores of individual sports medicine doctors revealed a low degree of correlation between websites raising concerns about their reliability [19]. When reviews were correlated with scores, very high and very low ratings were correlated with the patient's perception of physician skill and quality, which would seem difficult for patients to truly know [20]. Similarly, Web-based reviews of hand surgeons showed very positive reviews were related to perceived competence [21]. This raises concerns among physicians as to the accuracy and potential harm of negative reviews [6].

For providers, it is an uncomfortable position to be publically judged by patients, which could significantly damage a doctor's reputation and practice. Some physicians have responded with indifference or disdain for these reviews [12,22]. Others have embraced these ratings and made an effort to alter their behavior and practice to improve patient experience and scores [21]. When selecting a surgeon Web-based, polled patients listed Web-based reviews as a minor factor with insurance, office location, and hospital reputation as more important [23]. It seems most patients understand the limitations of Web-based reviews.

There are ethical pitfalls of Web-based reviews due to the possible financial gain from improved referrals. Most reviews appear genuine but careful study revealed a few anonymous reviews may be from the physicians themselves in an attempt to falsely raise their ratings [2]. Better regulation may be needed to prevent abuses by providers including asking or paying patients for positive reviews [23]. Some concerns have been raised by doctors that competitors or unhappy employees could easily pose as a patient and post a negative review [24].

Despite the negatives, Web-based reviews are likely to increase in importance and patient utilization [1]. Lee suggests Web-based reviews help improve physician-patient relationship by increasing transparency and trust [25]. Not only does this help patient decision making but also provides the physician with feedback that can improve patient experience.

Future Research

More research is needed to explore different physician quality metrics to see if there is an association with positive reviews. Determining more precisely what aspects of the health care system patient feedback can measure and improve is important. Prior studies explored many objective criteria such as mortality, infection, or readmission rates [4,6,7-10]. These are important metrics but do not necessarily reflect physician decision making. For example, a surgical site infection may occur despite excellent sterile technique and appropriate guidelines-based antibiotic practice. It raises questions as to whether physicians should be judged on outcomes or more on their decision making. Our study is the first to explore the link between Web-based reviews and probation status. Using probation as a proxy for clinical incompetence seems to be imperfect. When a physician is placed on probation, there are clearly issues to be remedied, but whether this negatively affects clinical care is difficult to determine. Whereas more research is needed to determine if probation is a reliable marker for clinical incompetence, it would seem to be a good indicator for poor physician decision making on some level. In addition, differing violations would seem to affect physician judgment and quality in diverse ways so lumping all probation cases has its limitations. More research is needed into which violations more severely affect physician quality.

Limitations

Our study has limitations. The timing of probation could be problematic as violations from many years ago could correlate poorly with recent reviews. However, a sensitivity analysis excluding 17 cases before 2005 showed similar results. We studied patient reviews of physicians practicing in California, so these results may not be generalizable to reviews from other states. Future research should be directed toward other populations to confirm our findings.

Conclusions

Web-based physician ratings were lower for doctors on probation indicating that patients may perceive a difference. Whereas statistically significant, the absolute difference was quite small. Physician rating websites have utility but are imperfect proxies for competence. Further research on physician Web-based ratings is warranted to understand what they measure and how they are associated with quality.

Acknowledgments

The Alafi Foundation provided funding for this research. Propublica.com provided the data regarding California physicians on probation.

Conflicts of Interest

None declared.

References

1. 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]
2. 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]
3. Fenton JJ, Jerant AF, Bertakis KD, Franks P. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. Arch Intern Med 2012 Mar 12;172(5):405-411. [doi: 10.1001/archinternmed.2011.1662] [Medline: 22331982]
4. 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]
5. 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]
6. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
7. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013;3(1):231-233 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
8. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012 Mar 12;172(5):435-436. [doi: 10.1001/archinternmed.2011.1675] [Medline: 22331980]

9. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
10. Glover M, Khalilzadeh O, Choy G, Prabhakar AM, Pandharipande PV, Gazelle GS. Hospital evaluations by social media: a comparative analysis of Facebook ratings among performance outliers. J Gen Intern Med 2015 Oct;30(10):1440-1446 [FREE Full text] [doi: 10.1007/s11606-015-3236-3] [Medline: 25749881]
11. Greaves F, Laverty A, Cano D, Moilanen K, Pulman S, Darzi A, et al. Tweets about hospital quality: a mixed methods study. BMJ Qual Saf Oct 2014;23(10):838-846. [doi: 10.1136/bmjqs-2014-002875] [Medline: 24748372]
12. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
13. Bolton JW. Varieties of clinical reasoning. J Eval Clin Pract 2015 Jun;21(3):486-489. [doi: 10.1111/jep.12309] [Medline: 25652845]
14. Fagerland MW. t-tests, non-parametric tests, and large studies--a paradox of statistical practice? BMC Med Res Methodol 2012 Jun 14;12:78 [FREE Full text] [doi: 10.1186/1471-2288-12-78] [Medline: 22697476]
15. Riemer C, Doctor M, Dellavalle RP. Analysis of online ratings of dermatologists. JAMA Dermatol 2016 Feb;152(2):218-219. [doi: 10.1001/jamadermatol.2015.4991] [Medline: 26677133]
16. 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]
17. 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]
18. 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]
19. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a sports medicine surgeon's "quality" in the modern era: an analysis of popular physician online rating websites. HSS J 2016 Oct;12(3):272-277. [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
20. Trehan SK, DeFrancesco CJ, Nguyen JT, Charalel RA, Daluiski A. Online patient ratings of hand surgeons. J Hand Surg Am 2016 Jan;41(1):98-103. [doi: 10.1016/j.jhsa.2015.10.006] [Medline: 26710742]
21. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? Results from an online-based cross-sectional study. J Med Internet Res 2016 Sep 19;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]
22. Ruan X, Bordelon GJ, Kaye AD. Online reviews of physicians. JAMA Dermatol 2016 Mar;152(3):350. [doi: 10.1001/jamadermatol.2015.6164] [Medline: 26964056]
23. Ziemba JB, Allaf ME, Haldeman D. Consumer preferences and online comparison tools used to select a surgeon. JAMA Surg 2017 Jan 04;152(4):410-411. [doi: 10.1001/jamasurg.2016.4993] [Medline: 28052162]
24. Samora JB, Lifchez SD, Blazar PE, American Society for Surgery of the Hand Ethics and Professionalism Committee. Physician-rating web sites: ethical implications. J Hand Surg Am 2016 Jan;41(1):104-10.e1. [doi: 10.1016/j.jhsa.2015.05.034] [Medline: 26304734]
25. Lee V. Transparency and trust - online patient reviews of physicians. N Engl J Med 2017 Jan 19;376(3):197-199. [doi: 10.1056/NEJMp1610136] [Medline: 28099823]

Abbreviations

SD: standard deviation

Edited by T Kool; submitted 10.02.17; peer-reviewed by T van de Belt, M Emmert; comments to author 11.03.17; revised version received 09.04.17; accepted 11.05.17; published 22.08.17.

Please cite as:
Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, Breyer BN
Web-Based Physician Ratings for California Physicians on Probation
J Med Internet Res 2017;19(8):e254
URL: http://www.jmir.org/2017/8/e254/
doi:10.2196/jmir.7488
PMID:28830852

Original Paper

Developments in the Frequency of Ratings and Evaluation Tendencies: A Review of German Physician Rating Websites

Stuart McLennan[1,2], MBHL, PhD; Daniel Strech[1], Dr med, Dr Phil; Swantje Reimann[1], Dr Phil

[1]Institute for History, Ethics and Philosophy of Medicine, Hannover Medical School, Hannover, Germany

[2]Institute for Biomedical Ethics, Universität Basel, Basel, Switzerland

Corresponding Author:
Daniel Strech, Dr med, Dr Phil
Institute for History, Ethics and Philosophy of Medicine
Hannover Medical School
OE 5450
Carl-Neuberg-Str. 1
Hannover, 30625
Germany
Phone: 49 5115326498
Fax: 49 5115325650
Email: strech.daniel@mh-hannover.de

Abstract

Background: Physician rating websites (PRWs) have been developed to allow all patients to rate, comment, and discuss physicians' quality online as a source of information for others searching for a physician. At the beginning of 2010, a sample of 298 randomly selected physicians from the physician associations in Hamburg and Thuringia were searched for on 6 German PRWs to examine the frequency of ratings and evaluation tendencies.

Objective: The objective of this study was to examine (1) the number of identifiable physicians on German PRWs; (2) the number of rated physicians on German PRWs; (3) the average and maximum number of ratings per physician on German PRWs; (4) the average rating on German PRWs; (5) the website visitor ranking positions of German PRWs; and (6) how these data compare with 2010 results.

Methods: A random stratified sample of 298 selected physicians from the physician associations in Hamburg and Thuringia was generated. Every selected physician was searched for on the 6 PRWs (Jameda, Imedo, Docinsider, Esando, Topmedic, and Medführer) used in the 2010 study and a PRW, Arztnavigator, launched by Allgemeine Ortskrankenkasse (AOK).

Results: The results were as follows: (1) Between 65.1% (194/298) on Imedo to 94.6% (282/298) on AOK-Arztnavigator of the physicians were identified on the selected PRWs. (2) Between 16.4% (49/298) on Esando to 83.2% (248/298) on Jameda of the sample had been rated at least once. (3) The average number of ratings per physician ranged from 1.2 (Esando) to 7.5 (AOK-Arztnavigator). The maximum number of ratings per physician ranged from 3 (Esando) to 115 (Docinsider), indicating an increase compared with the ratings of 2 to 27 in the 2010 study sample. (4) The average converted standardized rating (1=positive, 2=neutral, and 3=negative) ranged from 1.0 (Medführer) to 1.2 (Jameda and Topmedic). (5) Only Jameda (position 317) and Medführer (position 9796) were placed among the top 10,000 visited websites in Germany.

Conclusions: Whereas there has been an overall increase in the number of ratings when summing up ratings from all 7 analyzed German PRWs, this represents an average addition of only 4 new ratings per physician in a year. The increase has also not been even across the PRWs, and it would be advisable for the users of PRWs to utilize a number of PRWs to ascertain the rating of any given physician. Further research is needed to identify barriers for patients to rate their physicians and to assist efforts to increase the number of ratings on PRWs to consequently improve the fairness and practical importance of PRWs.

(J Med Internet Res 2017;19(8):e299) doi:10.2196/jmir.6599

KEYWORDS

physician rating websites; patient satisfaction

Introduction

Although the increasing focus on evidence-based medicine and quality improvement has led to much progress, there remains significant unwarranted variation among the medical treatments that are routinely used in practice and deficiencies regarding all of the key aspects of high-quality health care [1-3]. However, potentially because of a lack of publicly available health care quality information, the members of the public are often unaware of such variations and quality differences [4].

Typically grounded in the assumptions of a theoretical consumer choice model [4], public-reporting activities have been developed with the aim of providing quality information about organizations or individuals to the public [5-8]. Public-reporting activities have two key aims: (1) influencing patient decision making by increasing the chance that the patients who obtain information will choose better quality organizations or individuals [4,9] and (2) driving quality improvement by identifying aspects of care needing improvement so that changes can be made in practice [4,9].

One type of public-reporting activity that has been developed in recent decades is physician rating websites (PRWs), which allows patients to anonymously rate, comment, and discuss physicians' quality online as a source of information for others [10-13]. In addition to more than the 30 private PRWs internationally [14,15], an increasing number of public PRWs have been developed by governments and statutory health insurers. For instance, the United Kingdom launched the NHS Choices website in 2007 [16], which has evolved to allow patients to rate both physicians and hospitals, and Germany's largest public health insurer, Allgemeine Ortskrankenkasse (AOK), launched a similar website called Arztnavigator in 2010, which was rolled out nationwide in May 2011 [17].

Medical association representatives, however, have often been strongly opposed to the development of PRWs, referring to them as a "meaningless popularity contest" and expressing concerns that PRWs would be used for "doctor bashing" or defamation [18,19]. For example, the president of the German Medical Association responded in 2009 with regard to the planned introduction of the Arztnavigator by AOK by criticizing the "Marketing Antics" of AOK, describing PRWs as "platforms for denunciation" [19]. Furthermore, a number of shortcomings of PRWs have been identified, including incomplete lists of physicians, low number of physicians rated, and low number of ratings per physician that are overwhelmingly positive, which in turn has raised concerns about the representativeness, validity, and usefulness of information on PRWs [15,20]. Indeed, recent research has indicated that PRWs can influence patient decision making and have an impact on quality improvement [21,22]; however, the ability of PRWs to achieve these goals is somewhat dependent on PRWs having a sufficient number of ratings.

At the beginning of 2010, a study was conducted to examine the evaluation criteria, evaluation tendencies, and utilization of German PRWs not only to allow a factual discussion of the current status quo of PRWs but also to serve as a baseline to document future developments and changes [23]. To examine the frequency of ratings and evaluation tendencies, a random stratified sample of 298 physicians from the physician associations in the states of Hamburg and Thuringia was generated and searched for on 6 German PRWs (Imedo, Jameda, Docinsider, Esando, Medführer, and Topmedic). It was reported that between 75% and 98% of selected physicians could be identified on one of the PRWs; between 3% and 28% of physicians had been rated at least once; the average number of ratings per physician ranged between 1.1 and 3.9; the maximum number of ratings per physician ranged from 3 to 27; and the average converted standardized rating (1=positive, 2=neutral, and 3=negative) ranged from 1.1 to 1.5 [23].

A number of other previous research studies have also examined the frequency of ratings and evaluation tendencies. In terms of the number of physicians rated at least once on PRWs, other previous studies in Germany have reported that between 3% and 26% of a sample of physicians had been rated in 2009 [24], 37% in 2013 [11], and 50% in 2014 [25]. In addition, previous studies conducted in the United States have reported that 16% of physicians were rated on RateMDs between 2005 and 2010 [26], and 27% of a sample of physicians had been rated in 2009 [15]. In terms of the average number of ratings per physician, other previous studies in Germany have reported an average number of ratings per physician of 2.8 in 2013 [11] and 3.1 in 2014 [25]. Research studies conducted in the United States have found a similar average number of ratings per physician: 2.4 [27], 3.2 [26], 2.4 [15], and 2.7 [28]. Finally, in terms of the average rating on PRWs, other previous German studies showed that almost 80% of all ratings on the PRW called "Jameda" were from the two best rating categories in 2013 [11], and 86% of the ratings on the 5 main PRWs were favorable (with 75% assigned to the best rating category and only 5% to the worst category) in 2014 [25]; an analysis of 3000 narrative comments on Jameda also found that 80% of all comments were positive [13]. Studies in the United States have produced similar positive results [15,26,27,29].

To examine the developments in the frequency of ratings and evaluation tendencies on German PRWs, the results of the 2010 study will serve as a baseline for the re-examination of the same 6 German PRWs. In addition, AOK-Arztnavigator was included in this study to assess how it compares with the other PRWs. The objectives of this study were therefore to examine (1) the number of identifiable physicians on German PRWs; (2) the number of rated physicians on German PRWs; (3) the average and maximum number of ratings per physician on German PRWs; (4) the average rating on German PRWs; (5) the website visitor ranking positions of German PRWs; and (6) how these data compare with 2010 results.

Methods

Sample

Following the 2010 study, a random stratified sample of physicians was generated from the physician associations in the German federal states of Hamburg and Thuringia. The state of Hamburg is a major port city in northern Germany and has a total population of 1,787,408 million residents (valid December 31, 2015; [30]) and a total of 15,831 physicians (valid December 31, 2015; [31]). The state of Thuringia lies in east-central

Germany and has a total population of 2,154,816 million residents (1,091,735 million female; [32]) and a total of 12,530 physicians (valid December 31, 2015; [31]).

In October 2014, all physicians working in general medicine, obstetrics and gynecology, urology, and pediatrics were searched for on the websites of the Hamburg and Thuringia physician associations. From each specialty, a random sample was generated for each state, which comprised 50 physicians from general medicine, 33 physicians from obstetrics and gynecology, 33 physicians from pediatrics, and 33 physicians from urology. From the Thuringia physician association, the random sample comprised 50 of 976 general medical physicians, 33 of 289 obstetrics and gynecology physicians, 33 of 183 pediatric physicians, and 33 of 83 urology physicians. Therefore, the sample of 149 physicians selected for the study represented 9.7% of a total of 1531 physicians. From the Hamburg physician association, the random sample comprised 50 of 634 general medical physicians, 33 of 238 obstetrics and gynecology physicians, 33 of 123 pediatric physicians, and 33 of 71 urology physicians. Therefore, the sample of 149 physicians selected for the study represented 14% of a total of 1066 physicians.

The 6 PRWs (Imedo, Jameda, Docinsider, Esando, Medführer, and Topmedic) used in the 2010 examination were again selected to allow comparison. In addition, AOK-Arztnavigator was also included in this study to assess how it compared with the other PRWs. AOK, Germany's largest public health insurer, launched Arztnavigator nationwide in May 2011 after the data collection of the initial study. Selected physicians were therefore searched for on a total of 7 PRWs: Imedo, Jameda, Docinsider, Esando, Medführer, Topmedic, and AOK-Arztnavigator.

Data Collection

Between October and December 2014, every selected physician in the sample was searched for on the 7 PRWs. If a physician could not be found, this was recorded as "not found." If a physician could be found, the physician's rating and the number of ratings (if any) were recorded. On the PRW AOK-Arztnavigator, the results of the ratings are only published if there are at least five ratings. Consequently, data were recorded separately for physicians with more than 5 ratings and physicians with less than 5 ratings.

As the PRWs use different rating scales (percentage, school grade, and stars), the scales were recoded to standardize average ratings (see Table 1; [15,23]). Although recoding the rating scales results in a loss of richness, for reasons of comparability with the 2010 examination, this system was used again. However, to make the variation more transparent, original average ratings have also been listed.

Alexa Internet (www.alexa.com) was once again used to examine visitors to PRWs, compared with other websites. Founded in 1996, Alexa provides commercial Web traffic data and analytics. Traffic estimates are based on data from a global traffic panel and from websites that have chosen to install the Alexa script on their site and certify their metrics. The Alexa global traffic ranking is based on the estimated average of daily unique visitors and its estimated number of page views over the past 3 months relative to all other websites. In addition, Alexa provides a similar country-specific ranking, based on how a website ranks relative to other websites in a particular country over the past month [33]. The 7 PRWs were searched for on Alexa and their Germany-specific ranking recorded. Although AOK-Arztnavigator was not one of the PRWs examined in the first study in terms of frequency of ratings and evaluation tendencies, it was included in the first website visitor ranking table for comparison purposes.

Table 1. Recoding of original rating scales of physician rating websites (PRWs) to standardize ratings scale.

Physician rating websites and original rating scales	Recoding					
Docinsider						
6 star rating	0 (− −)	1	2	3	4	5 (++)
Recoding[a]	3	3	2	2	1	1
Imedo and Esando						
5 star rating	1 (− −)	2	3	4	5 (++)	
Recoding	3	3	2	1	1	
Medführer and AOK-Arztnavigator						
Percent rating	0-33.3		33.3-66.6		66.6-100	
Recoding	3		2		1	
Jameda and Topmedic						
German school grade rating	1 (++)	2	3	4	5	6 (− −)
Recoding	1	1	2	2	3	3

[a]Recoding: 1=positive, 2=neutral, and 3=negative.

Data Analysis

All statistical analyses were conducted using Statistical Package for the Social Sciences (SPSS version 24 for Windows, IBM Corporation). Descriptive statistics included means and standard deviations for continuous variables and percentages for categorical variables. Relative change percentages were included for all variables with data from both studies. Two PRWs (Jameda and Docinsider) offer users two options to provide feedback, which include providing a rating (school grade or stars) or only recommending the physician. The number of these recommendations was assigned to the "number of ratings" and counted toward a positive rating. On the PRW AOK-Arztnavigator, physicians with less than 5 ratings have no published overall rating; the number of these ratings were recorded and counted toward "rated physicians" and "average number of ratings per physician." To analyze differences between the two studies, chi-square tests were used for categorical data and *t* tests for continuously distributed data. The reanalysis of the 2010 data identified a number of minor errors in the results of the published 2010 study. These errors were corrected and data of this study compared with the corrected data rather than the published 2010 data.

Results

Overall results combining both federal states are presented in Table 2. For transparency purposes, the results for each federal state are presented in Multimedia Appendices 1 and 2 (see Multimedia Appendix 1 for Thuringia results; see Multimedia Appendix 2 for Hamburg results).

Identifiable Physicians

The proportion of physicians from the random sample that were able to be identified on the selected PRWs ranged between 65.1% (194/298) on Imedo to 94.6% (282/298) on AOK-Arztnavigator. This represents a decrease from the 2010 study, which ranged between 75.5% (225/298) on Medführer to 98.3% (293/298) on Jameda. Indeed, compared with the 2010 study, the portion of the physicians able to be identified significantly decreased on Imedo (χ^2_1=51 *P*<.001), Jameda (χ^2_1=27.3, *P*<.001), Docinsider (χ^2_1=9.4, *P*=.002), and Esando (χ^2_1=4.5, *P*=.03). However, the decrease of the overall portion of the sample (293/298, 98.3%) that was able to be identified on any of the PRWs compared with the 2010 sample (297/298, 99.7%) was insignificant (χ^2_1=2.7, *P*=.10).

Rated Physicians

The proportion of physicians from the sample that had been rated at least once ranged between 16.4% (49/298) on Esando to 83.2% (248/298) on Jameda. This represents an increase from the 2010 study, which ranged between 3.3% (10/298) on Medführer to 27.8% (83/298) on Imedo. Indeed, compared with the 2010 study, the portion of the physicians that had been rated at least once increased on all PRWs, with the exception of Imedo, and very significantly so on Jameda (χ^2_1=191.4, *P*<.001), Docinsider (χ^2_1=17.8, *P*<.001), Medführer (χ^2_1=239.6, *P*<.001), and Topmedic (χ^2_1=46.1, *P*<.001). The increase of the overall portion of the sample (285/298, 95.6%) that had been rated at least once on any of the PRWs compared with the 2010 study (193/298, 64.7%) was also highly significant (χ^2_1=89.4, *P*<.001).

Average and Maximum Number of Ratings

The average number of ratings per physician ranged between 1.2 (SD 0.5) on Esando to 7.5 (SD 6.7) on AOK-Arztnavigator. This represents an increase from the 2010 study, which ranged between 1.1 (SD 0.3) on Esando and 3.1 (SD 3.5) on Jameda for average number of ratings per physician. Indeed, all PRWs saw an increase in the average number of ratings per physician compared with the 2010 study, although the increase was found to be significant only for Medführer (t_{12}=−10.5, *P*<.001, 95% CI −2.936 to −1.933) and Imedo (t_{153}=−2.1, *P*=.04, 95% CI −0.722 to −0.021). However, the increase of the overall average number of ratings per physicians across all PRWs (5, SD 4.2) compared with the 2010 study (2.3, SD 2.8) was highly significant (t_{476}=−8.4, *P*<.001, 95% CI −3.312 to −2.057). The aggregated average number of ratings per physician on all PRWs was 27.2 ratings, compared with 11.2 in 2010. This represents an average addition of 4 new ratings per physician each year on the German PRWs over 4 years. The maximum number of ratings per physicians ranged from 3 (Esando) to 115 (Docinsider). This represents an increase from the 2010 study, which found that the maximum number of ratings ranged from 2 (Esando) to 27 (Docinsider).

Average Converted Standardized Rating

The average converted standardized rating (1=positive, 2=neutral, and 3=negative) ranged between 1.0 (SD 0.1) on Medführer to 1.2 (SD 0.4) on Jameda and Topmedic. This represents a further improvement toward "very good" from the 2010 study, which found a range between 1.1 (SD 0.4) on Imedo and Jameda to 1.6 (SD 0.7) on Medführer. Although the average converted rating improved on 4 PRWs (Docinsider, Esandoa, Medführer, and Topmedic) compared with the 2010 study, this improvement was significant only for Docinsider (t_{105}=4.0, *P*<.001, 95% CI 0.179-0.538) and Medführer (t_9=2.7, *P*=.03, 95% CI 0.089-1.090). Nevertheless, the improvement of the overall average converted rating across all PRWs (1.1, SD 0.2) compared with the 2010 study (1.2, SD 0.5) was highly significant (t_{255}=3.4, *P*=.001, 95% CI 0.053-0.200).

Website Visitor Ranking Positions

The visitor ranking positions of the selected PRWs in Germany on Alexa indicates that the use of such websites is not common, with only Jameda (position 317) and Medführer (position 9796) being placed among the top 10,000 visited websites in Germany (see Table 3). In comparison, the hotel rating site holidaycheck.de ranking position was 118, with google.de in position 1. Compared with baseline data, only Jameda and Topmedic increased their ranking position, with the rest being visited less.

Table 2. Overall ratings of physicians.

Overall ratings, N (%)=298/2597 (11)	Imedo[a]	Jameda[b]	Docinsider[c]	Esando[a]	Medführer[d]	Topmedic[b]	AOK-Arzt-navigator[d,e]	Overall
Identifiable physicians								
n (%)	194 (65.1)	260 (87.2)	229 (76.8)	234 (78.5)	231 (77.5)	281 (94.3)	282 (94.6)	293 (98.3)
2010 Baseline (%)	267 (89.6)	293 (98.3)	258 (86.6)	254 (85.2)	225 (75.5)	271 (90.9)	N/A	297 (99.7)
Relative change, %	−27	−11	−11	−8	3	4	N/A	−1
Pearson chi-square tests	χ^2_1=51, *P*<.001	χ^2_1=27.3, *P*<.001	χ^2_1=9.4, *P*=.002	χ^2_1=4.5, *P*=.03	χ^2_1=0.3, *P*=.56	χ^2_1=2.5, *P*=.12	N/A	χ^2_1=2.7, *P*=.10
Rated physicians								
n (%)	72 (24.2)	248 (83.2)	119 (39.9)	49 (16.4)	188 (63.1)	101 (33.9)	212 (71.1)	285 (95.6)
2010 Baseline (%)	83 (27.8)	80 (26.8)	72 (24.2)	36 (12.1)	10 (3.4)	32 (10.7)	N/A	193 (64.8)
Relative change, %	−13	210	65	36	1780	216	N/A	48
Pearson chi-square tests	χ^2_1=1.1, *P*=.30	χ^2_1=191.4, *P*<.001	χ^2_1=17.8, *P*<.001	χ^2_1=2.3, *P*=.128	χ^2_1=239.6, *P*<.001	χ^2_1=46.1, *P*<.001	N/A	χ^2_1=89.4, *P*<.001
Average number of ratings per physicians								
Mean (SD)	1.8 (1.1)	6.7 (8.1)	4.8 (12.4)	1.2 (0.5)	3.7 (1.2)	1.7 (1.0)	7.5 (6.7)	5.0 (4.2)
2010 baseline (SD)	1.4 (1.1)	3.2 (3.5)	2.8 (3.6)	1.1 (0.3)	1.3 (0.7)	1.5 (1.0)	N/A	2.3 (2.8)
Relative change, %	29	109	71	9	184	13	N/A	117
t test	t_{153}=−2.1, *P*=.04	t_{197}=−1.1, *P*=.27	t_{190}=−1.3, *P*=.20	t_{83}=−1.0, *P*=.32	t_{12}=−10.5, *P*<.001	t_{134}=−0.9, *P*=.40	N/A	t_{476}=−8.4, *P*<.001
95% CI	−0.722 to 0.021	−4.381 to 1.214	−4.883 to 0.970	−0.264 to 0.086	−2.936 to −1.933	−0.571 to 0.225		−3.312 to −2.057
Maximum number of ratings per physicians								
n	6	67	115	3	6	6	38	N/A
2010 Baseline	7	18	27	2	3	6	N/A	N/A
Relative change, %	−14	272	326	50	100	0	N/A	N/A
Average rating converted[f]								
Mean (SD)	1.1 (0.4)	1.2 (0.4)	1.1 (0.4)	1.1. (0.5)	1.0 (0.1)	1.2. (0.4)	1.1 (0.4)	1.1 (0.2)
2010 baseline (SD)	1.1 (0.4)	1.1 (0.4)	1.5 (0.7)	1.2 (0.5)	1.6 (0.7)	1.3 (0.5)		1.2 (0.5)
Relative change, %	0	9	−27	−8	−38	−8	N/A	−8
t test	t_{153}=−0.5, *P*=.65	t_{325}=−0.6, *P*=.53	t_{105}=4.0, *P*<.001	t_{84}=0.3, *P*=.80	t_9=3.0, *P*=.03).	t_{43}=1.3, *P*=.21	N/A	t_{255}=3.4, *P*=.001
95% CI	−0.152 to 0.095	−0.141 to 0.072	0.179 to 0.538	−0.180 to 0.234	0.089 to 1.090	−0.076 to 0.331	N/A	0.053 to 0.200
Average rating original (SD)	4.2 (0.7)	1.8 (1.0)	4.6 (0.9)	4.6 (0.9)	72 (6.2)	1.6 (0.9)	88 (15.1)	N/A

[a]1 to 5 star: 1 star worst rating, 5 stars best rating.

[b]School grade: 6 worst rating, 1 best rating.

[c]0 to 5 star: 0 star worst rating, 5 stars best rating.

[d]Percentage.

[e]No baseline data are given for AOK-Arztnavigator because it was not included in the first study.

[f]Recoding: 1=positive, 2=neutral, and 3=negative.

Table 3. Website visitor ranking positions.

Ranking	Imedo	Jameda	Docinsider	Esando	Topmedic	Medführer	AOK-Arztnavigator
Current ranking[a]	14,624	317	16,360	77,669	209,119	9796	52,925
Baseline ranking[b]	1472	1128	3073	8340	273,403	8340	38,407[c]
Relative change, %	893	−72	432	831	−24	18	38

[a]The ranking relates to Germany as on January 11, 2016.

[b]Values from first study as on March 7, 2011.

[c]Although AOK-Arztnavigator was not one of the PRWs examined in the first study, it was included in the website visitor ranking table for comparison purposes.

Discussion

This update of the frequency of ratings and evaluation tendencies of German PRWs has resulted in two key findings. First, although there has been an overall increase in the average number per physician of ratings on German PRWs, this increase has not been even across the PRWs. Second, the average rating of physicians has shown further improvement toward "very good."

Number of Ratings

It is generally assumed that PRWs will only be helpful for users, and fair for those who are rated, if there are a high number of ratings [15,20]. The overall increase in the number of ratings on German PRWs since 2010, both in terms of the number of rated physicians and the average number of ratings per physician, is therefore a positive development and one that is consistent with previous studies in Germany.

In terms of the number of physicians rated at least once, between 16.4% (49/298) and 83.2% (248/298) of the sample had been rated at least once, compared with between 3.3% (10/298) and 27.8% (83/298) in 2010. Other previous German studies have reported that between 3% and 26% of physicians had been rated at least once in 2009 [24], 37% in 2012 [11], and 50% in 2014 [25]. Although it is difficult to directly compare these figures, given the different sampling and time frames used, they do suggest an upward trend and are generally higher than those reported internationally [15,26,34]. All PRWs in our study, except for Imedo, saw an increase in the proportion of physicians rated at least once. However, the increase in the proportion of rated physicians was not even across the PRWs, with Jameda (248/298, 83.2%), AOK-Arztnavigator (212/298, 71.1%), and Medführer (188/298, 63.1%) having more rated physicians compared with Docinsider (119/298, 40%), Topmedic (101/298, 33.9%), Imedo (72/298, 24.2%), and Esando (49/298, 16.4%). Furthermore, the overall proportion of the sample that had been rated at least once on any of the PRWs increased to 95.6% (285/298) from 64.8% (193/298) in 2010.

Similarly, in terms of the average number of ratings per physicians on German PRWs, physicians had an average number of ratings between 1.2 and 7.5, compared with 1.1 to 3.1 in 2010. Other previous German studies have reported average number of ratings per physicians of 2.4 in 2013 [11] and 3.1 in 2014 [25]. Research in the United States have found similar average number of ratings per physician on PRWs [15,26-28,34]. Whereas all PRWs in our study saw an increase in the average number of ratings per physician, this increase was not even across PRWs, with AOK-Arztnavigator (average 7.5 ratings), Jameda (average 6.7 ratings), Docinsider (average 4.8 ratings), and Medführer (average 3.7 ratings) having on average more ratings per physician than Imedo (average 1.8 ratings), Topmedic (average 1.7 ratings), and Esando (average 1.2 ratings).

It appears, therefore, that there is a need to differentiate German PRWs. Whereas Jameda was slightly ahead of others German PRWs in terms of the number of ratings in 2010, the field was reasonably equally subdivided between different PRWs. However, in the subsequent 4 years, there has been a development with Jameda and the new AOK-Arztnavigator in particular, highlighting an increase in ratings more than the other PRWs. It remains to be seen whether the other PRWs will be able to increase their number of ratings in the future. However, it is noticeable how quickly AOK, Germany's largest public health insurer, has been able to establish AOK-Arztnavigator as one of the most used German PRWs since being introduced nationwide in May 2011. Two other large public health insurers, Techniker Krankenkasse (TK) and BARMER GEK, have also subsequently developed their own PRWs (TK-Ärzteführer and BARMER GEK-Arztnavi). AOK, TK, and BARMER GEK all utilize a central database known as "Weisse Liste," recruiting ratings from their insurees via their own platforms but pooling these ratings on the shared Weisse Liste. So, if a patient rates a physician on AOK-Arztnavigator, this rating will also appear on TK-Ärzteführer. Future updates are needed to assess whether this practice may allow the public health insurers to take a bigger share of the PRW ratings away from their smaller private competitors.

Whereas the overall increase in the number of ratings on German PRWs suggests that the practical importance of PRWs is increasing, the relatively low number of physician ratings indicates that PRWs are still used very little in Germany for posting ratings on current physicians. Despite the focus on informed and autonomous patients and the relatively high use of comparative quality information concerning other consumer services and products [35], the German public seem to be rather reluctant in contributing to comparative quality information on health providers.

However, currently there is limited research examining the reasons why patients are not rating their physicians on PRWs, and more research is needed regarding this issue. A recently published study by Patel et al [36] explored patients' views

regarding rating general practitioners on PRWs, within the context of other feedback methods available in England. Participants reported that they would not leave feedback on PRWs because of accessibility issues, privacy and security concerns, and because they felt that feedback left on a website may be ignored [36]. Hanauer et al [37] also asked participants in their 2012 US study to consider the implications of leaving negative comments about a physician. Participants reported being concerned that their identity could be disclosed (34%), and that the physicians may take action against them for leaving negative comments (26%) [37].

Average Rating

Whereas physician representatives were concerned before the implementation of PRWs that they would be primarily used for "doctor bashing" or defamation [18,19], these fears have proved to be unfounded. The average rating of physicians has further improved toward "very good," with the average converted standardized rating (1=positive, 2=neutral, and 3=negative) ranging from 1.0 to 1.2, compared with 1.1 to 1.6 in 2010. Other previous research has also found that the majority of ratings are overwhelmingly positive. In Germany, 86% of the ratings of the 5 main German PRWs were favorable in 2014 [25], whereas an analysis of 3000 narrative comments on Jameda from 2014 also found that 80% of all comments were positive [13]. Studies in the United States have produced similar positive results [15,26,27,29,34]. Such overwhelmingly positive ratings, however, raise concerns about the representativeness, validity, and usefulness of PRWs [15,20].

Whereas some form of trust is essential in all social relationships, it is particularly important when one finds themselves dependent on others for their well-being. Indeed, the need for trust is arguably greater in the health care setting than many other areas of life because of the ineradicable imbalances of power, knowledge, and vulnerability found there [38]. Given their position in society, physicians are the recipients of not only public trust but also of a close interpersonal trust by patients, who enter into the physician-patient relationship with the expectation that physicians will act competently and dutifully. Patients' willingness to disclose information about such a relationship is likely to be extremely low unless their expectations are far exceeded, or they feel that their trust has been violated in some way. Research concerning the rating of products on Amazon has reported such a "bimodal" trend, with "amateur reviewers" (those who review only occasionally) typically contributing a review only because of a strong reaction to a product either because they love it or hate it, and that for some, doing so is an almost a cathartic experience [39]. One would expect to see a similar trend on PRWs, and further research would be helpful to better understand why there are not more negative experiences reported on PRWs.

It is clear, however, that at least in some countries, the lack of negative reviews on PRWs is partly because of strict data protection laws and legal responses taken to such reviews by physicians and their advocacy groups [40]. Whereas most businesses (particularly small businesses such as physicians) are concerned about negative reviews and the impact these might have on their reputation, physicians have been particularly opposed to reviews of their services and often take negative reviews more personally than other business owners [41]. It has been argued that "medical narcissism" is a key reason that physicians find it so difficult to acknowledge and disclose medical errors, as such a disclosure can be too much of a challenge to their self-image of competence, control, and treatment-oriented focus [42]. A similar response may be a contributing factor behind many physicians' unwillingness to accept negative reviews on PRWs and their efforts to prevent transparency of patient experiences and satisfaction with their performance. Critical reviews on PRWs, however, are a (usually anonymous) type of "patient complaints," which are seen by many as an opportunity to learn and improve care [43]. Whereas there is evidence that some physicians do use reviews on PRWs to improve care [22], this opportunity is likely to be limited while patients are being encouraged not to post negative reviews on PRWs [12], and the negative reviews that are posted are legally challenged.

Limitations

This study has a number of limitations that should be taken into account when interpreting the results. First, the selection of German PRWs was not exhaustive; consequently, some PRWs that have gained importance since the 2010 study may not have been taken into account. Second, the fact that the sample was only taken from 2 states in Germany limits the generalizability of the results. Results in Thuringia and Hamburg, however, were very similar, and we have no reason to suspect other states in Germany would be significantly different. Third, the development of the frequency of ratings and evaluation tendencies is not longitudinal, as the same sample of physicians was not used in both studies. Fourth, as the PRWs or Alexa.com were not webcited in either study when data were collected, this prevents the results from being reproduced. Finally, it was not controlled for the time frame in which ratings were allowed to be published.

Conclusions

This update of the frequency of ratings and evaluation tendencies of German PRWs indicates that there has been an overall increase in the number of ratings on German PRWs, both in terms of the number of rated physicians and the average number of ratings per physician. This is a positive development and suggests that the practical importance of German PRWs is increasing. However, the overall average number of ratings per physician of all PRWs represents an average addition of only 4 new ratings per physician each year over the 4 years, which indicates that PRWs are still used very little in Germany for posting ratings on current physicians. However, without a higher number of ratings, the PRWs will continue to have a limited value. Further research is needed to identify barriers for patients to rate their physicians and to assist efforts to increase the number of ratings on PRWs, thereby improving the fairness and practical importance of PRWs. The increase in the number of ratings has also not been even across the PRWs. Given that physicians' ratings are currently spread out across PRWs in an uneven manner, it would be advisable for users of PRWs to utilize a number of PRWs when searching for a new physician. The implementation of a website using "meta-crawling" to pool

physicians' ratings across all PRWs may also be a helpful addition to the field to allow users to easily see all of a physician's ratings in one place. Future updates are also needed to assess whether the practice of using a central database may allow the public health insurers to take an even bigger share of the PRW ratings away from their smaller private competitors. However, if these smaller PRWs are unable to significantly increase their number of ratings in the future, consideration should be given to whether their continued existence in the German PRWs market is providing value or is, in fact, causing harm. Finally, the continued overwhelmingly positive ratings on German RWs have not allayed fears regarding the representativeness and validity of PRWs. Further research would be helpful to better understand why there are not more negative experiences reported on PRWs. Additionally, the medical profession itself should do more to ensure that patients are not being actively discouraged by physicians to post critical reviews, as they are a potentially important opportunity for physicians to learn and improve care.

Acknowledgments

The authors would like to thank Dr Susanne Wieschowski for her assistance with running statistical tests.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Ratings of physicians in Thuringia.

[PDF File (Adobe PDF File), 29KB - jmir_v19i8e299_app1.pdf]

Multimedia Appendix 2

Ratings of physicians in Hamburg.

[PDF File (Adobe PDF File), 29KB - jmir_v19i8e299_app2.pdf]

References

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press; 2001.
2. Institute of Medicine. Best care at lower cost: The path to continuously learning health care in America. Washington, DC: The National Academies Press; 2013.
3. Wennberg JE. Forty years of unwarranted variation--and still counting. Health Policy 2014 Jan;114(1):1-2. [doi: 10.1016/j.healthpol.2013.11.010] [Medline: 24373550]
4. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? Medical Care 2009;47(1):1-8. [doi: 10.1097/MLR.0b013e3181808bb5]
5. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? a review of the evidence. JAMA 2000;283(14):1866-1874. [Medline: 10770149]
6. Fung CH, Lim Y, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med 2008;148(2):111-123. [Medline: 18195336]
7. Berger Z, Joy S, Hutfless S, Bridges J. Can public reporting impact patient outcomes and disparities? a systematic review. Patient Educ Couns 2013;93(3):480-487. [doi: 10.1016/j.pec.2013.03.003] [Medline: 23579038]
8. Mukamel DB, Haeder SF, Weimer DL. Top-down and bottom-up approaches to health care quality: the impacts of regulation and report cards. Annu Rev Public Health 2014;35:477-497. [doi: 10.1146/annurev-publhealth-082313-115826] [Medline: 24159921]
9. 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]
10. Emmert M, Sander U, Esslinger A, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51:20. [doi: 10.3414/ME11-01-0045]
11. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013;15(8):e157. [doi: 10.2196/jmir.2655] [Medline: 23919987]
12. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24. [doi: 10.2196/jmir.2360] [Medline: 23372115]
13. Emmert M, Meierb F, Heidera A, Dürra 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;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]

14. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
15. Lagu T, Hannon N, Rothberg M, Lindenauer P. Patients´ evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010;25(9):942-946. [Medline: 20464523]
16. National Health Service. NHS. 2017. NHS choices URL: http://www.nhs.uk/pages/home.aspx [WebCite Cache ID 6oglxkLXV]
17. Weisse-liste. 2017. AOK- Arztnavigator URL: https://weisse-liste.arzt-versichertenbefragung.aok-arztnavi.de/ [accessed 2017-03-03] [WebCite Cache ID 6ogmSA7jx]
18. Carvel J. The Guardian. 2008. Patients to rate and review their GPs on NHS website URL: http://www.theguardian.com/society/2008/dec/30/doctors-rating-website-nhs [WebCite Cache ID 6ogmjtvkj]
19. Aekv-ebersberg. 2009. Ärztlicher Kreisverband Ebersberg Körperschaft des öffentlichen Rechts URL: https://www.aekv-ebersberg.de/aktuelles/105-baek-newsletter-2009-004-vom-25062009 [accessed 2017-03-03] [WebCite Cache ID 6ogmznsHJ]
20. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012;27(6):685-692. [Medline: 22215270]
21. 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]
22. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? results from an online-based cross-sectional study. J Med Internet Res 2016;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]
23. Strech D, Reimann S. Deutschsprachige Arztbewertungsportale: Der Status quo ihrer Bewertungskriterien, Bewertungstendenzen und Nutzung. Gesundheitswesen 2012;74:e61-e67. [doi: 10.1055/s-0031-1297254]
24. Emmert M, Maryschok M, Eisenreich S, Schöffski O. Arzt-Bewertungsportale im Internet – Geeignet zur Identifikation guter Arztpraxen? Gesundheitswesen 2009;71(4):e18-e27. [doi: 10.1055/s-0028-1103288]
25. Emmert M, Gerstner B, Sander U, Wambach V. Eine Bestandsaufnahme von Bewertungen auf Arztbewertungsportalen am Beispiel des Nürnberger Gesundheitsnetzes Qualität und Effizienz (QuE). Gesundh ökon Qual manag 2014;19(4):161-167. [doi: 10.1055/s-0033-1335899]
26. Gao G, McCullough J, Agarwal R, Jha A. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38. [doi: 10.2196/jmir.2003] [Medline: 22366336]
27. Ellimoottil C, Hart A, Greco K, Quek M, Farooq A. Online reviews of 500 urologists. J Urol 2013;189(6):2269-2273. [Medline: 23228385]
28. 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]
29. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
30. Statistisches Amt für Hamburg und Schleswig-Holstein. Statistik-nord. 2016. Bevölkerung in Hamburg 2015 URL: http://www.statistik-nord.de/fileadmin/Dokumente/Presseinformationen/SI16_131.pdf [WebCite Cache ID 6oieX961o]
31. Bundesaerztekammer. Bundesaerztekammer. 2015. Ärztinnen/Ärzte nach Landesärztekammern und Tätigkeitsarten URL: http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/Statistik2015/Stat15AbbTab.pdf [accessed 2017-03-04] [WebCite Cache ID 6oifDuiDR]
32. Thüringer Landesamt für Statistik. Statistik.thueringen. 2015. Bevölkerung der Gemeinden, erfüllenden Gemeinden und Verwaltungsgemeinschaften am 30 URL: http://www.statistik.thueringen.de/datenbank/TabAnzeige.asp?tabelle=gg000201 [accessed 2017-03-04] [WebCite Cache ID 6oifValjR]
33. Alexa. 2017. URL: http://www.alexa.com/about [accessed 2017-03-04] [WebCite Cache ID 6oifgfKhW]
34. Gray BM, Vandergrift JL, Gao GG, McCullough JS, Lipner RS. Website ratings of physicians and their quality of care. JAMA Intern Med 2015;175(2):291-293. [doi: 10.1001/jamainternmed.2014.6291] [Medline: 25437252]
35. Zhang KZ, Cheung CM, Lee MK. Examining the moderating effect of inconsistent reviews and its gender differences on consumers' online shopping decision. Int J Inf Manage 2014;34:89-98. [doi: 10.1016/j.ijinfomgt.2013.12.001]
36. 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;17(12):e276 [FREE Full text] [doi: 10.2196/jmir.4989] [Medline: 26681299]
37. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
38. McLennan S, Gillett G, Celi LA. Healer, heal thyself: health care workers and the influenza vaccination. Am J Infect Control 2008;36(1):1-4. [doi: 10.1016/j.ajic.2007.07.010] [Medline: 18241729]

39. Gilbert E, Karahalios K. Understanding Deja Reviewers. 2010 Presented at: Proceedings of the 2010 ACM conference on Computer supported cooperative work; February 06-10, 2010; Savannah, GA.
40. Atasoy S. SRF. Patienten benoten Schweizer Ärzte Montag URL: http://www.srf.ch/news/schweiz/patienten-benoten-schweizer-aerzte [accessed 2017-03-04] [WebCite Cache ID 6oieCb6cw]
41. Goldman E. Forbes.: Forbes; 2013. How doctors should respond to negative online reviews URL: https://www.forbes.com/sites/ericgoldman/2013/11/21/how-doctors-should-respond-to-negative-online-reviews/#2762a644a0e7 [WebCite Cache ID 6oldeS4cE]
42. Banja J. Medical Errors and Medical Narcissism. Boston: Jones and Bartlett Publishers; 2005.
43. Reader T, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. BMJ Qual Saf 2014;23(8):678-689. [doi: 10.1136/bmjqs-2013-002437]

Abbreviations

AOK: Allgemeine Ortskrankenkasse
PRW: physician rating website
TK: Techniker Krankenkasse

Edited by G Eysenbach; submitted 01.09.16; peer-reviewed by G Gao, S Bidmon, F Rothenfluh, M Emmert, M Lorenz; comments to author 31.12.16; revised version received 03.04.17; accepted 23.06.17; published 25.08.17.

Please cite as:
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;19(8):e299
URL: http://www.jmir.org/2017/8/e299/
doi:10.2196/jmir.6599
PMID:28842391

Original Paper

Web-Based Medical Service: Technology Attractiveness, Medical Creditability, Information Source, and Behavior Intention

Shan Huei Wang[1], PhD

Department and Graduate Institute of Business Administration, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Shan Huei Wang, PhD
Department and Graduate Institute of Business Administration
National Taiwan University
9th Floor, Building I, College of Management
No. 85, Section 4, Roosevelt Road
Taipei,
Taiwan
Phone: 886 2 33661058
Fax: 886 2 23625379
Email: d98741001@ntu.edu.tw

Abstract

Background: Web-based medical service (WBMS), a cooperative relationship between medical service and Internet technology, has been called one of the most innovative services of the 21st century. However, its business promotion and implementation in the medical industry have neither been expected nor executed. Few studies have explored this phenomenon from the viewpoint of inexperienced patients.

Objective: The primary goal of this study was to explore whether technology attractiveness, medical creditability, and diversified medical information sources could increase users' behavior intention.

Methods: This study explored the effectiveness of web-based medical service by using three situations to manipulate sources of medical information. A total of 150 questionnaires were collected from people who had never used WBMS before. Hierarchical regression was used to examine the mediation and moderated-mediation effects.

Results: Perceived ease of use (P=.002) and perceived usefulness (P=.001) significantly enhance behavior intentions. Medical credibility is a mediator (P=.03), but the relationship does not significantly differ under diverse manipulative information channels (P=.39).

Conclusions: Medical credibility could explain the extra variation between technology attractiveness and behavior intention, but not significant under different moderating effect of medical information sources.

(J Med Internet Res 2017;19(8):e285) doi:10.2196/jmir.8114

KEYWORDS

web-based medical service; technology attractiveness; medical creditability; information source; behavior intention

Introduction

With the increase in the aging population and rising Medicare expenditures, medical circles [1-3] and management practitioners are debating how people can use technology and the Internet to improve health services [4,5]. Web-based medical services (WBMS) such as eHealth, mHealth, and telehealth seem feasible solutions [6,7]. WBMS is considered as one of the most innovative services in medical technology in the 21stcentury [8]; however, promoting these services is still a challenge [5]. There are 3 impediments: perceived usefulness, behavior change, and medical law limitation.

Most WBMS focus on patients or potential patients with chronic conditions such as hypertension and diabetes. It is difficult for patients who do not use WBMS to see their value. The goal of WBMS is "prevention is better than cure." That means WBMS needs to take a long time to show its effectiveness before serious symptoms occur. In addition, most patients are accustomed to face-to-face clinic services. Older patients in particular do not want to communicate with health care via the Internet. Finally, medical advertisement and doctor-endorsement on public are

illegal for hospitals and related institutions, which makes word of mouth and noncommercial referrals by medical workers play an important role in WBMS.

A large body of literature has explored the relationship between the technology acceptance model and behavior intention in medical fields [2,7,9-13]. Results showed that perceived usefulness (PU) and perceived ease of use (PEOU) have positive impact on behavior intention. Other studies have investigated these relationships from the viewpoint of physicians and nurses [4,14-17]. Unlike other medical services, WBMS incorporates the interaction among medical workers and patients into the Web-based technology system to deliver medical services [18,19]. It is a patient-centered service designed to improve access to care for patients who need long term care or who have limited mobility [20], however, exploring from patient's viewpoint is still a neglected area.

Earlier studies have explored antecedent factors [14,16,21], and moderating factors which affect technology attractiveness, in terms of PU, PEOU, and behavior intention [22]. Previous studies claim that medical credibility plays an important role in WBMS [20,23]; however, few studies explore the mediating effects of medical credibility on this relationship. In addition, under the constraints of medical law, different sources of medical information may result in different moderating effects among technology attractiveness, medical credibility and behavior intentions on WBMS [24,25]. The purpose of this study was, therefore, to explore the relationship among technology attractiveness, medical credibility, information source, and behavior intention on WBMS service. The objectives of this study were to (1) examine the technology attractiveness of WBMS and behavior intention from inexperienced patients' view point, (2) examine the mediating effect of medical credibility, and (3) explore the moderated-mediation effect of different medical information sources.

This study begins with a review of the literature on WBMS, technology attractiveness, medical credibility, and medical information sources. Based on the literature review, this study formulates the hypotheses, describes the methods and sample, and presents the results. Finally, the theoretical contribution, managerial implications, and future directions for research are discussed.

Literature Review and Hypotheses

Web-Based Medical Service Model

Web-based medical service (WBMS) is defined broadly as the use of Information and Communication Technologies (ICT) to provide medical information and services, including telephone intervention, medical education, and timely medical consultations [6]. It is called eHealth and mHealth (mobile health) in the Journal of Medical Internet Research (JMIR) and tele-health in other journals such as New England Journal Medicine (NEJM) and the Journal of the American Medical Association (JAMA). It is a patient-centered and online to offline service that provides patients with timely medical consultations at home and is a self-management medical service [1,26]. It is also a promising strategy for improving heart failure outcomes by monitoring patients remotely; therefore, physicians can intervene early if there is evidence of clinical deterioration [27].

The service process of WBMS is as follows: a patient participating in the service is given a set of specifically designed equipment, including a mini-personal computer or mobile phone and several devices to measure physiological indicators, which are installed at his or her home. Patients are requested to measure several physiological indicators (eg, blood pressure, blood sugar, ECG, and SPO2), and these indicators are automatically uploaded to the database of the hospital via the Internet. If some of the indicators are outside the acceptable ranges set by physicians according to each patient's physical condition, nurses will then check relevant historical indicators and provide appropriate medical education and consultations immediately. If the patient's condition is not improving, then an early doctor visit or emergency services may be necessary.

Technology Attractiveness and Behavior Intention

In this study, technology attractiveness is defined as the degree of users' acceptance of new technology, in terms of PU and PEOU of WBMS. PU is referred as the degree to which a person believes that using a particular system would enhance his or her job performance, and PEOU is referred as the degree to which a person believes that using a particular system would be free of effort [7,9].

A number of studies have investigated the relationship between technology attractiveness and behavior intention of medical workers [4]. One stream focuses on physician's intention [4,14,16,17]; the other focuses on nurse's intention [14]. Wu et al [17] find that when making a decision to accept or reject a telemedicine technology, physicians appear to be fairly pragmatic, concentrating on the technology's usefulness rather than on its ease of use. Similarly, Ammenwerth et al [14] introduced a computer-based nursing documentation system and systematically evaluated its preconditions and effects in a pretest-posttest intervention study. The results showed that the amount of nurse's self-confidence when using computer is important before technology system implementation stage, but after implementation stage, the fit between nursing workflow and the functionality of the system were relatively more important. In sum, prior studies all support that PU and PEOU of computer and Internet access at workplace would be helpful in increasing medical workers' behavior intention [15,17]. However, unlike other medical service, WBMS is a 3-sided cooperative model, including health care worker, patient, and technology. Patients' acceptance will affect the success of WBMS in the early stages, but surveys from inexperienced patients' viewpoints are still scarce [2,26,28]. According to the above reasons, this study proposes following hypothesis:

> Hypothesis 1: There is a positive relationship between technology attractiveness and potential patient behavior intention.

The Mediation Effect of Medical Credibility

Previous studies focused on the antecedent factors of intention behavior (see the review article by Or and Karsh [21]), and moderating effects on technology attractiveness and behavior intention [22]. Or et al [29] summarize that technology

attractiveness, subjective norm, and healthcare knowledge can predict most of the variance in patients' acceptance of web-based self-management technology. Sun and Zhang [22] find that organizational, technological, and individual factors can moderate the relationship between technology attractiveness and behavior intention; however, few studies have examined the mediating effects on this relationship [9].

Ajzen and Fishbein [30] claim that a person who carries out a behavior intention is affected by subjective norm and attitude towards the behavior. However, attitudes do not fully mediate the relationship between technology attractiveness, in terms of PU and PEOU, on behavior intention [9]. That means that more significant factors such as medical credibility can mediate this relationship. However, studies have seldom noted this relationship. Medical credibility is defined as the willingness of patients to be vulnerable to the actions of medical service based on the expectation that the medical workers will perform a particular and trusted medical treatment important to them [31,32].

Unlike traditional medical service, WBMS requires the simultaneous cooperation of medical workers, patients, and the web-based technology system to deliver medical services [18,19]. However, most elderly patients are accustomed to face-to-face clinic services and they do not like communication over the Internet. McGrail et al [20] found that older patients are more likely to see a known medical provider, and patients from the lowest socioeconomic strata were the least likely to use a virtual medical service. Hence, earning patients' trust and strengthening the medical creditability of WBMS in the service delivery process is important.

There are 2 stages in the medical delivery process of WBMS. In the first stage, patients encounter the technology and perceive its potential function, convenience, and fairness [33]. After the patients have used the technology for a while, the medical workers will provide a monthly report that documents each day's physiological indicators, medical education, and suggestions offered by nurses and physicians. In the second stage, patients will evaluate the credibility of medical reports and the whole medical service delivery process of WBMS [28]. An earlier study supports that when patients have positive perception on medical credibility and trust this service, they are more willing to recommend it [34]. Hence, this study argues that a successful business model of WBMS needs both patients' confidence in the web-based technology system and their trust in medical service they contact via the Internet every day.

> Hypothesis 2: A medical creditability plays a mediator role between technology attractiveness and potential patient's behavior intention.

The Moderated Role of Medical Information Sources

A person who engages in a particular behavior or produces an intention is affected by his or her subjective norm and attitude towards the behavior. Subjective norm is defined as the way in which individuals decide to take a behavioral intention when facing social pressure to accept the opinion, which may come from family, friends, supervisors, and colleagues [30]. Fishbein and Ajzen [18] point out that individual behavior, such as assessing a product or service, is subject to the amount of information the person has obtained. In addition, the key person (or information source) has an important role in an individual's evaluation of information [35]. For example, studies have found that the way in which consumers choose and evaluate information on medicines can influence their attitudes to and use of those medicines [23,25]. Peterson et al [25] interviewed people who researched medicines online. Some of those participants preferred information provided by pharmaceutical companies, others preferred information from governments, organizations, and medical schools, and still others preferred word of mouth from other consumers who had already taken that medication. Koufopoulos et al [24] found that communities can have positive treatment outcomes, particularly for the management of chronic illness. Some researchers have concluded that patients trust physicians more than they trust online medical information, which they consult before talking with their physicians [23,36,37]. Accordingly, this study proposes that different information sources have different effects on the relationship among technology attractiveness, patients' trust in medical credibility, and behavior intention. This study investigates whether information sources have a moderated-mediation effect by word of mouth of friends, recommendations from medical workers, and advertisements from a nonofficial medical company.

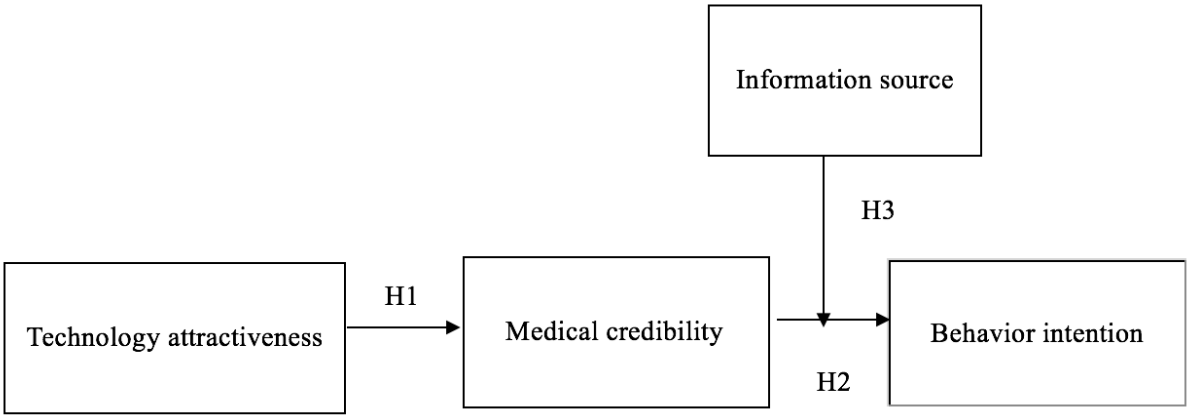
> Hypothesis 3: Medical information sources have a moderated-mediation role between technology attractiveness and potential patient's behavior intention.

Methods

Conceptual Framework

In order to explore (1) technology attractiveness of WBMS from potential patients' viewpoint, (2) the mediate effects of medical credibility, and (3) the moderated-mediation effect of medical information sources, the framework of this study was constructed as shown in Figure 1.

Figure 1. The framework of this study.



Sample Selection and Design

WBMS had been carried out for several years in Taiwan, but service providers wanted to know why it is still unfamiliar to and rejected by most of the population. This study collected 213 samples from several courses in Taipei; however, unlike previous studies, the purpose of this study was to investigate whether potential patients were willing to use or recommend WBMS to other potential patients. For this reason, the target sample had to meet the following criteria: (1) the participants had not used or heard of WBMS before the investigation; (2) they, their families, and friends were at high risk for high blood pressure, diabetes, and cardiovascular diseases; and (3) they had basic computer knowledge and Internet experience. Surveys with missing data were excluded, as were participants who did not meet the 3 criteria. Only 150 questionnaires were used in this study, 98 (65.3%) of which were completed by females and 52 (34.7%) were completed by males. The participants ranged in age from 26 to 50 years (mean 35). All participants used the Internet on an average of 3 to 4 times a week.

To examine the effect of different information source, participants were randomly divided into 3 groups. Group 1 (n=50) manipulated information source from friends or family, group 2 (n=50) had an information source from nonofficial medical company, and group 3 (n=50) had information from an official medical source, such as a physician. Randomization was done by a random number generator [38]. In addition, all participants were asked to log into the eHealth website [39] that explained how to operate the WBMS instrument and the benefits of WBMS in Figure 2. This study conformed to the consolidated standards of reporting trials eHealth checklist (Multimedia Appendix 1) [40] and CHERRIES checklist to improve the quality of questionnaires checklist (Multimedia Appendix 2) [41].

Figure 2. Web-based medical service.

Measures

To explore the intention of potential patients when they used the different information source, this study combined experimental design and questionnaires to collect data through 3 kinds of different situations. It was not possible to find a perfectly suited questionnaire to measure our hypotheses, hence appropriate amendments mainly based on past scholars' questionnaires were made. Furthermore, the adapted scales were originally in English, so previously established protocols of back translation were followed [42] to ensure that the translated Chinese questionnaire had similar meanings. Only 2 people who were bilingual in English and Chinese were hired, with one translating the English questionnaire into Chinese and the other cross-translating the items back into English. The researchers and the bilingual persons resolved any semantic inconsistencies and made minor adjustments in word choice before the survey was implemented. This study referred to Davis [9] questionnaire for testing technology attractiveness and behavior intention and followed Ajzen [35] and Gefen et al's [19] questionnaires for the items of medical credibility. Participants rated all items on a 5-point Likert scale ranging from 1 (''strongly disagree'') to 5 (''strongly agree''). More details are shown in Multimedia Appendix 3.

Data Analyses

This study used LISREL 8.54 (Scientific Software International, Inc. Skokie, IL, USA) to assess the convergent and discriminant validity of the constructs with a confirmatory factor analysis (CFA), and adopted maximum likelihood to rotate the constructs [43]. This study used multiple indexes to assess model fit: chi-square test (χ^2), root mean of squared error of approximation (RMSEA), and its confidence interval, comparative fit index (CFI), incremental fit index (IFI), and goodness-of-fit index (GFI) [44]. In addition, scholars pointed out that the RMSEA higher than 1.0 meant poor model fit, between 0.05 and 0.08 meant accepted range, and lower than 0.05 meant good model fit [44]. CFI, GFI, and IFI were between 0 and 1, and the minimum acceptable value of CFI, GFI and IFI were 0.9 [45]. This study also used SPSS version 10 (SPSS Inc, Chicago, IL, USA) statistical software to examine the hypotheses. To test Hypotheses 1 and 2, this study followed Baron and Kenny [46] procedures for the partial mediation test. A total of 3 conditions were applied: (1) a significant effect of technology attractiveness on patients' behavior intention, (2) a significant effect of technology attractiveness on medical credibility, and (3) a significant effect of medical credibility on potential patients' intentions when model considered technology attractiveness.

To test the moderated-mediation effects in the model (Hypothesis 3), this study followed 3 condition steps recommended by Muller et al [47], (1) a significant effect of technology attractiveness on patients' behavior intention, (2) a significant effect of technology attractiveness on medical credibility and a significant interaction between technology attractiveness and medical information source, and (3) a significant interaction between technology attractiveness and medical information source on potential patients' behavior intention and a significant interaction between medical credibility and information sources.

Results

Table 1 shows the means, standard deviations, and correlations of all dimensions. The correlations of 4 dimensions are highly relevant to each other, hence, this study used variance inflation factor (VIF) to examine the problem. None of the VIF was larger than 0.5, which means that there is no multicollinearity problem in this study [48].

Reliability and Validity

To make sure the questionnaire items reflect the true meaning of constructs, a series of reliability and validity tests were conducted. This study used CFA to examine convergent validity, and 2 items were deleted in this study. The overall model fit of the measurement model was examined from several fit indexes: chi-square test was significant (χ^2_{51}=89.6, P<.001), RMSEA was .07, GFI was .91, CFI was .93, adjusted goodness-of-fit (AGFI) was .857, and all indicators ensured good model fit in the measurement [49-52]. Furthermore, the average variance extracted [20] of each construct was larger than 0.5, a good convergent validity in the measurement model [43,53]. Finally, Fornell and Larcker [53] claimed that a good discriminant validity was evident when the Average Variance Extracted (AVE) was larger than the square root of the construct correlation (Φ2). All correlations in Table 2 were significant (P<.001). All together, these results provided evidences for the convergent and discriminant validity of the proposed model. In addition, the range of composite reliability [23] was 0.79, indicating good internal reliability of the questionnaire items [54].

Table 1. Means, standard deviations, and intercorrelations of all dimensions. All correlation values are significant.

Variables	1	2	3	4	5	Means	Standard deviation
Perceived usefulness	1					3.91	0.82
Perceived ease to use	.47	1				3.81	0.74
Medical credibility	.47	.37	1			3.92	0.63
Willing to use	.34	.31	.37	1		3.84	0.95
Recommend to other person	.35	.34	.36	.55	1	3.90	0.74

Table 2. Confirmatory factor analysis standardized loading.

Variables	Standardized loading[a]	Average variance extracted	Composite reliability
Perceived usefulness		.5	.79
PU1	.50		
PU2	.47		
PU3	.46		
PU4	.70		
Perceived ease to use		.5	.79
PEOU1	.64		
PEOU 2	.68		
PEOU 3	.93		
PEOU4	.80		
Medical credibility		.5	.79
PMP1	.73		
PMP 2	.62		
PMP 3	.61		
PMP 4	.75		

[a]All standardized loading values are significant.

Hypotheses Testing

This study used the recommendations of Baron and Kenny [46] and Muller et al [47] to examine the hypotheses. First, as reported in Model 1 of Tables 3 and 4, technology attractiveness were positively significantly related to potential patients' behavior intentions, PU and PEOU were positively related to willingness to use (beta=.29, SE=0.14, *P*=.04; beta =.36, SE=0.11, *P*=.002) and to recommend to other persons (beta =.39, SE=0.12, *P*=.002; beta =.32, SE=0.09, *P*=.001), so Hypothesis 1 was supported. Second, PU and PEOU were positively significantly related to medical credibility (beta =.38, SE=0.08, *P*=.001; beta =.30, SE=0.06, *P*=.001) in Model 2. Third, after considering the technology attractiveness, medical creditability, and potential patients' behavior intentions in Model 3, the coefficient of medical creditability was significant. Furthermore, PU and PEOU in Model 3 were significantly smaller than those in Model 1. Hence, medical credibility had partial mediating effect on technology attractiveness and potential patients' behavior intentions, so Hypothesis 2 was supported. This study also used the Sobel test [55] to examine the mediating effect. Sobel test assumed that the relationships between the dependent variable and independent variables were hypothesized to have an indirect effect due to the influence of mediating effect. More specifically, in a regression model that includes the mediator, the effect of the independent variables is reduced, and the effect of the mediator is still statistically significant. The result showed that the medical credibility played a mediation role on the relationship between technology attractiveness and willingness to use (Z=2.32, SE=0.02, *P*=.03; Z=1.91, SE=0.04, *P*=.06) and the relationship between technology attractiveness and willingness to recommend (Z=1.96, SE=0.05, *P*=.05; Z=2.08, SE=0.03, *P*=.04). Hence, the Sobel test provided a more robust examination, and Hypothesis 2 was supported. Finally, this study explored whether medical information sources played a moderated-mediation effect. The testing conditions 1 and 2 were same as above; this study only described condition 3 here. Model 3 of Tables 3 and 4 showed that the interaction between medical credibility and medical information sources on willingness to use (beta =.01, SE=0.12, *P*=.39) and willingness to recommend to others (beta =.00, SE=.01, *P*=.39) were not significant, so Hypothesis 3 was not supported. Hence, medical information sources did not play a moderated-mediation role on technology attractiveness and potential patients' behavior intentions when medical credibility existed.

Table 3. Hierarchical linear regression results for testing willing to use (mediating effect). Table entries corresponding to the predicting variables are unstandardized estimations of the fixed effects, two-tailed.

Variables[a]	Model 1 Willing to use beta (SE)	Model 2 Medical credibility beta (SE)	Model 3 Willing to use beta (SE)
Constant	1.26 (0.53)	1.16 (0.29)	.87 (0.56)
Independent variable			
Perceived usefulness	.29 (0.14)	.38 (0.08)	.16 (0.06)
Perceived ease to use	.36 (0.11)	.30 (0.06)	.27 (0.13)
Medical credibility			.30 (0.16)
Information source x Medical credibility			.01 (0.12)
r^2	.15	.38	.18

[a]n=150.

Table 4. Hierarchical linear regression results for testing recommend to others (mediating effect). Table entries corresponding to the predicting variables are unstandardized estimations of the fixed effects, two-tailed.

Variables[a]	Model 1 Recommend to others beta (SE)	Model 2 Medical credibility beta (SE)	Model 3 Recommend to others beta (SE)
Constant	.99 (0.43)	1.16 (0.29)	.64 (0.45)
Independent variables			
Perceived usefulness	.39 (0.12)	.38 (0.08)	.28 (0.13)
Perceived ease to use	.32 (0.09)	.30 (0.06)	.23 (0.10)
Medical credibility			.39 (0.12)
Information source x Medical credibility			.00 (0.01)
r^2	.24	.38	.28

[a]n=150.

Discussion

Principal Findings

The purpose of this study is to explore whether technology attractiveness, medical creditability, and different medical information sources could create different behavior intentions. The finding indicates that technology attractiveness must be significantly positive to enhance behavior intentions. In addition, medical credibility is a mediator but not significantly different under diverse manipulative information channels .

Comparison With Prior Work

Prior studies explored technology attractiveness and behavior intention from the viewpoints of physicians and nurses. This study finds a positive relationship between technology attractiveness and potential patients' behavior intentions. The result is completely consistent with Davis [9] TAM models. This finding indicates that even though potential patients have never used WBMS before, they are willing to share information about this medical service with their friends and family members, as they perceive the technology attractiveness of WBMS.

This study also finds that medical credibility plays a mediating role between technology attractiveness and patients' behavior intentions. This finding echoes the conclusions of McGrail et al [20] and Hesse et al [23] that patients have more trust in the medical credibility of face-to-face clinic service than face-to-machine, meaning that medical credibility has an important role in the service delivery process of WBMS.

This study also finds that attitudes do not fully mediate the relationship between technology attractiveness and behavior intention [9]. This study provides additional evidence that medical credibility is an important mediator. Finally, this study finds that different sources of medical information do not significantly moderate the relationship. This finding is not consistent with previous studies that emphasize the effects of different information sources on behavior intention [23,35]. Possible reasons are that the dependent variables of this study are willingness to use and willingness to recommend. The target group of WBMS is patients with chronic medical conditions, such as cardiovascular disease, hypertension, and diabetes. According to social network theory, inexperienced patients may not need this service, but their friends and family may need it. It is not costly to share this information, so the information

channel of WBMS may not be important for inexperienced patients after receiving the information.

Conclusions

This study makes several contributions. Numerous studies have used the technology acceptance model (TAM) to explore customers' intentions after new technology adoption, focusing on physicians' and nurses' intentions, but few studies have investigated the acceptance of web-based medical service by inexperienced patients. In addition, most extensive studies of TAM explore antecedent factors [14,16,21]. This study fills the research gap on mediation effect of TAM. The result shows that medical credibility is an important mediator in WBMS. Finally, this study is one of the few to explore whether differences in sources of medical information have a moderated-mediation effect on web-based medical service.

This study also offers some practical suggestions for telemedicine providers and hospitals wishing to promote WBMS. Patients are accustomed to face-to-face clinic services; they do not like communication via the Internet, especially if they are elderly. This study therefore suggests that telemedicine providers not only emphasize technology attractiveness but also the medical credibility of WBMS when introducing the service to new patients. WBMS combines online and offline services, and therefore, too much emphasis on the technology side of WBMS without taking into account the contributions of medical innovation and medical support may not appeal to the general population. This study suggests that the telemedicine providers and hospitals should pay attention to the convenience of technology (online) and medical credibility (offline) and educate patients on why prevention is better than treatment and why immediate medical intervention is important. Besides, most nonofficial medical companies think they are underperforming because they lack a physician's recommendation. This study finds that the medical information channel is not a significant variable in influencing inexperienced patients' behavior intentions. This study suggests that telemedicine providers should rethink their promotion strategy, rather than spend a lot of money on medical advertisement and doctor-endorsement fees. Finally, most of the target groups comprise patients with chronic disease. According to social network theory, the inexperienced patients may not need this service, but their friends and family may need the service. Telemedicine providers should not underestimate the power of social networks, especially the recommendations of inexperienced patients and the general population.

Limitations

This study uses website descriptions and experiment designs to measure the intentions of inexperienced patients; however, the pay issues and different charging mechanisms are not considered, which could be one of the reasons that make the service difficult to promote. After all, there are still differences between intention and actual use behavior. This study suggests future research should consider this potential problem. Besides, medical information sources such as government promotion or community outreach service have not been considered. Hence, this study suggests further research should consider another effect of different sources.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V1.6.1).

[PDF File (Adobe PDF File), 1MB - jmir_v19i8e285_app1.pdf]

Multimedia Appendix 2

Checklist for reporting results of Internet E-Surveys (CHERRIES).

[PDF File (Adobe PDF File), 39KB - jmir_v19i8e285_app2.pdf]

Multimedia Appendix 3

Questionnaires.

[PDF File (Adobe PDF File), 50KB - jmir_v19i8e285_app3.pdf]

References

1. Desai AS, Stevenson LW. Connecting the circle from home to heart-failure disease management. N Engl J Med 2010;363(24):2364-2367. [doi: 10.1056/NEJMe1011769]
2. Hung M, Jen W. The adoption of mobile health management services: an empirical study. J Med Syst 2012;36(3):1381-1388. [doi: 10.1007/s10916-010-9600-2]
3. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, et al. Telemedical interventional monitoring in heart failure (TIM-HF), a randomized, controlled intervention trial investigating the impact of telemedicine on mortality in

ambulatory patients with heart failure: study design. Eur J Heart Fail 2010;12(12):1354-1362. [doi: 10.1093/eurjhf/hfq199] [Medline: 21098580]
4. Chau P, Hu P. Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. Inf Manage 2002;39(4):297-311. [doi: 10.1016/S0378-7206(01)00098-2]
5. Wu J, Li H, Cheng S, Lin Z. The promising future of healthcare services: when big data analytics meets wearable technology. Inf Manage 2016;53(8):1020-1033. [doi: 10.1016/j.im.2016.07.003]
6. Perednia DA. Telemedicine technology and clinical applications. JAMA 1995;273(6):483-488. [doi: 10.1001/jama.1995.03520300057037]
7. Legris P, Ingham J, Collerette P. Why do people use information technology? a critical review of the technology acceptance model. Inf Manage 2003;40(3):191-204. [doi: 10.1016/S0378-7206(01)00143-4]
8. Chen Y, Ho Y, Huang H. Assessment of the clinical outcomes and cost-effectiveness of the management of systolic heart failure in Chinese patients using a home-based intervention. J Int Med Res 2010;38(1):242-252. [doi: 10.1177/147323001003800129] [Medline: 20233536]
9. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 1989;13(3):319-340. [doi: 10.2307/249008]
10. Delone W, McLean E. The DeLone and McLean model of information systems success: a ten-year update. J Manage Inform Syst 2003;19(4):9-30.
11. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
12. Martins J, Gonçalves R, Branco F. A full scope web accessibility evaluation procedure proposal based on Iberian eHealth accessibility compliance. Comput Hum Behav 2017;73:676-684. [doi: 10.1016/j.chb.2016.12.010]
13. Xiang J, Stanley SJ. From online to offline: exploring the role of e-health consumption, patient involvement, and patient-centered communication on perceptions of health care quality. Comput Hum Behav 2017;70:446-452. [doi: 10.1016/j.chb.2016.12.072]
14. Ammenwerth E, Mansmann U, Iller C, Eichstädter R. Factors affecting and affected by user acceptance of computer-based nursing documentation: results of a two-year study. J Am Med Inform Assoc 2003;10(1):69-84 [FREE Full text] [Medline: 12509358]
15. Chen I, Yang K, Tang F, Huang C, Yu S. Applying the technology acceptance model to explore public health nurses' intentions towards web-based learning: a cross-sectional questionnaire survey. Int J Nurs Stud 2008;45(6):869-878. [doi: 10.1016/j.ijnurstu.2006.11.011] [Medline: 17482191]
16. Lapointe L, Rivard S. Getting physicians to accept new information technology: insights from case studies. CMAJ 2006;174(11):1573-1578 [FREE Full text] [doi: 10.1503/cmaj.050281] [Medline: 16717265]
17. Wu J, Wang S, Lin L. Mobile computing acceptance factors in the healthcare industry: a structural equation model. Int J Med Inform 2007;76(1):66-77. [doi: 10.1016/j.ijmedinf.2006.06.006] [Medline: 16901749]
18. Fishbein M, Ajzen I. AAttitudes and voting behavior: an application of the theory of reasoned action. In: Progress in applied social psychology. Chichester: John Wiley & Sons; 1981.
19. Gefen D, Karahanna E, Straub D. Inexperience and experience with online stores: the importance of tam and trust. IEEE Trans Eng Manage 2003;50(3):307-321. [doi: 10.1109/TEM.2003.817277]
20. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: results of a patient survey and observational study. J Med Internet Res 2017;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
21. Or CK, 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]
22. Sun H, Zhang P. The role of moderating factors in user technology acceptance. Int J Hum Comput Stud 2006;64(2):53-78. [doi: 10.1016/j.ijhcs.2005.04.013]
23. 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]
24. Koufopoulos J, Conner M, Gardner P, Kellar I. A web-based and mobile health social support intervention to promote adherence to inhaled asthma medications: randomized controlled trial. J Med Internet Res 2016;18(6):122-125 [FREE Full text] [doi: 10.2196/jmir.4963] [Medline: 27298211]
25. Peterson G, Aslani P, Williams KA. How do consumers search for and appraise information on medicines on the Internet? a qualitative study using focus groups. J Med Internet Res 2003;5(4):e33 [FREE Full text] [doi: 10.2196/jmir.5.4.e33] [Medline: 14713661]
26. Akter S, D'Ambra J, Ray P. Development and validation of an instrument to measure user perceived service quality of mHealth. Inf Manage 2013;50(4):181-195 [FREE Full text] [doi: 10.1016/j.im.2013.03.001]
27. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029]
28. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635]

29. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336]
30. Ajzen I, Fishbein M. Attitude-behavior relations: a theoretical analysis and review of empirical research. Psychol bull 1977;84(5):888-918. [doi: 10.1037/0033-2909.84.5.888]
31. Hall MA, Camacho F, Dugan E, Balkrishnan R. Trust in the medical profession: conceptual and measurement issues. Health Serv Res 2002;37(5):1419-1439 [FREE Full text] [Medline: 12479504]
32. Thom DH, Stanford Trust Study Physicians. Physician behaviors that predict patient trust. J Fam Pract 2001 Apr;50(4):323-328. [Medline: 11300984]
33. Joshi K. A model of users' perspective on change: the case of information systems technology implementation. MIS Quart 1991;15(2):229-242. [doi: 10.2307/249384]
34. Thom DH, Hall MA, Pawlson LG. Measuring patients' trust in physicians when assessing quality of care. Health Aff 2004;23(4):124-132 [FREE Full text] [doi: 10.1377/hlthaff.23.4.124]
35. Ajzen I. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. J Appl Social Pyschol 2002;32(4):665-683. [doi: 10.1111/j.1559-1816.2002.tb00236.x]
36. Marrie RA, Salter AR, Tyry T, Fox RJ, Cutter GR. Preferred sources of health information in persons with multiple sclerosis: degree of trust and information sought. J Med Internet Res 2013;15(4):e67 [FREE Full text] [doi: 10.2196/jmir.2466] [Medline: 23635393]
37. Van DB, 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;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
38. Random. URL: https://www.random.org/ [accessed 2017-07-03] [WebCite Cache ID 6rgpWuKNE]
39. Telecare. URL: http://mohw.telecare.com.tw [accessed 2017-07-04] [WebCite Cache ID 6rh3lzCaY]
40. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
41. Eysenbach G. Improving the quality of Web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
42. Brislin RW. Translation and content analysis of oral and written material. In: Handbook of cross-cultural psychology. Boston: Allyn and Bacon; 1980:349-444.
43. Anderson J, Gerbing D. Structural equation modeling in practice: A review and recommended two-step approach. Psychol Bull 1988;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
44. Jöreskog K, Sörbom D. LISREL 8: Structural equation modeling with the SIMPLIS command language. Lincolnwood, IL: Scientific Software International, Inc; 1993.
45. Bentler P, Bonett D. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
46. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986;51(6):1173-1182. [doi: 10.1037/0022-3514.51.6.1173]
47. Muller D, Judd CM, Yzerbyt VY. When moderation is mediated and mediation is moderated. J Pers Soc Psychol 2005;89(6):852-863. [doi: 10.1037/0022-3514.89.6.852]
48. Berry W, Feldman S. Multiple regression in practice. Newbury Park, CA: Sage; 1985.
49. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
50. Marsh H, Hau K. Assessing goodness of fit: is parsimony always desirable? J Exp Educ 1996;64(4):364-390. [doi: 10.1080/00220973.1996.10806604]
51. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychol Methods 2002;7(1):64-82. [doi: 10.1037/1082-989X.7.1.64]
52. McColl-Kennedy J, Anderson R. Impact of leadership style and emotions on subordinate performance. Leadersh Q 2002;13(5):545-559. [doi: 10.1016/S1048-9843(02)00143-1]
53. Fornell C, Larcker D. Evaluating structural equation models with unobservable variables and measurement error. J mark Res 1981;18(1):39-50. [doi: 10.2307/3151312]
54. Hair J, Anderson R, Tatham R, Black W. Multivariate Data Analysis. Upper Saddle River, NJ: Prentice Hall; 1998.
55. Sobel M. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol Methodol 1982;13:290-312. [doi: 10.2307/270723]

Abbreviations

AGFI: adjusted goodness-of-fit
AVE: Average Variance Extracted

CFA: confirmatory factor analysis
CFI: comparative fit index
GFI: goodness-of-fit
IFI: incremental fit index
JMIR: journal of medical Internet research
JAMA: journal of the American medical association
NEJM: New England journal medicine
PU: perceived usefulness
PEOU: perceived ease of use
RMSEA: root mean of squared error of approximation
TAM: technology acceptance model
VIF: variance inflation factor
WBMS: Web-based medical services

Edited by G Eysenbach; submitted 30.05.17; peer-reviewed by JJ Wu, B Sapkota; comments to author 28.06.17; accepted 10.07.17; published 02.08.17.

Please cite as:
Wang SH
Web-Based Medical Service: Technology Attractiveness, Medical Creditability, Information Source, and Behavior Intention
J Med Internet Res 2017;19(8):e285
URL: http://www.jmir.org/2017/8/e285/
doi:10.2196/jmir.8114
PMID:28768608

Review

Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review

Simon Hoermann[1*], PhD; Kathryn L McCabe[1,2*], PhD; David N Milne[1], PhD; Rafael A Calvo[1], PhD

[1]Positive Computing Laboratory, School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia

[2]Psychiatry and Behavioral Sciences, University of California, Davis, CA, United States

[*]these authors contributed equally

Corresponding Author:
Rafael A Calvo, PhD
Positive Computing Laboratory
School of Electrical and Information Engineering
The University of Sydney
Building J03
Maze Crescent
Sydney, 2008
Australia
Phone: 61 2 9351 8171
Fax: 61 2 9351 3847
Email: rafael.calvo@sydney.edu.au

Abstract

Background: Synchronous written conversations (or "chats") are becoming increasingly popular as Web-based mental health interventions. Therefore, it is of utmost importance to evaluate and summarize the quality of these interventions.

Objective: The aim of this study was to review the current evidence for the feasibility and effectiveness of online one-on-one mental health interventions that use text-based synchronous chat.

Methods: A systematic search was conducted of the databases relevant to this area of research (Medical Literature Analysis and Retrieval System Online [MEDLINE], PsycINFO, Central, Scopus, EMBASE, Web of Science, IEEE, and ACM). There were no specific selection criteria relating to the participant group. Studies were included if they reported interventions with individual text-based synchronous conversations (ie, chat or text messaging) and a psychological outcome measure.

Results: A total of 24 articles were included in this review. Interventions included a wide range of mental health targets (eg, anxiety, distress, depression, eating disorders, and addiction) and intervention design. Overall, compared with the waitlist (WL) condition, studies showed significant and sustained improvements in mental health outcomes following synchronous text-based intervention, and post treatment improvement equivalent but not superior to treatment as usual (TAU) (eg, face-to-face and telephone counseling).

Conclusions: Feasibility studies indicate substantial innovation in this area of mental health intervention with studies utilizing trained volunteers and chatbot technologies to deliver interventions. While studies of efficacy show positive post-intervention gains, further research is needed to determine whether time requirements for this mode of intervention are feasible in clinical practice.

(J Med Internet Res 2017;19(8):e267) doi:10.2196/jmir.7023

KEYWORDS

chat; dialog system; remote psychotherapy

Introduction

Web-based technologies offer novel ways to deliver mental health interventions that, according to the literature, can be effective [1-3]. One form of online psychological intervention is synchronous text-based, one-to-one conversation (or "chat") in its different incarnations: SMS text messaging (short message service, mobile phone apps, or websites.

Previous systematic reviews by Barak and colleagues [1], as well as by Dowling and Rickwood [4], evaluated psychological

interventions delivered via text-based communication. Focused specifically on synchronous text-based communication, Dowling and Rickwood identified 6 studies meeting their study inclusion criteria, and they noted some evidence to support this mode of mental health intervention. However, they also reported that the small number and overall low quality of the studies indicated weak empirical support for synchronous text-based interventions. Since the writing of these reviews, new technologies have changed the landscape of such mental health interventions. In particular, the penetration of mobile phone technologies and improvements in artificial intelligence and natural language processing (NLP) techniques [5] means that with more tools available to deliver this mode of intervention, one could reasonably expect that the body of research investigating synchronous text-based online interventions have similarly evolved. In particular, the boom of messaging services among major social media companies and the more recent rally to build automated services are driving innovation in this space. Several start-ups companies have been established in this sector and have released products designed to provide mental health services using these new technologies, including, for example, 7cups.com and talklife.co. Although some of these new organizations may be carrying evaluations of their services, the constraints of the commercial world mean not much is known yet. However, the accelerated pace of innovation and uptake in this space requires a systematic evaluation and analysis of the research evidence on how these services are being applied and how well the application of these tools can support mental health and well-being. For example, determining the impact of automation (eg, application of NLP to generate text-based dialogue) on the efficacy of mental health interventions may have significant implications on the cost of providing and accessing mental health services. Studies that use sensor data or automatically generated summaries to drive conversation may also have significant impact on how mental health interventions can be delivered.

The modes used to deliver mental health support have been diversifying and are becoming increasingly technologically complex. At a simple level, chat-based texting has different incarnations: as an SMS service, a dedicated mobile app, or embedded within websites. These different methods of delivery allow for individuals to engage in one-to-one contact with a therapist, a peer, or in some cases even automated “relational agents,” where individuals converse with an artificial intelligence system [6,7]. The conversations may even use information acquired from sensors that track activity or sleep. Some of these innovations have been explored in recent systematic reviews of largely asynchronous applications of texting interventions in mental health [8] and public health more broadly. Asynchronous interventions are when clients do not have the expectation of receiving an immediate reply. These have been used to encourage medication adherence or treatment compliance, aftercare support (eg, [9-11]), deliver appointment reminders [12], and to monitor and capture mood or symptom change [13], especially in conditions with high relapse rates (eg, alcoholism and schizophrenia].

The potential benefits and limitations of synchronous text-based interventions (ie, a dialogue where an immediate reply is expected) still need to be considered. This article is part of a special issue on computing and mental health aimed at bringing both communities together. The computing literature tends to focus on feasibility studies, whereas mental health literature tends to apply better-known technologies to study intervention efficacy. Thus, there is a continuum of research studies examining synchronous text-based interventions from early feasibility studies to those that examine evidence of efficacy. The challenge of bringing these 2 bodies of literature together is addressed in the present systematic review of individual synchronous text-based psychological interventions. Thus, whereas the primary objective of this systematic review is to report studies evaluating the efficacy of such interventions, we also considered it valuable to the reader to include a summary of feasibility studies. Therefore, we provide a concise overview of the state of scientific research on synchronous text-based psychological interventions that we hope will guide researchers as well as mental health service providers to the most promising avenues for research and implementation.

Methods

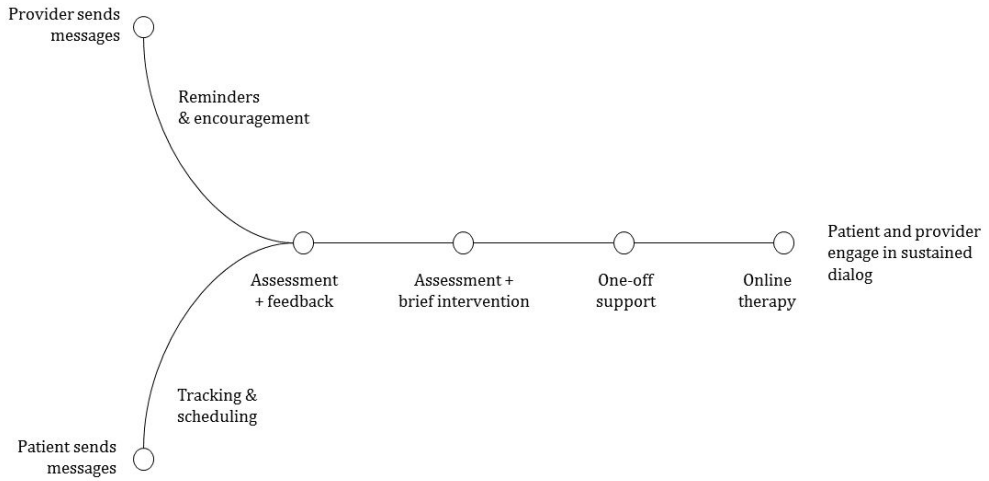
Objective

The objective of this study was to provide a systematic review and synthesis of the application of synchronous one-to-one text-based chat and texting in psychological interventions. The focus of this systematic review is the mode of delivery of the intervention independently of the client group and their impact on mental health outcomes. For the purpose of this review, “Web-based chat” is defined as text-based communication over the Web that allows participants to rapidly exchange messages, and thereby, mimics spoken dialog (eg, text communication part of Facebook messenger, WhatsApp, or Skype). Similarly “texting” also involves the exchange of text-based messages but by using mobile devices and cellular networks for transmission (eg, SMS text messaging or MMS multimedia messaging).

In Figure 1, we present a taxonomy to illustrate text-based communication pathways in psychological interventions. It ranges from leftmost (a) low volume communication reduced to a single message such as an appointment reminder or confirmation that augment the normal therapy to rightmost (b) high volume communication, where the entire intervention is carried out via text-based communication. For this systematic review, only studies that included at least several bidirectional message exchanges were included.

The study followed the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols guidelines [14,15]. The protocol including the exact search strategy for this systematic review was registered prospectively at PROSPERO (Registration Number: CRD420160494490).

Figure 1. Taxonomy of the use of text-based communication in mental health interventions – spectrum from low volume (left) to high volume text-based interaction (right).



Data Sources and Search Strategy

The search strategy included a combination of mental health and computing terminology, as well as keywords related to the type of intervention (eg, dialogue, chat, text, and messaging), intervention, and outcomes (eg, counseling, adherence, feasibility, mood, quality of life, and mental health). Preliminary searches of Medical Literature Analysis and Retrieval System Online (MEDLINE), produced a poor match for the study, and only a limited number of studies were identified. Therefore, the search strategy was revised and terms were expanded to include a broader range of mental health conditions in order to boost the number of studies retrieved. Keywords were collected from various sources including scientific literature, mental health websites, and the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM 5). Also included were search terms referencing situations that are associated with mental health and well-being, such as bullying or cyber bullying, abuse, violence, and stigma.

The revised search strategy was then adapted for application using OvidSP on the following databases: MEDLINE (1996-October 2016), PsycINFO (1987-October 2016), EMBASE (1974-October 2016), as well as the Cochrane Central Register of Controlled Trials (September 2016). Web of Science (1996-present), Scopus (1996-present), ACM Digital Library (1996-present), and IEEE Xplore (1996-present) searches were also performed. The final search strategy was based on the three main topic areas of the review: (1) psychological interventions, (2) use of text-based dialogue, and (3) application to mental health and well-being. Searches in MEDLINE, PsyInfo, and EMBASE were carried out using the advanced search functionality and included keyword truncations and mappings to subject heading (Medical Subject Heading [MeSH] and emtree) that were adapted for each database. Similar to other systematic reviews reported in this area of research [16,17], the reference lists of all studies meeting the inclusion criteria resulting from the database were hand searched to identify all references related to the topic.

The primary focus was psychological outcomes of synchronous text-based intervention, though some technical description, caregiver, or individual evaluations of the intervention were occasionally reported. Based on the taxonomy displayed in Figure 1, we included studies reporting feasibility, adherence, or acceptability of synchronous individual text-based interventions for mental health and well-being.

Selection Criteria

Studies were included if they were published in English and in peer-reviewed scientific journals or conference papers (common high quality publishing venues in engineering and computer science). Specific criteria relating to intervention type and outcome(s) are described below.

Included were studies where synchronous text-based dialogue was used in a psychological intervention, where text-based dialogue was the main intervention or as a control intervention in only one aspect or one arm of the study design. Broad definitions and search terms for mental health and well-being were used for inclusion criteria, all based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) listed psychiatric diagnoses such as addiction (eg, alcohol and cigarette smoking). Intervention terms were also searched, including promoting resilience, motivation, and positive psychological outcomes, as well as mainstream definitions of experiences that affect mental health and well-being.

We excluded studies if the intervention was not primarily based on synchronous text-based communication, such as when the main mode of communication was spoken communication (eg, telephone) or face-to-face communication (eg, video chat and in-person therapy session). Studies were also excluded if the format was not one-on-one, such as group dialogue (eg, a psychologist moderates an online group discussion in a chat-room). Additionally, studies were excluded when the intervention outcomes did not include mental health and well-being, including when studies were limited to health behaviors (eg, weight loss, hypertension reduction, and exercise support). Studies were also excluded if the intervention reported fewer than 5 participants, or utilized unidirectional (ie, only one

entity sends messages) or nonsynchronous (eg, email) communication.

Given the focus of the study was the mode of intervention (synchronous texting), the inclusion of a control condition in the study design was not a key requirement and was therefore not considered grounds for exclusion.

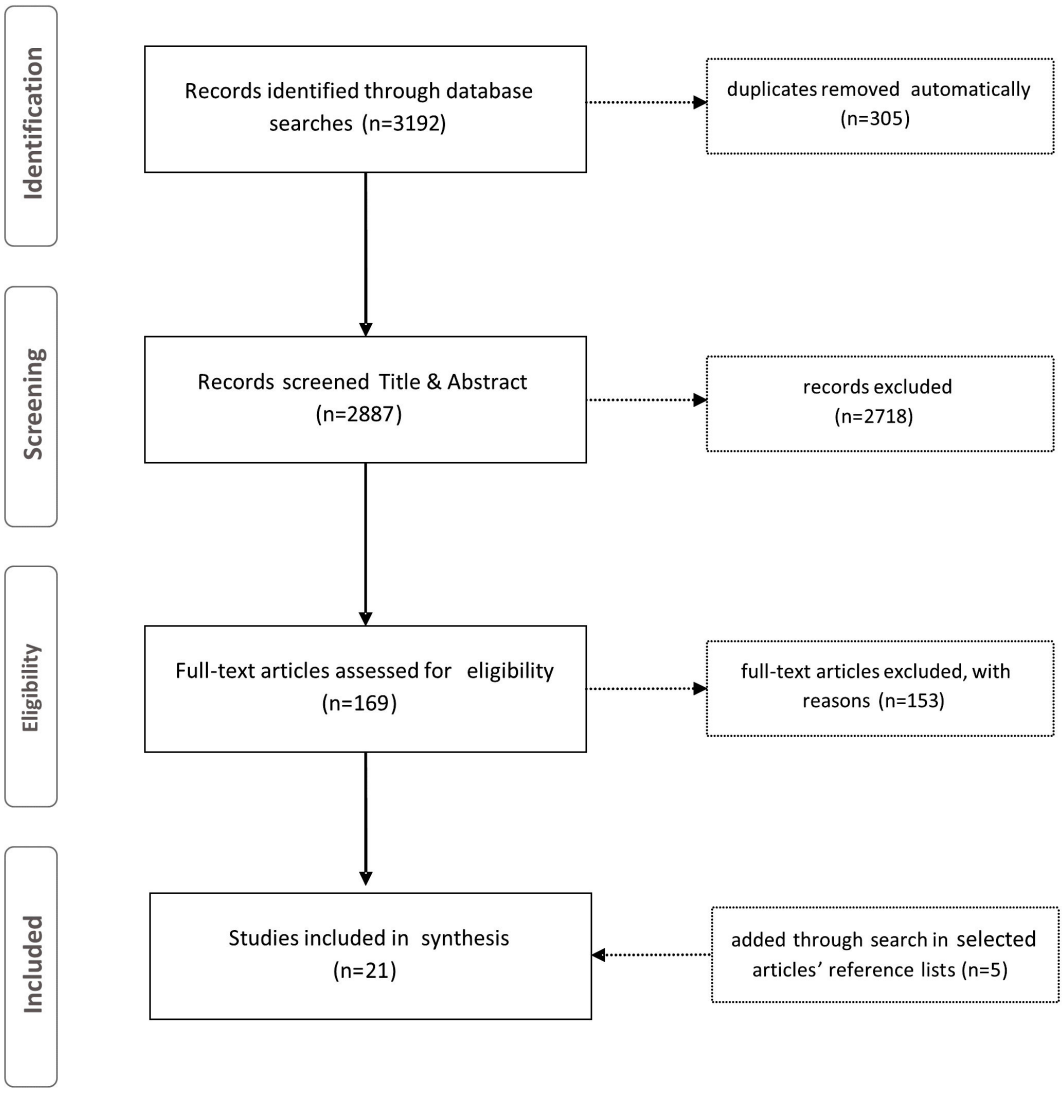
Studies from the database search were exported to systematic review software (Covidence) for ease of management. Duplicate records were deleted. An independent review of study titles and abstract was conducted for all studies by 2 authors (SH and KLM). Articles not meeting the stated inclusion or exclusion criteria were removed. Two authors (SH and KLM) independently conducted full text screening of studies to select the studies to be included. Finally, a reference list search of each of the studies selected from the previous step was conducted to identify additional studies. Additional studies were also independently evaluated by 2 authors (SH and KLM) against the inclusion and exclusion criteria. The same 2 authors extracted data from each study into a form based on systematic review literature. Where consensus was not established, additional reviewers (RAC and DNM) were enlisted to resolve selection disputes.

Results

Study Selection

A total of 3192 articles were retrieved from the following databases: MEDLINE (n=551), PsycINFO (n=1273), EMBASE (n=329), Web of Science (n=311), Scopus (n=271), Central (n=344), IEEE Xplore (n=43), and ACM DL (n=70). After automated identification of duplicates with Endnote X7, 2887 articles remained. Title and abstract-level screening resulted in the elimination of studies that did not report a psychological intervention, did not measure efficacy as a primary outcome, did not use individualized synchronous chat, or consisted of reviews, meta-analyses, dissertation, or opinion. In-depth review of the remaining 169 articles resulted in the elimination of 153 studies. Reference lists of the remaining 16 studies were searched, and a further 5 studies were included in the review. Figure 2 shows an overview of the process. Finally, intervention technology meeting inclusion criteria was identified outside of the database search. These 3 studies have been included in the feasibility studies summary and table (Multimedia Appendices 1 and 2).

Figure 2. Flow of information through different phases of systematic review.



Among the retrieved studies, there was significant heterogeneity in terms of mental health conditions, psychological outcomes, intervention design, and subjective ratings of mental health (eg, mood and symptom ratings). For these reasons, the results are presented in separate tables as well as a narrative analysis. Studies reporting efficacy are described in Multimedia Appendix 3, and studies reporting acceptability or feasibility are presented separately in Multimedia Appendix 1. Additionally, a narrative analysis of studies presented in Multimedia Appendix 1 can be found in Multimedia Appendix 2.

Efficacy Evaluation Studies

Studies S1-S14 [18-31] are described in Multimedia Appendix 3. Information (where available) is provided relating to the study: location, sample, experimental design, type of intervention, intervention duration, control or comparison condition, attrition, and the main results of the study.

In summary, the studies meeting inclusion criteria varied in the target populations, intervention sample size (eg, intervention group range: n=5-152), treatment type, treatment length, as well as specific outcome measures. The majority of studies were completed in Europe and the United Kingdom (S1, S3, S5-7, S9, S10, S13, and S14), with the remaining studies completed in North America (S2, S11, and S12) or Australia (S4 and S8). Participant ages ranged across the lifespan with 5 studies testing interventions in child or adolescents samples (S3, S5, S6, S8, and S10), 3 in adolescent or adults (S2, S4, and S14), and 5 studies focused exclusively on adults (S1, S7, S9, S11, and S12). Some studies were restricted to specific populations (eg, men who have sex with men (S11) or individuals with Autism Spectrum Disorder (ASD), or Attention Deficit Hyperactivity Disorder ([ADHD], S14). The treatment target for the psychological interventions varied and included specific interventions for individuals with depression (S7, S9, S10, and S11), anxiety (S2), substance use (alcohol S1, cannabis S13), or general mental health support (S3, S4, S5, S8, and S12). Intervention length or number of sessions was highly variable, ranging from studies reporting single session interventions (S2, S5, S6, and S8), between 2-5 sessions (S3, S4, and S10), 5-10 sessions (S1, S7, S11, and S13), 10 or more (S14), and 1 intervention lasting 12 months with access to regular text-based counseling (S9). For 1 study, it was difficult to assess the frequency of Web-based text counseling (S12).

In terms of the intervention effectiveness, 1 study was abandoned due to low recruitment rates (S3). Studies that included in their design a waitlist (WL) condition (S1, S7, S10, and S13) showed significant and sustained improvements for participants following the intervention condition compared with WL. S1 text-based chat therapy sessions showed improved outcomes versus WL condition, with significant reduction of large effect in alcohol consumption and improved Quality of Life Scale (QOLS) scores at 3 months and 6 months. For S7, 38.1% (43/113) allocated to text-based therapy intervention compared with 24% (23/97) allocated to WL met study recovery criteria (Beck Depression Inventory [BDI] score<10) at 4 months post-intervention, whereas at 8 months, 42.2% (46/109) versus 25.7% (26/101) met study recovery criteria. In S10, online text-based intervention showed greater improvement at 9 weeks and 4.5 months compared with WL participants. In S13, compared with WL, the selfhelp+chat intervention showed significant posttreatment reduction in cannabis (days of use per week), though groups did not differ on mental health outcome measures (5-item Mental Health Inventory [MHI-5]). Some studies did not include comparison conditions (S4, S11, and S14) and instead, examined change in psychological distress, life satisfaction as a function of the number of sessions completed (S4), or completed pre or post differences on measures of self-esteem (S14) or symptoms of depression or anxiety (S11).

Studies with comparison conditions (S1, S2, S3, S5, S6, S8, and S9) included treatment as usual (TAU) and/or specifically designed comparison conditions (eg, S9). TAU included telephone-based intervention (S5, S6, and S8) and face-to-face intervention (S2, S3, and S12). Compared with TAU, synchronous text-based interventions showed post treatment improvement that was not superior to the TAU condition (S2, S6, and S12). Compared with TAU, one synchronous text-based intervention showed post treatment improvement, though the TAU condition showed marginally greater impact (S8: telephone), and another showed the inverse with larger effect size reported for text-based intervention compared with TAU (telephone: S5).

Feasibility and Acceptability Studies

Information about each included study S15-S24 [32-41] is provided in Multimedia Appendix 1, and a more detailed description of each study is provided as supplementary information in Multimedia Appendix 2.

The wide range of the included feasibility and acceptability studies show that current research is not limited to evaluating Web-based synchronous text-based intervention as a direct substitute for face-to-face therapy. In fact, most studies made use of at least one additional technical possibility to support or augment interventions. For example, S24 stated that synchronous text-based interactions cover only half of the content of a regular intervention, they also successfully evaluated in their study the utilization of pre-developed responses that could be reused in conversation with different clients.

Other technological solutions that were part of studies included chatbots, dynamic customer personalized learning modules, and progress monitoring through Web-based questionnaires, as well as support for scheduling, customer psycho-education, and training. Interventions were mostly provided by professionals. However, novel approaches in synchronous text-based interventions included trained volunteers delivering basic and specific mental health services (S15-17), as well as fully automated interventions where a computer program was substituted for the therapist (S20). A promising approach was to make better use of the time that participants wait before they engage with a human therapist. For example, S18 and S19 showed that the time participants spend in a virtual waiting room could be used meaningfully for them to complete questionnaires. In the future and similar to the intervention described by Lindenburg and colleagues (see S21), automated concierge services that support participants or clients while they wait to access counseling may involve tools for pre-assessment

that could direct clients to psycho-educational information relevant to their needs. The use of Web-based chat services was mentioned to be the main preference for participants (S22) and was shown to have the potential to easily reach a geographically dispersed population in most studies. Web-based chat services may also help individuals to reduce the barrier of help seeking and motivate them to seek face-to-face professional help if required (S22).

Discussion

Principal Findings

There is no doubt that individualized chat-based psychological support is a growing area in the mental health care sector, and this review demonstrates a concomitant increase in the number of published studies reporting outcomes from individualized chat-based psychological therapies. Dowling and Rickwood [4] reported 6 in their systematic review, and we describe an additional 8 studies that have been published in the intervening 3 years and 7 reporting additional features of Web-based synchronous text-based interventions. Whereas Dowling and Rickwood [4] in their previous systematic review reported tentative support for Web-based chat counseling with evidence indicating it was as effective as face-to-face support [19,29], the authors noted limitations related to the overall poor quality of studies. Similarly, our findings are mixed, though overall positive. Compared with WL conditions, synchronous text-based interventions showed post treatment improvement, and synchronous text-based interventions appear to be as effective as TAU, though findings were somewhat varied. Mixed findings are likely to be explained, at least in part by the high variability in intervention design. Length (range: single session-1 year) and target (eg, general health and well-being, depression symptomatology, and ASD symptoms), as well as the type of intervention administered (eg, motivational interviewing and cognitive behavioral therapy), varied amongst the studies. For these reasons, though improvement on measures of outcomes was reported for most studies, it is difficult to say at this stage whether one-to-one synchronous text-based communication is preferentially suitable to one type or length of intervention, or mental health condition over others.

In this review, we note that the only trial that compared a synchronous text-based intervention with a face-to-face condition was abandoned because of low recruitment rates [20]. However, several studies reported no immediate advantage of Web-based text-based communication over self-help, enhanced support, or telephone intervention [18,23,26], whereas others showed significant benefits to Web-based, text-based communication above telephone delivered support [22]. All reported that Web-based support was significantly beneficial in comparison with WL [18,24,27,30], though others [26,29] reported no advantage of Web-based chat support interventions compared with TAU conditions.

Studies indicate that Web-based chat is acceptable and feasible as a mode of therapeutic support, with participants in some studies indicating a preference for this mode of interaction over telephone call. Interestingly, a study by Fukkink et al [23], which also reported that children rated the quality of chat conversations equally or higher than a telephone conversation, found that the average text-based session time was 24 min, almost three times as long as the telephone session (9.3 min). King and colleagues [25] also observed longer counseling times for their text-based intervention, with typical session times between 50-80 min for text-based counseling versus 45-60 min for telephone counseling. Despite this, counseling, satisfaction outcomes in this study were higher in the telephone condition, notwithstanding longer session time in the text-based condition. Increased session length and lower satisfaction may be a consequence of reduced availability of conversational cues during text-based communication, such as the absence of intonation and other nonverbal communication. Nevertheless, the extended times for Web-based chat interventions could also simply be a result of the slower communication channel (it generally takes longer to type than say something). This was shown by Rodda et al [41], who indicated that text-based exchanges generated approximately half the words of an oral conversation. Additionally Fukkink et al [23] found even stronger differences, with an average of 32 words a minute (chat) versus 143 words a minute in the telephone conversation. When taken together, if this mode of intervention delivery generates similar effectiveness but longer session times, generating fewer words, and in some groups, lower satisfaction, this draws into question the clinical practicality of this mode of delivery. Future studies will need to examine how and whether synchronous text-based interventions can best be utilized in the mental health sector and what additional innovations, or constraints may be required before this mode of intervention delivery is financially and logistically viable.

The studies described here have several limitations. Sample sizes were often too small to provide adequate power to studies to explore potential mediating and moderating effects such as other participant or intervention characteristics that may have impacted the outcomes. The ability to detect potential subgroups for whom the intervention may have been particularly effective was also hampered by small sample sizes. Attrition rates remain problematic as well as potential self-selection bias of samples and inadequate (or absent) comparison conditions. However, it should be noted that Crutzen, and colleagues [20] abandoned and then reported their study for this very reason. The lack of adequate documentation or statistical control of additional treatments participants sought out during the interventions is also of concern. The study by Kordy et al [26] is interesting for this reason. They report comparisons to a Web-based chat condition with a quite substantial TAU condition, as well as a structured intervention without a chat component. The heterogeneity of samples participating in studies is also a limitation that requires consideration. Whereas several studies state specific psychiatric or substance dependency criteria for inclusion, others recruited cross diagnostically and lacked adequate assessment of comorbid psychiatric conditions. Furthermore, self-report and Web-based completion of psychometric measures, as well as when and where measures were completed, was noted by Rodda and Lubman [41] as potential limitations of Web-based chat-based interventions that may influence the reliability and validity of study findings. Finally, the “dose” and intensity of Web-based chat interventions was poorly articulated by many studies and varied between the

studies with interventions ranging from a single session to 12 months. Follow-up time frames were also absent or inadequate for some studies that in turn limited the ability to examine the durability of any treatment outcome gains, as well as the generalizability of positive outcomes.

Limitations

It is important to note that some of the journals reporting synchronous text-based individualized interventions were not yet registered with the major databases. Therefore, they were not identified as part of the initial database search. Also, many innovations may have been published in technology conferences. Though the primary databases for these conference manuscripts were included (ie, ACM; IEEE), our search may not have captured other technically oriented manuscripts. The fact that these innovations do not appear in our review, even though the major engineering databases were searched, is likely because they lack the type of evaluations required in the mental health domain. The type of outcome measures of technologically driven papers is usually related to the accuracy of the algorithms and not mental health. More collaboration between computing scientists and mental health researchers is required to address this issue.

Another limitation impacting this area of research and that impacted our search yield in the current review relates to the lack of standard nomenclature to describe Web-based psychological interventions. This issue has been discussed in previous systematic reviews [1,4]. For example, when refining our database abstract or title or keywords search strategy, we found that including the keyword "support" in our search query would have resulted in 6759 instead of 591 returns in MEDLINE (on November 4, 2016). Thus, given the heterogeneity in terminology applied in this relatively new field of research, we suspect that despite our best effort, due to terminology used by some studies, these manuscripts may have been missed in our literature search. In order to minimize the likelihood of missing relevant research, we also reviewed reference lists of included manuscripts (ie, backward and forward reference search).

Future Directions

The authors expect a strong trend toward more Web-based mental health services delivered using synchronous text-based systems. This trend is already visible with the recent introduction of commercial platforms such as Koko (http://www.itskoko.com) or Ginger.io (http://Ginger.io) that for a fee provide psychological support and include individualized chat-based access to clinicians and in the case of Koko, moderated peer support for mental health problems. Other platforms are also beginning to offer infrastructure for chat-based services (eg, qntfy.com). As aforementioned, new services from the private sector populating the mental health space are largely unevaluated, and therefore, there is limited research evidence regarding how effectively these services support mental health and well-being. Despite the promise of this mode of intervention and commercial interest in its application, the evidence described here would seem to suggest that the clinical viability of this mode of delivery has yet to be proven.

The introduction of conversational agents might play an important role. In the study described by Gaffney et al [37], a conversational agent was trained to utilize principles of Method of Levels (MOL) therapy. The optimization and augmentation of human service through artificial intelligence is speculated to be a promising area of future research and commercial application [42,43]. Another example, 7Cups (www.7cups.com) has introduced a "therapist bot" that greets users and carries out basic assessments to aid in evaluating, based on algorithms, their needs and how best to connect them with a human therapist to chat with on the Web. However, future research still needs to address the suitability of this type of technology for mental health and well-being interventions. For instance, it remains unclear whether it is more suited to particular mental health problems, is best suited to short- or long-term intervention, or whether it is best applied as an adjunct or screening tool to streamline services.

Feasibility studies indicate that text-based interventions are well tolerated by users seeking help for mental health concerns. A significant advantage of this mode of intervention noted by studies was the availability and immediacy of support. Users may prefer immediate support when they seek it, even when trained volunteers or artificial agents are delivering this support. For these reasons, overall current research suggests that synchronous text-based communication between clients and intervention providers is feasible and acceptable for a wide range of users. However, as to whether these interventions are an adequate substitute for non-text based intervention cannot be answered based on the current level of evidence from studies of efficacy.

Conclusions

This review provides an evaluation of individual synchronous Web-based chat technologies as a mode of psychological intervention and support. Based on the current evidence of the application of this technology in this area of mental health research, we see tentative support for this mode of intervention. Interventions utilizing text-based synchronous communication showed better outcomes compared with WL conditions and overall equivalent (though variable) outcomes compared with TAU, and were at least as good as the comparison interventions. However, the issue of whether these technologies are cost effective in clinical practice remains a consideration for future research studies. New technologies are rapidly changing the area by increasing the capacity to automate some of the components of synchronous text-based intervention; however, our conclusions remain tentative because of the limited number of studies reporting randomized controlled trial (RCT) evaluation of interventions and adequate follow-up procedures. This review raises the need for more interdisciplinary work in evaluating new Web-based synchronous text-based technologies.

Acknowledgments

This project is funded by Australian Research Council Linkage Grant LP130100453. RAC is supported by Australian Research Council Future Fellowship FT140100824.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of feasibility studies.

[PDF File (Adobe PDF File), 48KB - jmir_v19i8e267_app1.pdf]

Multimedia Appendix 2

Narrative summary of feasibility studies.

[PDF File (Adobe PDF File), 70KB - jmir_v19i8e267_app2.pdf]

Multimedia Appendix 3

Summary of efficacy studies.

[PDF File (Adobe PDF File), 109KB - jmir_v19i8e267_app3.pdf]

References

1. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008;26(2-4):109-160. [doi: 10.1080/15228830802094429]
2. Griffiths KM, Christensen H. Review of randomised controlled trials of internet interventions for mental disorders and related conditions. Clin Psychol 2006;10(1):16-29. [doi: 10.1080/13284200500378696]
3. 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]
4. Dowling M, Rickwood D. Online counseling and therapy for mental health problems: a systematic review of individual synchronous interventions using chat. J Technol Hum Serv 2013;31(1):1-21. [doi: 10.1080/15228835.2012.728508]
5. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng 2017 Jan 30:1-37. [doi: 10.1017/S1351324916000383]
6. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 2005 Jun;12(2):293-327. [doi: 10.1145/1067860.1067867]
7. Bohannon J. Artificial intelligence. The synthetic therapist. Science 2015 Jul 17;349(6245):250-251. [doi: 10.1126/science.349.6245.250] [Medline: 26185240]
8. 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]
9. Keoleian V, Polcin D, Galloway GP. Text messaging for addiction: a review. J Psychoactive Drugs 2015 May;47(2):158-176 [FREE Full text] [doi: 10.1080/02791072.2015.1009200] [Medline: 25950596]
10. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
11. Robinson S, Perkins S, Bauer S, Hammond N, Treasure J, Schmidt U. Aftercare intervention through text messaging in the treatment of bulimia nervosa--feasibility pilot. Int J Eat Disord 2006 Dec;39(8):633-638. [doi: 10.1002/eat.20272] [Medline: 16937381]
12. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
13. Depp CA, Mausbach B, Granholm E, Cardenas V, Ben-Zeev D, Patterson TL, et al. Mobile interventions for severe mental illness: design and preliminary data from three approaches. J Nerv Ment Dis 2010 Oct;198(10):715-721 [FREE Full text] [doi: 10.1097/NMD.0b013e3181f49ea3] [Medline: 20921861]
14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [Medline: 19622551]

15. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;349:g7647 [FREE Full text] [Medline: 25555855]
16. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
17. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014 Sep 16;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
18. Blankers M, Koeter MW, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341. [doi: 10.1037/a0023498]
19. Cohen GE, Kerr BA. Computer-Mediated Counseling. Computers in Human Services 1999;15(4):13-26. [doi: 10.1300/J407v15n04_02]
20. Crutzen R, Bosma H, Havas J, Feron F. What can we learn from a failed trial: insight into non-participation in a chat-based intervention trial for adolescents with psychosocial problems. BMC Res Notes 2014 Nov 20;7:824 [FREE Full text] [doi: 10.1186/1756-0500-7-824] [Medline: 25409911]
21. Dowling M, Rickwood D. A naturalistic study of the effects of synchronous online chat counselling on young people's psychological distress, life satisfaction and hope. Counsell Psychother Res J 2015;15(4):274-283.
22. Fukkink RG, Hermanns JM. Children's experiences with chat support and telephone support. J Child Psychol Psychiatry 2009 Jun;50(6):759-766. [doi: 10.1111/j.1469-7610.2008.02024.x] [Medline: 19207634]
23. Fukkink R, Hermanns J. Counseling children at a helpline: chatting or calling? J Community Psychol 2009;37(8):939-948. [doi: 10.1002/jcop.20340]
24. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial. Lancet 2009 Aug 22;374(9690):628-634. [doi: 10.1016/S0140-6736(09)61257-5] [Medline: 19700005]
25. King R, Bambling M, Reid W, Thomas I. Telephone and online counselling for young people: a naturalistic comparison of session outcome, session impact and therapeutic alliance. Counsell Psychother Res J 2006 Sep;6(3):175-181. [doi: 10.1080/14733140600874084]
26. Kordy H, Wolf M, Aulich K, Bürgy M, Hegerl U, Hüsing J, et al. Internet-delivered disease management for recurrent depression: a multicenter randomized controlled trial. Psychother Psychosom 2016;85(2):91-98. [doi: 10.1159/000441951] [Medline: 26808817]
27. Kramer J, Conijn B, Oijevaar P, Riper H. Effectiveness of a web-based solution-focused brief chat treatment for depressed adolescents and young adults: randomized controlled trial. J Med Internet Res 2014 May 29;16(5):e141. [doi: 10.2196/jmir.3261] [Medline: 24874006]
28. Lelutiu-Weinberger C, Pachankis JE, Gamarel KE, Surace A, Golub SA, Parsons JT. Feasibility, acceptability, and preliminary efficacy of a live-chat social media intervention to reduce HIV risk among young men who have sex with men. AIDS Behav 2015 Jul;19(7):1214-1227 [FREE Full text] [doi: 10.1007/s10461-014-0911-z] [Medline: 25256808]
29. Murphy L, Parnass P, Mitchell DL, Hallett RH, Cayley P, Seagram S. Client satisfaction and outcome comparisons of online and face-to-face counselling methods. Br J Soc Work 2009 Apr 15;39(4):627-640. [doi: 10.1093/bjsw/bcp041]
30. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
31. Wentz E, Nydén A, Krevers B. Development of an internet-based support and coaching model for adolescents and young adults with ADHD and autism spectrum disorders: a pilot study. Eur Child Adolesc Psychiatry 2012 Nov;21(11):611-622. [doi: 10.1007/s00787-012-0297-2] [Medline: 22736195]
32. Baumel A. Online emotional support delivered by trained volunteers: users' satisfaction and their perception of the service compared to psychotherapy. J Ment Health 2015;24(5):313-320. [doi: 10.3109/09638237.2015.1079308] [Medline: 26485198]
33. Baumel A, Correll CU, Birnbaum M. Adaptation of a peer based online emotional support program as an adjunct to treatment for people with schizophrenia-spectrum disorders. Internet Interv 2016 May;4(Part 1):35-42. [doi: 10.1016/j.invent.2016.03.003]
34. Baumel A, Schueller SM. Adjusting an available online peer support platform in a program to supplement the treatment of perinatal depression and anxiety. JMIR Ment Health 2016 Mar 21;3(1):e11 [FREE Full text] [doi: 10.2196/mental.5335] [Medline: 27001373]
35. Dowling M, Rickwood D. Investigating individual online synchronous chat counselling processes and treatment outcomes for young people. Advances in Mental Health 2014;12(3):216-224. [doi: 10.1080/18374905.2014.11081899]
36. Dowling M, Rickwood D. Exploring hope and expectations in the youth mental health online counselling environment. Comput Human Behav 2016 Feb;55(Part A):62-68. [doi: 10.1016/j.chb.2015.08.009]

37. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. Behav Cogn Psychother 2014 Nov;42(6):731-746. [doi: 10.1017/S135246581300060X] [Medline: 23899405]
38. Lindenberg K, Moessner M, Harney J, McLaughlin O, Bauer S. E-health for individualized prevention of eating disorders. Clin Pract Epidemiol Ment Health 2011;7:74-83 [FREE Full text] [doi: 10.2174/1745017901107010074] [Medline: 21687562]
39. Pietrabissa G, Manzoni G, Algeri D, Mazzucchelli L, Carella A, Pagnini F, et al. Facebook use as access facilitator for consulting psychology. Aust Psychol 2015 Jul 20;50(4):299-303. [doi: 10.1111/ap.12139]
40. Reynolds Jr DJ, Stiles WB, Bailer AJ, Hughes MR. Impact of exchanges and client-therapist alliance in online-text psychotherapy. Cyberpsychol Behav Soc Netw 2013 May;16(5):370-377 [FREE Full text] [doi: 10.1089/cyber.2012.0195] [Medline: 23530546]
41. Rodda S, Lubman DI. Characteristics of gamblers using a national online counselling service for problem gambling. J Gambl Stud 2014 Jun;30(2):277-289. [doi: 10.1007/s10899-012-9352-7] [Medline: 23297169]
42. Milne D, Hoermann S, Calvo R. A trial of real-time text-based support for young people. In: CHI - Computing and Mental Health Workshop. 2016 Presented at: CHI 2016; 08 May 2016; San Jose, California, USA.
43. Naqshbandi K, Milne DN, Davies B, Potter S, Calvo RA, Hoermann S. Helping young people going through tough times: Perspectives for a peer-to-peer chat support system. In: OzCHI '16 Proceedings of the 28th Australian Conference on Computer-Human Interaction. New York, NY, USA: ACM; 2016 Presented at: OzCHI '16; November 29 - December 02, 2016; Launceston, Tasmania, Australia p. 640-642. [doi: 10.1145/3010915.3011848]

Abbreviations

TAU: treatment as usual
SMS: short message service
NLP: natural language processing
MMS: multimedia messaging service
MEDLINE: Medical Literature Analysis and Retrieval System Online
DSM: Diagnostic and Statistical Manual of Mental Disorders
ADHD: attention-deficit/hyperactivity disorder.
ASD: autism spectrum disorder.
QOLS: Quality of Life Scale.
BDI: Beck Depression Inventory.
MHI-5: 5-item Mental Health Inventory.

Edited by H Christensen; submitted 22.11.16; peer-reviewed by D Rickwood, B Puschner; comments to author 28.01.17; revised version received 05.04.17; accepted 16.05.17; published 21.07.17.

Please cite as:
Hoermann S, McCabe KL, Milne DN, Calvo RA
Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review
J Med Internet Res 2017;19(8):e267
URL: http://www.jmir.org/2017/8/e267/
doi:10.2196/jmir.7023
PMID:28784594

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org